The information-theoretic turn

James P. Blevins

University of Cambridge, UK

Over the past decade, information theory has been applied to the analysis of a successively broader range of morphological phenomena. Interestingly, this tradition has arisen independently of the linguistic applications of information theory dating from the 1950's. Instead, the point of origin for current work lies in a series of studies of morphological processing in which Kostić and associates develop a statistical notion of 'morphological information' based on 'uncertainty' and 'uncertainty reduction'. From these initial studies, analyses based on statistical notions of information have been applied to general problems of morphological description and typological classification, leading to a formal rehabilitation of the complex system perspective of traditional WP models.

Key words: *inflectional paradigms, word and paradigm morphology, information theory, morphological processing*

1 MORPHOLOGY AS A COMPLEX SYSTEM

From the outset of the western grammatical tradition, morphology was regarded as what is, in modern terms, called a 'complex system'. Early debates in the classical world addressed the nature of the observed complexity. Advocates of 'analogy' saw morphological systems as exhibiting 'organized complexity', in essentially the sense of Weaver (1948), emerging from regular patterns of accidence and cooccurrence. Proponents of 'anomaly' viewed the irregularity and randomness of morphological systems as exhibiting 'disorganized complexity'. Both conceptions were incorporated in a qualified form in what became the classical word and paradigm (WP) model. A high degree of system-level congruence was associated with inflectional patterns, while irregular and random variation was treated as characteristic of lexical inventories and, to a lesser extent, of word-formation. The organized complexity of inflectional systems lent itself to economical factorization. A collection of exemplary items enumerated the distinctive patterns of inflection in a language, while a set of diagnostic forms associated non-exemplary item with one of these patterns. The disorganized complexity exhibited by the word stock of a language and word formation processes were less amenable to economical description and were listed in grammars and dictionaries, marked for any irregular or exceptional properties.

Corresponding author: jpb39@cam.ac.uk

The classical description of morphological complexity survives in this exemplar-based form until the present day, remaining, as Hockett (1954: 386) remarks, "the traditional framework for the discussion of Latin, Greek, Sanskrit, and a good many more modern familiar languages". The cornerstone of this framework is a recognition that the organized complexity of inflectional systems derives from implicational relations. For Matthews (1991: 197), "[t]he most general insight" expressed in the traditional model "is that one inflection tends to predict another". Predictive relations are a precondition of the factorization into exemplary patterns and diagnostic forms. It is the interdependency of choices that allows items to be represented by diagnostic forms; if choices were independent, items would have to be represented by a full set of forms, since no forms would be predictable. The role of interdependencies gives rise in turn to the conception of an integrated morphological system that is more than a collection of individual forms and patterns. Weaver's characterization of organized complexity arising from "a sizable number of factors which are interrelated into an organic whole" (p. 539) thus contains an echo of the traditional notion that the words assigned to the inflectional paradigms and classes "are themselves the parts within a complex whole" (Matthews, 1991: 204).

The survival of the traditional model owed much to its practical utility. As acknowledged by Hockett (1954: 386), factoring morphological systems into patterns and diagnostic forms was particularly useful for describing flectional languages. The same factorization was of proven pedagogical value, as Matthews (1991) stresses.

> 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: 187f.)

Yet although a complex system perspective was clearly implicit in the descriptive practices of traditional accounts, the underlying conception was never explicitly formalized. Interdependencies were exhibited by the forms of exemplary items, or else encapsulated in 'inflection classes' that these items were taken to represent. Neither strategy provided a means of abstracting implicational dependencies from the sets of forms or exponents that exhibited them. The lack of an explicit model largely reflected the descriptive and pedagogical interests of traditional grammarians. However, the absence of an abstract model fostered the impression that classical WP approaches were of essentially practical value and too closely tuned to the properties of particular languages to be of general theoretical relevance. Within the grammatical traditions that grew up around individual languages and language families, there was often a consensus regarding issues such as the number of classes in a language or the choice of diagnostic forms. But there was no procedure for selecting diagnostic forms or for determining the informativeness of single forms or sets of forms. The

problem of selecting principal parts (or 'leading forms') came to be regarded by some as intrinsic to, or possibly even recalcitrant within, traditional models.

> One objection to the Priscianic model ... was that the choice
> of leading form was inherently arbitrary: the theory creates a
> problem which it is then unable, or only partly able, to resolve.
> (Matthews, 1972: 74)

The same fundamental issue recurred in a different guise in the problem of distinguishing valid from spurious analogies. Correspondences within an inflectional system were traditionally described in terms of analogical patterns. Following particularly Paul (1920), morphological analogy tended to be expressed in the form of 'four-part proportional analogies' of the form $a : b = c : x$. This deductive schema uses the relation between a and b to establish a pattern that is extended from a given c to identify an unknown x. But just as the selection of principal parts depends on the diagnostic value of a form, the validity of a deduction depends on the reliability of the relation between a and b.[1] Both relations depend in turn on a notion of 'informativeness' that measures how reliably one element identifies another. Yet informativeness is precisely what remains implicit in the traditional model.

2 REALIZATION MODELS

The exemplar-based structure of a traditional WP model led, in short, to an impasse. At the point where Hockett (1954) returned the model to the attention of the theoretical community, there was no formulation of the model that was stated independently of the description of a specific language. Robins (1959) subsequently sketched out the general architecture of a WP model and described a range of patterns that seemed particularly amenable to analysis in WP terms, but stopped short of formalizing any of the patterns explicitly. A general strategy for formalizing this tradition was first mentioned briefly in Chomsky (1965) and developed in more detail by Matthews (1965). Both accounts modelled the inflectional component of a WP grammar in terms of sets of morphosyntactic properties and rules that 'realize' them.

> The present paper is intended to supply a part of this formulation.
> It is restricted to inflectional problems alone: to be more precise,
> it deals with that subsection of the grammar (we will call it the
> inflectional component) which assigns a realization, or various
> alternative realizations, to each grammatical word. (Matthews,
> 1965: 142)

In the specific model that Matthews proposed, the paradigm of an item is represented abstractly by a set of cells, each consisting of a bundle of properties

---

1   See Kuryłowicz (1964) for a philological approach to the problem of 'founding' analogies.

(or 'features'). The lexical entry of an item specifies stem forms on which inflected forms are based. A set of REALIZATION rules 'interpret' the features in a cell by applying operations to a stem or to a form derived from the stem. A set of such rules realize the surface form associated with each of the cells in the paradigm of an item.

This process can be illustrated by the realization of the Finnish forms in Table 1. These forms consist of a lexical stem, *talo*, a plural marker *–i* (realized as *–j* intervocalically), and genitive and partitive case markers. The ending *–a* can be analyzed as a number-neutral partitive marker, while *–en* is a genitive plural marker. The ending *–n* can be treated either as a genitive singular marker or as a number-neutral genitive. For the sake of illustration, let us assume the second of these analyses.

Table 1. Partial paradigm of Finnish *talo* 'house' (Karlsson, 1999: 249).

	Sing	Plu
Gen	talon	talojen
Part	taloa	taloja

The rules that spell out partitive and genitive forms are given in Figure 1 below. The observation that the plural marker occurs closer to the stem than the case endings is expressed by organizing the rules into two ordered 'blocks'. The first block, A, contains the plural marker and the second block, B, contains the case exponents.

$$A: \begin{bmatrix} \text{PLU} \\ X \end{bmatrix} \rightarrow X + i$$

$$B: \begin{bmatrix} \text{PART} \\ X \end{bmatrix} \rightarrow X + a \quad \begin{bmatrix} \text{GEN} \\ X \end{bmatrix} \rightarrow X + n \quad \begin{bmatrix} \text{PLU} \\ \text{GEN} \\ X \end{bmatrix} \rightarrow X + en$$

Figure 1. Plural, partitive and genitive realization rules.

The application of these rules is illustrated in Figure 2. The partial feature bundles represent the features that distinguish the four genitive and partitive forms. The row below the bundles identifies the lexical stem, *talo* that underlies the surface forms. The bottom rows then exhibit the effect of the rules in 'blocks' A and B.

	$\begin{bmatrix} \text{SING} \\ \text{GEN} \end{bmatrix}$	$\begin{bmatrix} \text{SING} \\ \text{PART} \end{bmatrix}$	$\begin{bmatrix} \text{PLU} \\ \text{GEN} \end{bmatrix}$	$\begin{bmatrix} \text{PLU} \\ \text{PART} \end{bmatrix}$
Stem	talo	talo	talo	talo
'Block' A	talo	talo	taloi	taloi
'Block' B	talon	taloa	talojen	taloja

Figure 2. Spell-out of genitive and partitive forms of *talo.*

The central architectural property of this model is a clean separation between the principles that define property bundles and the rules that interpret those bundles.[2] Although the initial presentation in Matthews (1965) sketches out just a few 'proof of concept' analyses, the descriptive scope of the realizational model is demonstrated in the analysis of the Latin conjugational system in Matthews (1972). The monograph-length elaborations of realizational models in Anderson (1992), Aronoff (1994) and Stump (2001) subsequently established the relevance of this approach to the ananlysis of a larger typological sample of languages.

Yet this model formalizes merely one aspect of a traditional WP model. A guiding – but empirically problematic – assumption of the 'morphemic' model developed within the Bloomfieldian tradition is that the structure and meaning of complex forms could be defined by combining individually meaningful parts. The more flexible form-meaning relation of a realizational model accommodated a wide range of morphological patterns that exhibited mismatches between 'units of form' and 'units of meaning'. However, realizational models retained an essentially Bloomfieldian conception of morphological analysis as applying to words in isolation. This conception is stated explicitly in Anderson's (1992) definition of a paradigm.

> We can now define the PARADIGM of a lexical item in terms of its lexical stem set: an item's paradigm is 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 [i.e., realization rules JPB] of the language. (Anderson, 1992: 134)

A similar conception is common to all realizational models. The forms of a morphological system are, for the most part, defined independently, each representing a separate 'realizational pathway' from feature bundles to (morphotactically unstructured) surface forms. Some realizational models incorporated rules that would relate the realization of multiple forms. Initially proposed as "parasitic derivations" by Matthews (1972: 185) (subsequently as "morphological transformations" in Matthews (1991: 193)) and as 'rules of referral' in Zwicky (1985) and Stump (1993), these devices were invoked principally to capture patterns of syncretism.[3] As such, these devices linked the output, or realizational pathway of specified forms. However, the existence of individual correspondences between cells or derivations did not fundamentally shift the locus of realizational models from individual forms. They remained "that subsection of the grammar ... which assigns a realization, or various alternative realizations, to each grammatical word" (Matthews, 1965: 142).

---

2 A property that is later aptly termed 'The Separation Hypothesis' by Beard (1995).

3 Morphological transformations are proposed in the first edition of Matthews' *Morphology*, initially published in 1974; the page reference above is to the more widely available second edition.

3 IMPLICATIONAL STRUCTURE

By avoiding the agglutinative bias of Bloomfieldian approaches, realizational models developed a more flexible analysis of word structure. They did not, however, directly analyze the structure of morphological systems except insofar as this structure was derivative of properties of individual forms. Despite occasionally being termed 'extended WP models', realizational approaches were not designed to address the challenges that arose in formalizing the complex system perspective of a traditional WP model. In particular, the organized complexity that underlies the traditional factorization of morphological systems plays no role in a realizational model. A morphological system in which the realization of each cell of a paradigm is fully independent of the realization of every other cell is no less amenable to a realizational analysis than a system in which the choices are interdependent.

As observed by Carstairs (1983) in connection what he termed 'Paradigm Economy', a high degree of interdependence is characteristic and plausibly definitional of inflectional system. Some patterns of cooccurrence can be captured in a realizational model by cross-indexing cells and realization rules. Inflection classes are, following Chomsky (1965: 171) and Matthews (1965), often represented by 'class features' in realizational models. For example, associating the cells of the second declension noun DYNJA 'melon' with the index [DC II], allows this index to be 'spelled out' by the realization rules that introduce the inflectional endings in Table 2.

Table 2. Paradigm of Russian DYNJA 'melon' (Timberlake, 2004).

	Sing	Plu
Nom	dynja	dynji
Gen	dynji	dyn'
Dat	dyne	dynjam
Acc	dynju	dynji
Inst	dynej	dynjami
Loc	dynje	dynjax

Cross-indexing strategies are, however, less suitable for the representation of asymmetrical or directional implications. In addition to predicting a common stem *dyn'* and a set of exponents, the forms in Table 2 also imply information about each other. The traditional nominative singular citation form *dynja* is a highly reliable predictor of class. It is thus a useful principal part, but not uniquely so; the accusative singular *dynju* is equally informative. In contrast, the locative singular *dynje* could belong to the first declension, and the locative plural *dynjax* could belong to any of the three main declensions. To say that a form is diagnostic of class amounts to saying that it reliably predicts the form variation in other cells. Thus a diagnostic form such as *dynju* provides the basis for valid analogies: *dynju: dynja = njanju: x* defines the nominative singular form *njanja* of NJANJA 'nurse'. At the other extreme, analogies with the locative plural *dynjax* are maximally unreliable.

The information that forms provide about other forms is not a quirk of Russian but reflects the general implicational structure of inflectional systems. Even tighter patterns of mutual implication operate over Estonian declensions. For example, nearly every form in Table 3 identifies the full paradigm of PUKK 'trestle'. The deductions are set out in descriptions of Estonian ranging from schoolbooks to academy grammars, so they can be concisely summarized here. In a paradigm containing a strong vowel-final partitive singular such as *pukki*, the remaining singular forms are fully predictable. The short illative singular *pukki* is identical to the partitive, the nominative singular *pukk* is related by truncation, and the genitive singular *puki* is be a weak variant of the partitive singular. The 'stem partitive plural' *pukki* also reflects the default *i↔e* 'exchange pattern' for nouns with theme vowel *–i* (or *–e*). The long partitive plural *pukkisid* and genitive plural *pukkide* follow the general first declension patterns, adding *–sid* and *–de* to the partitive singular base. The nominative plural *pukid* follows the case-neutral declensional pattern in which *–d* is added to the genitive singular base. The remaining 'semantic' (or 'local') case forms are predictable from the genitive singular and plural forms. Since semantic case endings are invariant, stem selection represents the sole locus of variation.

Table 3. Paradigm of Estonian PUKK 'trestle' (Erelt, 2006).

	Sing	Plu
Nominative	pukk	pukid
Genitive	puki	pukkide
Partitive	pukki	pukkisid
Illa2/Part2	pukki	pukke
Illative	pukisse	pukkidesse
Inessive	pukis	pukkides
Elative	pukist	pukkidest
Allative	pukile	pukkidele
Adessive	pukil	pukkidel
Ablative	pukilt	pukkidelt
Translative	pukiks	pukkideks
Terminative	pukini	pukkideni
Essive	pukina	pukkidena
Abessive	pukita	pukkideta
Comitative	pukiga	pukkidega

Keying deductions to the partitive singular follows the traditional practice of identifying class from a distinguished 'diagnostic form'. However, this overstates the informational asymmetry between partitive singulars and other forms. A genitive singular such as *puki* is also diagnostic of a weakening declension. Although morphologically weak, the stem *puk* is phonetically long

(or 'second quantity' Q2). A Q2 genitive singular can only occur in a paradigm with an overlong ('third quantity' Q3) nominative and partitive singular. The genitive plural is also diagnostic, since the removal of the ending –*de* identifies the partitive singular base. Moreover, since all of the semantic case endings are invariant, each of the semantic case forms identifies the corresponding genitives. Hence, nearly every form is either diagnostic of class or identifies a form that is diagnostic. Other declension and conjugation classes exhibit other characteristic patterns of interdependency.[4]

A realizational analysis can describe these patterns by means of referral rules that cross-reference the realization of interrelated cells. However, the realizational goal of defining "the complete set of surface word forms" of an item that this achieves is subtly different from the objective of a traditional analysis. A classical WP model also aims to describe the deductive structure of a set of forms by characterizing the predictive value of individual forms and the degree of cohesion exhibited by the set.

Although realizational models do not represent deductive structure directly, they do not wholly dispense with implication, so much as shift the primary locus to the relation between feature bundles and the individual forms that realize them. This strategy exploits the closed, relatively uniform feature space of an inflectional system. Within this closed space, the features of paradigm cells can be defined independently of the forms that realize them. From the class of an open-class item, it is usually possible to determine the features that are distinctive for that class and predict the number of forms of the item. Apart from irregular items, paradigms are broadly comparable in size and structure within a word or inflection class.

The fact that realizational models are so finely tuned to the structure of inflectional systems creates difficulties in extending them to other types of patterns. In particular, realizational models are less applicable to the variable structure exhibited by 'families' of derivational forms. Processes that can change word class, valence or other intrinsic properties usually do not define a finite set of forms within a uniform feature space. From just the word class or inflection class of an item, one cannot in general predict the number and type of derivational formations in which it occurs. Given a list of derivational processes active in a language, it is of course possible to assign a uniform family of 'potential' forms to all of the members of a word class. Yet the uniformity achieved is deceptive, because it collapses a critical distinction between those forms that are established in a language and those that are merely possible in principle. The point is perhaps more obvious in connection with compounds. Of the infinitely many possible noun compounds in a language such as English only a comparatively small number are established, and a speaker cannot predict the set of established compounds containing an item from the item itself.

---

4 One description of the Estonian inflectional system is outlined in Blevins (2007, 2008). More traditional accounts, which differ in where they draw the line between classes and subclasses, are presented in Erelt *et al.* (1995), Mürk (1997), Saagpakk (2000), Erelt *et al.* (2000), Erelt (2003) and Erelt (2006).

Implicational models also exploit the structure of inflectional systems, but in a manner that generalizes more straightforwardly to other types of morphological patterns. The same closed space that permits the feature-form separation in a realizational model makes inflectional processes highly interpredictable. This interpredictability is reflected in the ways that morphological systems are characterized. Individual inflected forms are seldom described as established on their own. The availability of a form correlates instead with the productivity of a whole inflectional pattern. Conversely, notions like 'morphological gap', 'suppletion', and even, to a significant degree, 'syncretism' are mainly or exclusively applied to inflectional paradigms. Families of derivational forms do not usually have 'gaps' because they do not define implicational relations over a closed and uniform space that give rise to definite expectations about the existence of specific forms, even where the shape would be predictable. Suppletion likewise arises only where there are specific assumptions about the shape of particular members of a form set. Inflectional paradigms generate strong assumptions of this nature, but derivational families do not. Syncretism presupposes a similar structure, as syncretic patterns again imply the existence of an independent grid within which to locate the cells and/or forms.

Although implicational relations were traditionally recognized as most reliable in inflectional systems, they were also taken to be relevant, in a weaker form, in the domain of derivation. Indeed, the fundamental contrast between the organized complexity of inflection and the disorganized complexity of word formation could be attributed largely to the relative strength of implicational relations in the two morphological domains. All that is required was a means of formalizing implicational organization to the same precision as morphemic or realizational structure.

The Paradigm Structure Conditions of Wurzel (1984) represent the most systematic approach to the problem of formalizing the organization of inflection classes.

> Observation of complicated paradigms shows that implicative relations do not only obtain between one basic inflexional form, either lexical (*sobaka* [Russian 'dog']) or non-lexical (*Männer* [German 'men']), and all the other inflexional forms, but exist throughout the whole paradigm: all paradigms (apart from suppletive cases) are structured on the basis of implicative patterns which go beyond the individual word, patterns of varying complexity. Of particular complexity in this respect is, for example, the implicative pattern of the *i*-declension in Latin: /im/ in the A.Sg. ⊃ /ī/ in the Abl.Sg. ⊃ /īs/ in the A.Pl. ⊃ /ium/ in the G.Pl. ⊃ ... From the A.Sg. form we can derive, via a number of steps, all other forms, but not vice-versa. Since the implicative patterns determine the structure of the paradigms of a language we call them implicative PARADIGM STRUCTURE CONDITIONS (PSCs) (Wurzel, 1984: 208).

By capturing "implicative patterns which go beyond the individual word", PSCs described system-level patterns. Although this passage is phrased in terms of a process that "can derive, via a number of steps, all other forms", PSCs can also be interpreted non-derivationally. On this interpretation, PSCs function as 'paradigm admissibility conditions' that determine the class of wellformed paradigms of a language.

Yet the use of logical implication to express dependencies severely constrains the descriptive scope of PSCs. The use of the material conditional '⊃' to model predictive relations restricts the application of PSCs to exceptionless patterns. Some inflectional patterns are exceptionless, or nearly so. In Russian, a nominative singular in –a reliably predicts an accusative singular in –u (for all but indeclinable nouns). In Estonian, the nominative plural always corresponds to the genitive singular plus a final –d. However, other informative patterns are more probabilistic. A locative singular in –e does not identify the class of a second declension noun in Russian but reduces the space of alternatives by eliminating the third declension. Conversely a genitive singular in –i excludes the first declension. Hence a locative singular in –e and a genitive singular in –i are both partially informative; together they are as informative as a nominative singular in –a. A general analysis of the implicational structure of an inflectional system needs to be able to represent the partial information expressed by forms, and also the ways in which it can be combined.

To a large degree, the goal of representing exceptionless implications, like that of selecting fully diagnostic principal parts reflects pedagogical idealizations. Logical implication is manifestly the wrong tool to measure the predominantly statistical structure of inflectional systems. It is even less suited to the analysis of derivational patterns. This point is implicit in Matthew's (1991) earlier claim that "[t]he most general insight" of the traditional model was the recognition "that one inflection tends to predict another". However, this insight could only be fully exploited within a tradition that was already sensitive to the statistical properties of language.

4 INFLECTIONAL INFORMATION AND LOAD

Exactly such a tradition had grown up in the psycholinguistic literature on word processing. Within this literature, two branches could be distinguished, correlating fairly closely with the traditional division between inflection and derivation.

One line of research had investigated correlations between response latencies in visual decision tasks and various frequency measures related to inflected forms. The earliest studies established that the surface frequency of an inflected form in a corpus was negatively correlated with response latencies (Taft, 1979). The base frequency of a word (the summed frequency of its inflected variants) was later found to exhibit a positive correlation with response latencies (Baayen, Lieber, & Schreuder, 1997). The logarithm of the ratio between these measures (surface frequency and base frequency) again exhibited

a negative correlation (Hay, 2001). These and other studies confirmed the value of token frequencies – or other distributional factors that correlated with these frequencies – on predicting latencies in the processing of inflected forms.

Following Schreuder and Baayen (1997), a second line of research investigated the relevance of type frequency to response latencies in visual decision. A range of studies established that an increase in number of semantically transparent items in the morphological family of a form facilitated processing of the form (de Jong, 2002; Moscoso del Prado Martín, 2003; Moscoso del Prado Martín, Bertram, Häikiö, Schreuder, & Baayen, 2004a).

The design of these studies had not been guided by the assumptions of a traditional model, though the results were often compatible with this model. From a traditional perspective, it was unsurprising that type frequency effects would be more transparent in the domain of derivation. The relatively uniform structure of inflectional systems provides less scope for variation in lemma size for items belonging to a given word or inflection class. Hence distributional factors, such as token frequency (or 'contextual distinctiveness' in the sense of McDonald and Schillcock (2001)) is the primary locus of inflectional variation. In contrast, as shown by resources such as CELEX (Baayen, Piepenbrock, & van Rijn, 1995), the 'families' of forms associated with distinct derivational bases can vary by up to two orders of magnitude.

What was somewhat more surprising was that models proposed to account for response latencies also clarified the factors that underlay the descriptive success of traditional approaches. The central insight was developed in a series of studies that measured the information carried by case-number inflections in Serbian (Kostić, 1991, 1995; Kostić, Marković, & Baucal, 2003). These studies proceed from a surprisal-based perspective in which the information carried by an item, I_x, corresponds to the negative log of its probability $p(x)$ (i.e., the less likely x is, the more informative it is).

$$I_x = - \log_2 p(x)$$

However, they refine this measure by weighting the probabilities of items for the number of functions and meanings they express. Thus, for example, the form in –e realizes the dative and locative singular in Table 2. The dative in turn functions as an indirect object, as an object of adjectives, as a subject of impersonal and infinitive constructions, etc., while the locative has its own characteristic range of functions. The inflectional information of variants in -e, I_e, is calculated by first dividing the frequency of these variants by the number of meanings and functions R_e, to obtain the ratio $(p(e)/R_e)$, the average probability per meaning/function. This value is then weighted by the sum of the ratios for each of the exponents in Table 2, $(\Sigma_e p(e)/R_e)$ and the result log-transformed to give an information value in bits.

$$I_e = - \log_2 \left(\frac{p(e)/R_e}{\Sigma_e p(e)/R_e} \right)$$

Kostić et al. (2003) show that inflectional information correlates positively with response latencies and that the processing cost of an inflected variant is predicted by its frequency and functional load. Subsequent refinements consider other dimensions of variation in order to isolate the factors that contribute to processing cost.

These initial studies focussed on inflectional variants. To arrive at a unified measure of type and token frequency effects, Moscoso del Prado Martín, Kostić, and Baayen (2004b) proposed a generalized model with two central innovations. The first was a notion of 'paradigm' that encompassed derivational families as well as inflectional paradigms. The second was the use of a standard entropy measure (Shannon 1948) to calculate a frequency-weighted measure of morphological information. In the general case, where X is a random variable with values $x_1 \ldots x_n \in X$, the entropy of X, $H(X)$, is just the frequency-weighted sum of the surprisal of the values in X.

$$H(X) = - \sum_{x \in X} p(x) \log_2 p(x)$$

Moscoso del Prado Martín et al. (2004b) observed that paradigms could be regarded as random variables that took forms as values. The entropy of a paradigm could then be defined as below. Each form x in $\mathcal{P}$ has a probability of occurrence of $p(x|\mathcal{P}) \simeq F(x)/F(\mathcal{P})$, where $F(x)$ the surface frequency of x and $F(\mathcal{P})$ is the base frequency of P (the summed frequency of the forms in the paradigm).

$$H(\mathcal{P}) = - \sum_{x \in \mathcal{P}} p(x|\mathcal{P}) \log_2 p(x|\mathcal{P})$$

These generalizations allowed Moscoso del Prado Martín et al. (2004b) to subsume the family size information that correlates with type-frequency effects and the inflectional information that correlates with token-frequency effects under a single measure, which they termed the 'information residual' of a word.

Later studies, including Baayen, Feldman, and Schreuder (2006) and Baayen, Wurm, and Aycock (2008) suggested that these effects should in fact be kept apart. However, the usefulness of information-theoretic shift outlined in Moscoso del Prado Martín et al. (2004b) was confirmed in a range of subsequent studies, which are summarized and extended in Milin, Kuperman, Kostić, and Baayen (2009b). Perhaps the most striking result involved the relative entropy effect reported in Milin, Filipović Djurdjević, and Moscoso del Prado Martín (2009a), which showed that response latencies were positively correlated with the degree of divergence between the probability distribution of an inflected variant in the paradigm of an item and the probability distribution of the variant within the inflection class to which the item belongs.

5 INFORMATION-THEORETIC WP

It was the generalized information-theoretic conception of 'morphological information' developed in these processing studies that suggested a solution to the traditional problem of modelling implicational structure. The complexity of a morphological system could be represented in terms of the entropy of the system. Implicational relations could then be expressed in terms of entropy reduction. This perspective also provided a natural 'economy' measure for morphological systems, along with tools for measuring the complexity of a system or even comparing systems, as first noted in a purely linguistic context in Ackerman, Blevins, and Malouf (2009).

5.1 Morphological variation as uncertainty

To apply information theory to the analysis of inflectional systems, it is useful to refine slightly the perspective outlined above. Rather than regarding paradigms as random variables that take variants as values, one can treat each individual cell C as a random variable that takes as values the exponence patterns $x_1 \ldots x_n \in C$.[5]

$$H(C) = - \sum_{x \in C} p(x) \log_2 p(x)$$

The entropy of a cell is determined by the number of distinct exponence patterns and their distribution. For illustration, consider the schematic description of the Russian declensional system in Table 4. Each of the cells N.Sg ... L.Pl will be realized by a set of patterns of exponence. The genitive singular cell G.Sg is realized by the patterns {Xa, Xi}, the dative singular cell D.Sg by the patterns {Xu, Xe, Xi}, etc. Treating realizations as patterns rather than exponents permits a description of nonaffixal patterns (such as the grade alternations exhibited in the paradigm of Estonian PUKK in Table 3), though the patterns considered below are all affixal.

Table 4. Structure of Russian declensional paradigms.

	Sing	Plu
Nom	N.Sg	N.Pl
Gen	G.Sg	G.Pl
Dat	D.Sg	D.Pl
Acc	A.Sg	A.Pl
Inst	I.Sg	I.Pl
Loc	L.Sg	L.Pl

A worst-case estimation of cell entropy can be obtained from a grammar that lists all of the distinct patterns of exponence that realize each cell. The

5 Though Bonami and Luís (in press, 2013) argue that traditional cells may not always correspond to random variables, particularly in cases where individual exponents exhibit patterns of interdependence, so that forms do not instantiate single patterns.

information available from grammars is limited, in that they do not provide information about the distribution of inflectional variants. Hence when working from these sources, the frequency of distinct patterns cannot be estimated and there is no principled alternative to treating patterns as equiprobable. In this limiting case, the probabilities 'cancel out' and cell entropy reduces to the $\log_2$ of the number of patterns $|C|$.

$$H\,(C) = \log_2\,(|C|)$$

The uncertainty measured by entropy increases as a function of the number of alternatives and the uniformity of their distribution (since both factors make it more difficult to guess the value of a random variable). Hence the cell entropy determined by a grammar description is of value in defining the entropy ceiling for a cell.

Since a paradigm is a set of cells, the entropy of the paradigm is determined by the entropies of those cells. However, paradigm entropy is equal to the sum of the cell entropies only in the case where the cells are statistically independent, i.e., where there is no implicational relation between cells. This is precisely the case where a traditional factorization of paradigms is inapplicable. Hence, information-theoretic formalizations of traditional WP models implicitly assume what might be termed 'the joint entropy conjecture', according to which the entropy of a paradigm is always less than the sum of the entropies of the cells. Let P be a paradigm with the cells $C_1, \ldots, C_n$, which correspond to the sets of the exponent patterns $C_1, \ldots, C_n$. Then H (P), the entropy of P, is defined as the joint entropy of $H\,(C_1, \ldots, C_n)$:

$$H(C_1,\ldots,C_n) = -\sum_{x_1 \in C_1} \ldots \sum_{x_n \in C_n} p(x_1,\ldots,x_n)\log_2 p(x_1,\ldots,x_n)$$

where this joint entropy is strictly less than the sum of the entropies of the cells:

$$H(C_1,\ldots,C_n) < H(C_1) + \ldots + H(C_n)$$

5.2 Implicational structure as uncertainty reduction

Entropy measures the uncertainty associated with a cell or paradigm. Once an uncertainty estimation has been obtained, implicational structure can be modelled by uncertainty reduction. Consider first the application to proportional analogies. These are conventionally expressed in a shorthand $a : b = c : x$ that uses simple forms as proxies for feature-form pairs. In the analogy discussed above, $dynju : dynja = njanju : x$, grammatical features can be omitted because the forms *dynju* and *dynja* unambiguously realize the accusative and nominative singular in the paradigm of DYNJA 'melon'. A more explicit formulation of proportional analogies is given below, in which $C_{1_{\lambda_1}}$ and $C_{2_{\lambda_1}}$ are distinct cells in the paradigm of a lexeme λ_1 and $C_{1_{\lambda_2}}$ and $C_{2_{\lambda_2}}$ are corresponding cells in the paradigm of a lexeme λ_2.

$$\langle C_{1_{\lambda_1,}} \ a \rangle : \langle C_{2_{\lambda_1,}} \ b \rangle = \langle C_{1_{\lambda_2,}} \ c \rangle : \langle C_{2_{\lambda_2,}} \ X \ \rangle$$

The reliability of a proportion can be seen to depend on how well C_1 predicts C_2, either within a particular inflection class or the inflectional system as a whole. This relation can be directly formulated in information-theoretic terms as the question of how much uncertainty about C_2 remains given knowledge of C_1. Proportional analogies do not normally apply to all realizations of a given cell, but merely to those that conform to the pattern exhibited by the first term in the proportion. The reduction in the uncertainty of C_2 attributable to a cell with a particular realization, i.e., $C_1 = x$ is measured by the SPECIFIC CONDITIONAL ENTROPY $H(C_2 | C_1 = x)$. For example, the reliability of the earlier Russian proportion would correlate with the value of $H(\text{N.Sg}|\text{A.Sg} = Xu)$. This value lies between 0, where the knowledge that A.Sg = Xu eliminates all uncertainty about the realization of N.Sg, and $H(\text{N.Sg})$, where knowing that A.Sg = Xu preserves all of the uncertainty about N.Sg. The closer that $H(\text{N.Sg}|\text{A.Sg} = Xu)$ approaches 0, the more reliable the proportion.

A generalization of the specific conditional entropy measure likewise provides a means of selecting principal parts. The problem is traditionally posed as one of identifying diagnostic cells rather than, as with analogies, identifying cells that conform to a given pattern. Diagnosticity at the level of individual cells again corresponds to entropy reduction. The diagnosticity of cell C_1 relative to a cell C_2 correlates with the value of $H(C_2 | C_1)$ the conditional entropy of C_2 given C_1. This relation can be defined as below in terms of $p(y|x)$, the conditional probability of y given x.

$$H(C_2 | C_1) = - \sum_{x \in C_1} p(x) \sum_{y \in C_2} p(y|x) \ \log_2 p(y|x)$$

A cell C_1 is diagnostic for a cell C_2 to the extent that the value of $H(C_2 | C_1)$ approaches 0. A cell C_1 is diagnostic for a paradigm $\mathcal{P}$ to the extent that $H(C_i | C_1)$ approaches 0 for each $C_i \in \mathcal{P}$. The same notions generalize to multiple cells. A set of cells $C_1, \ldots, C_m$ is diagnostic for a cell C_n to the extent that the value of $H(C_n | C_1, \ldots, C_m)$ approaches 0. The cells $C_1, \ldots, C_m$ are diagnostic for a paradigm $\mathcal{P}$ to the extent that $H(C_i | C_1, \ldots, C_m)$ approaches 0 for each $C_i \in \mathcal{P}$.

The principal parts of an inflection or word class then comprise the cells that are diagnostic for any item in that class. In the limiting case, one cell may suffice. However, the goal of identifying a single or unique principal part or set of principal parts again reflects pedagogical idealizations. In the general case, multiple principal parts will be required, and different combinations of cells may be equally informative.[6] From an information-theoretic perspective, which views language as a communication channel, the availability of multiple solutions to the problem of deducing new forms does not give rise to nonoptimal redundancy, but to desirable robustness.

---

6 As Finkel and Stump (2007, 2009) note, there is a general trade-off between the size and uniformity of principal part inventories. Allowing the choice of cells to vary across inflection classes reduces the number of cells required for each class, given that the diagnostic value of cells varies across classes.

Paradigm structure constraints can also be recast in information-theoretic terms. The PSC described in the earlier passage by Wurzel (1984) is summarized below.

$$\langle \text{A.Sg}, Xim \rangle \supset \langle \text{Abl.Sg}, X\bar{\imath} \rangle \supset \langle \text{A.Pl}, X\bar{\imath}s \rangle \supset \langle \text{G.Pl}, Xim \rangle$$

The asymmetrical relation expressed by specific conditional entropy can capture the directionality in the claim that "[f]rom the A.Sg. form we can derive, via a number of steps, all other forms, but not vice-versa" (Wurzel, 1984: 208). The original claim will be true whenever all of the conditional entropy values below approach 0.

$$H(\text{Abl.Sg}|\text{A.Sg} = Xim), H(\text{A.Pl}|\text{Abl.Sg} = X\bar{\imath}), H(\text{G.Pl}|\text{A.Pl} = X\bar{\imath}s)$$

Using entropy in place of logical implication permits a generalization of PSCs, since this permits the formulation of probabilistic as well as exceptionless patterns.

5.3 Morphological complexity and typology

In sum, an information-theoretic framework provides a transparent formal basis for the various types of implicational relations proposed in traditional WP models. This approach has been applied to the analysis of paradigmatic structure in Uralic (Ackerman et al., 2009), Slavic (Sims, 2011), Romance (Bonami, 2012; Bonami & Luís, in press, 2013) and a diverse 10-language sample (Ackerman & Malouf, 2013).

In addition, information theory offers a fresh perspective on issues that had largely resisted formulation in traditional terms. Although paradigms and inflection classes are assumed to be cohesive in approaches that recognize them, traditional models provide no means of measuring or comparing morphological cohesiveness. The Paradigm Economy Principle proposed by Carstairs (1983) stakes out one extreme position by proposing that the number of classes in a system is bounded by the largest number of affixal realizations associated with any single cell. However, this principle ultimately constrains the description of affixal classes and provides no means of measuring morphological cohesion or the factors that contribute to it.

Information theory provides a number of applicable measures. A symmetrical measure of paradigm cohesion can be defined in terms of mutual information between cells. The mutual information of a pair of cells, $I(C_1; C_2)$ expresses their mutual dependence. This can be defined as below in terms of the entropy of either cell minus the entropy that remains given knowledge of the second cell.

$$I(C_1; C_2) = H(C_1) - H(C_1|C_2) = H(C_2) - H(C_2|C_1)$$

If the realization of two cells is independent, their mutual information will be 0. If they are completely interdependent, their mutual information will be identical to their common entropy value. Consequently, a measure of the cohesion of a paradigm, $\kappa(\mathcal{P})$, can be obtained summing the mutual information of each pair of cells in a paradigm and dividing the result by the sum of their joint entropies.

$$\kappa(\mathcal{P}) = \sum_{C_i, C_j \in \mathcal{P}} \frac{I(C_i; C_j)}{H(C_i, C_j)}$$

The more cohesive (i.e., interdependent) the cells of a paradigm are, the higher the value of κ will be. A traditional model requires only that the value be greater than 0, though a more refined model could posit a circumscribed range of variation.

This is precisely the claim expressed by The Low Conditional Entropy Conjecture of Ackerman and Malouf (2013). To represent the asymmetrical structure of inflectional dependencies, these analyses calculate the average conditional entropy for pairs of cells in paradigms drawn from a sample of 10 typologically diverse languages. As Ackerman and Malouf show, the values are not only lower than the averages of the original cell entropies, but also appear to fall within a narrow range.

There are two particularly encouraging aspects of this preliminary incursion into information-theoretic typology. The first is that the reduction in entropy is almost certainly underestimated by the fact that the initial calculations are based on descriptions of languages for which reliable frequency information is not available. Since estimations based on grammar descriptions define worst case entropy ceilings, the addition of frequency information is likely to produce further reductions in entropy. Conditional entropy is also compatible with a psychologically plausible model of paradigmatic complexity. Instead of modelling the raw uncertainty associated with a grammatical description of a system, it models just the uncertainty that a speaker must resolve in deducing a new form of a previously encountered item.

6 CONCLUSION

The research paradigms summarized in the preceding sections have undergone rapid expansion and many basic questions remain open. However, an information-theoretic approach to modelling morphological structure has already shown considerable promise in resolving longstanding challenges and in suggesting new lines of research.

This is not the first application of information theory to problems of linguistic analysis. The review of Shannon and Weaver (1949) in Hockett (1953) lists various linguistic applications of informational theory, and this perspective is incorporated into Hockett (1955). The distributional model developed by Harris (1951, 1991) was likewise guided by information theory.[7] Yet Hockett was mainly concerned with sound patterns and Harris with segmentation and arrangement. Furthermore, as Bloomfieldians, both focussed exclusively on the analysis of syntagmatic patterns.

---

7 See Chomsky (1957) for a critical assessment of Hockett (1955) and the general relevance of statistical models to grammar, and Goldsmith (2001) and Pereira (2002) for some recent reappraisals.

One reason why the reapplication of information theory may be particularly fruitful in the domain of morphology is that morphological variation – especially inflectional variation – is essentially finite. The speaker's choice between alternatives, whether in production or comprehension, is also strongly conditioned by lexical neighbourhood effects, distributional biases and a range of other factors. Hence the communicative function of morphological variation conforms well to the Shannon idealization of communication as "selecting from a set of possible messages".

> The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have *meaning*; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one *selected from* a set of possible messages. The system must be designed to operate for each possible selection, not just the one which will actually be chosen since this is unknown at the time of design. (Shannon, 1948: 22)

Current information-theoretic approaches to morphological analysis develop the idealization of variation as uncertainty and information as uncertainty reduction. This conception falls naturally within a larger intellectual tradition that regards knowledge and learning in terms of cue-driven uncertainty resolution.[8] At the same time, an information-theoretic perspective permits a clear statement of the assumptions that underlie the architecture of traditional approaches. The role of words in a classical WP model reflects the assumption that the uncertainty associated with a word is less than the sum of the uncertainty of its parts. The role of paradigms likewise reflects the assumption that the uncertainty associated with a paradigm is less than the sum of the uncertainty of the members of the paradigm. In each case, the reduction in uncertainty reflects interdependencies that can be modelled by conditional entropy. A general morphological model that develops these claims is outlined in Ackerman et al. (2009), Ackerman and Malouf (2013) and Blevins (to appear).

Within this resurgent tradition, there is a direct line of descent from information-theoretic models of language processing to analyses of language structure and complexity. In an inversion of the customary flow of influence from pure to applied domains, a series of pioneering studies (Kostić, 1991, 1995; Kostić et al., 2003) and their generalizations (Moscoso del Prado Martín et al., 2004b; Milin et al., 2009a) led to the revitalizion of an ancient perspective on morphological analysis.

---

8   See, e.g., Ramscar, Yarlett, Dye, Denny, and Thorpe (2010); Arnon and Ramscar (2012); Baayen, Milin, Filipović Đurđević, Hendrix, and Marelli (2011); Ramscar (2013).

REFERENCES

Ackerman, F., Blevins, J. P., & Malouf, R. (2009). Parts and wholes: Implicative patterns in inflectional paradigms. In J. P. Blevins & J. Blevins (Eds.), *Analogy in Grammar: Form and Acquisition*. OUP, 54–81.

Ackerman, F., & Malouf, R. (2013). Morphological organization: The Low Conditional Entropy Conjecture. *Language*, *89*, 429–464.

Anderson, S. R. (1992). *A-Morphous Morphology*. Cambridge: Cambridge University Press.

Arnon, I., & Ramscar, M. (2012). Granularity and the acquisition of grammatical gender: How order-of-acquisition affects what gets learned. *Cognition*, *122*, 292–305.

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

Baayen, R. H., Feldman, L. B., & Schreuder, R. (2006). Morphological influences on the recognition of monosyllabic monomorphemic words. *Journal of Memory and Language*, *53*, 496–512.

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

Baayen, R. H., Milin, P., Filipović Đurđević, D., Hendrix, P., & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Pscychological Review*, *118,* 438–481.

Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1995). *The CELEX Lexical Database* (Release 2, CD-ROM edn). Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA.

Baayen, R. H., Wurm, L. H., & Aycock, J. (2008). Lexical dynamics for low-frequency complex words: A regression study across tasks and modalities. *The Mental Lexicon*, *2*, 419–463.

Beard, R. (1995). *Lexeme-Morpheme Base Morphology: A General Theory of Inflection and Word Formation*. Cambridge: SUNY Press.

Blevins, J. P. (2007). Conjugation classes in Estonian. *Linguistica Uralica*, *43*, 250–267.

Blevins, J. P. (2008). Declension classes in Estonian. *Linguistica Uralica*, *44*, 241–267.

Blevins, J. P. (to appear). *Word and Paradigm Morphology*. Oxford: Oxford University Press.

Blevins, J. P., & Blevins, J. (Eds.). (2009). *Analogy in Grammar: Form and Acquisition*. Oxford: Oxford University Press.

Bonami, O. (2012). *Discovering implicative morphology*. Paper presented at the Les Décembrettes 8, Bordeaux.

Bonami, O., & Luís, A. (2013). *Causes and consequences of complexity in portuguese verbal paradigms*. Paper presented at the 9th Mediterranean Morphology Meeting, Dubrovnik.

Bonami, O., & Luís, A. (in press). Sur la morphologie implicative dans la conjugaison du portugais: Une étude quantitative. In J. L. Léonard (Ed.), *Morphologie flexionnel le et dialectologie romane: typologie(s) et modélisation(s)*. Leuven: Peters.

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

Chomsky, N. (1957). Review of *A Manual of Phonology* by C. F. Hockett. *International Journal of American Linguistics*, *23*, 223–234.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

de Jong, N. H (2002). *Morphological Families in the Mental Lexicon*. (Unpublished doctoral dissertation). University of Nijmegen.

Erelt, M. (Ed.). (2003). *Estonian Language*. Tallinn: Estonian Academy Publishers.

Erelt, M., Erelt, T., & Ross, K. (2000). *Eesti keele käsiraamat*. Tallinn: Eesti Keele Sihtasutus.

Erelt, M., Kasik, R., Metslang, H., Rajandi, H., Ross, K., Saari, H., Tael, K., & Vare, S. (1995). *Eesti Keele Grammatika* (Volume I: Morfoloogia). Tallinn: Eesti Teaduste Akadeemia Eesti Keele Instituut.

Erelt, T. (Ed.). (2006). *Eesti keele õigekeelsussõnaraamat*. Tallinn: Eesti Keele Sihtasutus.

Finkel, R., & Stump, G. (2009). Principal parts and degrees of paradigmatic transparency. In J. P. Blevins & J. Blevins (Eds.), *Analogy in grammar: form and acquisition* (pp. 13–53). Oxford: Oxford University Press.

Finkel, R., & Stump, G. T. (2007). Principal parts and morphological typology. *Morphology*, *17*, 39–75.

Goldsmith, J. A. (2001). On information theory, entropy and phonology in the 20th century. *Folia Linguistica*, *34*, 85–100.

Harris, Z. S. (1951). *Methods in Structural Linguistics*. Chicago: University of Chicago Press.

Harris, Z. S. (1991). *A Theory of Language and Information*. Oxford: Clarendon Press.

Hay, J. (2001). Lexical frequency in morphology: Is everything relative? *Linguistics*, *39*, 1041–1070.

Hockett, C. F. (1953). Review of The Mathematical Theory of Communication by Claude L. Shannon and Warren Weaver. *Language*, *29*, 69–93.

Hockett, C. F. (1954). Two models of grammatical description. *Word*, *10*, 210–231.

Hockett, C. F. (1955). *A Manual of Phonology*. Bloomington: Indiana University Publications in Anthropology and Linguistics 11.

Joos, Martin (Ed.). (1957). *Readings in Linguistics I.* Chicago: University of Chicago Press.

Karlsson, F. (1999). *Finnish: An Essential Grammar*. London: Routledge.

Kostić, A. (1991). Informational approach to processing inflectional morphology: Standard data reconsidered. *Psychological research*, *53*, 62–70.

Kostić, A. (1995). Informational load constraints on processing inflectional morphology. In L. B. Feldman (Ed.), *Morphological Aspects of Language Processing* (pp. 317–344). Hillsdale, NJ: Lawrence Erlbaum Associates.

Kostić, A., Marković, T., & Baucal, A. (2003). Inflectional morphology and word meaning: Orthogonal or co-implicative domains? In R. H. Baayen & R. Schreuder (Eds.), *Morphological Structure in Language Processing* (pp. 1–44). Berlin: Mouton de Gruyter.

Kuryłowicz, J. (1964). *The Inflectional Categories of Indo-European*. Heidelberg: Carl Winter.

Matthews, P. H. (1965). The inflectional component of a word-and-paradigm grammar. *Journal of Linguistics*, *1*, 139–171.

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.

McDonald, S. A., & Schillcock, R. C. (2001). Rethinking the word frequency effect: The neglected role of distributional information in lexical processing. *Language and Speech*, *44*, 295–323.

Milin, P., Filipović Djurdjević, D., & Moscoso del Prado Martín, F. (2009a). The simultaneous effects of inflectional paradigms and classes on lexical recognition: Evidence from Serbian. *Journal of Memory and Language*, *60*, 50–64.

Milin, P., Kuperman, V., Kostić, A., & Baayen, R. H. (2009b). Words and paradigms bit by bit: An information-theoretic approach to the processing of inflection and derivation. In J. P. Blevins & J. Blevins (Eds.), *Analogy in grammar: form and acquisition* (pp. 214–252). Oxford: Oxford University Press.

Moscoso del Prado Martín, F. (2003). *Paradigmatic structures in morphological processing: computational and cross-linguistic studies*. (Unpublished doctoral dissertation). University of Nijmegen.

Moscoso del Prado Martín, F., Bertram, R., Häikiö, T., Schreuder, R., & Baayen, R. H. (2004a). Morphological family size in a morphologically rich language: The case of Finnish compared to Dutch and Hebrew. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *30*, 1271–1278.

Moscoso del Prado Martín, F., Kostić, A., & Baayen, R. H. (2004b). Putting the bits together: An information-theoretical perspective on morphological processing. *Cognition*, *94*, 1–18.

Mürk, H. W. (1997). *A Handbook of Estonian: Nouns, Adjectives and Verbs*. Bloomington: Indiana University Uralic and Altaic Series 163.

Paul, H. (1920). *Prinzipien der Sprachgeschichte*. Tübingen: Max Niemayer Verlag.

Pereira, F. (2002). Formal grammar and information theory: together again? In B. E. Nevin & S. B. Johnson (Eds.), *The Legacy of Zellig Harris: Language and Information into the 20th Century* (Volume 2: Mathematics and computability of language) (pp. 13–32). Amsterdam: John Benjamins.

Ramscar, M. (2013). Suffixing, prefixing, and the functional order of regularities in meaningful strings. *Psihologija*, *46*(4), 377–396.

Ramscar, M., Yarlett, D., Dye, M., Denny, K., & Thorpe, K. (2010). The effects of feature-label-order and their implications for symbolic learning. *Cognititive Science*, *34*, 909–957.

Robins, R. H. (1959). In defence of WP. *Transactions of the Philological Society*, *58*, 116–144.

Saagpakk, P. F. (2000). *Estonian-English Dictionary* (3d ed.). Tallinn: Koolibri.

Schreuder, R., & Baayen, R. H. (1997). How complex simplex words can be. *Journal of Memory and Language*, *37*, 118–139.

Shannon, C. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*(3), 379–423.

Shannon, C. E., & Weaver, W. (1949). *The Mathematical Theory of Communication*. Champaign: University of Illinois Press.

Sims, A. (2011). Information theory and paradigmatic morphology. Presented at Workshop on Information-based Approaches to Linguistics, Boulder, CO.

Stump, G. T. (1993). On rules of referral. *Language*, *69*(3), 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.

Timberlake, A. (2004). *A Reference Grammar of Russian*. Cambridge: Cambridge University Press.

Weaver, W. (1948). Science and complexity. *American Scientist*, *36*, 536–544.

Wurzel, W. U. (1984). *Studien zur deutschen Lautstruktur*. Berlin: Akademie-Verlag.

Zwicky, A. M. (1985). How to describe inflection. In M. Niepokuj, M. Van Clay, V. Nikiforidou, & D. Feder (Eds.), *Society Proceedings of the Eleventh Annual Meeting of the Berkeley Linguistics* (pp. 372–386). Berkeley: Berkeley Linguistics Society.