J. Linguistics **42** (2006), 531–573. © 2006 Cambridge University Press
doi:10.1017/S0022226706004191 Printed in the United Kingdom

Word-based morphology[1]

JAMES P. BLEVINS

University of Cambridge

(Received 17 December 2004; revised 22 March 2006)

This paper examines two contrasting perspectives on morphological analysis, and considers inflectional patterns that bear on the choice between these alternatives. On what is termed an ABSTRACTIVE perspective, surface word forms are regarded as basic morphotactic units of a grammatical system, with roots, stems and exponents treated as abstractions over a lexicon of word forms. This traditional standpoint is contrasted with the more CONSTRUCTIVE perspective of post-Bloomfieldian models, in which surface word forms are 'built' from sub-word units. Part of the interest of this contrast is that it cuts across conventional divisions of morphological models. Thus, realization-based models are morphosyntactically 'word-based' in the sense that they regard words as the minimal meaningful units of a grammatical system. Yet morphotactically, these models tend to adopt a constructive 'root-based' or 'stem-based' perspective. An examination of some form-class patterns in Saami, Estonian and Georgian highlights advantages of an abstractive model, and suggests that these advantages derive from the fact that sets of words often predict other word forms and determine a morphotactic analysis of their parts, whereas sets of sub-word units are of limited predictive value and typically do not provide enough information to recover word forms.

1. Introduction

Morphological models tend to be classified in terms of the units that they treat as grammatically 'meaningful' and the way that they associate properties with these units. The post-Bloomfieldian model is regarded as 'morpheme-based', on the grounds that it associates grammatical properties with individual morphs. Realization-based models are described as 'word-based' because they associate properties with words. Yet models can also be classified MORPHOTACTICALLY, in terms of the status that they assign to these units. From a morphotactic perspective, a model is 'word-based' if it treats surface word forms as the basic elements of a system, and

[1] I would like to thank Farrell Ackerman, Emmon Bach, Juliette Blevins, Andrew Garrett, Alice Harris, James Kirby, Andrew Spencer and audiences at the University of York and at the University of Leipzig for comments, discussion and feedback on earlier versions of this paper. I am also grateful to the *JL* editors, Orin Gensler and two anonymous *JL* reviewers for suggestions and criticisms that have led to improvements in the current version.

https://doi.org/10.1017/S0022226706004191 Published online by Cambridge University Press

regards roots, stems and exponents as abstractions over a set of full forms. A model is 'root-based' or 'morph-based' if it assumes an inventory of morphotactically minimal forms, from which surface forms are 'built' or 'derived'.

The morphotactic assumptions of a model strongly influence the types of analysis that the model assigns. This influence is particularly salient in the treatment of morphological classes. In languages whose morphological systems are organized into inflectional classes, the shape of one or more word forms tends to identify the class of an item. Traditional models exploit this predictability by establishing a set of exemplary paradigms and representing individual items by diagnostic surface forms. Yet the properties of roots or exponents in isolation are rarely reliable indicators of inflection class. Hence models that represent items by underlying root forms must often introduce diacritic class properties to restore lost information about inflection class.

The morphotactic assumptions of a model have a parallel effect on the treatment of various types of lexical classes. Given a set of surface forms, it is often possible to identify a shared root, which identifies an individual lexeme, along with stem formatives and inflectional exponents, whose distribution is associated with particular lexical classes. Yet the choice of formatives and exponents is frequently not predictable from either the form or the properties of a root. A model that represents items by roots thus requires various lexical class features to provide 'assembly instructions' for choices that are underdetermined by the properties of lexical roots alone.

Morphotactic assumptions also underlie the problems raised by the patterns of stem syncretism that Matthews (1972) terms 'parasitic' and Aronoff (1994) calls 'morphomic'. A stem is MORPHOMIC in the intended sense if it underlies a number of surface forms, but does not realize any consistent set of grammatically meaningful properties in those forms. Matthews (1991) sets out the descriptive issues posed by morphomic stems in his discussion of participial stem syncretism in Latin.

> Consider next the opposition between the Future Participle (Active) and the Past Participle (Passive). For a Verb like AMO 'to love', the latter is based on a stem *amāt-* (Nominative Singular Masculine *amāt-u-s*). The former is based correspondingly on *amātūr-* (Nominative Singular Masculine *amātūr-u-s*). But what is the relation between them? In terms of formatives, the Future Active *amātūr-* seems to derive from *amāt-* by the addition of *-ūr-*. Or, as an ancient grammarian would have put it, *amātūrus* comes from *amātus* by the change of *-s* to *-rus*. But there is no sense in which the meaning of the Future Active Participle includes that of the Past Passive Participle. Formally, *amāt-ūr-* includes *amāt-*. But in meaning all they have in common is that both are Participles. (Matthews 1991: 200)

https://doi.org/10.1017/S0022226706004191 Published online by Cambridge University Press

The problem, in short, is one of assigning a discrete 'meaning' to *amāt*. This problem only arises if one regards *amāt* as a 'minimum unit' from which future active and past passive participles are CONSTRUCTED. This is precisely the viewpoint that Aronoff (1994: 167) adopts when he remarks that 'this stem can be defined in neither phonological nor semantic nor syntactic terms, but only in terms of which forms are built on it'. In a word-based model of the sort that Matthews (1991) attributes to 'an ancient grammarian', it is unproblematic that a common stem can be abstracted from morphosyntactically distinct participial paradigms. The full surface forms each realize a distinctive set of properties, and exhibit a correspondence in form that make future active participles reliable predictors of past passive participles, and vice versa. It is only when one attempts to define a stem ENTRY for forms like *amāt* that any problem of analysis arises.

The descriptive challenges raised by inflection classes, lexical classes and morphomic stems illustrate different facets of a single phenomenon. In each case, an analysis that takes a larger form as the basis for abstracting smaller forms avoids difficulties that arise if the smaller forms are taken as the basis for deriving the larger forms. Patterns of this sort lend a strong measure of support to Anderson's (1992: 369) suggestion that a word-based model which 'regards the grammar as a set of relations among full surface forms … may merit more consideration than it has sometimes received'. The present paper presents a sustained argument for this position.

2. TWO APPROACHES TO MORPHOLOGICAL DESCRIPTION

A morphological analysis can approach the patterns in a grammatical system from one of two directions. One type of analysis isolates recurrent bases and exponents within a system, encapsulates each of these elements in an individual rule or entry that represents their grammatical properties, and then derives surface word forms from these simple elements by rules or other combinatoric principles. A second type of analysis treats word forms as the basic units of a system, and classifies recurrent parts as abstractions over full forms. There is an obvious sense in which the first alternative is 'morph-based' and the second is 'word-based'. Nevertheless, to avoid overloading terms such as 'word-based', which already have a range of meanings, it will be useful to refer to models that are word-based in a morphotactic sense as ABSTRACTIVE, and root-based, stem-based or morph-based models as CONSTRUCTIVE. The constructive method is strongly associated with post-Bloomfieldian models, in which morphological analysis is essentially a process of segmentation and classification. The abstractive method is characteristic of the pre-Bloomfieldian tradition represented by Paul (1880), Saussure (1916) and Kuryłowicz (1964).

https://doi.org/10.1017/S0022226706004191 Published online by Cambridge University Press

2.1 *Constructive approaches*

For the most part, the contrast between constructive and abstractive perspectives is orthogonal to the properties that are usually taken to distinguish morphological models. Each of the models that Hockett (1954) identifies, namely 'item and arrangement' (IA), 'item and process' (IP), and 'word and paradigm' (WP), can be interpreted constructively. A constructive perspective is implicit in the idea that morphological analysis 'isolates minimum meaningful elements' and describes 'the arrangements in which the minimum meaningful elements occur' (Hockett 1947: 229). This description of a standard IA model applies without significant qualification to contemporary descendents, such as Lieber (1992). An IP model is similarly constructive when it regards a 'derived form' as consisting of 'one or more underlying FORMS to which a PROCESS has been applied' (Hockett 1954: 396). This remains true of more recent IP models, such as Steele (1995). Even 'realization-based' models are constructive in orientation, to the point that most contemporary 'word and paradigm' approaches are more accurately described as 'stem and paradigm' models. A constructive orientation is clearly reflected in Aronoff's (1994: 167) characterization of the 'third stem' in Latin 'in terms of which forms are BUILT on it' (emphasis added). A constructive perspective is also implicit in Anderson's (1992: 134) definition of 'an item's paradigm' as 'the complete set of surface word forms that can be projected from the members of its stem set by means of the inflectional Word Formation Rules of the language'. A similarly constructive orientation underlies the notion of a 'paradigm function' that maps 'the root of a lexeme' and 'a complete set of morphosyntactic properties' (Stump 2001: 43) onto the paradigm cell occupied by a surface form of the lexeme. There are important differences in the way that surface forms are derived in IA, IP and WP models. Nevertheless, each of these approaches takes at least some minimal forms as a point of departure for the derivation of larger units.

A further assumption that is strongly characteristic of constructive approaches is the idea that individual forms are derived in isolation from the other forms in a grammatical system. This assumption is clearly reflected in the purely syntagmatic structure of IA and IP models, as well as hybrid models such as Distributed Morphology (Halle & Marantz 1993). In a standard IA or IP model, word forms are 'assembled' or 'processed' one at a time, either by retrieving sub-word units from the lexicon or by applying rules that encapsulate these units. Since words represent the output of a morphological derivation, they have no status as permanent lexical units, and the derivation of a given form has no access to 'paradigmatic' information about other word forms.

Individual derivations are isolated to a similar degree in most realization-based approaches. In A-Morphous Morphology (Anderson 1992), inflected

https://doi.org/10.1017/S0022226706004191 Published online by Cambridge University Press

word forms are constructed by 'spelling out' the properties associated with syntactic preterminals, so that words again have no existence outside the syntactic context in which they are constructed. In principle, models that make use of 'rules of referral' (Zwicky 1985; Stump 1993, 2001) introduce a paradigmatic dimension. But much depends on the interpretation of these rules. If the expression '$\mathbb{R}(\mathcal{S})$' is taken to represent 'the realization of the property set $\mathcal{S}$', then the rule '$\mathbb{R}(\mathcal{S}) = \mathbb{R}(\mathcal{T})$' can be interpreted as defining $\mathbb{R}(\mathcal{S})$ in terms of $\mathbb{R}(\mathcal{T})$, the realization of the property set $\mathcal{T}$. Now if the set $\mathcal{T}$ contains all of the distinctive properties that would characterize the 'paradigm cell' or lexical entry of a word form φ, then one could interpret the realization rule paradigmatically, as a rule that sets the value of $\mathbb{R}(\mathcal{S})$ to whatever word form spells out $\mathcal{T}$, i.e. $\mathbb{R}(\mathcal{S}) = \mathbb{R}(\mathcal{T}) = \varphi$. But if, as proposed in Stump (1993: 450), 'realization rules … merely encompass the individual steps by which an individual word is built up from the root of its paradigm', then the introduction of referral rules will remain consistent with the assumption that forms are defined in isolation.[2]

The idea that morphological forms are derived in isolation may be regarded merely as a theoretical idealization. However, a substantial body of psycholinguistic research suggests that this idealization is psychologically implausible. It has been shown that the frequency of inflected forms and the size of morphological 'families' have a robust effect on lexical decision tasks in a range of languages. One line of research has investigated the correlation between response latencies for inflected forms and the token frequency of the elements of their inflectional paradigms (Taft 1979, Baayen, Lieber & Schreuder 1997, Hay 2001). A second line of research has demonstrated that the processing of an item is facilitated by the semantically transparent items that make up its morphological family (Schreuder & Baayen 1997, de Jong 2002, Moscoso del Prado Martín 2003). A third line of research attempts to provide a single information-theoretic measure that subsumes the token-frequency effects relevant to inflectional paradigms and the type-frequency effects relevant to morphological families (Kostić, Marković & Baucal 2003, Moscoso del Prado Martín, Kostić & Baayen 2004). There is ongoing debate about the best way to model these effects. Yet two things are clear from this research. The first is that the processing of any given form may be facilitated (or possibly inhibited) by other, related forms. The second is that these types of effects can only be understood if the related forms are available as elements of a speaker's mental lexicon.

[2] It is immaterial whether a stepwise referral rule is interpreted 'extensionally', by assigning $\mathbb{R}(\mathcal{S})$ the exponent that spells out $\mathbb{R}(\mathcal{T})$ or whether it is interpreted more 'intensionally', as a kind of 'metarule' (Matthews 1991: 201) that realizes $\mathbb{R}(\mathcal{S})$ by applying the rules that spell out $\mathbb{R}(\mathcal{T})$. On either interpretation, the application of a referral rule in the derivation of a given word form makes no essential reference to any other word form.

https://doi.org/10.1017/S0022226706004191 Published online by Cambridge University Press

2.2 *Abstractive approaches*

The mutual influence exerted by related forms fits well with a traditional conception which 'regards the grammar as a set of relations among full surface forms' (Anderson 1992: 369). Kuryłowicz (1949: 160) articulates this type of abstractive perspective when he remarks that 'the stem is thus a type of abstraction intended to embody the paradigm' ('Le thème est donc une sorte d'abstraction déstinée à résumer le paradigme'). The priority of surface forms is likewise explicit in Kuryłowicz's (1949: 159) description of the relation between declensional stems and case forms:

> Car la notion du thème est postérieure aux formes concretes composant le paradigme: on trouve le thème en dégageant les éléments communs à toutes les formes casuelles du paradigme (quand il s'agit de la déclinaison). P. e. *lup-us*, *-i*, *-o*, *-um*, *-orum*, *-is*, *-os* fondent le thème *lup-*.[3]

The claim that stems are founded on 'all of the case forms of a paradigm' ('toutes les formes casuelles du paradigme') also signals the relational character of an abstractive approach. This type of approach does not assume that the morphological analysis of a form can, in general, be given in isolation. An analysis may, in particular, depend on other forms in an inflectional paradigm. As this example also shows, a WP model (Robins 1959, Hockett 1967, Matthews 1972) may recognize roots and exponents as 'units of analysis' but these units are purely derivative. They are not 'meaningful' in isolation, and they are morphotactically dependent on surface word forms.

The morphotactic priority assigned to words in abstractive approaches is thus highly compatible with traditional WP models, which treat words as the 'minimum meaningful elements' of a grammatical system. Yet the central assumptions of an abstractive perspective are, in principle, independent of particular views of word-internal structure. An abstractive model is fully compatible with the practice of segmenting words into component morphs, as in IA models, provided that these morphs are regarded as abstractions over forms, not as the 'building blocks' from which the forms are constructed. The 'derivational history' represented by an IP analysis can likewise be interpreted as the record of a process that abstracts a base and exponents from a surface form. One might therefore ask whether constructive and abstractive perspectives are alternative conceptions of morphological analysis or whether they are just complementary 'modes'

[3] 'For the notion of the stem is dependent on the concrete forms composing the paradigm: one finds the stem in extracting the elements common to all the case forms of a paradigm (when dealing with declension). For example, *lup-us*, *-i*, *-o*, *-um*, *-orum*, *-is*, *-os* found the stem *lup-*.' [All translations are the author's. Any emphasis marked by small caps in quotations is repeated from the original passages unless they are explicitly identified as added.]

https://doi.org/10.1017/S0022226706004191 Published online by Cambridge University Press

of combination. Could one just regard a constructive perspective as representing a 'bottom-up' mode of combination, and an abstractive perspective as representing a 'top-down' mode?

The answer is 'no', because there are significant empirical differences between these perspectives, which derive from different assumptions about the basic 'units of storage' in the grammatical system, and, by extension, in the speaker's mental lexicon. Constructive approaches assume that the basic units of a grammatical system are segmentally minimal, and that the open-class lexicon consists, at least for the most part, of roots and exponents (or rules that encapsulate exponents), but not full word forms that contain these elements. An abstractive approach assumes that the lexicon consists in the main of full forms, from which recurrent parts are abstracted. These different 'ontological' commitments go together with different ideas about how forms are related. The notion of a DERIVATION, which builds larger units from smaller elements, is central to constructive approaches. In abstractive approaches, PREDICTABILITY is the key relation. There is no requirement that one form should underlie another; a derivational relation is just a limiting case in which one of two forms in a predictability relation is a proper part of the other.

Constructive strategies for representing morphological patterns using minimal elements and combinatoric principles are relatively familiar. Post-Bloomfieldian accounts break a grammatical system down into a stock of minimal 'underlying forms', together with a set of rules for 'assembling', 'processing' or otherwise combining these forms. Realization-based accounts, such as Anderson (1992), Aronoff (1994) and Stump (2001), likewise operate with an inventory of minimal roots (or stems), but encapsulate exponents in realization rules. The intuition behind this type of analysis is that the lexicon should be largely 'redundancy free', with predictable patterns expressed independently – and symbolically – by means of general combinatoric devices.

Traditional word-based models embody an essentially different, exemplar-based, intuition. The leading idea is that form variation within an inflectional system can be represented by exemplary patterns or 'paradigms' and that the forms of non-exemplary items can be deduced from principal parts that identify which pattern a given item follows. Unlike many contemporary approaches, traditional models do not impose a radical separation of 'data' and 'patterns', but represent the morphological patterns of a language by actual forms that display those patterns.

Thus, exemplary paradigms serve a dual role within this type of approach, as they specify the forms of particular lexical items at the same time that they provide a model for the inflection (or derivation) of new items. To extend exemplary patterns to new items, traditional grammars offer principles for associating these items with paradigms or, as it is often put, for assigning items to the inflection classes to which the paradigms belong. These

https://doi.org/10.1017/S0022226706004191 Published online by Cambridge University Press

principles tend to exploit the implicational structure of paradigms and use one or more principal parts to predict the class of an item. The predictive value of principal parts derives from the fact that in almost all inflection class systems, certain cells of a paradigm show more variation than others. The relative predictive value of an individual element is not due to any aspect of its form, such as the presence or absence of an exponent or whether it constitutes a 'base' for other forms. Instead, the predictive value of a form correlates with the amount of variation exhibited by the cell that the form realizes across the paradigms in a language. For example, if all of the dative plurals in a language are formed in the same way (as in the case of the Russian paradigms discussed in section 3.1 below), then the dative plural will be a poor, indeed useless, predictor of class and class-dependent form variation. The form of the nominative singular, on the other hand, is highly variable across classes in Russian and thus is a good predictor of class and of other patterns of class-specific variation.[4]

Taken together, exemplary paradigms and principal parts provide the information required to deduce new forms. The special status of an exemplary paradigm lies in the fact that it exhibits general patterns of inflection. A set of principal parts contributes item-specific word forms, which establish a link to the exemplary pattern that provides the model for the inflection of an item. Matching the principal parts of an item against cells of an exemplary paradigm establishes a correspondence between principal parts and their counterparts in the exemplary paradigm. New forms of an item are deduced by extending this correspondence analogically to other cells.

The view that analogy plays a central role in creative and productive language use is one of the central Neogrammarian tenets and is expressed particularly forcefully by Paul (1880: 112).

> Sehr bedeutend ist die schöpferische Tätigkeit des Individuums aber auch auf dem Gebiete der Wortbildung und noch mehr auf dem der Flexion … Besonders klar sehen wir die Wirkungen der Analogie bei der grammatischen Aneignung der Flexionsformen einer fremden Sprache. Man lernt eine Anzahl von Paradigmen auswendig und prägt sich dann von den einzelnen Wörtern nur soviel Formen ein, als erforderlich sind, um die Zugehörigkeit zu diesem oder jenem Paradigma zu erkennen. Mitunter genügt dazu eine einzige. Die übrigen Formen bildet man in dem

[4] Importantly, it is the cross-class variability of the nominative that determines its predictive value. The nominative singular is often of high predictive value in Indo-European languages. However, there is no principled reason why morphosyntactically 'unmarked' elements should be particularly informative in general. In Estonian, the nominative singular is normally among the least INFORMATIVE case forms, and in nearly all declensions it is the partitive and the genitive singular forms that are of the greatest predictive value (see section 3.2 below and Blevins 2005).

https://doi.org/10.1017/S0022226706004191 Published online by Cambridge University Press

> Augenblicke, wo man ihrer bedarf, nach dem Paradigma, d. h. nach Analogie.[5]

In this quotation, Paul sets out the basic organization of a traditional abstractive model. A set of exemplary paradigms exhibits the patterns of a language. Principal parts identify the class of non-exemplary items and thereby determine which paradigm will serve as a deductive pattern. New forms are obtained by extending the relation between exemplary cells and principal parts to other cell/form pairs, using what are known as 'proportional analogies'. The most familiar type of analogy is a 'four-part analogy', in which the relation between two exemplary cells, C_1 and C_2, provides a model for generalizing from a principal part P to the corresponding form X. These simple proportions are often formalized as '$C_1 : C_2 = P : X$', in which the forms C_1, C_2 and P are all given, and the analogical step involves 'solving for unknown X'. Proportional analogies clearly require at least three known forms to identify a fourth. The exemplary forms C_1 and C_2 exhibit a pattern, which is extended from the principal part P to the unknown form X. Without both C_1 and C_2, there is no pattern to extend, and without P no base for the extension.

There are important differences between proportional analogies and the types of combinatoric rules applied in constructive models. A pair of exemplary cells C_1 and C_2 may consist of a common root and distinct affixal exponents. But because analogical principles deduce rather than build new forms, they can exploit any predictive patterns and need not attach any significance to the segmentation invoked in a proportional analogy. Indeed, much of the initial appeal of analogy derived, as Morpurgo Davies (1998: 258f.) remarks, precisely from the fact that 'it offered an algorithm for a structurally based form of morphological segmentation, without making any claims about the segments in question'. Moreover, C_1 and C_2 may also exhibit stem alternations or exhibit any pattern that can be extended to another pair of forms. As a result, proportional analogies apply to cases in which a principled stem-exponent segmentation is difficult to motivate, as well as to cases in which the segments either do not correspond to general stems and exponents or else cannot be associated with consistent morphosyntactic properties.

Proportional analogies can also be applied to patterns of 'word formation', to define new items (or 'lexemes') that belong to the 'derivational

[5] 'The creative activity of the individual is also very important in the domain of word formation and even more so in the domain of inflection … We see the effect of analogy particularly clearly in the grammatical acquisition of the inflectional forms of a foreign language. One learns a number of paradigms by heart and then memorizes only as many forms of individual words as are necessary to recognize their affiliation to this or that paradigm. Now and again a single form suffices. One constructs the remaining forms at the moment they are needed according to the paradigm, i.e., by analogy.'

https://doi.org/10.1017/S0022226706004191 Published online by Cambridge University Press

paradigm' of an existing item. This flexibility points to an important difference between exemplar-based and realization-based approaches. Realization rules only operate over a closed feature space in which distinctive 'feature bundles' can be defined independently of the forms that spell them out. Inflectional systems tend to be closed in the relevant sense. One can identify the features that are distinctive for particular word classes and thereby identify the set of paradigm cells or morphosyntactic representations associated with items of a word class or subclass. Thus, for example, the two number values and the six case values in Russian determine the familiar 12-cell noun paradigms in section 3.1 below. Each distinctive number/case pair defines a distinct cell, which can then be interpreted by realization rules. Yet whereas inflectional paradigms exhibit a relatively high degree of uniformity, the size of derivational paradigms varies widely, as studies of morphological families have emphasized (de Jong 2002). Hence one cannot in general characterize the derivational paradigm of an item by defining a set of abstract feature bundles that are subsequently interpreted by spell-out rules.

However, proportional analogies can extend any exemplary pattern that provides a model for the deduction of new forms. Principles of analogy that apply to inflectional patterns can also be applied to exemplary relations between verbs and agent nominals, between nouns and diminutives, and, particularly, between nouns and compounds (Krott, Schreuder & Baayen 2001, 2002, Krott, Schreuder, Baayen & Dressler in press). Inflectional patterns are the focus of the present paper, in part because these patterns are most susceptible to a constructive realization-based analysis. But it should be borne in mind that the advantages of an exemplar-based model are at least as pronounced in the domain of derivation.

The Bloomfieldians were initially inclined to disregard this sort of traditional approach out of hand, as Hockett (1954) acknowledges. Yet Hockett himself came to believe first that 'WP deserves the same consideration here given to IP and IA' (1954: 386) and later that analogy offered the most PSYCHOLOGICALLY plausible basis for the recognition and production of novel forms.

> To cover the complex alternations of Yawelmani by principal-parts-and-paradigms would take much more space than is occupied … by the morphophoneme-and-rewrite-rule presentation. But there would be a net gain in realism, for the student of the language would now be required to produce new forms IN EXACTLY THE WAY THE NATIVE USER OF THE LANGUAGE PRODUCES OR RECOGNIZES THEM – by analogy. (Hockett 1967: 221, emphasis added)

The majority of post-Bloomfieldian accounts still tend to deprecate traditional methods of analysis and relegate analogy to the periphery of the grammatical system, often as an auxiliary strategy for describing

https://doi.org/10.1017/S0022226706004191 Published online by Cambridge University Press

nonproductive patterns. Despite generative claims about the 'inadequacy' or 'unclarity' (Chomsky 1975: 31) of notions like analogy, analogical extensions based on exemplary paradigms and leading forms provide a wholly effective procedure for deducing inflectional patterns in many languages. As Matthews (1991) remarks, it is precisely the effectiveness of the traditional model that accounts for its continued importance in pedagogical grammars.

> This is not only traditional, it is also effective. It seems unlikely that, if a structuralist method or a method derived from structuralism were employed instead, pupils learning Ancient Greek or Latin – or, for that matter, Russian, Modern Greek or Italian – would be served nearly so well. (Matthews 1991: 188)

Matthews' point also has a more general relevance. Contemporary theoretical approaches are, for the most part, indifferent to the task of providing broad descriptive coverage. The narrow focus of theoretical studies is often justified by a desire to isolate aspects of a grammatical system that are taken to be particularly interesting or revealing. There is usually an implicit assumption that the less interesting parts of a system would be amenable to a similar, if somewhat more tedious, analysis. Yet it is far from clear that an assumption of this sort is warranted in the domain of morphology. Morphological systems with a large and diverse inventory of lexical or inflection classes are invariably described in terms of exemplary paradigms and principal parts. The declensional system of Estonian and the conjugational system of Georgian provide clear examples. Traditional grammars and dictionaries provide comprehensive descriptions based on exemplars and analogy. There are no comparably broad descriptions based on constructive methods.

It is, of course, possible that this gap merely reflects a divergence of interests. But an examination of even parts of the morphological class systems in Estonian and Georgian presents challenges that have not been addressed seriously from a constructive standpoint, much less solved in any way. These challenges strongly suggest that the narrow descriptive focus of constructive models is not accidental, but instead reflects basic limitations of the constructive perspective. The conclusion accords with Hockett's (1967) view that constructive models merely provide a 'shorthand' that achieves a degree of succinctness when applied to certain types of patterns.

> A correct principal-parts-and-paradigms statement and a correct morphophoneme-and-rule statement subsume the same actual facts of an alternation, the former more directly, the later more succinctly. We ought therefore to be free to use the latter, provided we specify that it is to be understood only as convenient shorthand for the former. (Hockett 1967: 222)

https://doi.org/10.1017/S0022226706004191 Published online by Cambridge University Press

When applied to systems whose patterns cannot be readily encapsulated in independent rules, a constructive approach sacrifices not only succinctness but also basic descriptive coverage. The following sections elaborate this point by considering class systems of ascending complexity.

3. MORPHOTACTIC CLASSES

In many languages, items are organized into MORPHOTACTIC CLASSES on the basis of form or form alternations. Inflection classes are morphotactic in the sense that they contain items that exhibit common patterns of inflection. Many types of lexical classes or 'word types' are also based on form, at least in part. A recurrent property of morphotactic class systems is that one does not need to know all of the forms of an item to determine its class. The class of regular items can usually be identified from a small number of forms, and sometimes from a single form.[6] This predictability is the cornerstone of the traditional method of analysis, in which general patterns are expressed by exemplary paradigms and individual items are represented by a set of leading forms. On this conception, items are assigned to inflection classes on the basis of form, but inflection class membership is not, strictly speaking, a morphosyntactic 'property' of any item.

In constructive approaches, inflection classes are still form classes, but the association between items and classes is mediated by inflection class 'features'. The reason for this is that inflection class is not, in general, predictable from stem forms or from any morphosyntactic property of stems. Hence in an approach that represents open-class items by stem entries, rather than by word entries, it is necessary to introduce class features to cross-index stems and exponents.

3.1 *Declension classes in Russian*

Russian nouns are traditionally assigned to three or four declensions, depending on whether masculine and neuter nouns are consolidated into a 'macrodeclension'. Since either alternative can be used to illustrate the difference between abstractive and constructive approaches, the description in table 1 just adopts the four-class classification proposed in Corbett (1983).

As in many Slavonic languages, the nominative singular form of a Russian noun is a fairly reliable indicator of inflection class. A noun whose nominative singular ends in a 'hard' consonant belongs to Class I. A noun whose nominative singular ends in *-a* belongs to Class II, and one whose nominative singular ends in *-o* belongs to Class IV. Nouns whose nominative singular

[6] Though phonological similarities may determine a competing classification, as discussed by Pertsova (2004).

https://doi.org/10.1017/S0022226706004191 Published online by Cambridge University Press

	CLASS I		CLASS II		CLASS III		CLASS IV	
	SING	PLUR	SING	PLUR	SING	PLUR	SING	PLUR
NOM	zakon	zakony	škola	školy	kost'	kosti	vino	vina
GEN	zakona	zakonov	školy	škol	kosti	kostej	vina	vin
ACC	zakon	zakony	školu	školy	kost'	kosti	vino	vina
LOC	zakone	zakonax	škole	školax	kosti	kostjax	vine	vinax
DAT	zakonu	zakonam	škole	školam	kosti	kostjam	vinu	vinam
INST	zakonom	zakonami	školoj	školami	kostju	kostjami	vinom	vinami
	'law'		'school'		'bone'		'wine'	

Table 1
Exemplary noun paradigms in Russian (Corbett 1983: 36)

ends in a 'soft' consonant are divided between Classes I and III. If the noun is feminine, like KOST' 'bone', it belongs to Class III. If the noun is masculine, like SLOVAR' 'dictionary', it belongs to Class I. In a traditional analysis, the actual forms of an exemplary paradigm provide a base for a network of proportional analogies. To determine the accusative singular of a new noun, such as MUŠČINA 'man', one matches the nominative singular leading form *muščina* against the exemplary form *škola*, as in (1a), and then 'solves for *X*', as in (1b). For nouns that end in a soft consonant, one needs to know either the gender of the noun or else a pair of forms: the nominative singular, together with a genitive form or the locative, dative or instrumental singular.

(1) *Analogical deduction*
  (a) škola: školu = muščina: *X*
  (b) *X* = muščinu

There is good reason to believe that the forms of highly frequent nouns are stored, irrespective of their regularity (Baayen, McQueen, Dijkstra & Schreuder 2003), so it is clearly an idealization to identify a unique exemplary paradigm for each class and a single principal part for each noun, even in systems that can be described in such an efficient manner. Discrete proportional analogies can also be regarded as idealized symbolic representations of the way that patterns resident in a set of forms provide information about the shape of forms that fall outside the set.[7] But an abstractive approach is not committed to the claim that deductive processes are symbolic nor to the idea that sets of exemplary paradigms and principal parts

[7] For an explicitly formalized approach to the modelling of paradigmatic analogy, see Skousen (1989, 1992).

https://doi.org/10.1017/S0022226706004191 Published online by Cambridge University Press

are minimal. The key assumptions of an abstractive approach are that exemplary paradigms and principal part inventories contain word forms, and that grammatically distinctive patterns are resident in these actual forms. This perspective contrasts with the morphotactic and combinatoric assumptions of a constructive approach, on which individual nouns are represented by uninflected stems, and inflected noun forms are built from stems through the addition of inflectional exponents. Candidate analyses for the stems underlying the exemplary nouns in table 1 are given in (2a). It does not matter whether these analyses are interpreted as lexical entries, as in most IA models or as 'realizational pairs', which determine the 'spell-out' or 'selection' of a stem, as in Aronoff (1994).

(2) *Stem analyses*
   (a) ⟨[MASC, CL1], zakon⟩, ⟨[FEM, CL2], škol⟩, ⟨[FEM, CL3], kost'⟩, ⟨[NEUT, CL4], vin⟩
   (b) ⟨[MASC, CL2], muščin⟩

The stem forms in (2) correspond to the nominative singular, minus any inflectional ending. Since these stems ALL end in consonants, form no longer provides a basis for matching entries with inflectional patterns. One can invoke gender at this point, as many grammars do, and associate masculine nouns with Class I exponents, feminines with Class II and III exponents, and neuters with Class IV exponents. However, as Corbett (1983) shows, gender is not a reliable predictor of inflection class in Russian. One systematic mismatch between gender and class arises in Class II. Although this class is predominantly feminine, it also contains masculines such as MUŠČINA 'man' in (2b). The nominative singular *muščina* is diagnostic of Class II membership. But neither the form *muščin* nor the gender features in (2b) identify MUŠČINA as a Class II noun. The inadequacy of a classification based on gender reflects the fact that gender provides a secondary cue of class membership in Slavonic, which serves to reinforce a system of form classes.

To compensate for the fact that stems lose class information that is represented by the form of the nominative singular, constructive approaches tag stems with inflection class 'properties', like the features 'CL1, CL2', etc. in (2). The rules or entries that introduce exponents must be tagged by the same features, so that this cross-referencing determines the inflection of a noun. This strategy reencodes characteristics that are EXHIBITED by forms within a traditional analysis by means of formal morphosyntactic properties. The use of diacritic properties to cross-index entries and classes yields a taxonomic treatment of inflection classes, giving rise to the issues discussed in Carstairs (1983), Carstairs-McCarthy (1991, 1994) in connection with 'paradigm economy'.

However, the most significant aspect of this solution from the present perspective is that it flattens the implicational structure of the Russian

https://doi.org/10.1017/S0022226706004191 Published online by Cambridge University Press

system. A traditional analysis exploits – and, indeed, depends on – the predictive value of the nominative singular. An inflectional system can only be factored into a set of exemplary paradigms and leading forms if some collection of forms smaller than the full set implies the other forms. Otherwise the distinction between exemplary paradigm and sets of leading forms collapses. A constructive approach that 'derives' the nominative singular in the same way as other case forms (at least in Classes II and IV) replaces an asymmetrical implicational relation by a symmetrical indexing relation. It might seem that one can reintroduce an implicational relation on the exponents in each class. Yet even this cannot be achieved without further artifice, such as the introduction of 'zero' exponents, given that it is the LACK of a case exponent on the nominative singular that predicts Class I exponents.

In sum, a constructive analysis disrupts the implicational structure of noun declensions in Russian because the implicational patterns are word-based and cannot be reconstructed in terms of stems or exponents in isolation. The use of class features largely restores the information that is lost by removing class-identifying exponents from noun stems. The 'feature overhead' of this strategy is kept in check to some degree by the relative simplicity of the Russian declensional system.[8] However, the consequences of using 'morphosyntactic' properties to encode form characteristics or even 'assembly instructions' emerge clearly in languages which have a richer set of morphotactic classes and more intricate patterns of word-based implication.

3.2 *Grade-alternating noun declensions in Saami and Estonian*

The inflectional systems of Saami and Estonian provide the first illustration. These systems exhibit word-level patterns of interpredictability that cannot be expressed in terms of stems or exponents alone. Moreover, sets of interpredictable forms comprise 'pure' form classes that do not correspond to any natural morphosyntactic class. Hence no feature-based description of the patterns offers any improvement over an exemplary set of forms that exhibits the patterns.

3.2.1 *Noun declensions in Saami*

The comparatively simple declensional paradigms in (Northern) Saami provide a useful point of departure. Saami declensions are based on two principal parts: the genitive and nominative singular. Nouns can be assigned

[8] Though see Kirby (2005) for a more detailed treatment of the complexity of noun declensions in Czech.

https://doi.org/10.1017/S0022226706004191 Published online by Cambridge University Press

	WEAKENING		STRENGTHENING	
	SING	PLUR	SING	PLUR
NOMINATIVE	**bihttá**	*bihtát*	**bargu**	*barggut*
GENITIVE/ACCUSATIVE	*bihtá*	*bihtáid*	*barggu*	*bargguid*
ILLATIVE	**bihttái**	*bihtáide*	**bargui**	*bargguide*
LOCATIVE	*bihtás*	*bihtáin*	*barggus*	*bargguin*
COMITATIVE	*bihtáin*	*bihtáguin*	*bargguin*	*bargguiguin*
ESSIVE	**bihttán**		**bargun**	
	'piece'		'work'	

Table 2
Exemplary first declension paradigms in Saami (Bartens 1989)

to three classes, based on syllable count and stem alternations. For present purposes, it will suffice to restrict attention to the first declension, in which the genitive singular has an even number of syllables and the nominative and genitive singular exhibit a grade contrast.[9] In paradigms that show 'weakening' gradation, the nominative contains a geminate consonant that identifies it as strong and the genitive singular contains a short counterpart that identifies it as weak. In paradigms that show 'strengthening' gradation, the genitive singular is strong and the nominative singular is weak. Thus the opposition between the strong nominative *bihttá* and the weak genitive *bihtá* in table 2 illustrates weakening gradation, while the contrast between the strong genitive *barggu* and the weak nominative *bargu* shows the strengthening pattern.

First declension nouns exhibit three patterns of generally symmetrical cell-based syncretism (which are distinguished in table 2 by text boxes, italicized and bold forms). First, the form of the nominative singular predicts and is predicted by the form of the illative singular and essive.[10] A strong nominative singular, such as *bihttá*, correlates with a strong illative singular and essive. Conversely, a weak nominative singular, such as *bargu*, correlates with a weak illative singular and essive. Second, the form of the genitive singular predicts and is predicted by the locative and comitative singular and all plural forms. Third, the comitative singular and the locative plural are always identical (a pattern that is also extended to possessive paradigms).

[9] In the second declension, the genitive singular has an odd number of syllables. The genitive may be either strong or weak, but the nominative is always weak. In the third declension, the principal parts are based on distinct stems.

[10] In the second and third declension, the illative singular is based on the genitive singular.

https://doi.org/10.1017/S0022226706004191 Published online by Cambridge University Press



Table 3
Implicational structure of first declension paradigms in Saami

These patterns of implication are schematized in table 3, in which form interpredictability is indicated by double arrows and form identity is marked by an equal sign.[11] As table 3 indicates, the implicational patterns in Saami declensions refer to 'paradigm cells', not to stem grade or to other independent stem properties. The grade of the genitive singular and nominative singular forms correlates with the grade of each of the interpredictable elements.

Although it is possible to abstract strong and weak stems from first declension paradigms, no inflected form is consistently strong or consistently weak. Hence there is no predictive, or indeed descriptive, value in abstracting 'strong' and 'weak' or 'direct' and 'oblique' stems. Any description that begins with abstract stems must tag individual nouns with diacritic class features to indicate whether they follow a weakening or strengthening pattern. But these patterns are distinguished essentially by whether their nominative or genitive singular is strong. So the greater abstraction achieved by referring to stems rather than to paradigm cells is immediately undone by the use of features that covertly reestablish the original association between forms and cells.

As tables 2 and 3 also indicate, the sets of mutually predictable forms have no evident properties in common that could be exploited in a rule-based description. The nominative singular, illative singular and essive constitute no recognizable morphosyntactic class. The forms that are predictable from the genitive singular are even more heterogeneous. The syncretism between comitative singular and locative plural cannot be regarded as any kind of neutralization either. In short, the classes identified in tables 2 and 3 are pure form classes, which cannot be described more concisely in terms of any substantive properties. It is possible to tag the members of these form classes

[11] Intuitively, the arrows in table 2 and below represent implicational dependencies. Formally, pairs of forms linked by arrows can be understood as having a **mutual information** value approaching 1 (i.e. as being maximally informative about each other), or as having a **conditional entropy** approaching 0 (i.e. as being maximally correlated within a paradigm). For detailed discussion of these notions, see Cover & Joy (1991) or Manning & Schütze (1999).

https://doi.org/10.1017/S0022226706004191 Published online by Cambridge University Press

with diacritic features, but this strategy does not avoid listing the members of each class, since the diacritic tags must be assigned on an element-by-element basis.

3.2.2 *Noun declensions in Estonian*

Like their Saami counterparts, noun declensions in Estonian are based mainly on two principal parts, usually the genitive and partitive singular, and can be assigned to classes based on prosodic structure and inflectional exponence. The number of classes varies considerably across accounts, though a conservative estimate would put the actual number of declensions somewhere between four (Blevins 2005) and seven (Erelt, Erelt & Ross 2000).[12] As in Saami, grade alternations are characteristic of the first declension. In productive 'quantitative' alternations, strong forms contain an 'overlong' (Q3) syllable (which, following Viks (1992), is marked by an grave accent before the Q3 syllable). The initial Q3 syllable of a strong form contrasts with the non-overlong (Q1 or Q2) initial syllable of a weak form. There are also non-productive 'qualitative' alternations, in which the strong grade preserves a segment that is lost or modified in the weak grade. A weak form that has been reduced to a monosyllable by a process of qualitative gradation shows a kind of 'compensatory' lengthening to Q3, to satisfy the minimal word constraint of Estonian.[13]

In table 4 below, the nouns KOOL 'school' and KUKK 'rooster' illustrate quantitative weakening patterns, PIDU 'party' shows a qualitative pattern, and PESA 'nest' exhibits no grade alternation.

The Estonian declensional system in table 4 extends the Saami pattern in a pair of relevant respects. First, the Estonian case system is larger and more differentiated, with 14-odd cases divided between a predominantly fusional grammatical subsystem and more agglutinative semantic subsystem. Second, the patterns of word-based implication are more intricate, as one can see by considering the semantic case forms of the noun KOOL 'school' in table 4. The inessive plural *`koolides* is fully representative of the semantic case forms as a whole. This form is based on the genitive plural *`koolide*, which in turn is based on the partitive singular *`kooli*. But the partitive singular itself contains the strong stem *`kool*, which realizes the nominative singular. Hence the form [[[*`kool*]*i*]*de*]*s* contains as many as three layers of 'parasitic' stems.

[12] Descriptions that assign nouns to classes based on the morphotactically minimal nominative singular tend to recognize a large number of 'word types' – as many as 400, in the case of Saagpakk (2000). The vast majority of these types merely represent 'choices' in the inflection of a noun that are not predictable from its minimal form.

[13] An open-class word in Estonian consists minimally of a foot, which must contain at least a disyllable or a Q3 monosyllable. For discussion, see Prince 1980, Lehiste 1997, Ehala 2003, Viitso 2003a, Blevins 2005.

https://doi.org/10.1017/S0022226706004191 Published online by Cambridge University Press

https://doi.org/10.1017/S0022226706004191 Published online by Cambridge University Press

		NONE		QUANTITATIVE				QUALITATIVE	
	GRADE	SING	PLUR	SING	PLUR	SING	PLUR	SING	PLUR
GRAMMATICAL	NOMINATIVE	*pesa*	**pesad**	*ˈkool*	**koolid**	*ˈkukk*	**kuked**	*pidu*	**ˋpeod**
	GENITIVE	**pesa**	*pesade*	**kooli**	*ˋkoolide*	**kuke**	*ˋkukkede*	**ˋpeo**	*pidude*
	PARTITIVE	*pesa*	*pesasid*	*ˈkooli*	*ˋkoolisid*	*ˋkukke*	*ˋkukkesid*	*pidu*	*pidusid*
	STEM PARTITIVE		*pesi*		*ˋkoole*		*ˋkukki*		—
	SHORT ILLATIVE	*ˋpessa*		*ˈkooli*		*ˋkukke*		*ˋpittu*	
SEMANTIC	ILLATIVE	**pesa**sse	*pesade*sse	**kooli**sse	*ˋkoolide*sse	**kuke**sse	*ˋkukkede*sse	**ˋpeo**sse	*pidude*sse
	INESSIVE	**pesa**s	*pesade*s	**kooli**s	*ˋkoolide*s	**kuke**s	*ˋkukkede*s	**ˋpeo**s	*pidude*s
	ELATIVE	**pesa**st	*pesade*st	**kooli**st	*ˋkoolide*st	**kuke**st	*ˋkukkede*st	**ˋpeo**st	*pidude*st
	ALLATIVE	**pesa**le	*pesade*le	**kooli**le	*ˋkoolide*le	**kuke**le	*ˋkukkede*le	**ˋpeo**le	*pidude*le
	ADESSIVE	**pesa**l	*pesade*l	**kooli**l	*ˋkoolide*l	**kuke**l	*ˋkukkede*l	**ˋpeo**l	*pidude*l
	ABLATIVE	**pesa**lt	*pesade*lt	**kooli**lt	*ˋkoolide*lt	**kuke**lt	*ˋkukkede*lt	**ˋpeo**lt	*pidude*lt
	TRANSLATIVE	**pesa**ks	*pesade*ks	**kooli**ks	*ˋkoolide*ks	**kuke**ks	*ˋkukkede*ks	**ˋpeo**ks	*pidude*ks
	TERMINATIVE	**pesa**ni	*pesade*ni	**kooli**ni	*ˋkoolide*ni	**kuke**ni	*ˋkukkede*ni	**ˋpeo**ni	*pidude*ni
	ESSIVE	**pesa**na	*pesade*na	**kooli**na	*ˋkoolide*na	**kuke**na	*ˋkukkede*na	**ˋpeo**na	*pidude*na
	ABESSIVE	**pesa**ta	*pesade*ta	**kooli**ta	*ˋkoolide*ta	**kuke**ta	*ˋkukkede*ta	**ˋpeo**ta	*pidude*ta
	COMITATIVE	**pesa**ga	*pesade*ga	**kooli**ga	*ˋkoolide*ga	**kuke**ga	*ˋkukkede*ga	**ˋpeo**ga	*pidude*ga
		‘nest’		‘school’		‘rooster’		‘party’	

Table 4
Exemplary first declension nouns in Estonian (Blevins 2005)

As in Saami, one could in principle recognize a root *`kool*, an inflectional stem *`kooli* and a plural stem *`koolide*. However, the limitations of a stem-based type of strategy are even clearer in Estonian than in Saami. To begin with, one would also require a weak stem, such as *kooli*, to realize the genitive singular and underlie the singular semantic case forms. But the resulting stem inventory is highly heterogeneous, containing some elements, such as the nominative and genitive singular, that feed derivational processes and others that are confined to the inflectional system. Moreover, a stem inventory consisting entirely of full words highlights the fact that the principal function of abstract stems is to avoid direct relations between inflected word forms. But the need to avoid relations between words is an artifact of a constructive perspective. More specifically, this reflects the view that forms within a paradigm are related by shared patterns of derivation, that is, that they are all built from a common inventory of minimal parts. Since words are regarded as output forms, they cannot constitute parts of other derived forms.

The main objection to abstract stems is not merely that they reflect a tacit theoretic bias, but that the abstraction involved in defining these stems discards critical information. As in Saami, no case forms in table 4 are consistently strong or consistently weak, and, indeed, no implicational patterns are based on independent stem properties.[14] Instead, the grade of a case form correlates with the grade of the corresponding principal part. If the partitive singular is strong, so are the forms that it implies. More generally, the implicational patterns that relate the forms of a first declension paradigm hold between paradigm cells, as schematized in table 5.[15]

A strong vowel-final partitive singular implies the full paradigm of a regular first declension noun. This pattern is discussed in detail in Blevins (2005) but a brief summary will give the general picture. A vowel-final partitive immediately identifies a noun as belonging to the first declension. This partitive also implies a nominative singular without the theme vowel, provided that the implied form is a metrical foot and thereby satisfies the minimal word constraint. A strong vowel-final partitive is identical to the short illative provided that it satisfies the prosodic conditions on the short illative form (discussed in footnote 14 above). Since no regular first declension

[14] Although the short illative singulars in table 4 all contain an initial Q3 syllable, the prosodic constraint on productive short illatives is met by any form that constitutes a trochaic prosodic word. In disyllables, this constraint is only satisfied by forms in which an initial Q3 syllable 'functions as a metrical foot' (Lehiste 1997: 11). In first declension paradigms containing a quadrisyllabic partitive singular, such as *koridori* 'corridor', the sequence of two light trochaic feet (*ko.ri*).(*do.ri*) also forms a trochaic word, so the short illative singular is identical to the partitive singular.

[15] The implicational relation between the partitive singular and nominative singular is asymmetrical in general given that the theme vowel is not recoverable from a consonant-final nominative. However, in open-class first declension paradigms, a vowel-final nominative singular is identical in form to the partitive (and genitive) singular.

https://doi.org/10.1017/S0022226706004191 Published online by Cambridge University Press

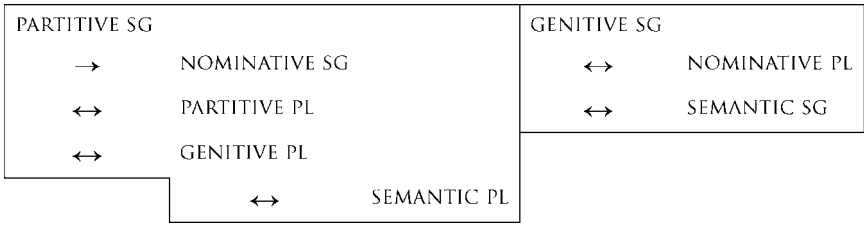


Table 5
Implicational structure of first declension paradigms in Estonian

noun is strong throughout its paradigm, a strong vowel-final partitive implies a weak genitive singular. A strong vowel-final partitive plural also implies a partitive plural in *-sid* (as well as a fusional 'stem' partitive plural and a genitive singular in *-de*. The genitives in turn imply the semantic case forms.

Abstract stems, on the other hand, are highly uninformative in this regard; indeed, the more abstract, the less informative they are. A constructivist bias has led many modern reference and theoretical descriptions of the language – including the authoritative *Eesti keele grammatika* (Erelt, Kasik, Metslang, Rajandi, Ross, Saari, Tael & Vare 1995: 333ff.) – to classify nouns based on their nominative singular. Yet for nouns with a 'truncated' nominative singular, such as KOOL or KUKK, the choice of theme vowel in the partitive and genitive singular is not predictable from the nominative 'stem'.[16] Since the other case forms all depend on the choice of theme vowel, they are also not predictable. A constructive analysis that begins with the stem thus requires class features to encode each of the four possible theme vowels: *i*, *e*, *a* or *u*. Unlike the conjugation vowels in Romance, theme vowels are associated with individual nouns in Estonian and are not markers of inflection class. Hence any 'theme-vowel classes' will be independent of inflectional patterns, so that the number of classes will be the product of 'theme-vowel classes' and inflection classes.

The description of declensional patterns can be greatly simplified if one begins with a vowel-final form. The fact that the partitive singular is clearly the diagnostic form of the first declension makes a form such as *`kukke* an obvious candidate. However, this form in isolation does not identify declension class, since a strong disyllable may belong either to the first or second declension. In the first declension, a strong partitive alternates with a

[16] Estonian preserves traces of an ancestral vowel harmony system, reconstructed for proto-Finnic (Viitso 2003b: 173) or even Uralic (Laakso 2001: 83), and *-i* functions as the default theme vowel in some subclasses. Yet these factors do not provide a reliable basis for predicting the theme vowel from a consonant-final nominative.

https://doi.org/10.1017/S0022226706004191 Published online by Cambridge University Press

weak genitive, as illustrated by the pairs *`kukke ~ kuke* and *`kooli ~ kooli* in table 4. The second declension contains a productive subclass of nouns, such as AASTA 'year', which are strong throughout their paradigm. Like all second declension nouns, these nouns mark the partitive singular by an exponent *-t* that distinguishes it from the genitive singular, as in the pair *`aastat ~ `aasta*. The second declension also contains a closed class of nouns, such as HAPE 'acid', which preserve an older strengthening grade pattern in which a weak partitive singular in *-t* alternates with a strong genitive, as in the pair *hapet ~ `happe*. Just as a strong vowel-final partitive is diagnostic of the first declension, a strong genitive is a marker of the second declension. But a strong stem by itself does not identify class, and a weak form such as *kuke* or *kooli* is even less informative in isolation.

	Q3	NON-Q3	Q3	NON-Q3	Q3	NON-Q3	Q3	NON-Q3
GEN SG		pesa		kooli		kuke	`peo	
PART SG		pesa	`kooli		`kukke			pidu
ILLA2 SG	`pessa		`kooli		`kukke		`pittu	
	'nest'		'school'		'rooster'		'party'	

Table 6
Stem inventories of first declension nouns

The basic class of a first declension noun could in principle be identified by a pair of stems, but even an expanded inventory of abstract stems is less informative than a single partitive singular principal part. The reason for this is that the contrast between the first declension nouns in table 4 is not due to differences in their form inventories, but to different associations between these forms and paradigm cells. As illustrated in table 6, the basic form inventory of a grade-alternating noun like KUKK is similar to that of the non-alternating noun PESA. Both nouns have forms with an initial Q3 syllable, *`kukke* and *`pessa*, which contrast minimally with forms that have an initial non-Q3 syllable, *kuke* and *pesa*. The difference between the role of *`kukke* and *kuke* in the paradigm of KUKK and the role of *`pessa* and *pesa* in the paradigm of PESA is attributable to the paradigm cells that these forms realize, not to any properties they have as forms. This is precisely the information that is lost when a set of abstract stems is extracted from a paradigm of word forms. The qualitative grade alternation in the paradigm of PIDU nicely reinforces this point. Qua forms, *`pittu* and *pidu* clearly pattern with *`kukke* and *kuke*. But within the morphological system, *pidu* and *`peo* pattern with *`kukke* and *kuke*, respectively, and this parallel is established by the paradigms cells they realize, and is not due to independent properties that they have as forms.

https://doi.org/10.1017/S0022226706004191 Published online by Cambridge University Press

3.3 *Form classes and the 'duality of patterning'*

The broader point illustrated by the stem inventories in table 6 is just that abstract stems are less informative than the word forms from which they are abstracted. This lost information must then be reintroduced, and in constructive approaches it is usually diacritic class features that carry the burden of representing inflectional choices that are underdetermined by an abstract stem. The resulting 'feature overhead' will tend to be inversely proportional to the informativeness of the basic lexical units. Within a traditional WP model, principal parts identify the class of an item, and thereby associate it with exemplary paradigms that permit the analogical deduction of other forms. So diacritic features have no part to play in marking class or guiding realization relations. If a constructive analysis retains the FORM of principal parts but discards information about the paradigm cells that they realize, it will require features to represent class-specific patterns of form variation. One can continue this process further by reducing principal part inventories to minimal – and minimally informative – nominative singular forms. But this requires additional features to represent stem vowels, along with any patterns of exponence that correlate with the choice of stem vowel.

Constructive analyses often carry this process still further, by isolating all recurrent form elements. The morphotactically simple systems above present no great difficulties of segmentation for such an approach. Even the 'fusional' case forms in Estonian can be segmented into basic stems, such as *`kukk*, *kuk*, etc., and the stem vowels *a*, *e*, *i*, and *u*. But since it is COMBINATIONS of these elements that are distinctive, the resulting 'segments' defy classification in isolation. Hence having obtained an inventory of minimal elements, a constructive approach can neither associate a constant meaning with the elements nor recover the original word forms. So a system of class features is invoked to reconstitute the larger forms from which they have been abstracted. The nature and limitations of this strategy are then explicit. A root- or stem-based analysis of the declensional system of Estonian, Saami or Russian may succeed in minimizing the segmental representation of items, but at the cost of smuggling morphotactic information into the morphosyntax in the guise of class features. Diacritic features of this sort are clear symptoms of 'overextraction', in which a morphological system has been broken down into elements which do not function as 'signs' within the system, and which have little or no predictive value.

At a more global level of comparison, there is a fairly direct trade-off between the number of forms in an abstractive model and the complexity of the features in a constructive approach. These choices reflect the fundamental differences in perspective discussed in section 1 above. Constructive models tend to assume a minimized lexicon and encapsulate patterns, insofar

https://doi.org/10.1017/S0022226706004191 Published online by Cambridge University Press

as possible, in independent rule systems. Abstractive models are exemplar-based, and exhibit patterns by giving sets of actual forms that instantiate those patterns. The traditional preference for exhibiting patterns rests on the existence of form classes that resist description in terms of substantive features. The noun declensions in section 3.2 above illustrate various patterns of this sort. Both principal parts of a Saami noun both underlie sets of forms that do not comprise natural classes. Even if one principal part is identified as a default form with an 'elsewhere' distribution, the other principal part will still underlie a recalcitrant form class. The syncretism between the comitative singular and locative plural form represents another pattern that cannot be explicated in feature terms. First declension paradigms in Estonian are rife with form-based patterns. Among the forms based on the partitive singular are the genitive plural and the short illative singular, while the genitive singular underlies the nominative plural and all of the singular 'semantic' cases.

The distribution of declensional stems in Saami and Estonian cannot be attributed to any common set of substantive morphosyntactic properties that one might associate with the forms that they underlie. These declensional systems are not unique in this regard, and realization-based models have long recognized the existence of MORPHOMIC stems. A solution that is often proposed within this literature involves introducing a system of 'stem indices' which may be 'selected' by particular cells. In the model of Stump (2001), the alternating stems in Saami and Estonian might be designated by properties such as 'weak' and 'strong'. Within other approaches, such as Aronoff (1994) or Brown (1998), terms such as 'first stem' and 'second stem' express more explicitly the diacritic character of these indices. A stem index preserves the appearance of a realization relation between properties and forms, by allowing a strong stem to be treated as a kind of 'spell-out' of the index 'strong' and a weak stem as the spell-out of the index 'weak'.

Yet this type of technical solution merely relocates the original problem. The fact that weak and strong stems do not underlie coherent sets of forms entails that indices like 'weak' and 'strong' cannot be assigned to paradigm cells on the basis of substantive properties. Instead, the association between cells and indices must be listed, on what amounts to a cell-by-cell basis. This solution serves to underscore the challenge posed by morphomic stems, and by form classes more generally. Classes of forms that share a common set of properties may be described efficiently by features that characterize their common properties. But by definition, pure form classes cannot be characterized in this way (since if they DID share substantive morphosyntactic properties, they would not be pure form classes). Incorporating form indices into feature descriptions recasts the problem of defining the distribution of morphomic forms into a problem of determining the distribution of morphomic indices. Just listing the paradigm cells (or morphosyntactic

https://doi.org/10.1017/S0022226706004191 Published online by Cambridge University Press

representations) that are associated with a particular stem index represents no improvement over listing the forms themselves. In short, form classes resist a feature-based analysis because they EXHIBIT a pattern that cannot be STATED more efficiently in symbolic terms.

The relatively simple noun declensions in Russian, Saami and Estonian highlight a contrast between the 'information content' that can be associated with an inflected word as a whole and the information that can be assigned to its parts. A constructive analysis of Russian can break noun forms down into roots and formatives and associate grammatical properties with these parts, but it cannot recover the original forms without the aid of diacritic assembly instructions. A constructive analysis can likewise segment the forms of Estonian noun paradigms and assign the parts to separate inventories of stems and exponents. As in Russian, one cannot reconstitute the original forms from these minimal elements. One can also not distribute the grammatical properties of the original forms over their parts, nor shift the implicational structure of the paradigm onto these smaller units. The contrast in the informativeness of words and formatives reflects what Hockett (1958) called the 'duality of patterning'. A traditional insight, which Hockett later came to appreciate (Hockett 1967), is that duality arises principally at the level of word forms, which are often meaningful units composed of individually meaningless formatives.

In some systems, it is true that formatives may realize stable properties in all of the contexts in which they occur. Yet this can be seen to be a limiting rather than a normative case, and in many systems it is only recently morphologized formatives that can be described in this way. The 'agglutinative' semantic case suffixes in Estonian are, for example, the most recently morphologized case markers. The last four cases in table 4 above (terminative, essive, abessive and comitative) are not even fully integrated into the morphological system, in that they do not trigger agreement on dependent adjectives. For some cases, including comitative *-ga*, 'the former postposition is attested as a free word in early Estonian literary records' (Grünthal 2003: 50). As one moves inwards, through layers of morphomic stems, the association between formatives and properties becomes progressively more tenuous, to the point that an individual formative comes to distinguish certain words from others 'without having any meaning of its own' (Hockett 1958: 575).

4. Conjugational patterns in Georgian

Although post-Bloomfieldian models tend to regard non-meaning-bearing formatives as deviations from an agglutinative ideal, morphosyntactic opacity is highly characteristic of stem systems in many languages. The conjugational system of modern Georgian provides a clear case in point. The root and stem formatives of a verb can be identified by comparing whole

https://doi.org/10.1017/S0022226706004191 Published online by Cambridge University Press

word forms, and other words can again be predicted from diagnostic forms. However, separate inventories of roots, formatives or even partial stems do not allow one to reconstitute word forms, let alone predict the form of other words. A constructive analysis in terms of derivations from minimal elements thus leads to spurious 'unpredictability' and 'redundancy' in much of the system.

4.1 *Overview*

To understand the context in which these patterns arise, it will be useful to summarize briefly the organization of the Georgian conjugational system, and the structure of the Georgian verb.

Georgian contains four basic conjugation CLASSES, which are true form classes, defined in terms of future and aorist principal parts (Harris 1981: 260), but which correlate roughly with valence. Class 1 consists mainly of transitive verbs. Class 2 contains mostly unaccusative intransitives, which are often related to Class 1 transitives, and sometimes assigned to passive and inceptive/inchoative subclasses (Tschenkéli 1958: 257). Class 3 mainly contains active unergative intransitives. Class 4 consists of 'indirect' verbs, which exhibit thematic 'inversion' in all series.

There are also three inflectional SERIES in Georgian, which contain sets of 'tense' paradigms, traditionally termed SCREEVES. Series I contains present and future screeves, Series II contains aorist screeves, and Series III contains inverse screeves, which are residually perfect in form but often evidential in meaning. There may be hundreds of inflected forms of a verb, distributed over between 9 and 11 screeves. The organization of series and screeves is outlined in table 7.[17]

SERIES	SUBSERIES	NONPAST	PAST	SUBJUNCTIVE
I	Present	Present Indicative	Imperfect	Subjunctive Present
	Future	Future Indicative	Conditional	Subjunctive Future
II	Aorist	—	Aorist Indicative	Optative
III	Perfect	Present Perfect	Pluperfect	Subjunctive Perfect

Table 7
Feature classification of Georgian screeves

[17] This classification departs from Aronson (1990) in substituting 'subjunctive' for his label 'modal' and in treating the present perfect as a nonpast rather than a past screeve. A nonpast analysis captures the sense of 'completed action with present relevance' that Tschenkéli (1958: 493) attributes to this form, though nothing hinges on this point.

https://doi.org/10.1017/S0022226706004191 Published online by Cambridge University Press

The morphosyntactic complexity of the conjugational system is matched by the morphotactic complexity of individual verb forms. Hewitt (1995: 117) identifies 11 'position class' elements:[18]

> The following are the morphemes that may occur in a Georgian verb-form, though not necessarily simultaneously: 1. Preverb(s), 2. Pronominal Agreement-Prefix, 3. Version-Vowel, 4. Root, 5. Causative Suffix(es), 6. Inceptive/Passive Marker, 7. Thematic Suffix, 8. Perfect/Stative Marker, 9. Imperfect Marker, 10. Mood-Vowel, 11. Pronominal Agreement-Suffix(es).

The distribution of these elements is conditioned by a variety of different factors. Some elements, notably the 1st and 2nd person agreement markers, are relatively uniform, at least within a given class or series. Certain elements, such as the inceptive/passive marker, characterize particular classes, while others, such as the thematic suffix, tend to occur within particular series. Other elements, such as the 'mood' or 'screeve' vowel and the 3rd person markers, vary across screeves.[19] In effect, the screeve vowel and 3rd person markers serve as 'characteristic suffixes' (Aronson 1990: 470) that identify an individual screeve within the full conjugational paradigm of a verb. Other elements, such as the root and preverb, vary on an item-by-item basis. The choice of still other elements, notably the 'version' vowel, appears to be conditioned by lexical factors in some cases, and determined by inflectional considerations in other cases.

4.2 *Class 1 stem structure*

One can immediately see the general challenge that confronts a constructive description of Georgian in terms of diacritic verb classes. With at least 11 dimensions of variation, defined by the 11 position classes, and up to a dozen different choices for some dimensions, the class system required to derive word forms from minimal units would be vast. It is plain why descriptive and pedagogical grammars present full word forms organized into exemplary paradigms, and it is hard to understand why anyone acquainted with the complexities of the system would pursue any other kind of description.

To move from an a priori discussion to a more concrete illustration, it is useful to consider the interaction of a manageable number of elements. The schema in (3) gives the structure of simple word forms, which have no

[18] Even this class structure suppresses a certain amount of inessential variation and morphotactic detail.

[19] In the following descriptions, the term 'screeve vowel' is applied to for what Hewitt (1995: 117) terms a 'mood-vowel' and 'thematic suffixes' is used for what Aronson (1990: 40) terms 'present/future stem formants'.

https://doi.org/10.1017/S0022226706004191 Published online by Cambridge University Press

derivational exponents and no agreement markers. The subscripts in (3) indicate the correspondence to the position classes identified in Hewitt (1995: 117).

(3) *Structure of minimal verb forms*
Preverb$_1$–Version-Vowel$_3$–Root$_4$–Thematic-Suffix$_7$–Screeve-Vowel$_{10}$

The words in table 8 all conform to the schema in (3). To abstract away from class-specific variation and the effects of 'inversion' (Harris 1981), these forms are drawn from 'direct' (Series I and II) screeves of Class 1 verbs. To suppress agreement markers, the forms in table 8 select the literally unmarked subject (2nd person singular) and object (3rd person) agreement properties.

ROOT	PRESENT INDICATIVE	FUTURE INDICATIVE	AORIST INDICATIVE	2SGS/3O
t'ex	t'ex	**gat**'ex	gat'ex**e**	'smash'
c'er	c'er	**da**c'er	dac'er**e**	'write'
šen	**a**šen**eb**	**a**ašen**eb**	aašen**e**	'build'
tav	**a**tav**eb**	**ga**ataveb	gaatav**e**	'end'
p'at'iž	p'at'iž**eb**	**da**p'at'iž**eb**	dap'at'iž**e**	'invite'
xat'	xat'**av**	**da**xat'av	daxat'**e**	'paint'
k'er	k'er**av**	**še**k'erav	šek'er**e**	'sew'
c'on	c'oni	**a**c'oni	ac'on**e**	'weigh'
targmn	targmni	**gada**targmni	gadatargmn**e**	'translate'
tb	**a**tb**ob**	**ga**atbob	gaatb**e**	'warm'
sp'	sp'**ob**	**mo**sp'ob	mosp'**e**	'ruin'
dg	dg**am**	**da**dgam	dadg**i**	'place'
cv	**i**cv**am**	**ča**icvam	čaicv**i**	'put on'

Table 8
Class 1 verb forms in Georgian (Tschenkéli 1958: section 18, Hewitt 1995)

For ease of comparison, the groups of rows contain forms with a common thematic suffix. Moreover, as one moves from left to right within a row, elements that are not present on a form to the left are marked in bold. Thus, the Present column indicates thematic suffixes and version vowels, the Future column indicates preverbs, and the Aorist column marks screeve vowels.

Consider the task of deriving word forms from the roots in table 8, bearing in mind that this represents a narrowly circumscribed part of a more general

https://doi.org/10.1017/S0022226706004191 Published online by Cambridge University Press

task. The present indicative differs from the root in that it may contain a thematic suffix and/or a 'neutral' version vowel. Neither choice is predictable from the form of the root. Hewitt (1995: 170) describes the choice of neutral version vowels as 'lexically determined', and elsewhere elaborates on their role and distribution.[20]

> Georgian has a number of vowels which stand immediately before verb roots to convey certain aspects of meaning. The basic system of oppositions, known as 'version' … is seen most clearly in Series I and II of the transitive verbs. The neutral version is so-called because it adds nothing to the verb's fundamental meaning. As we have seen, some transitive verbs require an *a*-vowel before their root, whilst others take a zero marker in place of the *a*-vowel. Apart from verbs in *-eb* derived from nouns or adjectives, which always take the *a*-vowel, THERE IS NO WAY OF PREDICTING WHETHER A VERB'S NEUTRAL VERSION WILL BE IN A- OR ZERO. (Hewitt 1996: 52, emphasis added)

Whether a verb occurs with a version vowel is then not predictable in general, even within the class of verbs that take the thematic suffix *-eb*, as Hewitt (1995: 170) again acknowledges when he remarks that 'the vast majority of roots with Thematic Suffix *-eb*, however, have their Neutral Version in *a-*'. In table 8, for example, *a-* occurs in *ašeneb* and *ataveb*, but not in *p'at'ižeb*.

However, even if neutral version vowels were predictable within the *-eb* subclass, the thematic suffix on which the prediction depends is not itself determinable from the root. Thematic suffixes are characteristic of Series I, leading Aronson (1990: 40) to term them 'present/future stem formants'.[21] Although there is agreement that *-eb* is the most frequent, Hewitt (1995) reports that

> The Thematic Suffixes in use today with Transitive Verbs are: *-eb*, *-ob*, *-av*, *-am*, *-en*, and *-i*. Those in this list are widely used … (Hewitt 1995: 143)

Hence grammatical descriptions of Georgian list the thematic suffix that occurs with Series I forms of Class I verbs. Indeed, the classification of verbs according to thematic suffix is a basic organizational principle of Georgian grammars, such as Tschenkéli (1958) and Hewitt (1995). There is no suggestion anywhere that this choice can be determined from the shape or properties of a root.

One might ask whether the unpredictability of version vowels and thematic suffixes merely favours a constructive analysis that derives word forms

[20] See Gurevich (2006) for a detailed discussion of the form and function of the version system in Georgian.

[21] Though Hewitt (1995: 143) remarks that 'most Transitives keep their Thematic suffix also in the Perfect series'.

https://doi.org/10.1017/S0022226706004191 Published online by Cambridge University Press

from stems, as in Anderson (1992), over an approach that derives words from roots, as in Stump (2001). However, the unpredictability of the preverbs that mark future and aorist forms in table 8 raises parallel problems for stem-based alternatives. Preverbs are, if anything, LESS predictable than neutral version vowels and thematic suffixes. Preverbs originated as directional particles, and retain a deictic sense with motion verbs in Georgian (Tschenkéli 1958: section 9.2). However, they have acquired a grammaticalized function as markers of perfectivity or future tense, and in this use they are associated with individual verbs. The lexical nature of this association is strongly emphasized in the following passages.

> Eine Gesetzmässigkeit für den Gebrauch eines bestimmten Präverbs, also eine Regel dafür, ob bestimmte Präverbien mit bestimmten Verben zusammen gebraucht werden und mit welchen, gibt es in der georgischen Sprache ebensowenig wie in anderen Sprachen, die Aspekte mittels Präverbien ausdrücken. Es ist daher unumgänglich notwendig, EIN VERB IMMER ZUSAMMEN MIT SEINEM VOLLENDUNGSPRÄVERB ZU LERNEN … Die meisten transitiven Verben haben ihr bestimmtes Vollendungspräverb. Welches Präverb jeweils die Perfektivierung bewirkt, muss mit Hilfe des Wörterbuches festgestellt oder aus der Praxis erlernt werden …[22] (Tschenkéli 1958: 85)

> It is unfortunately hardly ever the case that one can predict with which of the preverbs any individual verb will be prefixed, but, once learnt, the relevant preverb will at least remain the one that is employed in all tense-mood forms outside the present sub-series … (Hewitt 1996: 71f.)

Neither the preverb, version vowel nor thematic suffix of a Series I form is predictable from the root. Nor is the preverb predictable from a stem containing the root and any other formative. In fact, the one relatively predictable element in (3) is the vowel *-e*, which is the regular screeve vowel that marks 1p or 2p forms in 'weak' aorist indicative screeves.[23] It is simply not clear how any constructive approach would go about describing the patterns in table 8, which are essentially trivial in the context of the full conjugational system of Georgian. It does not matter whether the approach is based on realization rules (Anderson 1992), paradigm functions (Stump 2001) or

[22] 'A regular law governing the use of a particular preverb, that is, a rule for whether certain preverbs are used together with particular verbs, and with which, is as lacking in the Georgian language as in other languages that express aspect by means of preverbs. It is thus absolutely essential ALWAYS TO LEARN A VERB TOGETHER WITH ITS PERFECTIVE PREVERB … Most transitive verbs have their own particular perfective preverb. Which preverb effects the perfectivization in each case must be determined with the help of a dictionary or learned through practice …'

[23] The factors that condition the occurrence of the 'strong' aorist ending *-i* are discussed in Tschenkéli (1958: section 18).

https://doi.org/10.1017/S0022226706004191 Published online by Cambridge University Press

post-Bloomfieldian morphemes (Halle & Marantz 1993). If one invokes class features to guide the 'derivation' of the forms in table 8, each root must be tagged with 'features' that encode (i) whether the root occurs with a neutral version vowel, and if so, which vowel or vowels, (ii) whether the root occurs with a thematic suffix, and if so, which suffix or suffixes, and (iii) whether the root occurs with a preverb, and if so, which preverb or preverbs. One could in principle construct such an analysis, but the class system would be vast and unilluminating, and the features in question would again amount to blatant assembly instructions.[24]

In constructive approaches that associate roots or stems with lexeme indices (Stump 2001) or other types of indices, it might seem that an indexing strategy could express an appropriate notion of 'lexical association'. For example, the verb 'to paint' would be assigned the lexeme index XAT'VA (the imperfective masdar citation form), the verb 'to end' would be assigned the index TAVEBA, and so on. The same index would then be associated with the set of lexically determined formatives: preverbs, neutral version vowels, thematic suffixes, etc. The set of elements indexed by XAT'VA would include those in (4a), and the set indexed by TAVEBA would include those in (4b), with subscripts again indicating the provisional 'position-class' of an element.

(4) *Lexical association via lexeme indexing*
(a) ⟨[XAT'VA], da—⟩$_1$, ⟨[XAT'VA], xat'⟩$_4$, ⟨[XAT'VA], —av⟩$_7$
(b) ⟨[TAVEBA], ga—⟩$_1$, ⟨[TAVEBA], a—⟩$_3$, ⟨[TAVEBA], tav⟩$_4$, ⟨[TAVEBA], — eb⟩$_7$

The elements in (4) can be regarded as realizational pairs assigned to rule 'blocks' or as affixal entries assigned to position class 'slots'. On either interpretation, these elements will define the word forms *daxat'av* and *gaataveb*, and a similar strategy will define the other forms in table 8. Yet storing 'lexically determined' parts of a word separately, along with information about their arrangement, is no different from storing the word itself. Precisely the same segmental material must be listed. In a traditional analysis, this material is assigned to a paradigm cell. In an index-based alternative, it is encapsulated in item-specific rules or entries. The same information could also be represented by complex lexical structures, of the sort proposed by Ackerman & Webelhuth (1998) for phrasal verbs, with separate 'fields' for roots and stem formatives. These differences are just a matter of implementation, as the same content is listed in each case.

[24] It is striking that theoretical descriptions of Georgian, which are nearly all constructive, tend, with the exception of Harris (1981, 1985), to focus on the agreement system, which shows the most uniform patterns of exponence.

https://doi.org/10.1017/S0022226706004191 Published online by Cambridge University Press

In short, one can only 'derive' word forms from roots or stems by stipulating the choice of elements in (3) on an item-by-item basis. These additional stipulations, however they are expressed, sacrifice whatever economy might be achieved by starting with roots or stems in the first place. Consider, for example, just the patterns that the root *xat'* participates in the dictionary of Tschenkéli (1960–1974: 2306). This root always occurs with the thematic suffix *-av*, but it allows a range of different combinations of preverbs and version vowels. When *xat'* occurs with no version vowel, it allows seven different preverbs. With version vowel *i-*, *xat'* occurs with five of these preverbs. With version vowel *u-*, *xat'* occurs with a different set of four preverbs, and with neutral version vowel *a-*, it occurs with yet a different set of two preverbs.[25]

The pattern in Georgian is similar to Estonian, but spans a larger combinatoric space. A Georgian root provides minimal information about the choice of formatives, whose own grammatical properties do not determine their distribution. Just as noun stems fail to predict theme vowels in Estonian, Georgian roots are not reliable predictors of thematic suffixes, and roots and suffixes do not predict preverbs or version vowels. Although one can isolate recurrent elements of form, these elements do not realize constant properties, and are of limited predictive value.

4.3 *Diagnostic forms and analogical deduction*

On the other hand, the lexical associations of the elements in (3) are naturally accommodated within an abstractive account that begins with words, and regards roots and exponents as the ENDPOINT of a morphological analysis. This approach is typical of descriptive and pedagogical grammars of Georgian. This traditional analysis is illustrated by Tschenkéli (1958), which describes the Georgian conjugational system in terms of 'basic tense forms' ('Grundzeitformen'):

> Dazu kommt noch, dass Präsens und Futur im Georgischen GRUNDZEITFORMEN sind, von denen im grossen und ganzen … die andern Zeitformen abgeleitet werden. (p. 86)
>
> Das Futur eines transitiven Verbs wird aus dem Präsens gebildet, indem man vor die Präsensform das entsprechende Vollendungspräverb setzt. (p. 82)
>
> Der Aorist gilt als eine der Grundzeitformen im Georgischen, d.h. vom ihm werden … andere Zeiten abgeleitet. Es ist deshalb notwendig, sich beim Lernen eines georgischen Verbs nicht nur das Präsens und Futur, sondern auch den Aorist zu merken … Zur Bildung des AORISTSTAMMES

[25] I am grateful to Alice Harris for drawing these patterns to my attention.

https://doi.org/10.1017/S0022226706004191 Published online by Cambridge University Press

werden die im Präsens bzw. Futur auftretenden Verbsuffixe *-i*, *-av*, *-am*, *-eb*, *-ob* … weggelassen.[26] (pp. 158f.)

These passages identify the future, present and aorist indicative as basic forms, from which other forms may be obtained. Although Tschenkéli (1958) selects forms with 1sg subjects and 3sg objects as diagnostic, other forms are equally informative. Some are even more informative, given that the 3sg and 3pl subject markers are 'characteristic suffixes' (Aronson 1990: 470). A collection of 3sg subject forms would be suitably diagnostic, though the 3pl subject form from the aorist screeve is of additional use in distinguishing the class of a verb (Harris 1981: 260).

SCREEVE	SUBJ AGR	*c'er*	*šen*	*xat'*	*tb*	*cv*
Present Indicative	3sg	c'ers	ašenebs	xat'avs	atbobs	icvams
Future Indicative	3sg	dac'ers	aašenebs	daxat'avs	gaatbobs	čaicvams
Aorist Indicative	3pl	dac'eres	aašenes	daxat'es	gaatbes	čaicves
Perfective Masdar	—	dac'era	ašeneba	daxat'va	gatboba	čacma
		'write'	'build'	'paint'	'warm'	'put on'

Table 9
Diagnostic forms of Class 1 verbs

Table 9 lists sets of diagnostic forms for a representative group of Class 1 verbs. By comparing these forms, an abstractive analysis can readily identify roots and other formatives. In Class 1, preverbs distinguish future and present forms. Hence a comparison of the present and future forms in table 9 immediately identifies each of the preverbs. Thematic suffixes distinguish future from aorist forms, so a comparison of future/aorist pairs identifies the thematic suffixes in table 9. To identify version vowels and roots, one must also consider a form in which version vowels either drop out or alternate with other formatives. For verbs with no neutral version vowel, the root corresponds to the formative shared by the present, future and aorist forms. Thus comparing present, future and aorist forms identifies the roots *c'er* and *xat'*. For verbs with a vowel-initial present form, it is useful to examine the masdar (or, as discussed in section 4.4 below, a form from the perfect series), as version vowels are lost in the masdar. Thus the fact that the preradical

[26] 'In addition, the present and future in Georgian are basic tense forms, from which the other tense forms are by and large derived.' (p. 86) 'The future of a transitive verb is built from the present by placing the corresponding perfective preverb in front of the present form.' (p. 82) 'The aorist has the status of a basic tense form in Georgian, that is, from it … other tenses are derived. It is therefore necessary in learning Georgian to take note not only of the present and future but also the aorist … In the formation of the aorist stem, the verbal suffixes *-i*, *-av*, *-am*, *-eb*, *-ob* … which appear in the present or the future, as the case may be, are removed.' (pp. 158f.)

https://doi.org/10.1017/S0022226706004191 Published online by Cambridge University Press

formatives *a-* and *i-* are lost in the masdars *ašeneba*, *gatboba* and *čacma* identifies *a-* and *i-* as version vowels in the indicative forms based on the roots *šen*, *tb* and *cv*. The form of the roots in table 9 also helps to explain why roots are of such limited predictive value. Georgian verb roots, like roots in many languages, have no positive morphological or prosodic characteristics, and are merely the remnant that is left when all other exponents are removed.

In short, a comparison of the diagnostic forms in table 9 identifies their component formatives, but individual formatives do not provide the information required to reconstitute the forms. Pairs of forms also serve to identify the remaining forms. Present and aorist indicatives of a Class 1 verb immediately determine the corresponding future indicative. The present and future or future and aorist are of similar predictive value, under reasonable assumptions about speakers' ability to recognize preverbs, thematic suffixes, screeve vowels and roots in these forms.

Further, as table 10 illustrates, a set of diagnostic forms also identifies the class of a verb.

SCREEVE	CLASS 1	CLASS 2	CLASS 3	CLASS 4
Present Indicative	xat'avs	rčeba	cek'vavs	uq'vars
Future Indicative	daxat'avs	darčeba	icek'vebs	eq'vareba
Aorist Indicative	daxat'es	darčnen	icek'ves	—
	'hide'	'remain'	'dance'	'love'

Table 10
Class-defining diagnostic forms

Harris (1981: 260) defines verb classes in terms of the shape of present, future and aorist forms. Class 1 and 3 verbs mark 3sg and 3pl subjects by *-s* and *-en* in present and future screeves, and mark 3pl subjects by *-es* in the aorist, all of which contrast with the Class 2 and 4 markers. Class 1 and 3 differ in the way that they distinguish future from present screeves. Whereas Class 1 futures are marked by a preverb, the roots of Class 3 verbs are circumfixed by *i—eb* in the future screeve. Class 2 and 4 differ in a number of respects. In particular, Class 2 verbs mark 3sg and 3pl subjects by *-a* and *-an* in future screeves, and mark 3pl subjects by *-nen* in the aorist. Class 4 verbs distinguish future screeves from the present by the version vowel *e-* and mark 3sg subject forms by *-a* in the future. Class 4 paradigms are often defective, as in the case of SIQ'VARULI 'love' in table 10, which lacks an aorist series and expresses the past by the imperfect Series I screeve.

In conjunction with a set of exemplary paradigms, the diagnostic forms in table 9 and 10 permit the analogical deduction of new forms. For example, given the future form *daxat'avs* in table 10, the remaining forms in the future

https://doi.org/10.1017/S0022226706004191 Published online by Cambridge University Press

	1SG	1PL	2SG	2PL	3
1SG	—	—	dagmalav	dagmalavt	davmalav
1PL	—	—	dagmalavt	dagmalavt	davmalavt
2SG	dammalav	dagvmalav	—	—	damalav
2PL	dammalavt	dagvmalavt	—	—	damalavt
3SG	dammalavs	dagvmalavs	dagmalavs	dagmalavt	damalavs
3PL	dammalaven	dagvmalaven	dagmalaven	dagmalaven	damalaven

Table 11
Exemplary future indicative paradigm of Class 1 MALVA 'hide'

indicative paradigm of XAT'VA can be defined analogically from an exemplary Class 1 paradigm, such as the future indicative paradigm of MALVA 'hide' in table 11. Present and aorist indicative forms of XAT'VA can be defined analogically from exemplary present and aorist paradigms, and a similar strategy applies to each of the screeves in table 7. Moreover, this type of traditional analysis highlights the fact that the primary part-whole relation in the Georgian conjugational system holds between word forms and their paradigms.

4.4 *The morphomic structure of Series III*

A constructive approach might yet regard the implicational relations between screeves as fortuitous patterns that arise through the use of a restricted set of common formatives. However, the patterns of stem formation in Perfect (Series III) screeves confirm that the relations identified by traditional grammars are systematic and not accidental. As in Estonian – or Latin (Matthews 1972, 1991; Aronoff 1994) – the forms of Series III screeves are based on morphomic stems. Indeed, Series III screeves exhibit a particularly intricate pattern, in which the choice of 'parasitic' base is conditioned by verb class and valence.[27] Series III forms in Class 1 and 3 show the simplest pattern. As Aronson (1990) notes, the present perfect and pluperfect stems in these classes (his 'I. conjugation' and 'III. conjugation') are based on future and aorist indicative forms:

> The stem of the present perfect is based on the future stem for both I. and III. conjugation verbs. All [version] vowels and all person number markers are dropped. (p. 268)
>
> All III. conjugation verbs and all I. conjugation verbs (except those with [version] vowel *a-* and [thematic suffix] *-eb*) have as the stem of the

[27] The largely obsolete subjunctive perfect follows similar patterns but is omitted for the sake of brevity.

https://doi.org/10.1017/S0022226706004191 Published online by Cambridge University Press

pluperfect the 3sg. aorist minus any [version] vowel and/or person markers. (p. 270)

Table 12 shows the relation between 3sg future and aorist forms and the corresponding present perfect and pluperfect forms. In the perfect, version vowels are replaced by *u-* in 3p forms and by *i-* in other forms. Thematic suffixes are generally replaced by *-ia* in the perfect, though there is prescriptive pressure to retain the suffix *-av* in forms like *dauxat'avs* (Hewitt 1995: 263).

	CLASS 1		CLASS 3			CLASS 1	CLASS 3
FUTURE	gaaɣebs	daxat'avs	ik'ivlebs	igorebs	AORIST	gaaɣo	ik'ivla
PERFECT	gauɣ**ia**	da**u**xat'**ia**	**u**k'ivl**ia**	**u**gor**ia**	PLUPERFECT	gae**ɣa**	**e**k'ivl**a**
	'open'	'paint'	'scream'	'roll'		ope**n**'	'scream'

Table 12
Formation of Series III screeves in Class 1 and 3 (Aronson 1990: 268–271)

The pluperfect is marked by the version vowel *e-* and takes the aorist indicative endings, which in some verb types follow a formative *-in* (Aronson 1990: 270, Hewitt 1995: 268). Most importantly, the patterns in table 12 are again purely formal, as the grammatical meaning of the future and aorist bases is not preserved in the perfect and pluperfect forms that they underlie.

Class 2 (and Class 4) verbs show an even more strikingly morphomic pattern. In these classes, the base of a Series III form is conditioned by the valence of a verb. Class 2 verbs that select just a subject are termed ABSOLUTE and those that select a subject and direct object are termed RELATIVE. Series III forms of absolute Class 2 verbs are based on the perfect participle:

> The present perfect is formed by adding the auxiliary verb 'be' to the stem of the perfect participle … The pluperfect is formed by adding the aorist of 'be' to the perfect participle stem. (Aronson 1990: 301f.)

However, Series III forms of relative Class 2 verbs are based on the masdar (verbal noun):

> The stem of the perfect series (both present perfect and pluperfect) of relative II. conjugation verbs bears little resemblance to the stem of the corresponding absolute forms; it is the same as the stem of the corresponding VERBAL NOUN minus the final *-a* of such verbal nouns. (Aronson 1990: 302)

https://doi.org/10.1017/S0022226706004191 Published online by Cambridge University Press

ABSOLUTE ('MONOPERSONAL')			RELATIVE ('BIPERSONAL')		
PERF PART	dač'erili	c'akceuli	MASDAR	dasc'reba	gaxdoma
PERFECT	dač'eril**a**	c'akceul**a**	PERFECT	dasc'reb**ia**	gaxdom**ia**
PLUPERFECT	dač'eril**iq'o**	c'akceul**iq'o**	PLUPERFECT	dasc'reb**oda**	gaxdom**oda**
	'be caught'	'fall down'		'attend'	'become'

Table 13
Formation of Series III screeves in Class 2 (Hewitt 1995: 307–325)

The striking correspondence between participles and masdars and the corresponding 3sg perfect and pluperfect forms is set out in table 13. A perfect participle transparently predicts the perfect and pluperfect form of an absolute Class 2 verb, and the masdar predicts the forms of a relative Class 2 noun. Series III forms with 1p and 2p arguments are marked by additional suffixal exponents, reflecting morphologized forms of the copula. But these patterns are again the types of regular exponence that are appropriately represented within exemplary paradigms.

Table 14 provides an overview of the implicational structure of Series III screeves. The implications in Classes 1 and 3 are unidirectional because theme vowels and neutral version vowels are not recoverable from Series III forms. The implications in Classes 2 and 4 are symmetrical because Series III forms mutually predict the corresponding perfect participle and masdar.

CLASS 1 & 3			CLASS 2 & 4		
FUTURE IND			FUTURE IND		
→	PERFECT		→	PERF PART	
→	AORIST IND			↔	ABSOLUTE (PLU)PERFECT
	→	PLUPERFECT	→	MASDAR	
				↔	RELATIVE (PLU)PERFECT

Table 14
Implicational structure of Series III screeves in Georgian

The implicational patterns in table 14 are exhibited by exemplary verb paradigms. As in the simpler Russian case forms in section 3.1 above, exemplary paradigms in Georgian provide a base for the deduction of Series III forms by applying traditional proportional analogies.

A constructive approach that attempts to define Series III forms directly from roots and component formatives must replicate the structure of their perfect participle and masdar bases, along with any irregularities preserved

https://doi.org/10.1017/S0022226706004191 Published online by Cambridge University Press

by Series III forms.[28] The generalizations that would be missed by this approach are again similar in principle to those exhibited by Estonian (and Latin), but somewhat more complicated in their details. Appealing to abstract stems does not sever the link between the Series III forms and their morphomic bases, but instead disguises this relation by mediating it through a level of description that serves solely to avoid direct relations between word forms.

4.5 *Summary*

Georgian conjugations, like Saami and Estonian declensions, exhibit a general pattern that might best be described as 'pseudo-agglutination'. Given a form of a Georgian verb, one can identify the sub-word formatives that distinguish it from other forms in the verb's conjugational paradigm. So procedures of segmentation can be applied to words to obtain the minimal 'units of form' of which they are composed. Yet problems arise when one attempts to classify the resulting segments. The grammatical properties of a word form cannot be fully allocated to its parts, nor can patterns of word-based implication be shifted onto these smaller units. Assigning segments to separate inventories of roots and formatives is particularly counter-productive, as sub-word units do not even carry enough information to reconstitute the original word forms.

By comparing word forms, an abstractive approach – like the speakers that the approach is meant to model – can identify the parts of verb forms and deduce the shape of novel forms. An abstractive approach avoids the problem of reconstituting forms from their less informative parts by retaining word forms, and using them as the basis for generalization and analogy.

5. Conclusion

The patterns described above illustrate a type of asymmetry that recurs in many morphological systems. It is often the case that larger units unambiguously predict smaller units, whereas the smaller units are of more limited predictive value. A great deal of the 'technical overhead' of morphological approaches developed since Bloomfield's time does little more than compensate for the descriptive limitations of the roots, exponents and other 'recurrent partials' that are isolated in constructive descriptions. From

[28] For example, the Class 1 aorist *čatvala* has a medial *-a* which is not present in the regular future form *čatvlis* but which is preserved in the pluperfect *čaetvala*. The paradigms of Class 4 verbs are typically defective and irregular, but often maintain a transparent relation between their masdars and Series III forms, as described in Hewitt (1995: section 4.7.4).

https://doi.org/10.1017/S0022226706004191 Published online by Cambridge University Press

an abstractive standpoint, these techniques merely represent strategies for reconstituting words, which are the basic units in a grammatical system.

It is significant that many of the challenges faced by a constructive account derive from the assumption that a speaker, having identified the parts of a word form, then proceeds to discard the original word. The idea that speakers 'optimize' their mental lexicon by storing only the parts of complex forms in turn reflects more general assumptions about the lexicon being largely 'redundancy free'. An examination of complex morphological systems suggests the need to reconsider these assumptions and the conception of grammatical analysis that underlies them.

Even if one accepts that it is a desirable goal to minimize the amount of redundant information in the mental lexicon, it is far from clear that a constructive approach contributes to this goal. In a morphotactically simple language, such as English, a word-based lexicon that includes plural nouns might be regarded as containing redundant information because each (regular) plural form exhibits a uniform stem + suffix structure. From an inventory of noun stems and a list of the phonologically conditioned forms of the default plural exponent one can predict regular plural forms in English. However, English is a poor model for inflectional systems in general. In systems of the complexity of Estonian and Georgian, an inventory of roots and exponents plainly does NOT allow one to predict inflected noun or verb forms. But if the particular choices and arrangements of formatives are not predictable from the formatives themselves, it follows that the combinatoric information expressed by whole word forms is not redundant and hence that representing the morphological system by minimal formatives also does not contribute to the reduction of lexical redundancy. So even if the reduction of redundancy is regarded as desirable, constructive approaches do not contribute to this goal in systems of any complexity.

More fundamentally, work on the effects of 'morphological families' (summarized in section 2.1 above) indicates that the mental lexicon is not in fact organized in a way that would reduce 'combinatoric redundancy'. This line of research has produced a substantial and growing body of evidence that supports the traditional view that the mental lexicon is to a large degree word-based. It likewise identifies factors such as frequency rather than morphological regularity as decisive in determining whether a particular item is stored as part of a speaker's mental lexicon. The robust effects of morphological families not only call into question the a priori idealization of a redundancy-free lexicon but also undermine the broader conception of grammatical analysis that has dominated much modern theorizing. The idea that a speaker's knowledge can be modelled by formal grammars or automata (Chomsky 1956) has had a profound influence on the development of contemporary grammatical theories. Particularly within formal approaches, it has come to be regarded as self-evident that the central task of grammatical analysis involves the 'generation' of novel combinations from a lexicon of

https://doi.org/10.1017/S0022226706004191 Published online by Cambridge University Press

simple items. Yet traditional models of morphological analysis and exemplar-based models of psychological processing suggest a different formal metaphor. From the perspective of these approaches, a speaker's lexical knowledge corresponds more to a large relational database than to a general-purpose grammar or automaton in which lexical storage is in some sense more 'costly' than computational operations.

The notions of 'economy' and 'redundancy' that arise within this conception of grammar shift the focus of inquiry from symbol-counting metrics onto information-based measures of the novelty of a pattern (Bochner 1993, Kirby forthcoming). This general shift in perspective in turn suggests a formal reassessment of traditional abstractive models. In effect, a traditional WP model traces implicational 'paths' through the network of dependencies in an inflectional system by isolating the deductive patterns that are fully reliable. For a variety of practical purposes it may be useful to abstract away from patterns that either express tendencies or else depend on the interaction of multiple factors. However, there is mounting evidence that a speaker's knowledge of a range of linguistic phenomena is at least in part probabilistic in nature (Bod, Hay & Jannedy 2003) and, more specifically, that 'the role of probability in morphology is far more pervasive than standard textbooks on morphology would lead one to believe' (Baayen 2003: 233). From this general standpoint, the traditional goal of characterizing inferences with no uncertainty comprises just a part of the general task of characterizing a speaker's knowledge, which will also include more statistical or probabilistic inferences about form dependencies within a system. A general measure of form interpredictability can then be given in terms of information-theoretic notions such as 'mutual information' or 'conditional entropy' (Cover & Joy 1991).

The idea that lexical knowledge resides in a database that represents information about encountered forms has obvious points of contact with connectionist models (Bybee 1985), particularly with those models that been applied to complex morphological patterns (Thyme 1993). There is no reason to assume that lexical and morphological knowledge is in any sense unique, and exemplar-based phonological models have been proposed in which the word is the basic unit of storage (Johnson 2005a, b). The notion of morphological analysis as a pattern-matching task also shares a common guiding intuition with exemplar-based models in other domains of grammar, including construction-based approaches to syntax and compounding (Goldberg 1995, Booij 2005) and usage-based models of lexical acquisition (Tomasello 2003).

There is a particularly close parallel between the role of words in traditional accounts of morphological systems and the role of constructions in traditional analyses of syntax. The organization of a traditional grammar recognizes that certain generalizations apply over domains smaller than the phrase, and for these generalizations the 'word is a more stable and solid

https://doi.org/10.1017/S0022226706004191 Published online by Cambridge University Press

focus of grammatical relations than the component morpheme' (Robins 1959: 128). But grammars also recognize that languages contain distinctive phrasal and clausal units, and for the characterization of these syntactic patterns, the construction is a more stable and solid focus than the component word. In each domain, one expects the units of a speaker's internal grammar to correspond to linguistic forms that are informative and sufficiently frequent. There is no compelling reason to believe that the speaker, in abstracting patterns from the speech stream, is driven to isolate minimal units of form, without regard to their information value. Yet this is essentially the conception of morphological analysis embodied in constructive models. A return to the abstractive perspective of traditional grammars thus provides a better description of word-based patterns in complex inflectional systems while offering a more plausible basis for psychological models.

REFERENCES

Ackerman, F. & Webelhuth, G. (1998). *A theory of predicates*. Stanford, CA: CSLI Publications.

Anderson, S. R. (1992). *A-morphous morphology*. Cambridge: Cambridge University Press.

Aronoff, M. (1994). *Morphology by itself: stems and inflectional classes*. Cambridge, MA: MIT Press.

Aronson, H. I. (1990). *Georgian: a reading grammar*. Columbus, OH: Slavica Publishers.

Baayen, R. H., Lieber, R. & Schreuder, R. (1997). The morphological complexity of simple nouns. *Linguistics* **35**. 861–877.

Baayen, R. H. (2003). Probabilistic approaches to morphology. In Bod, Hay & Jannedy (eds.), 229–287.

Baayen, R. H., McQueen, J. M., Dijkstra, T. & Schreuder, R. (2003). Frequency effects in regular inflectional morphology: revisiting Dutch plurals. In Baayen & Schreuder (eds.), 355–390.

Baayen, R. H. & Schreuder, R. (eds.) (2003). *Morphological structure in language processing*. Berlin: Mouton de Gruyter.

Bartens, H.-H. (1989). *Lehrbuch der saamischen* (*lappischen*) *Sprache*. Hamburg: Helmut Buske Verlag.

Blevins, J. P. (2005). Word-based declensions in Estonian. In Booij, G. & van Marle, J. (eds.), *Yearbook of Morphology* 2005. Dordrecht: Springer. 1–25.

Bochner, H. (1993). *Simplicity in generative morphology*. Berlin: Mouton de Gruyter.

Bod, R., Hay, J. & Jannedy, S. (2003). *Probabilistic linguistics*. Cambridge, MA: MIT Press.

Booij, G. (2005). Compounding and derivation: evidence for Constructional Morphology. In Dressler, W. U., Kastovsky, P. & Rainer, F. (eds.), *Morphology and its demarcations*. Amsterdam: Benjamins. 109–132.

Brown, D. (1998). Stem indexing and morphonological selection in the Russian verb: A Network Morphology account. In Fabri, R., Ortmann, A. & Parodi, T. (eds.), *Models of inflection*. Tübingen: Niemeyer. 196–224.

Bybee, J. 1985. *Morphology: a study of the relation between meaning and form*. Amsterdam: Benjamins.

Carstairs, A. (1983). Paradigm economy. *Journal of Linguistics* **19**. 115–125.

Carstairs-McCarthy, A. (1991). Inflection classes: two questions with one answer. In Plank, F. (ed.), *Paradigms: the economy of inflection*. Berlin: Mouton de Gruyter. 213–253.

Carstairs-McCarthy, A. (1994). Inflection classes, gender, and the principle of contrast. *Language* **70**. 737–788.

Chomsky, N. (1956). Three models for the description of language. *Institute of Radio Engineers Transactions on Information Theory* **II-2**. 113–124.

Chomsky, N. (1975). *The logical structure of linguistic theory*. Chicago, IL: University of Chicago Press.

https://doi.org/10.1017/S0022226706004191 Published online by Cambridge University Press

Corbett, G. (1983). *Hierarchies, targets and controllers: agreement patterns in Slavic*. London: Croom Helm.
Cover, T. M. & Joy, A. T. (1991). *Elements of information theory*. New York: John Wiley & Sons.
de Jong, N. H. (2002). *Morphological families in the mental lexicon*. Ph.D. dissertation, University of Nijmegen.
Ehala, M. (2003). Estonian quantity: implications for moraic theory. In Nelson, D. & Manninen, S. (eds.), *Generative approaches to Finnic and Saami linguistics*. Stanford, CA: CSLI Publications. 51–79.
Erelt, M. (ed.) (2003). *Estonian language*. Tallinn: Estonian Academy Publishers.
Erelt, M., Kasik, R., Metslang, H., Rajandi, H., Ross, K., Saari, H., Tael, K. & Vare, S. (1995). *Eesti keele grammatika: morfoloogia*. Tallinn: Eesti Teaduste Akadeemia Eesti Keele Instituut.
Erelt, M., Erelt, T. & Ross, K. (2000). *Eesti keele käsiraamat*. Tallinn: Eesti Keele Sihtasutus.
Goldberg, A. (1995). *A Construction Grammar approach to argument structure*. Chicago, IL: University of Chicago Press.
Grünthal, R. (2003). *Finnic adpositions and cases in change*. Helsinki: Société Finno-Ougrienne.
Gurevich, O. (2006). *Constructional morphology: the Georgian version*. Ph.D. dissertation, University of California, Berkeley.
Halle, M. & Marantz, A. (1993). Distributed morphology and the pieces of inflection. In Hale, K. & Keyser, S. J. (eds.), *The view from Building 20*. Cambridge, MA: MIT Press. 111–176.
Harris, A. C. (1981). *Georgian syntax: a study in Relational Grammar*. Cambridge: Cambridge University Press.
Harris, A. C. (1985). *Diachronic syntax: the Kartvelian case*. New York: Academic Press.
Hay, J. (2001). Lexical frequency in morphology: is everything relative? *Linguistics* **39**. 1041–1070.
Hewitt, B. G. (1995). *Georgian: a structural reference grammar*. Amsterdam: John Benjamins.
Hewitt, B. G. (1996). *Georgian: a learner's grammar*. London: Routledge.
Hockett, C. F. (1947). Problems of morphemic analysis. *Language* **23**. 321–343. [Reprinted in Joos (1957), 229–242.]
Hockett, C. F. (1954). Two models of grammatical description. *Word* **10**. 210–231. [Reprinted in Joos (1957), 386–399.]
Hockett, C. F. (1958). *A course in modern linguistics*. New York: MacMillan.
Hockett, C. F. (1967). The Yawelmani basic verb. *Language* **43**. 208–222.
Johnson, K. (2005a). Resonance in an exemplar-based lexicon: the emergence of social identity and phonology. In *UC Berkeley Phonology Lab Annual Report* 2005. 95–128. [To appear, *Journal of Phonetics*.]
Johnson, K. (2005b). Decisions and mechanisms in exemplar-based phonology. In *UC Berkeley Phonology Lab Annual Report* 2005. 289–311. [To appear in Sole, M. J., Beddor, P. & Ohala M. (eds.), *Experimental approaches to phonology*. Oxford: Oxford University Press.]
Joos, M. (ed.) (1957). *Readings in linguistics I*. Chicago, IL: University of Chicago Press.
Kirby, J. (2005). *Implicational structure in Czech nominal morphology*. MA dissertation, University of California, San Diego.
Kirby, J. (forthcoming). Minimal redundancy in word-based morphology. In Ackerman, F., Blevins, J. P. & Stump, G. T. (eds.), *Patterns in paradigms*. Stanford, CA: CSLI Publications.
Kostić, A., Marković, T. & Baucal, A. (2003). Inflectional morphology and word meaning: orthogonal or co-implicative cognitive domains. In Baayen & Schreuder (eds.), 1–43.
Krott, A., Schreuder, R. & Baayen, R. H. (2001). Analogy in morphology: modeling the choice of linking morphemes in Dutch. *Linguistics* **39.1**. 51–93.
Krott, A., Schreuder, R. & Baayen, R. H. (2002). Analogical hierarchy: exemplar-based modeling of linkers in Dutch noun-noun compounds. In Skousen, R., Londsdale, D. & Parkinson, D. B. (eds.), *Analogical modeling: an exemplar-based approach to language*. Amsterdam: John Benjamins. 181–206.
Krott, A., Schreuder, R., Baayen, R. H. & Dressler, W. U. (in press). Analogical effects on linking elements in German. *Language and Cognitive Processes*.
Kuryłowicz, J. (1949). La nature des procès dits 'analogiques'. *Acta Linguistica*. 121–138. [Reprinted in Hamp, E., Householder, F. W. & Austerlitz, R. (eds.) (1966), *Readings in linguistics II*. Chicago, IL: University of Chicago Press. 158–174.]
Kuryłowicz, J. (1964). *The inflectional categories of Indo-European*. Heidelberg: Carl Winter.

https://doi.org/10.1017/S0022226706004191 Published online by Cambridge University Press

Laakso, J. (2001). The Finnic languages. In Dahl, Ö & Koptjevskaja-Tamm, M. (eds.), *Circum-Baltic languages*, vol. I: *Past and present*. Amsterdam: John Benjamins. 179–212.
Lehiste, I. (1997). *Search for phonetic correlates in Estonian prosody*. In Lehiste, I. & Ross, J. (eds.), *Estonian prosody: papers from a symposium* Tallinn: Institute of Estonian Language. 11–33.
Lieber, R. (1992). *Deconstructing morphology*. Chicago, IL: University of Chicago Press.
Manning, C. D. & Schütze, H. (1999). *Foundations of statistical natural language processing*. Cambridge, MA: MIT Press.
Matthews, P. H. (1972). *Inflectional morphology: a theoretical study based on aspects of Latin verb conjugation*. Cambridge: Cambridge University Press.
Matthews, P. H. (1991). *Morphology*. Cambridge: Cambridge University Press.
Morpurgo Davies, A. (1998). *Nineteenth-century linguistics*. London: Longman.
Moscoso del Prado Martín, F. (2003). *Paradigmatic structures in morphological processing: computational and cross-linguistic studies*. Ph.D. dissertation, University of Nijmegen.
Moscoso del Prado Martín, F., Kostić, A. & Baayen, R. H. (2004). Putting the bits together: an information-theoretical perspective on morphological processing. *Cognition* **94**. 1–18.
Paul, H. (1880). *Prinzipien der Sprachgeschichte* (8th edn., 1968). Tübingen: Max Niemayer Verlag.
Pertsova, K. (2004). *Distribution of genitive plural allomorphs in the Russian lexicon and in the internal grammar of native speakers*. MA thesis, University of California, Los Angeles.
Prince, A. (1980). A metrical theory for Estonian quantity. *Linguistic Inquiry* **11**. 511–562.
Robins, R. H. (1959). In defence of WP. *Transactions of the Philological Society*. 116–144.
Saagpakk, P. F. (2000). *Estonian-English dictionary*. Tallinn: Koolibri.
Saussure, F. de. (1916). *Cours de linguistique générale*. Paris: Payot.
Schreuder, R. & Baayen, R. H. (1997). How complex simplex words can be. *Journal of Memory and Language* **37**. 118–139.
Skousen, R. (1989). *Analogical modeling of language*. Dordrecht: Kluwer.
Skousen, R. (1992). *Analogy and structure*. Dordrecht: Kluwer.
Steele, S. (1995). Towards a theory of morphological information. *Language* **71**. 260–309.
Stump, G. T. (1993). On rules of referral. *Language* **69**. 449–479.
Stump, G. T. (2001). *Inflectional morphology: a theory of paradigm structure*. Cambridge: Cambridge University Press.
Taft, M. (1979). Recognition of affixed words and the word frequency effect. *Memory and Cognition* **7**. 263–272.
Thyme, A. (1993). *A connectionist approach to nominal inflection: paradigm patterning and analogy in Finnish*. Ph.D. thesis, University of California, San Diego.
Tomasello, M. (2003). *Constructing a language: a usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
Tschenkéli, K. (1958). *Einführung in die georgische Sprache*. Zurich: Amirani Verlag.
Tschenkéli, K. (1960–1974). *Georgisch-deutsches Wörterbuch*. Zurich: Amirani Verlag.
Viitso, T.-R. (2003a). Structure of the Estonian language: phonology, morphology and word formation. In Erelt (ed.), 9–92.
Viitso, T.-R. (2003b). Rise and development of the Estonian language. In Erelt (ed.), 130–230.
Viks, Ü. (1992). *Morphological dictionary of Estonian: introduction and grammar*. Tallinn: Institute of Language and Literature.
Zwicky, A. M. (1985). How to describe inflection. In Niepokuj, M., Van Clay, M., Nikiforidou, V. & Feder, D. (eds.), *Proceedings of BLS 11*. Berkeley: Berkeley Linguistics Society. 372–386.

Author's address: Research Centre for English and Applied Linguistics, University of Cambridge, Faculty of English Building, 9 West Road, Cambridge CB3 9DP, U.K.
E-mail: jpb@jpblevins.com

https://doi.org/10.1017/S0022226706004191 Published online by Cambridge University Press