

Morphological splits and subsystems

Jim Blevins (jpb39@cam.ac.uk)

Li8: Morphology, 2018-02-26

One motive for the post-Bloomfieldian model consisted, that is to say, in a genuinely factual assertion about language: namely, that there is some sort of matching between minimal 'sames' of 'form' (morphs) and 'meaning' (morphemes). *Qua* factual assertion this has subsequently proved false: for certain languages, such as Latin, the correspondence which was envisaged apparently does not exist ... One is bound to suspect, in the light of such a conclusion, that the model is in some sense wrong. (Matthews 1972: 124)

Interpretative challenges for models of morphemic analysis

What is the correspondence between units of meaning and form?

- ▶ Recall the many-many patterns exhibited by the ancient Greek verb form *elelykete* 'you have unfastened' (Matthews 1991: 173) [L4].

e	le	lý	k	e	te
		PERF		PAST IND	2ND PLU
PAST				ACTIVE	

- ▶ This problem is highly recalcitrant for morphemic models.
- ▶ A model can either (i) minimize the ambiguity of an analysis or (ii) minimize the size of units, but it cannot do both at once.

Separationist morphology (Beard 1995)

- ▶ For any given word class or inflection class, it is possible to identify the morphosyntactic features that are distinctive for that class.
- ▶ For each feature, it is possible to identify admissible values.
- ▶ Multiplying out all of the possible feature-value pairs determines an abstract paradigm structure, defined **independently** of any item.

In short, the theory of syntactic features developed earlier can incorporate the traditional paradigmatic treatment directly. The system of paradigms is simply described as a system of features, one (or perhaps some hierarchical configuration) corresponding to each of the dimensions that define the system of paradigms. Interpretive phonological rules, some quite specific, some of considerable generality, then operate on the phonological matrix of the lexical entry, giving, finally, a phonetic matrix. (Chomsky 1965: 172)

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)

Distinctive nominal features in English

- ▶ In English, the distinctive inflectional features for nouns are just:
 - ▶ PERSON, with the values {1ST, 2ND, 3RD},
 - ▶ NUMBER, with the values {SING, PLU}
- ▶ All nouns are 3rd person and there is no lexical case marking.
- ▶ Hence the general 2-cell noun paradigm structure in English is defined entirely by the two number values:

[SING] [PLU]

Stem entries in English

- ▶ The ‘inherent’ or ‘intrinsic’ features that are associated with all forms of a given item are associated with the lexical entry of the item.
- ▶ In addition, entries specify the basic root or stem of an item:

$$\left[\begin{array}{c} 3RD \\ /baik/ \end{array} \right] \quad \left[\begin{array}{c} 3RD \\ /naif/ \end{array} \right] \quad \left[\begin{array}{c} 3RD \\ /ʃip/ \end{array} \right] \quad \left[\begin{array}{c} 3RD \\ /mæn/ \end{array} \right]$$

Paradigm structure in English

- ▶ The paradigm of a particular item are then defined by combining the distinctive features from the abstract structure with the 'inherent' features contained in the lexical entry of the item.
- ▶ Hence the example nouns have the following 2-cell paradigms:

3RD	3RD	3RD	3RD
SING	SING	SING	SING
/baɪk/	/naɪf/	/ʃɪp/	/mæn/
3RD	3RD	3RD	3RD
PLU	PLU	PLU	PLU
/baɪk/	/naɪf/	/ʃɪp/	/mæn/

Simple rules of exponence

- ▶ The surface form associated with a paradigm cell is then defined by **realisation** or **spell-out** rules that interpret features in the cell.
- ▶ Realization rules only apply if the required features are present.
- ▶ If no rules apply, the cell is realized by the basic stem of the entry.

English plural exponence rule (Matthews 1991: 175)

$$\begin{bmatrix} \text{PLU} \\ X \end{bmatrix} \rightarrow X + z$$

Realization of surface forms of nouns in English

Singulars

$\begin{bmatrix} 3RD \\ SING \\ /baɪk/ \end{bmatrix}$	$\begin{bmatrix} 3RD \\ SING \\ /naɪf/ \end{bmatrix}$	$\begin{bmatrix} 3RD \\ SING \\ /ʃɪp/ \end{bmatrix}$	$\begin{bmatrix} 3RD \\ SING \\ /mæn/ \end{bmatrix}$
---	---	--	--

- ▶ No rule applies, so the singular is realized by the stems.

Regular plurals

$\begin{bmatrix} 3RD \\ PLU \\ /baɪk/ \end{bmatrix}$	\rightarrow	$/baɪk + z/$	$(\leftrightarrow [baɪks])$
--	---------------	--------------	-----------------------------

Inflection class ‘features’

- ▶ But what prevents ‘overapplication’ to nouns with irregular plurals?
- ▶ As it stands, the plural rule applies to the irregular nouns and, incorrectly, defines /naɪf + z/, /ʃɪp + z/ and /mæn + z/.
- ▶ (We don’t want to prevent this entirely, since this is exactly what happens in cases of ‘overregularization’ in learners.)
- ▶ Realisation-based models use ‘inflection class features’ (Chomsky 1965; Matthews 1965) to cross-reference words and entries.

Class-indexed stem entries in English

- ▶ The ‘inflection class features’ DCI, DCII and DCIII are associated with the lexical entry of items with irregular plurals.
- ▶ But nouns with regular plurals remain **unspecified** for class.

$\left[\begin{array}{c} 3RD \\ /baik/ \end{array} \right]$	$\left[\begin{array}{c} DCI \\ 3RD \\ /naif/ \end{array} \right]$	$\left[\begin{array}{c} DCII \\ 3RD \\ /ʃip/ \end{array} \right]$	$\left[\begin{array}{c} DCIII \\ 3RD \\ /mæn/ \end{array} \right]$
---	--	--	---

Realization of surface forms of irregular nouns in English

- ▶ The appropriate rule now applies to each of the irregular nouns:

$$\left[\begin{array}{l} \text{DCI} \\ 3^{\text{RD}} \\ \text{PLU} \\ /naif/ \end{array} \right] \rightarrow \mathbf{voi(/naif/)} + z/ \quad (\leftrightarrow [naivs])$$

$$\left[\begin{array}{l} \text{DCII} \\ 3^{\text{RD}} \\ \text{PLU} \\ /fip/ \end{array} \right] \rightarrow /fip/ \quad (\leftrightarrow [fip])$$

$$\left[\begin{array}{l} \text{DCIII} \\ 3^{\text{RD}} \\ \text{PLU} \\ /mæn/ \end{array} \right] \rightarrow \mathbf{abl(/mæn/)} \quad (\leftrightarrow [mæn])$$

Disjunctive rule ordering (Anderson 1986; Stump 2001)

- ▶ The original unindexed plural rule still applies to regular nouns.

$$\left[\begin{array}{c} 3RD \\ PLU \\ /baik/ \end{array} \right] \rightarrow /baik + z/ \quad (\leftrightarrow [\text{baiks}])$$

- ▶ None of the indexed rules applies to regulars because regulars lack the class features that those rules realise.
- ▶ The requirements of the unindexed rule are met by the irregular nouns, because they all properly contain the features in the rule.
- ▶ However application of the unindexed rule to the irregulars is **blocked** by a convention that the most specific rule must apply.

From process to realization

- ▶ An interpretive, 'spell-out-based' treatment of word structure allows realizational models to describe non-affixal patterns.
- ▶ The separation between forms and their realizations also permits many-many relations between properties and exponents.
- ▶ The problem of motivating/evaluating segmentations is solved, but by dispensing with **any** model of word-internal structure.

Distributive morphology (Halle & Marantz 1993)

- ▶ The model of 'Distributive Morphology' is a morphological approach that implements morphemic analysis in realizational terms.
- ▶ Most of the core tenets of the approach reflect more general syntactocentric assumptions about grammatical organization.
- ▶ The main morphological claim is that the feature bundles spelled out by realization rules have a persistent status as 'morphemes'.

Partial paradigm of Finnish TALO 'house' (Karlsson 1999: 249)

- ▶ The grammatical case forms of TALO illustrate the interaction of **intrinsic** ('within-block') and **extrinsic** ('cross-block') constraints:

	Sing	Plu
Nominative	talo	talot
Genitive	talon	talojen
Partitive	taloa	taloja

- ▶ In general, case markers must follow the plural marker *-i/j*.
- ▶ However, in the nominative plural, fusional *-t* must preempt *-i/j*.

Partial stem entry for TALO

- ▶ The minimal entry will specify word class features and a stem form:

$$\begin{bmatrix} N \\ \text{talo} \end{bmatrix}$$

Plural and grammatical case realization rules

$$\begin{array}{l}
 \text{A: } \left[\begin{array}{c} \text{PLU} \\ X \end{array} \right] \rightarrow X + i \quad \left[\begin{array}{c} \text{PLU} \\ \text{NOM} \\ X \end{array} \right] \rightarrow X + t \\
 \text{B: } \left[\begin{array}{c} \text{PART} \\ X \end{array} \right] \rightarrow X + a \quad \left[\begin{array}{c} \text{GEN} \\ X \end{array} \right] \rightarrow X + n \quad \left[\begin{array}{c} \text{PLU} \\ \text{GEN} \\ X \end{array} \right] \rightarrow X + en
 \end{array}$$

- ▶ Plural *-i/j* and nominative plural *-t* occur in the same block, A.
- ▶ Hence when rules 'compete', the more specific rule applies.

Spell-out of grammatical case forms of TALO

	$\begin{bmatrix} \text{SING} \\ \text{NOM} \end{bmatrix}$	$\begin{bmatrix} \text{SING} \\ \text{GEN} \end{bmatrix}$	$\begin{bmatrix} \text{SING} \\ \text{PART} \end{bmatrix}$	$\begin{bmatrix} \text{PLU} \\ \text{NOM} \end{bmatrix}$	$\begin{bmatrix} \text{PLU} \\ \text{GEN} \end{bmatrix}$	$\begin{bmatrix} \text{PLU} \\ \text{PART} \end{bmatrix}$
Stem	talo	talo	talo	talo	talo	talo
A				talot	talo <i>i</i>	talo <i>i</i>
B		talon	talo <i>a</i>		talo <i>jen</i>	talo <i>ja</i>

- ▶ These blocks of rules define the grammatical case forms of TALO.
- ▶ But what would happen if *-t* (and/or *-i/j*) occurred in block B?

How (and why) to distinguish inflection from derivation?

- ▶ Inflection and derivation are traditionally distinguished along three structural dimensions. Consider English *work-er-s*.

Morphotactic Inflectional occurs 'outside' derivation: [[[work]-er]-s].

Morphosyntactic Derivation can 'change' properties; inflection merely 'adds' to the properties of a stem; : -er: $V \rightarrow N$; -s: +PLU.

Paradigmatic inflectional paradigms are more regular and transparent than derivational families. (Compare the regularity of plural paradigms against the variation in derived nominals)

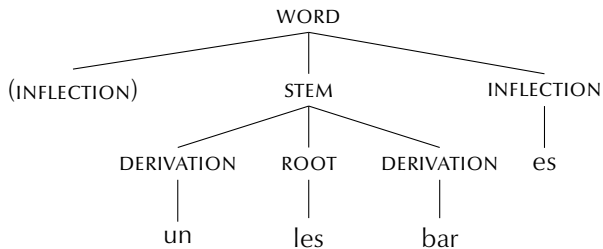
(Morpholexical) Derivation defines new items, inflection defines forms.

Some derived and inflected forms of German **LESEN** 'to read'

- ▶ *les-* 'read' (VERB ROOT)
- ▶ *les-t* 'read.PRES.INDIC-2PL' (V INFLECTION)
- ▶ *les-bar* 'readable' (V→A EXOCENTRIC DERIVATION)
- ▶ *un-lesbar* 'unreadable' (A→A ENDOCENTRIC DERIVATION)
- ▶ *unlesbar-es* 'unreadable-NEUT.NOM.SG' (N INFLECTION)

Hierarchical structure?

Candidate morphotactic structure of *unlesbares* 'unreadable'



Some characteristic properties of inflection

- ▶ Inflection is monotonic, adding but not changing properties.
- ▶ Inflection is relevant for syntactic dependencies (Anderson 1982).
- ▶ Inflection tends to be productive and semantically transparent.

Inflectional paradigms

- ▶ The inflected variants of an item tend to define a closed, uniform property space and a mostly transparent semantic space.
- ▶ For an item of a given word class or inflection class, it is possible to specify the distinctive properties and the number of forms.
- ▶ Notions ‘morphological gap’, ‘suppletion’, and even, to a large degree, ‘syncretism’ apply almost exclusively to inflection because they reflect expectations based on this uniform structure.
- ▶ Inflected forms are almost never described as ‘established’.

The Paradigm Economy Principle

- ▶ The number of inflectional classes in a system is bounded by the number of realizations of the maximally allomorphic paradigm cell.

Paradigm economy provides at least a partial answer to a question which, so far as I can discover, has not been asked before – a question about how, in any inflected language, the inflexional resources available in some word-class or part of speech are distributed among members of that word class. (Carstairs 1983: 116)

when we apply the traditional notion of ‘paradigm’, we find that the actual total of paradigms is at or close to the minimum logically possible with the inflexional resources involved, and nowhere approaches the logical maximum (Carstairs 1983: 127).

Contrasts between inflectional and derivational exponents

- ▶ Derivational elements tend to be prosodically larger, even satisfying the minimum word size of a language, with more specific meanings.
- ▶ Could this be related to the observation that derivational elements often look more like morphologized compounds, whereas the inflectional elements look like morphologized function words?
- ▶ Although less 'grammatical', the meanings of derivational formations are often much more lexically idiosyncratic.

Contrasts between paradigms and derivational families

- ▶ The size of derivational families can vary by orders of magnitude.
- ▶ In the English component of CELEX (Baayen et al. 1995), the noun *MAN* has a family size of 187 and *TWEET* a family size of 1.
- ▶ The semantic relations between members of a derivational family (forms sharing a common base) are also highly variable.
- ▶ Even in the case of highly productive derivational patterns, like the alternation between verbs and agentive nominals in *-er*, the meaning relations exhibit idiosyncracies.
- ▶ The most statistically prevalent interpretations of *worker* is not ‘one who works’ but ‘a member of a social class’ or ‘a type of bee’.

Psychological correlates of derivational family size

- ▶ The emergence of corpus-based psycholinguistics has revealed a variety of recurrent morphological patterns and robust effects.
- ▶ Variation in family size is a robust predictor of responses in a range of psycholinguistic tasks (e.g., visual word recognition/naming).
- ▶ Family size is a type count and family effects are predominantly sensitive to the size of lexical inventories (Schreuder & Baayen 1997; Jong et al. 2000; Moscoso del Prado Martín et al. 2004).

Agentive and abstract nominals

Effect	Suffix	Stems	Examples
$V \rightarrow N$	<i>-er</i>	any	baker, complainer, manager, receiver
	<i>-ant</i>	Latinate	attendant, contestant, dependent, inhabitant
$A \rightarrow N$	<i>-ness</i>	any	fairness, redness, tenderness, strangeness
	<i>-ity</i>	Latinate	agility, gravity, insanity, reality, curiosity

Deverbal and deadjectival forms

Effect	Suffix	Stems	Examples
V → A	<i>-able</i>	any	approachable, believable, breakable, livable
A → V	<i>-ize</i>	Latinate	civilize, legalize, tenderize, westernize, winterize
V → N	<i>-ability</i> (?)	any	readability, portability, affordability, believability

- ▶ cf. *geschwind* 'swift' ~ *Geschwindigkeit* 'speed', but **geschwindig*; *schlaflos* 'sleepless' ~ *Schlaflosigkeit* 'sleeplessness', but **schlaflosig*

Productivity, opacity and sub-lexicons

- ▶ There is a number of different ways to analyze the variation exhibited by the derivational system of English:
 1. Distinguish Latinate from general (or Germanic) sub-lexicons.
 2. Distinguish morphological 'strata' corresponding to these lexicons.
 3. Distinguish productive formations from inventories of fixed forms.

Constructive approaches to word formation

- ▶ It is conventional to think of complex forms as being 'built' from smaller parts, starting with a stem or root, and adding affixes.
- ▶ But the structure of a complex form can also be exhibited by comparing it with other forms that partially contrast with it.
- ▶ What kinds of patterns could distinguish these alternatives?

A determinate structure for productive ('level 2') affixation

- ▶ Affixes can usually be associated with a fixed base and output category, which jointly determine a fixed order of combination.
- ▶ Given that *grammatical* is an adjective:
 - ▶ *un-* can be added to form the derived adjective [_A un [_A grammatical]]
 - ▶ *-ness* can be added to form the derived noun [_N [_A grammatical] ness]
 - ▶ Both can be added to form the noun [_N [_A un [_A grammatical] ness]]
 - ▶ The prefix **must** combine before the suffix. Why?

Bracketing paradoxes with 'level 1' derivation

- ▶ The determinate base and output associated with affixes appear to determine a unique analysis for most complex English words.
- ▶ However, some simple and familiar examples pose a problem.
- ▶ *ungrammaticality* can't be derived like *ungrammaticalness* because
 1. *-ity* is a 'Level I' suffix and *un-* is a 'Level II' prefix (Kiparsky 1982),
 2. Hence 'Level I' *-ity* must apply before 'Level II' *un-*,
 3. But *-ity* combines to form the derived **noun** [_N [_A grammatical] ity],
 4. Thus *un-* can no longer apply, because it combines with an **adjective**.

Bracketing paradoxes with compounds

- ▶ The form *transformational grammamarian* appears to have the compound structure [transformational grammarian].
- ▶ But the expression doesn't refer to a kind of grammarian, but rather to a practitioner of a particular approach to grammar.
- ▶ This corresponds to the 'ian-form' of *transformational grammar*.
- ▶ The form *baroque flautist* doesn't refer to a kind of flautist, but to someone who plays a particular type of flute.
- ▶ So this compound corresponds to the 'ist-form' of *baroque flute*.

Analogical structure of 'bracketing paradoxes'

- ▶ There are technical approaches to these types of mismatches.
- ▶ But another type of solution exploits traditional notions of analogy, developed most clearly in the work of Paul (1920).
- ▶ A traditional 4-part analogy takes the form $a : b = c : X$.
 - ▶ The terms a and b define a pattern, and
 - ▶ c provides a base for extending the pattern and 'solving for X '.

Analogizing established lexical patterns (Spencer 1988)

- ▶ Analogizing from established patterns describes mismatches:
 1. *grammatical* : *grammaticality* = *ungrammatical* : *X*
(*X* = *grammaticality*)
 2. *flute* : *flautist* = *baroque flute* : *X*
(*X* = *baroque flautist*)
 3. *grammar* : *grammarian* = *transformational grammar* : *X*
(*X* = *transformational grammarian*)
 4. *linguistics* : *linguist* = *psycholinguistics* : *X*
(*X* = *psycholinguist*)
- ▶ Why do analogies seem to work in these cases? Does this suggest anything about the relation between 'rules' and analogies?

- Anderson, S. R. (1982). Where's morphology? *Linguistic Inquiry* **13**, 571–612.
- Anderson, S. R. (1986). Disjunctive ordering in inflectional morphology. *Natural Language and Linguistic Theory* **4**, 1–32.
- Baayen, R. H., Piepenbrock, R. & van Rijn, H. (1995). *The CELEX Lexical Database*. University of Pennsylvania, Philadelphia, PA.: Linguistic Data Consortium, release 2, cd-rom edition.
- Beard, R. (1995). *Lexeme-Morpheme Base Morphology: A General Theory of Inflection and Word Formation*. Cambridge: SUNY Press.
- Carstairs, A. (1983). Paradigm economy. *Journal of Linguistics* **19**, 115–125.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Halle, M. & Marantz, A. (1993). Distributed Morphology and the pieces of inflection. In Hale, K. & Keyser, S. J. (eds.), *The view from Building 20*, Cambridge, MA: MIT Press, 111–176.
- Jong, N. H. d., Schreuder, R. & Baayen, R. H. (2000). The morphological family size effect and morphology. *Language and Cognitive Processes* **15**, 329–365.
- Karlsson, F. (1999). *Finnish: An Essential Grammar*. London: Routledge.
- Kiparsky, P. (1982). Lexical phonology and morphology. In *Linguistics in the Morning Calm*, Seoul: Hanshin, 3–91.

- Marchand, H. (1960). *The Categories and Types of present-day English Word-Formation: A Synchronic-Diachronic Approach*. University of Alabama Press.
- 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.
- Moscoso del Prado Martín, F., Kostić, A. & Baayen, R. H. (2004). Putting the bits together: An information-theoretical perspective on morphological processing. *Cognition* **94**, 1–18.
- Paul, H. (1920). *Prinzipien der Sprachgeschichte*. Tübingen: Max Niemayer Verlag.
- Schreuder, R. & Baayen, R. H. (1997). How complex simplex words can be. *Journal of Memory and Language* **37**, 118–139.
- Spencer, A. (1988). Bracketing paradoxes and the English lexicon. *Language* **64**, 663–684.
- Stump, G. T. (2001). *Inflectional Morphology: A Theory of Paradigm Structure*. Cambridge: Cambridge University Press.