Atomistic and holistic exponence in Information-Based Morphology

Berthold Crysmann
Laboratoire de linguistique formelle, CNRS

Olivier Bonami
Laboratoire de linguistique formelle, U Paris-Diderot

Proceedings of the 24th International Conference on
Head-Driven Phrase Structure Grammar

University of Kentucky, Lexington

Stefan Müller (Editor)

2017

Stanford, CA: CSLI Publications

pages 141–161

Keywords: Inflection, Exponence, Realisational Morphology, Estonian, Swahili

Crysmann, Berthold & Olivier Bonami. 2017. Atomistic and holistic exponence in Information-Based Morphology. In Stefan Müller (ed.), *Proceedings of the 24th International Conference on Head-Driven Phrase Structure Grammar, University of Kentucky, Lexington*, 141–161. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2017.8.

Abstract

In this paper we discuss two contrasting views of exponence in inflectional morphology: the atomistic view, where content is associated individually with minimal segmentable morphs, and the holistic view, where the association is made for the whole word between complex content and constellations of morphs. On the basis of data from Estonian and Swahili, we argue that an adequate theory of inflection should be able to accomodate both views. We then show that the framework of Information-based Morphology (Crysmann and Bonami, 2016) is indeed compatible with both views, thanks to relying on realisation rules that associate *m* units of forms with *n* units of content.

1 Introduction

A core concern of any theory of inflectional morphology is to capture the fact that the same exponents may be used in different ways in different contexts. Relevant phenomena are both wide-spread and varied. In this paper we shall examine the following two cases: (i) parallel exponence, where the same shapes in different positions realise related but distinct property sets, and (ii) 'gestalt exponence' (Blevins et al., 2016), where the cooccurrence of two (or more) exponents in a word realises a property that neither realises in isolation.[1] As we shall show, the first case is best conceived in terms of an *atomistic* view of exponence, which establishes correspondences between function and minimal segmentable morphs, whereas the second one is best understood in *holistic* terms, where form and function are established rather at the level of the whole word. Building on previous work in Information-based Morphology (Crysmann and Bonami, 2016), we suggest that both views can be reconciled under a single formal approach to morphology that relies crucially on underspecification in inheritance hierarchies of typed feature structures, and show how this conception improves over other realisational approaches to inflection.

Section 2 presents the Estonian and Swahili data that we will use to motivate the appeal of holistic and atomistic views of morphology, and then discusses how these views are conceived as irreconcilable opposites in the extant literature. In Section 3 we outline Information-based Morphology, a framework that is actually agnostic towards holistic versus atomistic views — in other words, both analyses with a holistic and an atomistic flavour can be expressed in this framework. Section 4 then presents

[†]Versions of this work were presented at the *Analysis of Morphological Systems* (AnaMorphoSys) conference (Lyon, June 2016), the 24th International Conference on HPSG (Lexington, July 2017), the DELPH-IN summit (Oslo, August 2017), as well as part of a tutorial on IbM at Frankfurt university. We would like to thank the respective audiences for their comments and suggestions, in particular Farrell Ackerman, Emily Bender, Jim Blevins, Aaron Broadwell, Dan Flickinger, Raphael Finkel, Elaine Francis, Jean-Pierre Koenig, Laura Michaelis, Frank Richter, Manfred Sailer, Andrea Sims, Géraldine Walther and Gert Webelhuth. Furthermore, we are also indebted to the comments from the anonymous reviewers of the HPSG and AnaMorphoSys conferences. This work was partially supported by a public grant overseen by the French National Research Agency (ANR) as part of the "Investissements d'Avenir" program (reference: ANR-10-LABX-0083).

[1]Other phenomena that exhibit the very same general properties include polyfunctionality, where identical forms express different function (Spencer and Stump, 2013; Ackerman and Bonami, inpress), variable placement, where one exponent realising one property set occurs in different linear positions depending on the morphosyntactic context (Stump, 1993; Crysmann and Bonami, 2016); and exuberant exponence, where some property is marked over and over again by the same forms (Harris, 2009; Crysmann, 2014).

appropriate analyses of the Estonian and Swahili data that demonstrate precisely that feature of the framework.

2 Data

2.1 Estonian

Noun declension in Estonian has served as the primary piece of evidence to argue that form-function correspondences are better understood in holistic terms, that is, established in terms of relations between fully inflected words, rather than in atomistic terms, involving the combination of sub-word units (Blevins, 2005, 2006; Blevins et al., 2016).

NOKK 'beak'			ÕPIK 'workbook'			SEMINAR 'seminar'		
	SG	PL		SG	PL		SG	PL
NOM	nokk	nokad	NOM	õpik	õpikud	NOM	seminar	seminarid
GEN	noka	nokkade	GEN	õpiku	õpikute	GEN	seminari	seminaride
PART	nokka	nokkasid	PART	õpikut	õpikuid	PART	seminari	seminarisid

Table 1: Partial paradigms exemplifying three Estonian noun declensions (core cases; Blevins, 2005)

As illustrated in Table 1, morphological marking of number (SG/PL) and core cases (NOM/GEN/PART) clearly provides distinct forms for all six paradigm cells (modulo syncretism between two cell in the SEMINAR class), but the individual devices used to express the distinctions do not align well with the functional distinctions they are supposed to express. On the side of pure exponence, we find several devices: presence vs. absence of an inflection class-specific theme vowel (*-a/-u/-i*), which segregates the nominative singular from all other forms, suffixation of case/number markers, which is sometimes identical across inflection classes (e.g. NOM.PL *-d*), and sometimes not (e.g. GEN.PL *-d/-t*). Similarly, while one might be tempted to further decompose e.g. the genitive plural marker *-de/-te* there is no constant plural form or corresponding singular form on which this decomposition could be modelled.

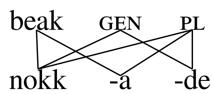

Figure 1: *m*:*n* relations in Estonian

Finally, the NOKK class displays an alternation between geminated and non-geminated stems, which witnesses an alignment with case that is the exact opposite in the singular and the plural. As summarised by the diagram in Figure 1 for *nokk-a-de* 'beak.GEN.PL', although words can readily be segmented into morphs, no morph exclusively expresses a single property, and conversely, no property is exclusively expressed by a single morph.

Thus, while individual formal devices can clearly be identified, association with function must be established at a level that involves combinations of forms. For Estonian, a holistic, or constructional (Gurevich, 2006) view appears therefore inevitable.

2.2 Swahili

In contrast to Estonian core cases, which are encoded in a highly opaque fashion, Swahili is much more transparent, thereby being compatible with an atomistic view that associates function more directly with individual exponents that serve to express them. However, if a holistic view can shed light on systems like Estonian that are not fully amenable to an atomistic analysis, one might wonder whether a more elegant model of morphology might not be arrived at by generalising all form/function relations to the level of the morphological word. Such an approach has been pursued, e.g. by Koenig (1999) who has proposed an essentially word-based constructional analysis of (part of) the Swahili position class system.

The phenomenon of parallel exponence, however, resists such a mode of analysis. Swahili verbs can inflect for both subject and object agreement, inserting exponents into different templatic slots (1). As these examples illustrate, in many cases, position, rather than shape, disambiguates which grammatical function is coded.

(1) a. ni-ta-wa-penda
1SG-FUT-3PL-like
'I will like them.'

b. wa-ta-ni-penda
3PL-FUT-1SG-like
'They will like me.'

While choice of morphosyntactic properties, and therefore, forms, are independent for both functions, it is clear from Table 2 that pairings of form and function draw largely on the same inventories. As a result, an analysis that wants to capture

PER	GEN	SUBJECT SG	SUBJECT PL	OBJECT SG	OBJECT PL
1		ni	tu	ni	tu
2		u	m	ku	wa
3	M/WA	a	wa	m	wa
	M/MI	u	i	u	i
	KI/VI	ki	vi	ki	vi
	JI/MA	li	ya	li	ya
	N/N	i	zi	i	zi
	U	u	—	u	—
	U/N	u	zi	u	zi
	KU	ku	—	ku	—

Table 2: Swahili person markers (Stump, 1993)

this generalisation must permit the reuse of the same morphological resources for different purposes within the same word, which necessitates reifying correspondences between shapes and partial morphosyntactic description—precisely what a holistic approach avoids doing.

2.3 Discussion: atomistic vs. holistic approaches to morphology

To summarise our presentation of the basic data, we can conclude that a credible morphological theory must afford ways to accommodate both atomistic and holistic analyses within the same formal system, rather than enforce one view or the other. Before we present such a theory, it is worth being more precise about the opposition between atomistic and holistic views, and showing how it connects to other metatheoretical distinctions.

For simplicity let us limit our attention to systems that are agglutinative enough that a segmentation of words into individual morphs is feasible and consensual. Given such a system, a purely atomistic view of exponence licenses/introduces/describes each morph through a separate mechanism, be it a morphemic lexical entry, a rule of exponence, or some other device. This contrasts with a purely holistic view, whereby constellations of co-occurring exponents are licensed/introduced/described simultaneously by a single mechanism, be it a rule, a schema, a construction, or some other device, like analogy.

It should be clear that, for many systems, both views may lead to a reasonable enough analysis. Sometimes an atomistic view will look more elegant because it allows for a more economical description, the distribution of exponents being largely orthogonal; sometimes a holistic view will look more elegant because there are many interdependencies between the distribution of exponents. In that sense, the two systems showcased above are extremes where one or the other view seems particularly unfit because it fails to capture some important generalization. It should also be clear that the distinction we are making is at least in part indifferent to the canonicity of exponence. Zero and cumulative (a.k.a. fused) exponence do not speak in favor of either view, as they do not create dependencies between the distribution of morphs. Widespread extended (a.k.a. multiple) and overlapping exponence are often used to argue in favor of a holistic view, although modern realisational approaches have developed means of dealing with such situations in an atomistic fashion, notably through the mechanism of rule blocks (Anderson, 1992; Stump, 2001), which localizes exponence strategies to a single set of paradigmatic alternatives. What is remarkable about Estonian declension is the combination of overlapping exponence and morphomic distribution (Aronoff, 1994), which leads to a situation where no insight is gained by describing the distribution of exponents individually.

The distinction we are making, we argue, is not reducible to one of the prevalent existing distinctions between morphological frameworks. It is separate from Stump's (2001) celebrated bi-dimensional opposition between lexical vs. inferential and incremental vs. realisational approaches: arguably, all the frameworks described by Stump, and more generally all morphological frameworks in the generative tradition, are committed to an atomistic view of inflection, although they differ vividly in the way they implement such a view.[2] In this they contrast with so-called 'word-based' (Ford et al., 1997; Blevins, 2006, 2013) or 'construction-based' (Gurevich, 2006; Booij, 2010; Harris, 2012) approaches.

We contend that our distinction does not reduce either to Blevins's (2006) con-

[2] Technically, Paradigm Function Morphology (Stump, 2001, 2016; Bonami and Stump, 2016) could accomodate holistic analyses through dedicated statements of the paradigm function appealing simultaneously to individual exponents in multiple rule blocks. But to the best of our knowledge such analyses have never been entertained, and it remains to be seen whether this can be done in an insightful fashion without appealing to some mechanism of underspecification that the framework is lacking.

trast between *constructive* and *abstractive* approaches. A constructive approach takes abstract morphological objects (morphemes, stems, lexemes, rules, etc.) as primitives from which surface words are derived, whereas an abstractive approach takes words as primitives from which other morphological entities may (but need not) be abstracted. Although most constructive approaches happen to presuppose an atomistic view, it is not incoherent to entertain holistic analyses within a constructive approach. In fact, the framework of Information-based Morphology that we will present in the next section is compatible with a constructive interpretation, but can accomodate fully holistic analyses, as we will see.

Let us finally note that the framework of HPSG is itself compatible with both atomistic and holistic views. The vast majority of extant proposals presuppose an atomistic view, independently of whether they implement an Item and Arrangement (Emerson and Copestake, 2015), Item and Process (Koenig, 1999; Sag, 2012), or Realisational approach (Erjavec, 1994; Crysmann, 2003; Bonami and Boyé, 2006; Crysmann and Bonami, 2012); for most authors this is related to the assumption that inflection, like derivation, operates on the basis of cascades of recursive rules. Notable exceptions are Krieger et al.'s (1993) early paradigm-based approach, and the analysis of Swahili conjugation entertained by Koenig (1999, 170–173). It is only with the advent of Information-based Morphology (Bonami and Crysmann, 2013; Crysmann and Bonami, 2016), where a single rule of exponence may introduce a discontinuous sequence of morphs, that holistic analyses have become a realistic large scale possibility.

3 Information-based Morphology

In this section, we shall present the basic architecture of Information-based Morphology (IbM), an inferential-realisational theory of inflection (cf. Stump, 2001) that is couched entirely within typed feature logic, as assumed in HPSG (Pollard and Sag, 1987, 1994). In IbM, realisation rules embody partial generalisations over words, where each rule may pair *m* morphosyntactic properties with *n* morphs that serve to express them. IbM is a morphous theory (Crysmann and Bonami, 2016), i.e. exponents are described as structured morphs, combining descriptions of shape (=phonology) and position class. As a consequence, individual rules can introduce multiple morphs, in different, even discontinuous positions. By means of multiple inheritance hierarchies of rule types, commonalities between rules are abstracted out: in essence, every piece of information can be underspecified, including shape, position, number of exponents, morphosyntactic properties, etc.

In contrast to other realisational theories, such as Paradigm Function Morphology (Stump, 2001) or A-morphous Morphology (Anderson, 1992), IbM does away with procedural concepts such as ordered rule blocks. Moreover, rules in IbM are non-recursive, reflecting the fact that inflectional paradigms in general constitute finite domains. Owing to the absence of rule blocks, IbM embraces a strong notion of Panini's Principle or the Elsewhere Condition which is couched purely in terms of informational content (=subsumption) and therefore applies in a global fashion.

3.1 Inflectional rules as partial abstraction over words

From the viewpoint of inflectional morphology, words can be regarded as associations between a phonological shape (PH) and a morphosyntactic property set (MS), the latter including, of course, lexemic information. This correspondence can be described in a maximally holistic fashion, as shown in Figure 2. Throughout this section, we shall use German (circumfixal) passive/past participle (*ppp*) formation, as witnessed by *ge-setz-t* 'put', for illustration.

$$\begin{bmatrix} \text{PH} & \langle \textit{gesetzt} \rangle \\ \text{MS} & \left\{ \begin{bmatrix} \text{LID} & \textit{setzen} \end{bmatrix}, \begin{bmatrix} \text{TMA} & \textit{ppp} \end{bmatrix} \right\} \end{bmatrix}$$

Figure 2: Holistic word-level association between form (PH) and function (MS)

Since words in inflectional languages typically consist of multiple segmentable parts, realisational models provide means to index position within a word: while in AM and PFM ordered rule blocks perform this function, IbM uses a set of morphs (MPH) in order to explicitly represent exponence. Having both morphosyntactic properties and exponents represented as sets, standard issues in inflectional morphology are straightforwardly captured: cumulative exponence corresponds to the expression of *m* properties by 1 morph, whereas extended (or multiple) exponence corresponds to 1 property being expressed by *n* morphs. Overlapping exponence finally represents the general case of *m* properties being realised by *n* exponents. Figure 3 illustrates the word-level $m : n$ correspondence of lexemic and inflectional properties to the multiple morphs that realise it. By means of simple underspecification, i.e. partial description, one can easily abstract out realisation of the past participle property, arriving at a description of circumfixal realisation.

Word:

$$\begin{bmatrix} \text{PH} & \langle \textit{gesetzt} \rangle \\ \text{MPH} & \left\{ \begin{bmatrix} \text{PH} & \langle \textit{ge} \rangle \\ \text{PC} & \textit{-1} \end{bmatrix}, \begin{bmatrix} \text{PH} & \langle \textit{setz} \rangle \\ \text{PC} & \textit{0} \end{bmatrix}, \begin{bmatrix} \text{PH} & \langle \textit{t} \rangle \\ \text{PC} & \textit{1} \end{bmatrix} \right\} \\ \text{MS} & \left\{ \begin{bmatrix} \text{LID} & \textit{setzen} \end{bmatrix}, \begin{bmatrix} \text{TMA} & \textit{ppp} \end{bmatrix} \right\} \end{bmatrix}$$

Abstraction of circumfixation ($1 : n$):

$$\begin{bmatrix} \text{MPH} & \left\{ \begin{bmatrix} \text{PH} & \langle \textit{ge} \rangle \\ \text{PC} & \textit{-1} \end{bmatrix}, \begin{bmatrix} \text{PH} & \langle \textit{t} \rangle \\ \text{PC} & \textit{1} \end{bmatrix}, \ldots \right\} \\ \text{MS} & \left\{ \begin{bmatrix} \text{TMA} & \textit{ppp} \end{bmatrix}, \ldots \right\} \end{bmatrix}$$

Figure 3: Structured association of form (MPH) and function (MS)

Direct word-based description, however, does not easily capture situations where the same association between form and content is used more than once in the same word, as we have seen in the case of Swahili (Stump, 1993; Crysmann and Bonami, 2016). Similar problems arise in the case of exuberant exponence, as witnessed by Batsbi (Harris, 2009; Crysmann, 2014). By way of introducing a level of R(EALISATION) R(ULES), reuse of resources becomes possible. Rather than expressing the relation between form and function directly on the word level, IbM assumes that a word's description includes a specification of which rules license the realisation between form and content, as shown in Figure 4.

Realisation rules (members of set RR) pair a set of morphological properties to be expressed, the morphology under discussion (MUD) with a set of morphs that re-

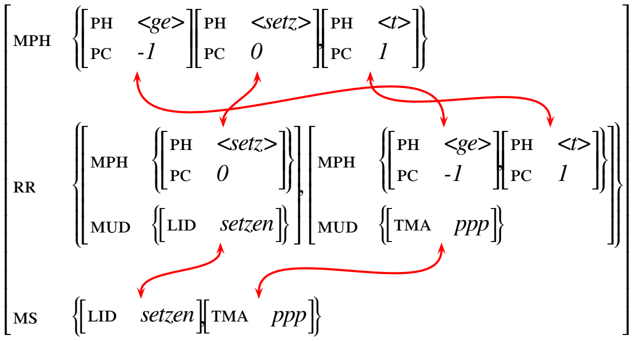

Figure 4: Association of form and function mediated by rule

alise them (MPH). A simple principle of morphological well-formedness (Figure 5) ensures that the properties expressed by rules add up to the word's property set and that the rules' MPH sets add up to that of the word, thereby ensuring a notion of 'Total Accountability' (Hockett, 1947) without relying on a 1 : 1 correspondence between form and content.

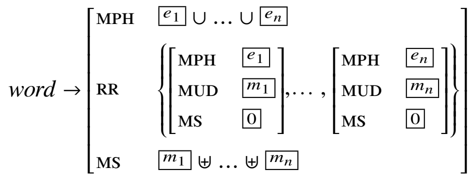

Figure 5: Morphological well-formedness

Realisation rules conceived like this essentially constitute partial abstractions over words, stating that some collection of morphs jointly expresses a collection of morphosyntactic properties. In the example in Figure 4, we find that realisation rules thus conceived implement the $m : n$ nature of inflectional morphology at the most basic level: while permitting the representation of classical morphemes as 1 : 1 correspondences, this is but one option. The circumfixal rule for past participal inflection directly captures the 1 : n nature of extended exponence.

3.2 Levels of abstraction

The fact that IbM, in contrast to PFM or AM, recognises $m : n$ relations between form and function at the most basic level of organisation, i.e. realisation rules, means that morphological generalisations can be expressed in a single place, namely simply as abstractions over rules. Rules in IbM are represented as typed feature structures organised in an inheritance hierarchy, such that properties common to leaf types can be abstracted out into more general supertypes. This vertical abstraction is illustrated in Figure 6. Using again German past participles as an example, the commonalities that regular circumfixal *ge-...-t* (as in *gesetzt* 'put') shares with subregular *ge-...-en* (as

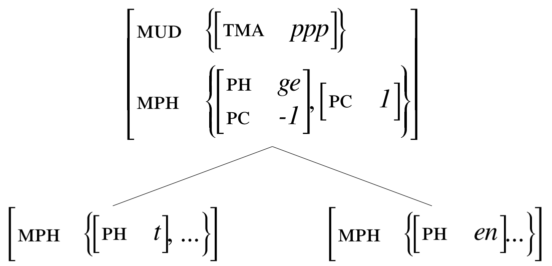

Figure 6: Vertical abstraction by inheritance

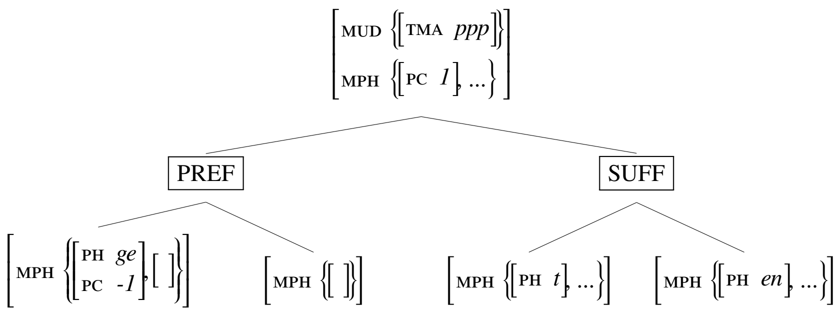

Figure 7: Horizontal abstraction by dynamic cross-classification

in *geschrieben* 'written') can be generalised as the properties of a rule supertype from which the more specific leaves inherit. Note that essentially all information except choice of suffixal shape is associated with the supertype. This includes the shared morphotactics of the suffix.

In addition to vertical abstraction by means of standard monotonic inheritance hierarchies, IbM draws on Online Type Construction (Koenig and Jurafsky, 1994): using dynamic cross-classification, leaf types from one dimension can be distributed over the leaf types of another dimension. This type of horizontal abstractions permits modelling of systematic alternations, as illustrated once more with German past participle formation:

(2) a. **ge**-setz-**t** 'set/put'
b. über-setz-**t** 'translated'
c. **ge**-schrieb-**en** 'written'
d. über-schrieb-**en** 'overwritten'

In the more complete set of past participle formations shown in (2), we find alternation not only between choice of suffix shape (*-t* vs. *-en*), but also between presence vs. absence of the prefixal part (*ge-*).

Figure 7 shows how Online Type Construction enables us to generalise these patterns in a straightforward way: while the common supertype still captures properties true of all four different realisations, namely the property to be expressed and the fact that it involves at least a suffix, concrete prefixal and suffixal realisation patterns are segregated into dimension of their own (indicated by PREF and SUFF).

Systematic cross-classification (under unification) of types in PREF with those in SUFF yields the set of wellformed rule instances, e.g. distributing the left rule type in PREF over the types in SUFF yields the rules for *ge-setz-t* and *ge-schrieb-en*, whereas distributing the right type in PREF gives us the rules for *über-setz-t* and *über-schrieb-en*, which are characterised by the absence of the participial prefix.

3.3 The atomistic/holistic divide in IbM

An interesting feature of the formal device of underspecification is that it is largely agnostic as to the distinction between what Blevins (2006) calls a *constructive* view of morphology, where words are derived from minimal elements, and what he calls an *abstractive* view, where words are taken as prior, and entities such as stems and affixes, to the extent that they are useful analytic devices, are higher-level abstractions over words.[3] Nodes in the inheritance hierarchy are nothing more than generalisations on the distribution of recurrent partials, i.e. useful abstractions from surface word-sized Saussurean signs. Because inheritance is monotonic — there are no defaults, unlike what happens in Network Morphology (Brown and Hippisley, 2012) and Construction Morphology (Booij, 2010)) —, the hierarchy can be seen both from a top-down point of view, as a way of encoding optimally constraints on exponence, and from a bottom-up point of view, as an explicit representation of relations of similarity and difference between words.

From the constructive point of view, it is important that the full hierarchy can be deduced from a partial hierarchy through the use of online type construction (Koenig, 1999): this means that only those realisation rules that include some constraint not inherited from supertypes need to be explicitly listed, rather than inferred from the shape of the system, by means of systematic intersection of leaf types from each dimension (boxed). Such inferrable types are indicated by dashed lines, as shown in Figure 8. From an abstractive point of view, on the other hand, the leaf types in the hierarchy are ontologically prior, as they constitute the directly observable associations between content and form.

The fact that rule inheritance hierarchies can be interpreted either in constructive or abstractive terms makes it very natural to accomodate both atomistic and holistic analyses within the same framework, as we will see in the next section.

4 Analysis

We now turn to the analysis of Swahili and Estonian, which, as we noted before, mark two extreme points on the atomistic vs. holistic cline.

[3]Blevins introduces the notion of an abstractive approach in the context of the study of the implicative structure of paradigms (Wurzel, 1984), arguing that segmentation is of little help to study that structure. Most work claiming the label 'abstractive' pursues the same agenda (e.g. Ackerman and Malouf 2013; Bonami and Beniamine 2016; Sims 2015). We contend however that the idea of an abstractive approach to morphology applies beyond the domain of implicative structure, and is relevant even for the analysis of agglutinative systems where segmentation is not disputed.

4.1 Swahili

The analysis in this subsection essentially rehearses the proposal in Crysmann and Bonami (2016), mainly serving the purpose of contrasting the advantages of an atomistic analysis of this system, compared to the holistic approach required by Estonian.

By way of illustration, Figure 8 provides a partial description of parallel exponence.

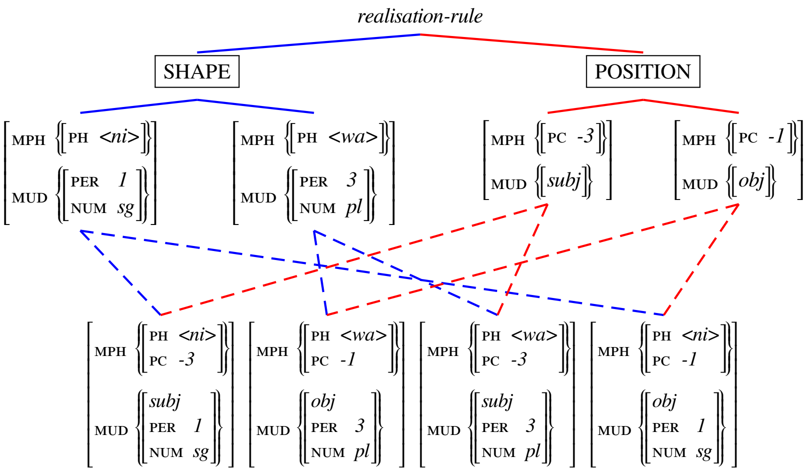

Figure 8: Rule type hierarchy for Swahili parallel position classes (Crysmann and Bonami, 2016)

Types in the SHAPE dimension on the left pair shapes (phonology of morphs) with person, number and gender properties, whereas the two types in the POSITION dimension specify position class information for subject vs. object agreement. Systematic intersection of leaf types (one each from either dimension) yields the fully expanded set of rules, effectively distributing positional marking of grammatical function over the exponents. In order to derive a morphologically wellformed, fully inflected word, every element of the morphological property set must be realised by some realisation rule, i.e. each member of the property set must be "consumed" by some MUD element of exactly one rule. As a result, rule type hierarchies constitute a repository of recipes that can be referred to more than once, e.g. for subject and object agreement.

4.2 Estonian

As we have seen in the discussion in section 2.1, association between form and function cannot be easily broken down to any specific exponents, but generally has to take into consideration combinations of stem alternation, theme vowels, and suffixation. I.e. it is only the specific combination of these marking devices that identifies any specific cell in the paradigm. Thus, rather than organising the hierarchy of realisation rule types in terms of morphosyntactic properties, we shall primarily partition it in terms of marking strategies, identifying three cross-classifying dimensions for stem selection, theme vowel selection and suffixation, as depicted in Figure 9. In the interest of readability, we represent the overall type hierarchy without the type constraints associated to the nodes. See the sub-hierarchies below for full detail.

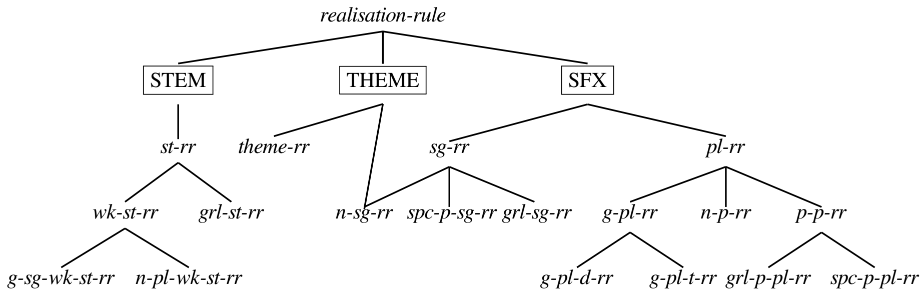

Figure 9: Hierarchy of rule types for Estonian

4.2.1 Suffixation

Probably the most straightforward observation regarding the Estonian data is that inflection in the plural uniformly involves suffixation, whereas in the singular it only sporadically does so. Moreover, plural inflection uniformly features a theme vowel. Taking stem selection into account, plural forms are thus tri-morphic, a generalisation captured by the top-most plural rule type *pl-rr* in Figure 10, which pairs the morphosyntactic property with a constraint on the number and position class indices of the exponents. Subtypes of *pl-rr* then constrain the shape of the exponents by case. In the case of the genitive and partitive, leaf types expand the partial shape descriptions, depending on inflection class.

In the singular, by contrast, we find much more morphotactic variation: while most singular forms are bi-morphic (*grl-sg-rr*), consisting of a stem and a theme vowel only, nominative singular is systematically monomorphic (*n-sg-rr*), featuring a bare stem. Quite idiosyncratic is the marking for partitive singular in the *õpik*-class, which is tri-morphic, involving the suffix *-t*.

As a consequence of this heterogeneity, the type *sg-rr* is largely underspecified. Its subtypes enumerate the three patterns, providing a general bi-morphic pattern (*grl-sg-rr*), which merely specifies morphotactics, a monomorphic pattern for the nominative (*n-sg-rr*), and the exceptional pattern for the partitive *õpik*-class. Note that Panini's Principle will force the use of the two more specific patterns where appropriate, owing to the fact that the description of MUD in *grl-sg-rr* unilaterally subsumes those in either *n-sg-rr* or *spc-p-sg-rr*.

4.2.2 Theme vowel selection

We have so far assumed without further discussion that inflection class information is represented as part of the morphosyntactic property set. Indeed, being lexemic in nature, this information is best tied to the equally lexemic specification of stem alternants. Since inflection class not only governs allomorphic alternation of inflection markers, but also systematically determines the shape of theme vowel, we represent

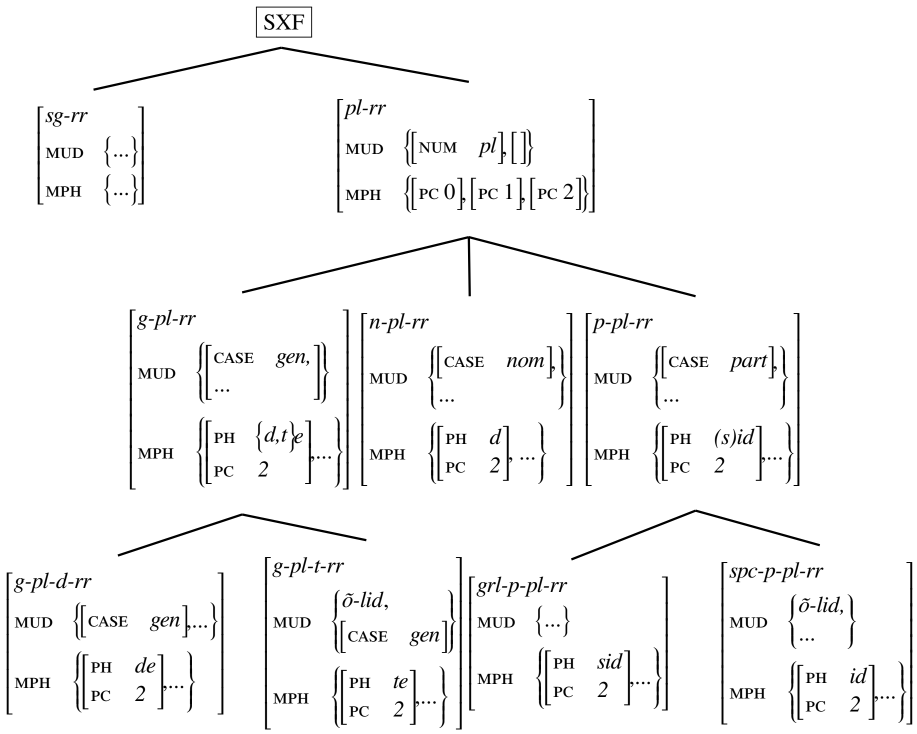

Figure 10: Sub-hierarchy of suffixation rule types for Estonian (plural)

Estonian inflection classes by means of a hierarchy of typed feature structures, as shown in Figure 12.

While every *lid* (=lexemic identity) has a ST(EM) value and a theme vowel specification as appropriate features, subtypes of *lid* determine the exact nature of that vowel. Nouns of class *n-lid* (e.g. *nokk*) display an alternation between a strong stem and an alternate weak stem. Therefore, we declare this type as having an additional appropriate feature WK-ST, to host the specification of the alternate weak stem.

The standard theme selectional rule (*theme-rr*) simply targets the TH feature of the *lid* and inserts its value as the phonology of a morph in position 1. Note that this theme selection rule is very similar to stem selection rules, which also typically just pick up some lexemically specified phonology and insert it in a morphotactic position.

The reason why we use a special rule to insert the theme vowel, rather than making it a property of the stem's phonology is two-fold: first, its form is highly systematic, and second, the presence vs. absence of the theme vowel helps to mark an inflectional contrast. While generally there is an overt theme vowel, the nominative singular of all three paradigms is always a bare stem, devoid of both inflectional suffixes and the theme vowel. The rule type *n-sg-rr* captures this case, restricting the MPH set to be monomorphic (=bare stem). Note again that this rule type will preempt by virtue of Panini's Principle the use of the general *theme-rr*, due to the fact that the latter properly subsumes the former in its MUD description.

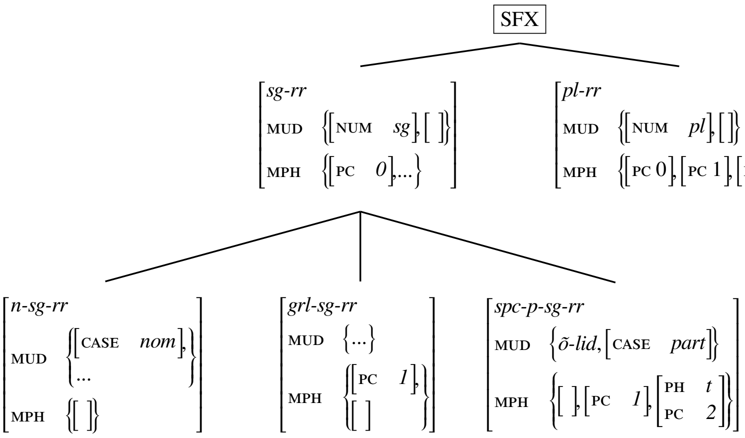

Figure 11: Sub-hierarchy of suffixation rule types for Estonian (singular)

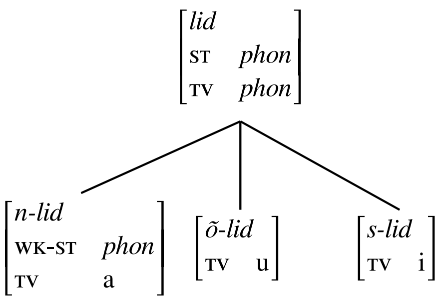

Figure 12: Hierarchy of *lid* types for Estonian

4.2.3 Stem selection

The last piece of inflection we need to address is stem selection: as depicted in Figure 14, the major split in the stem selection rules concerns the generic rule type *grl-st-rr* vs. the subtree under *wk-st-rr*.

While the general stem selection *grl-st-rr* picks out the ST feature and inserts it as a morph in position 0, its sister type *wk-st-rr* selects the alternate weak stem instead, restricting application to *nokk*-type nouns (*n-lid*). The two subtypes of *wk-st-rr* further restrict the applicability of this rule by means of enumerating the paradigm cells to which this alternate stem selection rule can be applied. Since Paninian competition is defined over leaf types (see Crysmann and Bonami, 2016; Crysmann, 2017),[4] application of the general stem selection rule is only preempted in two cells of *nokk*-class nouns, i.e. the nominative plural and the genitive singular. Thus, the general rule takes care not only of *õpik*-class and *seminar*-class nouns, but it also fills most of the cells of *nokk*-class nouns, thereby acting as a true default.

[4]More precisely, Panini's principle regulates competition between rules. Since IbM builds on Online Type Construction (Koenig and Jurafsky, 1994), rule instances (as opposed to types) must be maximally specific types w.r.t. all dimensions, i.e. Paninian competition is computed amongst leaf types of the fully expanded type hierarchy (cf. Crysmann, 2003). This distinction, while important in the general case, happens to be immaterial here, so it is sufficient to consider leaf types within each dimension separately.

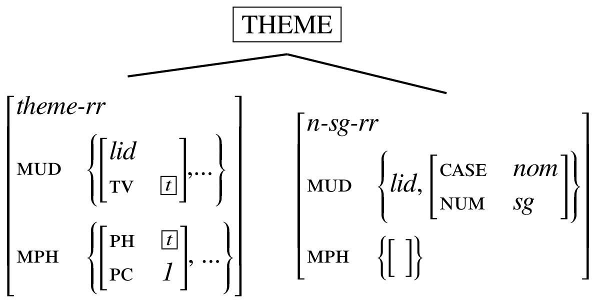

Figure 13: Theme rule type

4.2.4 Putting things together

We have argued that the system of Estonian core cases calls for a holistic approach and we have suggested that IbM is capable to do that, while at the same time its system of typed feature structure inheritance, and in particular Koenig/Jurafsky-style dynamic cross-classification will permit to squeeze out partial generalisations. So far, we have focussed on the latter aspect, laying out the organisation of partial description by means of rule types organised into the three dimension for stem selection (STEM), theme vowel selection (THEME), and suffixation (SFX). We shall now show how the constraints in the three dimensions interact to derive some interesting cases.

To start with, let us consider some tri-morphic plural, e.g. the nominative plural of *nokk*. Given the type hierarchy of rule types in Figure 9, any well-formed inflectional rule needs to inherit from exactly one leaf type in each of the three dimensions, as dictated by Online Type Construction (Koenig and Jurafsky, 1994). The inflectional rule suitable to derive this cell can be inferred by means of unifying the types *n-pl-rr* (from SFX), *theme-rr* (THEME), and *n-pl-wk-st-rr* (from STEM), yielding the fully expanded rule in (3) deriving *nokad*.

(3) $$\textit{theme-rr} \ \& \ \textit{n-p-rr} \ \& \ \textit{n-pl-wk-sg-rr} \equiv \begin{bmatrix} \text{MUD} & \left\{ \begin{bmatrix} \textit{n-lid} & \\ \text{WK-ST} & \boxed{s} \\ \text{TH} & \boxed{t} \end{bmatrix}, \begin{bmatrix} \text{CASE} & \textit{nom} \\ \text{NUM} & \textit{pl} \end{bmatrix} \right\} \\ \text{MPH} & \left\{ \begin{bmatrix} \text{PH} & \boxed{s} \\ \text{PC} & \textit{0} \end{bmatrix}, \begin{bmatrix} \text{PH} & \boxed{t} \\ \text{PC} & \textit{1} \end{bmatrix}, \begin{bmatrix} \text{PH} & \textit{d} \\ \text{PC} & 2 \end{bmatrix} \right\} \end{bmatrix}$$

Intersection of e.g. the genitive plural rule type *g-pl-d-rr* with THEME and STEM types will only yield successful unification with *theme-rr* and *grl-st-rr*, deriving e.g. *nokkade* and *seminaride*, i.e. any instance where the standard stem is selected. The expanded rule is given in (4).

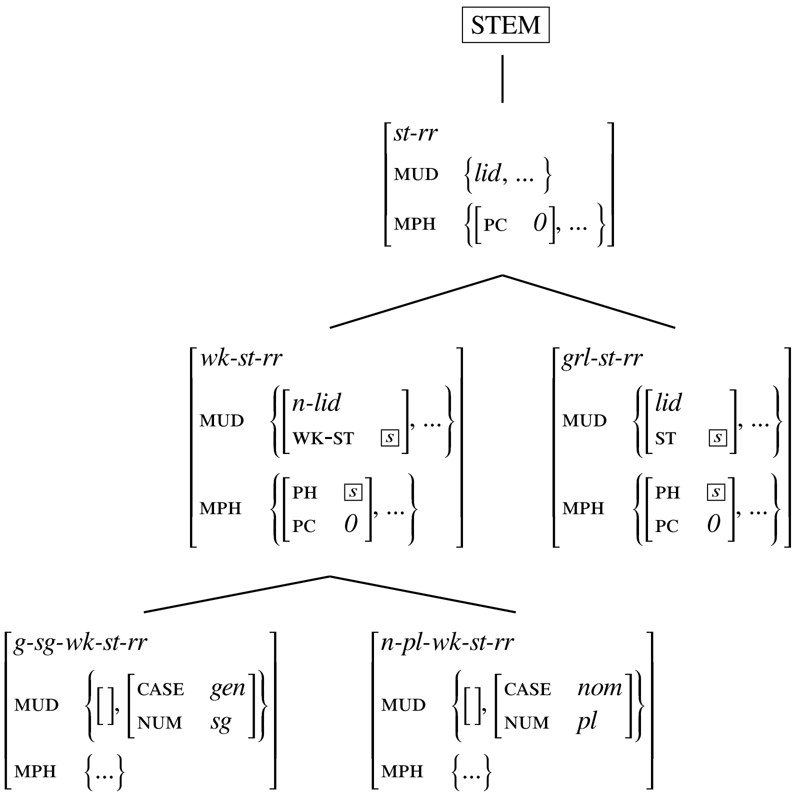

Figure 14: Hierarchy of stem selection rule types for Estonian

(4) *theme-rr* & *grl-pl-d-rr* & *grl-st-rr* $\equiv \begin{bmatrix} \textsc{mud} & \left\{ \begin{bmatrix} \textsc{st} & \boxed{s} \\ \textsc{th} & \boxed{t} \end{bmatrix}, \begin{bmatrix} \textsc{case} & \textit{gen} \\ \textsc{num} & \textit{pl} \end{bmatrix} \right\} \\ \textsc{mph} & \left\{ \begin{bmatrix} \textsc{ph} & \boxed{s} \\ \textsc{pc} & \textit{0} \end{bmatrix}, \begin{bmatrix} \textsc{ph} & \boxed{t} \\ \textsc{pc} & \textit{1} \end{bmatrix}, \begin{bmatrix} \textsc{ph} & \textit{de} \\ \textsc{pc} & \textit{2} \end{bmatrix} \right\} \end{bmatrix}$

Turning to singular patterns, let us consider the partitive, as witnessed by *nokka*, *õpikut* and *seminari*. Intersection of leaf types yields two solutions compatible with the partitive, both of which inherit from the general theme and stem selection rule types. The variation lies with the singular SFX rules: choosing the more general type *grl-sg-rr* yields the expanded bi-morphic singular rule in (5) for e.g. *nokka* and *seminari*, whereas choice of *spc-sg-rr* yields the class-specific tri-morphic rule for the partitive singular *õpikut*, as given in (6).

(5) *theme-rr* & *grl-sg-rr* & *grl-st-rr* $\equiv \begin{bmatrix} \textsc{mud} & \left\{ \begin{bmatrix} \textsc{st} & \boxed{s} \\ \textsc{th} & \boxed{t} \end{bmatrix}, \begin{bmatrix} \textsc{num} & \textit{sg} \end{bmatrix} \right\} \\ \textsc{mph} & \left\{ \begin{bmatrix} \textsc{ph} & \boxed{s} \\ \textsc{pc} & \textit{0} \end{bmatrix}, \begin{bmatrix} \textsc{ph} & \boxed{t} \\ \textsc{pc} & \textit{1} \end{bmatrix} \right\} \end{bmatrix}$

(6) *theme-rr* & *spc-sg-rr* & *grl-st-rr* ≡

$$\left[\begin{array}{ll} \text{MUD} & \left\{ \left[\begin{array}{ll} \multicolumn{2}{l}{\textit{õ-lid}} \\ \text{ST} & \boxed{s} \\ \text{TH} & \boxed{t} \end{array}\right], \left[\begin{array}{ll} \text{CASE} & \textit{part} \\ \text{NUM} & \textit{sg} \end{array}\right] \right\} \\ \\ \text{MPH} & \left\{ \left[\begin{array}{ll} \text{PH} & \boxed{s} \\ \text{PC} & \textit{0} \end{array}\right], \left[\begin{array}{ll} \text{PH} & \boxed{t} \\ \text{PC} & \textit{1} \end{array}\right], \left[\begin{array}{ll} \text{PH} & \textit{t} \\ \text{PC} & \textit{2} \end{array}\right] \right\} \end{array}\right]$$

Finally, we shall look at the nominative singular (*nokk, õpik, seminar*). Choosing a leaf type from each dimension, we get the result in (7), i.e. intersection of the general stem selection rule type with *n-sg-rr*, a rule type that is linked to both the $\boxed{\text{THEME}}$ and the $\boxed{\text{SFX}}$ dimensions, thereby trivially satisfying Online Type construction with respect to these dimensions. them.

(7) *n-sg-rr* & *grl-st-rr* ≡

$$\left[\begin{array}{ll} \text{MUD} & \left\{ \left[\begin{array}{ll} \text{ST} & \boxed{s} \\ \text{TH} & \textit{phon} \end{array}\right], \left[\begin{array}{ll} \text{CASE} & \textit{nom} \\ \text{NUM} & \textit{sg} \end{array}\right] \right\} \\ \\ \text{MPH} & \left\{ \left[\begin{array}{ll} \text{PH} & \boxed{s} \\ \text{PC} & \textit{0} \end{array}\right] \right\} \end{array}\right]$$

One may wonder what the result would be, if we had chosen instead the unification of *grl-st-rr* with the general theme (*theme-rr*) and singular suffixation (*grl-sg-rr*) rule types, which is indeed the description given in (5). In fact, this description per se happens to be compatible with the nominative singular. However, since (5) is in Paninian competition with the more specific rule in (7), its application will be preempted.[5]

Having seen how the proposed IbM theory of Estonian derives specific patterns, is is worth taking stock of what has been achieved: the approach we have taken is obviously holistic in that stem selection, theme selection and suffix selection recipes can only jointly pair function and form. The holistic nature of Estonian core cases is also revealed by the fact that the top-down organisation of the hierarchy is more form-driven, than content-driven. And we shall not forget that Paninian competition plays a crucial role in fixing specific form-function correspondences based on paradigmatic contrast, which must count as a systemic property as well. Despite all that it is clear that even seemingly opaque systems can be meaningfully decomposed in a theory that derives realisation rules from underspecified partial descriptions organised in a hierarchy of typed feature structures.

5 Conclusion

In this paper we have shown how Information-based Morphology can accomodate inflection systems lying at extreme ends of a gradient of morphological opacity by deploying either atomistic or holistic analyses.

[5] In IbM, preemption is performed by a closure operation on leaf types, enriching the more general description with the complement of the specific description. See Crysmann and Bonami (2016) and Crysmann (2017) for details.

The crucial contrast between the two proposed analyses is that rules for Swahili express 1 : 1 relations between morphs and partial property sets, while rules for Estonian express globally a word-level *m:n* relation between a sequence of exponents and a property set. We contend that these are necessary features of adequate analyses of these two systems. In the case of Swahili, no word-level constraint can capture the fact that the same affixes play double duty as subject and object markers — hence the word-level analysis proposed by Koenig (1999) is sub-optimal, and abstraction of realisation rules of sub-word relevance is crucial. At the other end of the spectrum, in Estonian, simultaneous introduction of all morphs is the formal rendering of the idea of 'gestalt exponence' (Blevins et al., 2016) — words are segmentable, but content is attributed to combinations of morphs rather than individual morphs. Note that adopting such a 'gestalt' view in no way precludes identifying generalisations across words where they are relevant. For instance, the fact that plural marking is always manifested at the right edge of the word in the Estonian dataset is captured by a general type linking the expression of plural to position 2 without constraining its shape. In this sense the approach is close in spirit to Berkeley Construction Grammar, where generalisations hold at variable levels of granularity.

Although these two analyses purposefully showcase (sub)systems that constitute polar opposites, nothing in the formal setup we assume entails that a system may contain only word-level or only morph-level rules: indeed, outside the domain of core cases, Finno-Ugric Estonian is rather of the agglutinative type. This opens up the possibility of capturing appropriately diverse combinations of opaque and transparent corners of an inflection systems, and hence helps provide a formally sound typological characterisation of exponence systems, rather than assume a 'one size fits all' view of morphological modeling that masks diversity. In particular, it is notable that the framework allows for the definition of a classical morpheme — a 1 : 1 association between a morph and a property set — where it is useful, without forcing its universal adoption, even in the analysis of the same system. In contrast to morpheme-based theories, this 1 : 1 relation does not enjoy any special formal status compared to $m : n$: it just happens to have a very simple and straightforward specification. Likewise, fully holistic analyses can and should be used when appropriate, but this does not preclude the explicit formulation of partial generalizations on the distribution of exponents.

It is worth noting that the ability to address the whole spectrum of morphological opacity is intimately tied to two central design features of Information-based Morphology: the recognition of positionally-indexed morphs, and the use of monotonous multiple inheritance hierarchies of rules of exponence. These design properties constitute on of the central innovations of IbM, and set it apart both from previous HPSG approaches to inflection and from other inferential-realisational frameworks. Although they were initially introduced to address the conceptually separate issue of variable morphotactics, these two ingredients are key to allowing a view of exponence as a partially underspecified description of $m : n$ relations between form and content, within which atomistic and holistic views of the world turn out to be compatible with each other.

References

Ackerman, Farrell and Bonami, Olivier. inpress. Systemic polyfunctionality and morphology-syntax interdependencies. In Andrew Hippisley and Nikolas Gisborne (eds.), *Defaults in Morphological Theory*, Oxford: Oxford University Press.

Ackerman, Farrell and Malouf, Robert. 2013. Morphological organization: the low conditional entropy conjecture. *Language* 89, 429–464.

Anderson, Stephen R. 1992. *A-Morphous Morphology*. Cambridge: Cambridge University Press.

Aronoff, Mark. 1994. *Morphology by itself*. Cambridge: MIT Press.

Blevins, James P. 2005. Word-based declensions in Estonian. In Geert E. Booij and Jaap van Marle (eds.), *Yearbook of Morphology 2005*, pages 1–25, Springer.

Blevins, James P. 2006. Word-based morphology. *Journal of Linguistics* 42, 531–573.

Blevins, James P. 2013. Word-based morphology from Aristotle to modern WP. In Keith Allan (ed.), *Oxford Handbook of the History of Linguistics*, pages 375–396, Oxford: Oxford University Press.

Blevins, James P., Ackerman, Farrell and Malouf, Robert. 2016. Morphology as an adaptive discriminative system. In Heidi Harley and Daniel Siddiqi (eds.), *Morphological Metatheory*, pages 271–302, Amsterdam: John Benjamins.

Bonami, Olivier and Beniamine, Sarah. 2016. Joint predictiveness in inflectional paradigms. *Word Structure* 9(2), 156–182.

Bonami, Olivier and Boyé, Gilles. 2006. Deriving inflectional irregularity. In *Proceedings of the 13th International Conference on HPSG*, pages 39–59, Stanford: CSLI Publications.

Bonami, Olivier and Crysmann, Berthold. 2013. Morphotactics in an information-based model of realisational morphology. In Stefan Müller (ed.), *Proceedings of HPSG 2013*, pages 27–47, Stanford: CSLI Publications.

Bonami, Olivier and Stump, Gregory T. 2016. Paradigm Function Morphology. In Andrew Hippisley and Gregory T. Stump (eds.), *Cambridge Handbook of Morphology*, pages 449–481, Cambridge: Cambridge University Press.

Booij, Geert. 2010. *Construction morphology*. Oxford: Oxford University Press.

Brown, Dunstan and Hippisley, Andrew. 2012. *Network Morphology: a defaults based theory of word structure*. Cambridge: Cambridge University Press.

Crysmann, Berthold. 2003. *Constraint-based Coanalysis. Portuguese Cliticisation and Morphology–Syntax Interaction in HPSG*. Saarbrücken Dissertations in Computational Linguistics and Language Technology, No. 15, Saarbrücken: Computational Linguistics, Saarland University and DFKI LT Lab.

Crysmann, Berthold. 2014. Deconstructing Exhuberant Exponence in Batsbi. Paper presented at the 16th International Morphology Meeting, Budapest, May 29–June 1.

Crysmann, Berthold. 2017. Inferential-realisational morphology without rule blocks: an information-based approach. In Nikolas Gisborne and Andrew Hippisley (eds.), *Defaults in Morphological Theory*, pages 182–213, Oxford: Oxford University Press.

Crysmann, Berthold and Bonami, Olivier. 2012. Establishing order in type-based realisational morphology. In Stefan Müller (ed.), *Proceedings of HPSG 2012*, pages 123–143, Stanford: CSLI publications.

Crysmann, Berthold and Bonami, Olivier. 2016. Variable morphotactics in Information-based Morphology. *Journal of Linguistics* 52(2), 311–374.

Emerson, Guy and Copestake, Ann. 2015. Lacking Integrity: HPSG as a Morphosyntactic Theory. In Stefan Müller (ed.), *Proceedings of the 22nd International Conference on Head-Driven Phrase Structure Grammar, Nanyang Technological University (NTU), Singapore*, pages 75–95, Stanford, CA: CSLI Publications.

Erjavec, Tomaž. 1994. Formalizing realizational morphology in typed feature structures. In Gosse Bouma and Gertjan van Noord (eds.), *Proceedings of the fourth CLIN meeting*, pages 47–58, Groningen.

Ford, Alan, Singh, Rajendra and Martohardjono, Gita. 1997. *Pace Panini*. Berlin: Peter Lang.

Gurevich, Olga I. 2006. *Construction Morphology: the Georgian version*. Ph. D.thesis, University of California, Berkeley.

Harris, Alice C. 2009. Exuberant exponence in Batsbi. *Natural Language and Linguistic Theory* 27, 267–303.

Harris, Alice C. 2012. Meaning resides in fully inflected forms: The Georgian "unwillingness" construction. In *Proceedings of the 8th International Morphology Meeting*.

Hockett, Charles F. 1947. Problems of Morphemic Analysis. *Language* 23, 321–343.

Koenig, Jean-Pierre. 1999. *Lexical relations*. Stanford: CSLI Publications.

Koenig, Jean-Pierre and Jurafsky, Daniel. 1994. Type underspecification and on-line type construction. In *Proceedings of WCCFL XIII*, pages 270–285.

Krieger, Hans-Ullrich, Pirker, Hannes and Nerbonne, John. 1993. Feature-based allomorphy. In *Proceedings of the thirty-first meeting of the Association for Computational Linguistics*, pages 140–147.

Pollard, Carl and Sag, Ivan A. 1987. *Information-based syntax and semantics*. Stanford: CSLI Publications.

Pollard, Carl and Sag, Ivan A. 1994. *Head-driven Phrase Structure Grammar*. Stanford: CSLI Publications; The University of Chicago Press.

Sag, Ivan A. 2012. Sign-Based Construction Grammar: An Informal Synopsis. In Hans Boas and Ivan A. Sag (eds.), *Sign-Based Construction Grammar*, pages 69–202, Stanford: CSLI Publications.

Sims, Andrea. 2015. *Inflectional defectiveness*. Cambridge: Cambridge University Press.

Spencer, Andrew and Stump, Gregory T. 2013. Hungarian pronominal case and the dichotomy of content and form in inflectional morphology. *Natural Language and Linguistic Theory* 31, 1207–1248.

Stump, Gregory T. 1993. Position classes and morphological theory. In Geert E. Booij and Jaap van Marle (eds.), *Yearbook of Morphology 1992*, pages 129–180, Kluwer.

Stump, Gregory T. 2001. *Inflectional Morphology. A Theory of Paradigm Structure*. Cambridge: Cambridge University Press.

Stump, Gregory T. 2016. *Inflectional paradigms*. Cambridge: Cambridge University Press.

Wurzel, Wolfgang Ulrich. 1984. *Flexionsmorphologie und Natürlichkeit. Ein Beitrag zur morphologischen Theoriebildung*. Berlin: Akademie-Verlag, translated as Wurzel (1989).

Wurzel, Wolfgang Ulrich. 1989. *Inflectional Morphology and Naturalness*. Dordrecht: Kluwer.