Introduction: Advances in Morphology

A Summary

Jan Don & Martin Everaert

In many domains of linguistics, theoretical differences have led to entrenchment and a certain degree of fragmentation. Morphology seems to be different. Theoretical positions differ substantially, but the differences never get in the way of informing oneself about the reasons for adhering to a different framework, making use of it. In this volume the following frameworks are discussed: a-morphous morphology (Anderson 1992), word and paradigm morphology (Blevins 2016), paradigm function morphology (Stump 2001, 2016), onomasiological approaches (Dokulil 1962; Štekauer 1998), construction morphology (Booij 2010), lexical semantic framework (Lieber 2004, 2016), cognitive grammar (Langacker 1987), and neo-constructionist approaches (Hale & Keyser 2002) such as distributed morphology (Embick 2015).

This volume thus contains a wealth of theoretical approaches, methodologies, and descriptive issues: a fitting tribute to Laurie Bauer, who made it his hallmark to serve the linguistic community with a broad range of textbooks, monographs, and research articles. Probably the most prominent contributions of Bauer as a morphologist (rather than as a researcher of variants of English) are those addressing compounding (Bauer 2017), conversion (Bauer 2018), and productivity (Bauer 2001). These are also the topics that recur in his work over a long period of time. His contributions are characterized by empirical rigor and a critical attitude towards foundational issues that are often taken for granted.

Bauer's influence is felt in many of the contributions in this volume, not only in a direct way through references to his publications, but also more indirectly in the choice of topics, the questions raised, and the general investigative approach. Bauer has played a central role in both illuminating the complex internal structure of words, and in disseminating research on this topic to general audiences around the world.

Before turning to some more general comments in Section 2, Section 1 will offer a brief overview of the contributions, highlighting the most important issues they address, to which we will return in Section 2. We have done so alphabetically, thus sidestepping the volume structure introduced by the editors.

1 Summarizing the Contribution

Artemis Alexiadou discusses the causative alternation in Greek. The anticausative in Greek is morphologically marked as either non-active (Nact; marked) or active (Act; unmarked). She addresses the question of whether in the anticausative variant there is a systematic relationship between the aspectual properties of the verb and the morphology with which it appears. The paper builds on the observation from Alexiadou et al. (2015) that the presence/absence of a prefix is a determining factor: the intransitive variant of unprefixed verbs bears Act morphology and prefixed verbs always bear NAct morphology. She addresses the question of why the prefix would influence the voice morphology of the intransitive variant. Alexiadou explains that this is the result of reanalysis that involves both a change in the status of the prefix as well as changes in the voice morphology of Greek, whereby NAct signals detransitivization. Importantly, Alexiadou shows that a similar situation seems to hold for Romance languages like Catalan and French. Building on a typological distinction proposed by Laurie Bauer, a change is documented from a more lexeme-like type of prefix system (Bauer 2003) to a system where prefixes are void of, for example, spatial semantic content, functioning as transitivity markers, so it seems. This leaves open the question of whether the observed generalization extends to other language families as well.

Stephen Anderson observes that in structuralist and early-generative grammars it is generally assumed that words are structured concatenations of minimal signs (morphemes). If words are formed by concatenation of morphemes, the operations creating words ought to be strictly monotonic (non-decreasing) in mathematical terms: the addition of a morpheme ought to have no effect other than to add further material to the form and content of the base. This paper addresses a specific case of a morphological operation altering the base in a non-monotonic way: the apparent addition of a formal marker – for instance, the middle/reflexive marker *-st* in Icelandic (Anderson 1990) – that is associated with the deletion of content, and not with the addition of semantic content. He discusses ways of dealing with the non-monotonic nature of detransitivizing constructions in the derivational morphology, but rejects them. His conclusion is that the most natural semantic analysis involves the deletion of a predicate and that such a deletion of semantic material constitutes an argument against the classical notion of morphemes as purely additive elements.

Mark Aronoff's paper discusses an example of a free morpheme grammaticalizing into a bound element. Concretely it is about the morpheme *free*, frequently occurring in words that could be viewed as either compounds or derivations. Laurie Bauer proposed criteria to distinguish between the second element of a compound and a derivational suffix (Bauer 2005). Aronoff

assumes, following up on this and Olsen (2014), that if the element is a suffix, it will (consistently) have a meaning that differs from the meaning of the element used as free form. The paper carefully describes the methodology by which Aronoff seeks to empirically establish whether *free* in *cordfree* has a meaning different from the dictionary meaning of *free*. Aronoff also explains how this would only work if the affix *-free* is different from the established suffix *-less*, an example of ecological niche differentiation in which each element survives by finding its own linguistic ecological niche (Aronoff 2016).

In her paper, **Heike Baeskow** argues that conversion should not be seen as zero-derivation, transferring a lexical item from one category into another without an (overt) derivational affix, but as a case of event-schema metonymy as envisaged in the cognitive grammar framework (Dirven 1999). Recently, Laurie Bauer revived this idea of conversion as metonymy by offering new justification for it (Bauer 2018). Central is the assumption that the participant of an event who metonymically represents the event as a whole is selected from a set of competing equals based on 'prominence'. She suggests that event-schema metonymy, based on cross-categorial contiguity, could be viewed as prominence inducing. Given that one participant becomes the metonymic vehicle to represent the whole, one could see this as constituting the attentional centre.

In her contribution **Juliette Blevins** continues her earlier work on the reconstruction of Proto-Basque. This paper takes the derivational patterns hypothesized for Proto-Basque in Blevins (2018) as her starting point and focuses on the following formatives: root-extension **-r*; acategorial **s-*; nominalizating **-s*; nominal **ha-*; and collective **hi-*, all part the Proto-Basque lexeme:

(i) $_4[(\{*ha\text{-}, *hi\text{-}\})_3[_2[(*s\text{-})_1[_0[root]_0(*\text{-}r)]_1]_2(*\text{-}s)]_3]_4$

This Proto-Basque reconstruction is distinct from earlier proposals in terms of its phoneme inventory, cluster phonotactics and stress pattern, and is also hypothesized to have undergone regular sound changes. The paper adds to the suggestion, earlier formulated in Blevins (2018), that Proto-Basque is distantly related to Proto-Indo-European.

Bożena Cetnarowska explores the distribution of "phrasal nouns" in Polish (also found in Greek); constructions in which a head noun is followed by another noun in the genitive case ($[[V]_N + N_{GEN}]$), which have phrasal properties but are still considered lexemes (Grzegorczykowa & Puzynina 1998; Szymanek 2010). She discusses how these phrasal nouns are in competition with well-known synthetic compounds $[N + [V]_N]$. Cetnarowska shows how these synthetic compounds and phrasal nouns "coexist" in (roughly) synonymous environments but are sometimes in competition. For instance, synthetic compounds with an event reading do not occur in texts,

while phrasal nouns do. Among other things, she also shows how analogy plays an important role in the creation of deverbal compounds, and argues that low-level construction schemas postulated within the framework of construction morphology (Booij 2010) can be employed to model such analogical compound formation.

Yu-Ying Chuang, Kaidi Lõo, James Blevins, and **Harald Baayen** argue, based on Estonian, for a morphological model in which the word is the basic cognitive unit over which paradigmatic analogy operates to predict form and meaning of novel forms. In other words, they advocate morphology without morphemes as minimal signs. Instead there are "lexomes," which are essentially the same elements that Laurie Bauer calls "morphemes" in his book on English morphology (Bauer 1983). These lexomes realize bundles of inflectional features at the form level. This study wants to show how the declensional system of Estonian (Blevins 2008) can be computationally modelled based on the principles of word and paradigm morphology (Blevins 2016), in which a full paradigm is deduced from only a small number of forms. The Linear Discriminative Learning model (Baayen et al. 2018) is non-decompositional at the form level, but analytical at the semantic level. That is, it is analytical in the sense of the methods used in distributional semantics, in which meaning is built up by summing the semantic vectors, in this case simulated vectors of its constituent lexomes. The model, it is argued, yields highly accurate results for both word comprehension and production.

In their paper **Wolfgang Dressler, Sonja Schwaiger,** and **Jutta Ransmayr** argue that the concept of 'word-formation family' is important to understand the organization of the lexicon. As an homage to Lauri Bauer's work on evaluative morphology (Bauer 1997), they focus on diminutives, more specifically on diminutive compounds. The question they ask is how morphosemantic subfamilies, different in their lexical semantic characterization, have an influence on the distribution and expansion of complex words. Their starting point is the relevance of the common semantic features that link the constituents of compounds, wanting to show the importance of conceptual similarities (in the sense of Coseriu 1978) in the organization of compounding. Results showed a substantial difference between the two diminutive forms concerning the formation of subfamilies. The first and most striking difference they noted is that nearly all formal diminutive families consist of more than one subfamily. Moreover, the degree of morphosemantic transparency between the diminutive suffix and its immediately preceding simple root shows a tendency of more transparent diminutive compounds being left-branching and more opaque compounds being right-branching. The authors were not able to characterize the subfamilies with strictly semantic features, which leads them to conclude that linguistic productivity depends largely on conceptual

productivity, including also pragmatic (i.e. non-semantic) characteristics, indicating that the subfamilies are conceptual rather than semantic.

The phenomenon of dvandva compounds is central in **Antonio Fábregas**'s paper. The starting point is an observation from Laurie Bauer that dvandva interpretations – coordinative compounds that express the aggregation of two distinct entities in a collectivity – are not generally available in the languages of the world (Bauer 2008). Given that noun-noun compounding allows virtually any type of interpretation, such a restriction is puzzling. Fábregas observes that also in Spanish dvandva compounds are distributionally quite restricted. Used as heads, they are unproductive and restricted to proper names; used as modifiers, they are dependent on the semantic nature of the head noun that combines with them. His explanation is built on a neo-constructionist formalization (Hale & Keyser 2002), assuming a structural layering of the noun phrase that defines the semantics that can be attached to it.

Lívia Körtvélyessy and **Pavol Štekauer** address onomatopoeia as instances of sound symbolism. They observe that in the Slavic linguistic tradition the concept of onomatopoeia applies only to a word that directly imitates the sound of an extra-linguistic reality, while in the Anglo-Saxon tradition the interpretation is more broadly for words that are related to sounds. Analyzing the word-formation processes based on (a phonic characterization of) onomatopoeia in Slovak and English (Körtvélyessy 2020), this paper tries to identify similarities and differences between these languages. They focus on the semantics of these word-formation processes, offering an onomasiological and cognitive analysis of the semantic shifts resulting from the onomatopoeia-based word-formation processes. Their analysis shows that onomatopoeia doesn't behave differently in terms of its word-formation characteristics, concluding that the only unique feature of onomatopoeia is its semiotic nature.

In her paper, **Rochelle Lieber** discusses why frameworks making use of model theoretic semantics have difficulty addressing questions of lexical semantics, more specifically issues accounting for the pervasive polysemy of deverbal nominalizations. Lieber argues that (lexical) semantics cannot be reduced to the Fregean concept of "reference": the relationship between a linguistic expression and the conditions under which the utterance of that expression would be true. She argues that addressing the semantics of natural language necessitates an eye for the conceptual basis of meaning (Frege's "sense") – everything about the meaning of an expression that is not captured by its reference. Lieber illustrates this with a discussion of the work of Pross (2019), who combines distributed morphology with formal semantics in the analysis of the German nominalization *Bemalung* (painting). She shows the limitations of such an approach and argues that a referential theory is inevitably going to need to be supplemented with elements with a conceptual basis. This is precisely what the lexical semantic framework advocated in Lieber

(2004, 2016) gives us. An important feature of that model is that both roots and affixes have the same semantic representations, allowing it to account for handling affixal polysemy in a unified way.

Central in the paper of **Ingo Plag, Sonia Ben Hedia, Arne Lohmann**, and **Julia Zimmermann** is the observation that the acoustic properties of plurals and genitive-plurals (*boys* vs. *boys'*) differ. Such a case shows unexpected effects of morphology on phonetic realization, and is, according to the authors, unexpected in certain morphophonological models or speech production models. They report on an experiment showing that the duration of plural *-s* is significantly shorter than the genitive + plural *–s*, arguing that this refutes a purely structural morphological approach or an approach in which the durational differences are related to prosodic structure. The observation fits morpheme/word-based frequency hypotheses, they postulate, just as a complexity hypothesis – the processing of a more complex morphosyntactic feature specification slows down the production of the exponent of that feature specification – is supported. They also entertain the idea that the differences they found are the result of an orthographic effect.

Franz Rainer's contribution examines metaphorical change in word-formation, more specifically the case of instrument nouns. He observes that there is little evidence for metaphorical extension in the case of the agent-instrument polysemy of Latin *-tor* (Rainer 2011), and asks himself whether this could be different for the case of Latin *-one* in its Romance successors. His "fact-finding" mission leads Rainer to conclude that, though there are etymological uncertainties, the conclusion seems to be warranted that instrumental *-one* was transmitted directly from Latin to Romance. On the basis of the broad spectre of languages, Rainer concludes that Latin heritage and local analogy go a long way in explaining the Romance facts.

The presence of "affixoids," a morphological category in between stems and affixes, in a number of Modern Greek varieties, is the central topic in **Angela Ralli**'s contribution to this volume. It is generally assumed that the existence of affixoids depends on the morphological type of a particular language. She argues that affixoids do indeed exist in morphologically rich languages with rich compounding and affixal derivation, assuming that in these languages the word-formation processes are stem-based (in line with Kastovsky 2009). Modern Greek is such a language developing affixoids derived from different sources, both stems and affixes. Clearly the process of affixoid creation is a diachronic one, but it prompts the creation of a synchronically relevant category (in line with Kenesei 2007).

Andrew Spencer discusses "uninflectedness" – lexemes/words that are uninflectable/uninflecting – noting that any part of speech can show uninflectability. Furthermore, uninflectability may be related to a particular position in a construction, such as the non-head of a compound, as noted by

Bauer (2017). Spencer adopts a lexicalist model of morphosyntax and a paradigm-based model of lexical relatedness, excluding models for which the notion "uninflectedness" would be either undefined or entirely derivative (such as distributed morphology). He argues that uninflectability poses the problem of how to define what it is that is "lexically inserted" (i.e. the lexicon–syntax interface) and how to account for it in a formal grammar. Spencer shows that uninflectability can be straightforwardly accounted for in the framework of paradigm function morphology, assuming that the root form is lexically inserted and that uninflectable items have neither a "content" paradigm nor a "form" paradigm (cf. Stump 2016).

Gregory Stump's contribution discusses what he calls the Monomorphemic Affix Assumption: the assumption that affixes are by definition monomorphemic. Laurie Bauer was among the first theoretical linguists who questioned this assumption (Bauer 1988). Stump discusses examples of "hypothetical" affixes that are a conflation of two affixes: asymmetrical patterns of paradigmatic opposition; single affixes standing in paradigmatic opposition to a sequence of affixes; classes of affixes that overlap in form and content; an affix's alignment depending on the presence or absence of another affix; cases in which a part of the content jointly realized by two affixes is attributable to neither affix on its own; and cases in which the appearance of an affix is sensitive to the presence of a more peripheral affix. On the basis of these observations he argues that a rule of "affix conflation" – that is, unification of two simple rules into a more complex one – is needed and can be formulated in the framework of paradigm function morphology (Stump 2016). Stump suggests that we should even go a step further, taking cases of circumfixation as cases of a rule combination similar to that of function composition, what he calls "synflation" (referring to Bauer's [1988] "synaffixes"). Both cases, however, defy the Monomorphemic Affix Assumption.

Salvador Valera's contribution is a study of the semantic patterns found in noun/verb conversion in English. The literature offers a small set of patterns, and in this paper the distribution of these patterns is studied on the basis of a stratified sample of lemmas extracted from the British National Corpus. The study wants to answer the question of whether the semantic pattern in the case of denominal conversion (Bauer et al. 2013) differs from cases of denominal verbal affixation, and whether there is reason to extend the set of semantic patterns. In the case of the semantic patterns, "performative" and "instrumental" relevant differences were found in conversion cases compared to affixation cases. The study indicates that there is reason to revise the set of semantic categories listed in the literature for noun/verb conversion. Some categories, such as "privative," seem to be less likely to occur than assumed, while other semantic categories are not listed but do seem to occur, such as "effected."

Petra Vogel puts names in "vague language" (Channell 1994) central in her paper. She shows which words/phrases are used as placeholders when a speaker doesn't know, has forgotten, or does not wish to use a name. The title of the paper gives examples from German (*dingsbums*) and English (*thingy*) which can refer to a person's name or place name; in the case of German a compound, and in the case of English a derivation. The paper discusses a sample of twenty-nine languages and discusses placeholders for names as opposed to (inanimate) items in these languages, focusing on structural and semantic properties. These placeholders are phrases (like *who is (s)he* or *this person*), compounds (*dingsbums*), derivations (*thingy*), and simple words (French *machine*). She concludes that more than half of the languages in her sample use phrases but *no* specific lexemic placeholders. Some languages (like Amharic) only show complex placeholders, namely phrasal compounds, compounds, or derivations. Languages like French and Spanish prefer simple words. At the semantic level, phrases are neutral with regards to a degrading or "belittling" dimension of the referent, whereas word placeholders often have a negative meaning.

2 Themes and Perspectives

2.1 Synchrony-Diachrony

Morphology is a field in which synchronic work and diachronic work inform each other and are not seen as separate enterprises, at least to a lesser extent than in other subdisciplines, as this volume illustrates. In the case of **Rainer**'s Latin heritage of nouns in *-one* in the Romance languages, the study concerns language change over a longer period (centuries). **Aronoff**'s paper addresses the synchronic manifestation of language change, the birth of a suffix. The grammaticalization process he describes reminds us of the work of Labov, taking the study of linguistic change, and in particular linguistic change in process, as central (Labov 1987) in the linguistic enterprise. **Blevins**'s paper is a pure case of the reconstruction of a proto-language. **Ralli** discusses dialectal differences from a historical perspective, and discusses the position that there is no strict dividing line between synchrony and diachrony (Stevens 2005; Lightfoot 2011), a position to which she herself does not adhere.

2.2 Methodology

The papers show a diversity of methodological tools to create their empirical basis. No methodology has a privileged status, whether acquired by experimentation or the use of corpora, or using native speakers' judgments in assessing linguistic structures. Which research methodology is appropriate

depends on the questions you have. However, in morphology, corpora, and dictionaries as a special manifestation of a corpus, are used quite often, as becomes clear from the title of Part III of this volume. Still, in many papers other research tools are also used.

Aronoff's paper can be read as a methodological manual on how to study language change in progress. He explains how he combined the online version of the *Oxford English Dictionary* and intuitive judgments to create data that he could check through web searches (carefully checking and curating the examples).

Apart from a corpus, **Vogel** used a questionnaire distributed to native speakers of a substantial group of typologically diverse languages. This questionnaire was supplemented by consulting dictionaries. **Dressler et al**. took their examples from the Austrian Media Corpus, but added their intuitive judgment by scoring the morphosyntactic transparency between base and the derivation on a Likert scale of 1–10.

Chuang et al. created a corpus of a completely different type. A set of word forms was manually created on the basis of the intuitive judgments of a native speaker, based on the information available in a dictionary. Subsequently this data set was used for a computational learning model to see whether this model correctly mimics the production and comprehension of words.

Plag et al. took observations based on the Buckeye Corpus of Spoken American English as their point of departure, in order to create an experiment in which (newly created) sentences were read aloud and their phonetic manifestations measured.

Cetnarowska used the National Corpus of the Polish Language (and dictionaries), but also web searches. Likewise, **Körtvélyessy and Štekauer** used the Corpus of Contemporary American English and several English dictionaries, a Slovak dictionary based on the Slovak National Corpus, and several specialized (web-based) dictionaries. **Valera**'s article made use of data derived from the British National Corpus, but also made use of the *Oxford English Dictionary* and intuitive judgments. **Baeskow** and **Lieber** built their work on data published in research articles, supplemented with data from the Corpus of Contemporary American English and, in the case of **Baeskow**, the iWeb Corpus.

Given their historical work, **Blevins**, **Ralli**, and **Rainer** used etymological/diachronic dictionaries and historical grammars. **Blevins**'s reconstructional work is based on the traditional methods of historical linguistics, using the comparative method and internal reconstruction.

Alexiadou, **Anderson**, **Fábregas**, **Spencer**, and **Stump** built their work primarily on data published in research articles and grammars, sometimes supplemented with data based on personal observations (or web-based observations).

2.3 Language Diversity

In many subdisciplines in theoretical linguistics the scope of languages discussed is substantial. This also holds true for morphology in general, and this volume in particular. **Vogel et al.** discuss twenty-nine languages from a variety of language families in a comparative perspective: Afro-Asiatic, Austro-Asiatic, Japonic, Koreanic, Niger-Congo, Sino-Tibetan, Tai-Kadai, Uralic, and Indo-European. **Rainer** also takes a comparative approach, taking examples from a wide variety of languages, but in this case they all belong to the Romance language subfamily: Romanian, Italian, Sardinian, Rheto-Romance, French, Occitan, Catalan, Spanish, and Portuguese.

Several papers discuss a diverse set of languages, but only because they are relevant for the theoretical point they want to make: **Anderson** discusses Icelandic, Faroese, and Navajo in detail, and gives examples from Danish, Norwegian, Swedish, Panare, Kadiweu, San Ildefonso Tultepec Otomi, Neverver, and Spanish; **Stump**'s paper is based on a discussion of Latin, Sanskrit, and Fula; **Spencer** uses data from English, German, and Japanese, just as **Cetnarowska** is using Polish and Japanese, **Baeskow** English and Mandarin Chinese, and **Lieber** English and German.

Dressler et al. study the commonalities and differences between two very closely related variants of what one could call the "same" language: Standard German and Austrian Standard German. A similar "micro-variational" perspective is taken by **Ralli**, who discusses Standard Modern Greek and several dialects and dialectal groups: Peloponnesian, Heptanesian, Constantinopolitan, and Lesbian, among others. **Körtvélyessy and Štekauer**'s paper is a contrastive study, comparing Slowak to English.

Fábregas's paper discusses only one language, Spanish, but he does so to address the question of why a certain phenomenon – in this case dvandva compounds – is so rare in languages like Spanish, English, and German, for instance, but not in other languages (such as Japanese). **Alexiadou** starts out with Greek, but given her hypothesis, explores whether her hypothesis would also hold for Romance. Primarily mono-language contributions are **Aronoff** (English), **Plag et al.** (English), **Chuang et al.** (Estonian), and **Blevins** (Basque).

2.4 Themes and Perspectives

Any attempt at grouping the papers of this volume into different themes runs the risk of either becoming empty because the chosen themes are so broad that it becomes vacuous (e.g. "morphology"), or ending up with almost as many themes as there are papers in the collection, which obviously misses the point of any categorization. Because of this difficulty, we have not made any attempt