MORPHOLOGICAL EMERGENCE
Wc: 9980 (excl. references)

Morphological Emergence

Authors

Péter Rácz[1], Janet B. Pierrehumbert[1,3,4], Jennifer B. Hay[1,2] and Viktória Papp[1,2]

Affiliations

1 New Zealand Institute of Language Brain and Behaviour, University of Canterbury
2 Department of Linguistics, University of Canterbury
3 Department of Linguistics, Northwestern University
4 Northwestern Institute on Complex Systems

Abstract

The chapter discusses natural language morphology in the light of more recent, emergent approaches. It starts out with the observation that morphology – contrary to what classical works would claim – is overlapping and gradient and then goes on to discuss available evidence on why this is the case. This evidence suggests that morphology is better understood as a system that emerges through the interaction of basic cognitive principles with ongoing language experience.

Keywords: morphology, phonology, rich memory models, chunking, NLP, iterated learning, sociolinguistics, corpus linguistics, language change, frequency

Introduction

The discipline of linguistic morphology is the study of word-structure, specifically of 'systematic covariation in the form and meaning of words' (Haspelmath & Sims 2010, p.2.) Classic structuralist and generative approaches to morphology made strong assumptions about the decomposability of words into minimal meaningful parts, or *morphemes*. They sought simple rules (such as affixation, compounding, or conversion rules) to explain how new words that express new composite meanings can be created.

Connecting the decomposition of known words to the possibilities for forming new words is an important insight of this approach. However, it does not capture the full complexity of morphology. Morphology is messy. It reflects a set of associations between meaning and form, but these associations can be partial or overlapping. Bybee (1985), for example, has an extensive cross-linguistic analysis of verbal inflection. She shows that it is usually not possible to delineate morphemes as one-to-one relationships of form to function. Typically, either one form is associated with several functions or one function is associated with several forms, and the possible sets of combinations can overlap in complex ways. It is also often impossible to tell where one morpheme ends and another starts. Furthermore, associations between form and function are gradient. The word *stick* is more evident in *dipstick* than in *slapstick*. The suffix *-ize* is more separable in *industrialize* than in *organize*. This gradience can be thought of in terms of association strengths amongst pairs or sets of words: *industrialize* is strongly linked to *industrial*, but weaker semantic relatedness has the result that *organize* is less associated with *organ*. Because different people know different words, this means that people's propensity to analyze existing forms as complex, and the probability that they will generalize patterns to create new forms, are functions of the content and structure of their individual lexical systems. Morphological systems, then, emerge in individuals as learned generalizations across lexical items. The content, frequency, and structure of lexical items, as well as the inter-relationships amongst them, thus play a central role in what is learned.

Many cognitive, linguistic and social factors are at play in influencing the types of morphological generalizations available to individuals, and those that they choose to operationalize in their own speech. This speech forms the input for new generations to learn from. The iteration of morphologically structured speech across speakers and generations results in the emergence in language communities of morphological patterns that are learnable and productive. However, these patterns are not perfectly stable, but instead change on historical time scales, with some trajectories of change showing striking similarities from one language to the next. To explain the systematic covariation in the form and meaning of words, we are thus led to engage many different types of data at many different time scales. Probabilistic models are crucial to integrate these sources of data, and relate effects at one time scale to effects at another.

This paper outlines some of the micro-forces which are at play in shaping the emergence and evolutions of morphological systems. Together, we argue, these forces work together to mold morphological structures. Our key assumptions about the key components and processes underpinning the emergence of morphological systems are as follows.

- Morphology is defined by associations between meaning and form. These associations can be partial, overlapping, and gradient.

- Morphological patterns are learned as generalizations over memories of words.

- There is considerable individual and social variation in morphological knowledge and use.

- Transmission across individuals and generations is imperfect. Over time, imperfect transmission can create or destroy morphological patterns.

- The production and perception of morphology is highly sensitive to frequency and predictability in the linguistic context, in a way that shapes trajectories of language change.

- A complex interplay of learnability, generalizability, frequency and social dynamics together shape the emergence, evolution, and production of morphological patterns.

In this article we step through each of these assumptions in turn. Together, they exert many micro-pressures on the communicative system, and together jointly influence the emergence of morphology as a self-organizing system. The chapter is structured as follows. First, we look at the basic properties of natural language morphology with an emphasis on its gradience. We then go on to overview the role of learning and generalization and available evidence on how they take place. This is followed by a discussion of computational models and the role of individual and social variation. We then talk about how structure can arise through transmission and investigate the roles of frequency and contextual probability. This is followed by brief conclusions.

The Explanandum: What is Morphology

Cognitive models of the lexicon describe it as a network, in which the words are nodes, and the links are relations between words. Broadly speaking, words are linked if they are similar, but the concept of similarity has many different dimensions. Words may be similar in how they sound (*candle*, *canter*), in what they mean (*guru*, *mentor*), or in both (*electric, electricity*). The mental lexicon is a dynamic structure in which nodes and links compete for activation. Those that are rarely used can be erased and new ones can be added. As we will see, emergentist models of morphological productivity and morphological change crucially relate to this fact. Current understanding of how the lexical network functions in linguistic systems is one of the triumphs of 20th century psycholinguistics (c.f. Miller 1995; Dell et al. 1997; Magnuson et al. 2003, 2007). The modern theory of morphology focuses on alignments of form and meaning in lexical networks.

Speakers of a language have strong intuitions about what counts as a good word. These intuitions are based on statistical generalizations over words that individuals already know, and over the relationships amongst these words. (Hay 2002; Frisch et al., 2001; Pierrehumbert, 2003). People can coin novel words that recombine pre-existing phonological elements in ways that are reasonably likely, but just happen to be absent in the pre-existing lexicon. For example, *Zampy* or *Plike* would be good possibilities for the English name of a new product. However *Shkmokp* would be less acceptable, because it begins and ends with sequences of consonants that do not occur at the edges of other English words. However, the dominant source of neologisms is not outright coinages, but rather the recombination of elements of other words that share both form and meaning. Linguistic morphology recognizes a distinction between compounding (as in *bird, birdhouse*) and affixation (as in *constitute, constitution*); in compounds, both parts can occur separately as lexical items, whereas affixes occur only as parts of words and never alone. But this distinction can become fuzzy when we consider word formation patterns such as blending (*work* + *alcoholic* → *workaholic*). For all such patterns, a central issue is *productivity* (or the propensity of word parts to be used in the creation of new words). Because productivity reveals the existence of cognitive abstractions, explaining productivity is a central issue in morphological theory.

The strength of a morphological pattern depends on the number of words that exhibit it, and on the degree to which it is evident in these words. Strength can translate to productivity (the ability to extend to novel elements), acceptability (by speakers of the language) and overgeneralization (in first language learning). For *agglutinative* morphological patterns (patterns formed by concatenating meaningful parts), it is practical to resort to the traditional concept of the morpheme in operationalizing this approach. In this case, the structure of the word partitions its phonological transcription, and the strength of each morpheme boundary in the transcription corresponds to the statistical strength of the evidence that a boundary is present at that location. Many factors in individual words affect this strength. These include their junctural phonotactics (i.e. the sound patterns around the potential boundary), the frequency relationships between the word and its parts, and the semantic and phonological transparency between the whole word and its parts. There is abundant empirical evidence about the factors affecting the strength of a morphological pattern. This evidence is surveyed by a number of recent review articles which we summarize below.

Hay & Baayen (2005) review available evidence in favor of the gradience of linguistic morphology. They also show that the productivity of an affixation pattern can be gradient and that multiple patterns can compete for the same lexeme. (The German plural of *Park* can be both *Parks* and *Pärke*.) In cases where there are multiple potential patterns, there can sometimes be statistically subtle semantic differences between different forms, which are accessible to, and utilized by, the language user (c.f. Ramscar et al. in press). Whether people interpret a word form as a single form or an affixed form is also gradient. The processing of individual forms (the parsing of specific compounds and suffixed forms) is affected by the observed token frequencies of word forms, including stems and affixes –as well as the ratios of these frequencies. However, the productivity of a morphological schema is strongly influenced by type frequency – the number of types that fit the schema.

Pierrehumbert (2003) reviews research on the learning of phonetic and phonological knowledge and then goes on to propose a model that is able to accommodate frequency information and gradience in processing. This model is richly detailed, so that it can store information like the frequency with which the individual language user encounters given forms. It uses overlapping representations, so that one does not have to break word forms into smaller units to sort out form-to-function mappings (hence we do not have to commit ourselves over the issue whether the *-s* at the end of *sleeps* marks person or number). Gradual updating of individual mental representations through language experience leads to feedback loops in the linguistic community as people communicate with each other over time.

Rácz et al. (forthcoming) give a detailed account of corpus-linguistic work on morphological processes, with an emphasis on the connections between the patterning of a morphological class in corpora and its productivity. Pierrehumbert (2010) provides an extensive discussion of both experimental results on the relevant factors in morphological processes and the picture of the lexicon these results suggest. A central conclusion of these authors is that gradience is an integral property of morphology.

Although lexicon-internal factors are powerful predictors of productivity, they are not the only predictors. Some morphological patterns rise in productivity despite modest beginnings with just a few attested types. One reason is that the productivity of a morpheme can also change in response to factors external to the lexicon (*exogenous* factors). Events in the world can affect what people chose to discuss, making some morphemes much more useful, or much less useful, than they were before. For example, the increasing productivity of *-genic* in words such as *carcinogenic* is probably driven in part by recent scientific discoveries. Novel word-formation patterns can also be adopted by social groups as an expression of social identity. An example of both of these factors is the word *selfie,* designated as word of the year by Oxford Dictionaries in the year in 2013. The tremendous rise in frequency of the word *selfie* owes a lot to the popular custom of sending self-portraits using cameras and mobile devices, as well as to the availability of these along with high bandwidth network connections. In this sense, the behavior of *selfie* is noteworthy, even if it can be recognized as the extension of the productive *–ie* suffix (as in words like *freebie, quickie, sharpie, roomie, sweetie*, etc.) These observations contribute to the picture of morphological systems as emergent, and we return them below in the context of morphological emergence at the population level.

Morphological Learning and Generalization in Individuals

Morphological patterns are generalizations that emerge over lexical items that are learned and stored by the individual. In some theories these patterns are epiphenomenal, whereas in others they are learned as such. We will return to the relevance of this distinction later on. The statistical distribution of any particular morphological pattern in the lexicon is a good predictor of the likelihood that it will be used to create new forms (Hay & Baayen (2005)). In particular, patterns that are represented by very many transparently derived forms in the lexicon are more likely to be adopted for new word formation than patterns represented by few clear types. These findings mean that morphology emerges within individuals in the course of acquiring a lexicon. The task of the learner is to acquire the lexicon, to generate generalizations over the lexicon, and to infer which can be most naturally extended. While these on the surface may seem like distinct tasks, they are intricately interwoven, and are likely to all fall out from the same general learning principles.

We intuitively understand the lexicon to be a storehouse of words, but the concept of *word* becomes vexed when we consider languages such as Finnish and Turkish, whose highly productive morphological systems result in words that might be translated as phrases in other languages (Creutz & Lagus 2007). Even for English, experiments on the processing of phrases reveal statistical patterns at multi-word scales. For example, Tily et al. (2009) investigate the dative alternation in English. Dative alternation refers to the variation in how English can express the relations between the roles of a ditransitive verb, as in *Gaff gave Deckard the Unicorn figure* versus *Gaff gave the Unicorn figure to Deckard*. They find a correlation between the acoustic features of spontaneous speech and the probability of given syntactic constituents. They use the Switchboard corpus of spoken American English to model probabilities of dative alternants and find that the less likely alternants (in a given construction) are spoken less fluently, with more hesitations and longer word durations. This suggests that the probability of syntactic constructions affects lexical processing, which, in turn, suggests that these probabilities are available for the speakers as a part of their linguistic competence.

Such results lead to a picture in which people encode and remember chunks of speech that vary in size. The smallest would be analyzed in the classical theory as mono-morphemic words, whereas the largest might be four morphemes or words, or even more. The processes that project structure for morphologically complex words, and the processes that decompose phrases into words, have many formal parallels in this approach. The similarity of word-based and morpheme-based language models is acknowledged in speech and language engineering (Hirsimäki et al., 2009). The models proposed for

Turkish (Hakkani-Tür et al., 2002) and Hungarian (Kornai, 1994) apply essentially the same principles that are used for languages with very little morphological structure to languages with very complex structure. The mechanisms that allow an algorithm to segment the speech signal into words are the same ones that allow it to recognize morphological structure on the word-level.

There are many parallels between parsing word boundaries and parsing affix boundaries. For example Hay (2003) shows that, in English, segmental transitional probabilities, which are strong cues of word boundaries, are exploited by listeners in positing boundaries between stems and affixes as well. Across languages, the same semantic content is distributed differently over words versus affixes. For instance, the Norwegian indefinite article is a separate word – just like in English -- while the definite article is suffixed to the stem. Certain phonological processes that operate within words also apply across word boundaries: Hungarian obstruent clusters undergo progressive voicing assimilation whether or not there is an intervening word boundary. This casts a new light on the observation by Carstairs-McCarthy et al. (2004), namely, that it is to some degree inexplicable why some languages have complex morphology while others have none. If the same principles are responsible for word-level and sub-word-level segmentation, *agglutinating* and *inflecting languages* (which have rich morphological structure) and *isolating* languages (which express grammatical relations through word order, for instance) are not radically different anymore.

An important point to take away from computational models is that storing some long sequences of morphemes or words during learning does not mean storing all long sequences. Long sequences are stored only if encountered enough times to be encoded and remembered; in some models they are stored only if they exhibit idiosyncratic properties and have been encountered often enough for these to be statistically learned. Long sequences that are not stored as part of the lexicon are processed via a compositional process that refers to their parts (Hirsimäki et al., 2009). We assume that such variable-scale processes are active from very early stages of language development, and work by *bootstrapping*; low levels of language exposure make available structures and relations that promote subsequent elaboration of the lexical system.

Infants are predisposed to attend to speech, and this attention is mediated by social factors such as joint attention (Smith et al., 2011). They can segment a continuous stream of language input into potential words, and apply implicit knowledge of the ambient language long before they can produce many words (see review in Daland & Pierrehumbert 2011). While research on the range of factors that impinge on early word learning has proliferated, two central aspects of word learning in children have received little attention. First, the widely noted difference between passive vocabulary (words that are recognized) and active vocabulary (words that are produced) is even bigger for children than for adults. Adding a new word to the active vocabulary is a very difficult task for children. Stokes and colleagues (2010; 2012; in press) have suggested that novel labels that are comprised of familiar phonological strings in the ambient language are more easily produced than labels that lack these features. As the lexicon grows, these effects seem to weaken, but in novel word learning experiments, children of 9-11 years are still well below adult levels of performance (Kan & Windsor, 2010). It thus appears that the role of the existing lexicon may be even more central to the question of active new-word production and creation than it is to passive word learning.

Children also differ from adults in their responses to the variability of morphology. A much-studied case is learning tense marking for English verbs. Early in development, children produce both irregular (*run, ran*) and regular (*jump, jumped*) marking correctly. After a time, there is a tendency towards over-regularization (*run, runned*). This period of inaccuracy is followed by development of the adult pattern, in which accurate production of irregular forms co-exists with high productivity for the most productive patterns (Rumelhart & McClelland, 1987). Overgeneralization is investigated in detail in recent work by Ramscar and colleagues (Ramscar et al., 2012, 2010), applying insights of learning theory to human language, and to the acquisition of morphological forms. Assuming a theoretical point of view with focus on the link between acquisition and learning on the one hand and expectation on the other, Ramscar investigate the well-documented course of acquisition of irregular plurals (such as *mice*), for which children go through a period of over-regularizing (**mouses*),.

During the middle of the learning trajectory, the form of regularizations is not the same across all children and is variable in use. This has been shown by research on language learning by adult and child learners (MacWhinney, 1983; Hudson Kam & Newport, 2005). The results suggest that individual differences in regularization reflect more general individual differences. These include learning differences, working or short-term memory differences, as well as differences in language experience and vocabulary size. A further source of individual differences in regularization and productivity patterns, we believe, is individual differences in existing vocabulary, and in the strength of the lexical representations and networks for words generated by prior learning. As outlined above, it appears that the strength of lexical representations for words (or chunks) is partially dependent on recurring word fragments in the input stream. This implies that children might regularize word chunks in a different way than the adults in their speech communities.

Beckner (2013, p.86) gives examples of holophrasis from the speech of a young child (age indicated in years and months in brackets). These examples suggest that, at this age, the child accesses and inflects the sequences holistically.

(1) *That what he look likes*. (2;10)
(2) *cool offed* (2;10 - 2;11)
(3) *come offed again* (3;0 – 3;1)
(4) *Why my mama miss mes*? (3:1)
(5) *stand upped* (4;9)
(6) *make sures*, *make sured* (4;11-5;2)

The variability in learning by children is a potential mechanism for new words to enter the community lexicon. Although over-regularization of morphology may appear to simplify the language, in the long run it may support complexity by creating the potential for more productively derived combinations. Research on morphological learning by children has concentrated on learning of major morphological patterns, such as plural or past-tense marking, by younger children. Learning by pre-teens and adolescents is less well studied. These groups have large enough vocabularies to support many less-productive morphological patterns, their cognitive systems differ from those of adults, and they form peer groups with distinctive linguistic patterns. Thus, learning and innovation by these groups has the potential to play an important role in the long-term dynamics of the language. Lastly, people learn new words throughout adulthood; the lexicon is the probably the most plastic part of the linguistic system. This fact is evidenced by the rapid profusion of complex words in some semantic fields that people only master as adults, such as biomedical terms. Taken together, these findings flag the heterogeneity of the individuals in the linguistic community as a major issue in the emergence of morphological patterns.

Computational models of learning

Although morphological patterns are complex and gradient, they are highly learnable, as evidenced by the ability of native speakers to understand, remember, and use new words in their language. In the speech and language engineering community, there has been a substantial effort to develop algorithms with the same capabilities, in order to process novel words and word sequences in applications such as text mining and automatic speech recognition. This effort in computational modeling is complemented by similar efforts with the cognitive science community, which have the different goal of making exact predictions and theory comparisons. The interpretation of computational models can be challenging because of the great differences in architecture between computers and the brain. However, the most successful models support broad conclusions about the learnability and robustness of morphological systems.

Before introducing specific models, let us define some terminology.

- Supervised versus unsupervised learning. An algorithm that achieves supervised learning has access to a set of data that is labeled with a correct analysis. If the input to the algorithm is unlabeled, then the algorithm is unsupervised. If the input includes some features that are indirectly associated with the correct analysis, the algorithm is semi-supervised.

- Batch processing versus incremental processing. A batch processing algorithm takes a large data set as input, and generates an optimal analysis of the whole set. An incremental algorithm begins with an analysis of a small set of data, and then updates the analysis as more data arrive.

- Word token versus word types. Word tokens are individual examples of words in running speech or text. Word types are words in the lexicon as distinguished by elements of form and meaning. The type-frequency of a morphological pattern is the number of distinct words it occurs in, regardless of how often any specific word may occur. We will see that this is a crucial distinction when we address the role of word frequency in change.

Supervised versus unsupervised models

The algorithms of most interest for understanding morphological emergence are unsupervised or semi-supervised algorithms. The labeled data used in fully supervised algorithms include information that is not available to language learners, who of course lack telepathic insight into the abstractions proposed by linguistic experts. Additionally, labeled data sets created by experts typically incorporate assumptions that are inconsistent with gradient word-based theories of morphology, such as the assumption that every word can be exhaustively decomposed into one or more morphemes. Incremental processing is relevant because the learners bootstrap their lexical systems from low levels of language exposure. Generalization of major morphological patterns is evident even in very small children, as reviewed in the last section. The most successful algorithms generalize over word types, reflecting empirical findings that the number of distinct word types exhibiting a morphological pattern is a dominant factor in its productivity. However, word token frequencies are also relevant because of their consequences for incremental learning. In the extreme, a rare word or word sequence may occur so rarely that it is simply unattested in the training set available to the human or computational learner. Frequent words are learned better than rare words, but not in proportion to their frequency: highly predictable information has less impact on memory than less predictable information.

A breakthrough in computational morphology was the development of artificial neural networks, or PDP (parallel distributed processing) models in the 1980s (see Rumelhart &

McClelland 1987; MacWhinney et al. 1989). These algorithms are incremental and moderately supervised. They do not use data labeled with abstract structures, but they do learn by comparing predictions at each point with the correct answer for the prediction. For example, MacWhinney et al. (1989) predicts which form of the German definite article a noun will take. At each stop in the training, the output pattern is compared to the correct pattern. The model succeeds in capturing major generalizations about gender, number, and case without explicitly representing these abstract classes. It also resembles humans in generalizing to new contexts and new words, and in displaying gradience in these generalizations. The ability of PDP models to capture the way similarity and frequency interact in generalization statistics is a signal of the approach's success.

The level of supervision employed in PDP models is high, compared to what children receive while learning a language. While adults may sometimes provide the correct form in response to an incorrect form, such feedback is sporadic. In many engineering applications, no supervision is available. These considerations led to the development of unsupervised algorithms that exploit general statistical properties of lexical systems.

Goldsmith (2001) describes an unsupervised morphological learning model that deduces the morphological inventory of a language from a large corpus, using concepts from information theory. Specifically it assumes that a lexicon can be characterized as a codebook of morphemes, and that this codebook is efficient in the sense that it provides a minimum description length compression of the training corpus. The model performs well on an English data set in finding morpheme boundaries drawn from theoretical analyses.

Creutz & Lagus (2007) present the model family *Morfessor* for the unsupervised induction of agglutinative morphology from raw text data. The goal, as for Goldsmith (2001), is the automatic segmentation of words in a given text to their morpheme-level parts. The algorithm combines a minimum description length principle with a very simple word grammar (prefixes precede stems, which precede suffixes). It scales to large datasets and outperforms the Goldsmith (2001) algorithm on English and Finnish test data.

Goldwater et al. (2011) propose a modeling framework that considers a crucial property of natural languages. This is that, in a corpus of a natural language, the distribution of word frequencies is closely approximated by a power law. They argue that if this statistical property of a corpus is not taken into account, a learning algorithm using the corpus as training data can draw false inferences on its structure. They suggest a more complex damping of word frequency using a two-stage modeling framework, where a generator component generates words and an adaptor component assigns frequency, providing a more parsimonious account of parsing morphological structure.

Batch versus incremental learning

All three models just discussed are not cognitively realistic because they are batch learners. The Morfessor algorithm for Finnish is trained on 16,000,000 words; for a child to do the same calculation, he or she would need to memorize about two years of speech input in unanalyzed form before attempting any analysis (c.f. Daland & Pierrehumbert 2011). A cognitively interesting point is that the models tend to characterize frequently co-occurring words as single units. Treating frequent co-occurrences as single units is an error with respect to gold standards of morphological segmentation. However, it is consistent with theoretical models in psycholinguistics, such as (Ellis, 1996), and with experimental evidence (Tremblay & Baayen, 2010). In particular, it closely approximates the tendency of humans to analyze frequent collocations as single, indivisible chunks (Blumenthal-Dramé, 2013).

Such results suggest that the lexical systems of languages are indeed shaped by the coding pressures that provide the foundation for these algorithms.

The unsupervised models just discussed above are informative about natural linguistic processes because they reflect certain aspects of human learning. They do not capture, however, many findings about categorization. Experimental work on human categorization, exemplified by the contributions of Rosch (1973), Medin & Schaffer (1978), and Nosofsky (1988), has shown that the distribution of previously encountered instances of stimuli plays a crucial role when humans put novel items into categories. The classification of a novel item is a function of the number of previously encountered instances that are nearby (similar) in the relevant featural space. The instance-based theories of categorization developed to account for these findings have been adopted in linguistics, where they provide a flexible schematic architecture for exploring the internal structure of categories and the interactions of multiple factors in defining them. This approach initially gained traction in phonetics, through the work of Goldinger (1996) and Johnson (2005). The Tilburg Memory Based Learner (Timbl, Daelemans et al. 2007) is an implementation of a similar, *nearest neighbor-based* algorithm in morphology. It interprets words as consisting of a number of features and predicts an unknown feature of the word based on the (known) behavior of other words which are most similar to it in terms of known features. This implementation is successfully employed to model the choice of linking morphemes in Dutch nominal compounds by Krott et al. (2001). Skousen (2002) discusses a comparable language-specific model. His analogical model relies on rich memory storage of exemplars which are defined by a set of features. A given target word is compared to the pre-existing exemplars using a subset of user-defined features and exemplars that are most similar to the target – based on the subset of features – are most likely to serve as basis for analogical classification. The implications of token frequency for instance-based learning are already delineated in MacWhinney (1978). This model connects the empirical finding that infrequent irregular forms are over-regularized more frequently than common irregular forms to the fact that irregular forms must be memorized, and the strength of the memory for a form depends on how often it is encountered.

Previously encountered instances are not only relevant in assigning category membership to novel forms. As suggested by the morphological learning models we discussed above, stored instances can also influence morphological parsing. The model by Baayen et al. (1997) makes the prediction that the perceived *parsing* of a word form depends not only on the frequency of the given word form, but also on the ratio of the word form's frequency to the frequency of its parts. This assumption is further substantiated in Hay (2003). For example, the extent to which people are likely to parse a word form as a compound or as a simplex word depends on how often they encountered the parts of the compound versus the whole compound in the past. The dual route access model proposed by Baayen et al. (1997) allows for whole-word access or parsing into parts, and is sensitive to the frequencies and distributions of the relevant word forms. More recently, Baayen (2010) and Baayen et al. (2011) have shown that the predictions of the dual route model are substantially equivalent to a different model which morphological boundaries are not explicitly represented. Their naive discriminative learner acquires associations between word forms and word meanings from language experience, and gradient effects fall out from the way that these associations are activated during processing. The model is incremental. However, it is important to note that it has a relatively high level of supervision, because the system of semantic categories (corresponding to the hidden layer learned in MacWhinney et al. 1989) is supplied as input to the learning process. It remains to be seen how well this approach will extend to the more realistic scenario in which semantic categories and form-meaning associations are learned concurrently.

Lexical learning models, relying on instance-based generalization, have also been proposed for word segmentation and the development of the lexicon (Li, Farkas & MacWhinney, 2004; Monaghan & Christiansen, 2010; Blanchard, Heinz & Golinkoff, 2010).

Crucially, the success of all the models reviewed here hinges on the availability of a vast amount of information to the learner. In batch processing and instance-based models, the learner stores millions of words and word sequences. In a dual route model, the learner keeps track of the frequencies of complex forms as well as the frequencies of stems and affixes. Presumably, this is not through a 'counter in the head', as Baayen (2010) puts it, but rather through intrinsic mechanisms of memory formation, such as connection strength. In Baayen's naive discriminative learner, associations between forms and meanings are incrementally updated through experience. Processing such vast amounts of information is what leads to fine-grained patterns, including gradient effects and detailed differences amongst languages and dialects.

Individual and Social Variation

We have argued that morphology is learnable from the input, and arises as a set of associations across lexical items. Different speakers of the same language may have different lexical systems, as a consequence of encountering and learning different words. If morphology emerges as a generalization over known words, then - because individuals know different words - they are also like to have different morphologies. A number of research studies have confirmed this prediction, finding differences by age, sex, education, and vocabulary level (Frisch et al., 2001; Plag et al., 1999; Baayen, 1994; Baayen & Neijt, 1997; Keune et al., 2006). These results only scratch the surface of questions surrounding heterogeneity, however. The interplay amongst social, linguistic and cognitive factors in determining morphological productivity and creativity is not well understood.

There is work conducted (so far mostly on English), in which systematic attention paid to the role of speaker variables in morphological variation and change, such as gender, social status, education, age, and geographic region of the individuals of people whose language comprises the corpus. This work shows that morphological choices and behaviors are not simply epiphenomena of vocabulary differences, but themselves reflect and carry social meaning.

The body of work done from the early 2000s on English diachronic morphology (and morphosyntax) by Nevalainen, Säily and Raumolin-Brunberg investigated the role of social variables in the *-ity/-ness* variation, the *-(e)th/-(e)s* shift, the *subject ye/you* shift and the *mine, thine / my, thy* shift (e.g. Raumolin-Brunberg, 2006). Based on their results, women seem to have led all of these changes, even though there was a disproportionately low number of identifiable women leaders in the processes. When comparing leaders across all the changes, Raumolin-Brunberg (2006) found that no more than one person was leading more than one change, which suggests that speakers may not generally be characterized as advanced or conservative in their language use, as all authors in her work proved to be selectively adopting the changes. Besides gender, the effects of genre or text type (e.g. Devitt, 1989), social network, social status and region (e.g. Nevalainen & Raumolin-Brunberg, 2003) were also investigated. From text types it is also sometimes possible to ascertain whether the change came from above (from literate and educated writers, in legal texts, etc.) or from below (colloquial genres, personal correspondence, comedies, etc.) Kytö (1993) found evidence that the *-(e)th/-(e)s* shift was an oral change as it appeared earlier in plays, private letters, and trial transcripts before in more literate texts in the south. Nevalainen & Raumolin-Brunberg (2003) observed a similar change from below in the case of the *ye/you* shift. Register played a role when the change was in its early stages as the incoming *you* was used among intimates, and the recessive *ye* by more distant correspondents. However, in later stages of the change the fact that *you* was favored in the capital region, and by women in particular, was more heavily weighted than its register associations. This is corroborated by Raumolin-Brunberg (2005) who shows that the diffusion of the *subject you* was a change from below in terms of social awareness, because *you* was preferred in oral genres and informal registers in the earliest time periods. The results also indicate that the change started in the middle ranks, and women led the change in its critical period of diffusion.

The interaction between gender and education, unsurprisingly, is clearly visible in Säily & Suomela's findings (Säily, 2008; Säily & Suomela, 2009): the productivity of the 'learned' *-ity* is significantly lower in women's use in the 17th-18th c. than men's, while there is no gender difference in the use of *-ness*. This is explained by women's restricted access to education, which was then necessary for a full command of the intricacies of the incoming form *-ity*.

As discussed in detail in Rácz et al. (forthcoming), morphological change is best tracked using comparable corpora from evenly distributed time periods. If such corpora are not available, that is, if the accessible texts differ in length, register, and time of origin, one can still resort to various statistical methods explored in detail by Hilpert and Gries (Hilpert & Gries, 2009; Gries & Hilpert, 2008; Hilpert, 2011; Gries & Hilpert, 2012).

That divergence and structured variation can also emerge through social pressure has also been shown in the laboratory. Roberts (2008) discusses an experimental paradigm where participants have to learn a simple artificial language and use it to trade resources with each other via a computer game. The game is played in groups of four with participants pairing up into two teams. In a given round, a participant is randomly assigned to chat with another participant and exchange resources. The participant does not know, however, whether the other is on their team or on the opposing team. Roberts shows that if team play is encouraged (for instance, by rewarding exchanges with a member of your team and penalizing exchanges with a member of an opposing team), participants will eventually innovate novel forms of the words of the nonce language in order to recognize their team-mates.

The study of individual and social variation across time is crucial to our understanding of morphological variation and the emergence of morphology on the population level and on longer time scales. Available evidence suggests that individuals can have different morphological systems, and that these emerge based on the way language is transmitted amongst individuals in the course of linguistic interactions.

Structure through Transmission

Morphology is acquired and created through the iterated transmission of words between individuals and generations. The transmission process is not perfect, and this imperfection in transmission appears to itself be a driving force in the shaping of morphology.

The drive to attend to and imitate words is fundamental to the convergence of linguistic norms in linguistic communities. As a result, lexical representations and processing mechanisms have consequences at longer time scales, as the output from one generation of speakers serves as the input for the next generation (Komarova & Nowak, 2001; Kirby & Christiansen, 2003). Systematic biases in replication can systematically shape the lexical system (Griffiths & Kalish, 2007). Transmission errors (or mismatches between the input to the subject and their output) can introduce variation into the system. These can cause overall simplification of the system (Komarova & Nowak, 2001) but can act as one source of innovation and structure (Kirby, Cornish & Smith, 2008).

Transmission errors and phonological biases lead to the loss of verbal conjugation in most Germanic languages, including English, Norwegian, or Swiss German. The analytic framework proposed by Komarova & Nowak (2001) deals with this scenario of incomplete learning of a system – in this case, the inflectional suffixation of verbs. Errors in transmission can also lead to structural reanalysis, which in turn defines new combinatorial possibilities. Natural language examples of restructuring and recombination include the productivity of the suffixal neologism *–oholic* (as in *chocoholic, shopaholic)*, coming from *alcoholic*, as well as English back-formations like *baby-sit*. German provides examples of `front-formations', where an adjective is re-interpreted as a participle. For instance, the adjective *bescheuert* (`idiotic') gave rise to the novel verb *bescheuern* (`to put someone in an unfortunate situation').

Ongoing research on the evolutionary origins of language (cf. Arbib, this volume) is largely based on the notion that linguistic evolution is cultural evolution, and that the principles that operate in human cognition take hold in language as well (cf. Beckner & Bybee, this volume). In turn, research on language evolution will rely on experimental designs in which these cognitive principles are shown to affect language structure.

One modeling-based approach to the emergence of inflectional morphology is the Naming Game paradigm (Steels, 2005). The Naming Game provides a concrete, if simplified, mechanism for a shared lexicon to emerge across speakers; according to the game rules, speakers invent and remember new word for a meaning until a shared name that both know is found. In applying this paradigm to explain the word-internal structures of morphology, Steels argues that the main benefit of having a natural language grammar is reducing the computational complexity of semantic interpretation and increasing the chance of communicative success. These aims make it possible for a population of agents to develop a grammar during communication. He frames such a grammar using a model where agents from a population are randomly paired up and then observe the same event. The agents interpret the event, assign structure to it, and then the speaker uses its grammar to communicate the event to the listener. If the utterance used by the speaker matches the utterance that the listener assumed to describe the event, communication was successful. Steels models this interaction using Fluid Construction Grammar, a flexible construction grammar formalism.

Beuls et al. (2012) propose a model for the emergence of *internal agreement* using the naming game paradigm. Grammatical agreement takes place when two linguistic units, typically two words, share certain features with each other, such as *gender* or *number*. They focus on internal agreement, one type of agreement that holds across constituents of a nominal phrase (NP), signaling that these constituents belong together. Gender- and number-based internal agreement is typical in Indo-European languages, like French or German.

The authors adopt the theory, advocated by Corbett (2006), amongst others, that internal agreement starts out with markers which group entities based on semantic content (so that, for example, women are *feminine* and men are *masculine*), and that these markers gradually lose their semantic content as they *emancipate* – lose links with their original semantic referents – and reduce in form. This results in a system of agreement in which the markers are arbitrary. This means that the function of the markers (such as marking femininity) is not recoverable from their forms and that word membership in a given category (such as masculine or feminine grammatical gender) is established through convention. (As we know from *Jules et Jim*, the sun is masculine and the moon is feminine in French, but it is the other way around in German.) Beuls et al. (2012) echo Steels (2005) in arguing that such a system would develop to decrease ambiguity and the related cognitive effort in communication.

Through a series of simulations, the authors show that the use of arbitrary agreement markers (which they call ‘stickers’) decreases cognitive effort during communication. These markers group the attributes of one object together, so that ‘big’, ‘red’ and ‘ball’ get one marker and ‘black’, ‘round’, and ‘chair’ get another. Furthermore, communication becomes more efficient through a decrease in both cognitive effort and inter-agent variation if the markers are based on the semantic content associated with the perceived scenes and if there is a bias in favor of more frequent markers.

The model is certainly a simplified one. It focuses on one type of agreement, internal agreement, one that delineates phrase structure. It has no perceptual or word level variation, so that the two agents always see exactly the same thing and, apart from constituent structure, they can faultlessly communicate to each other what they see – disregarding synonyms and homonyms, for instance. It also avoids a syntactic component, so that word order can never become a possible, competing marker of phrase structure (as it is in English and Mandarin Chinese, for instance). Nonetheless, it provides evidence for a plausible account of the emergence of internal agreement, one that is compatible with grammaticalization-based accounts of the development of *inflectional morphology*. Inflectional morphology is a term applied to word-level variation related to expressing grammatical function, such as internal agreement (inflection in English would include verb past tense and nominal plural formation).

The self-organization of *case systems* is investigated in a paper by van Trijp (2012). Case is usually interpreted as a grammatical function that is obligatorily assigned to nouns and pronouns to mark grammatical roles that largely map to thematic roles, such as *subject*, *object*, and *indirect object* in German or *agent*, *receiver*, *instrument*, or *goal*, etc. in Hungarian (Blake, 1994; É Kiss, 2002). van Trijp uses a naming game to investigate the reason why a case system might emerge in a language, the way a population can self-organize a case system, and, finally, the basis for the differences we observe across case systems. (One example of these differences could be that German uses the *accusative* case to mark goal of movement, whereas Serbian, a Slavic language, uses the *dative*.)

van Trijp (2012) uses a series of simulations in which dyads from a population of agents participate in a naming game where they both describe a scene (for instance, a boy going towards a girl) and then exchange information. The listener has to arrive at a parse of the sentence used by the speaker that matches the scene observed by the listener. So, if the listener sees the boy going towards the girl, it interprets the parse of the speaker’s description that gives it this event structure as the correct one (discarding a parse where the girl is going towards the boy, for instance). Cognitive effort can be calculated based on the number of re-parsings it takes for the listener to get the correct interpretation. Again, the agents interpret the scene in exactly the same way and they are able to communicate without any flaws. Van Trijp shows that a language with case marking entails less cognitive effort than one which has none, given that no other means of marking event structure are available. He also shows that abstract case marking, where cases generalize over roles in the event structure (thematic roles, such as ‘agent’, ‘goal’, or ‘beneficiary’) fares better than an idiosyncratic case system which applies a new case to any given situation. This is true both in terms of cognitive effort, coherence, systematicity, expressivity, and learnability.

The paper also offers a model of the emergence and variation of a case system, based on essentially the same principles. Like comparable work in the Naming Game paradigm, it gives a simplified account of morphological emergence. It does, however, give an essential answer to the question why case systems exist in the first place, given that some languages rely heavily on the lexis and prior knowledge of the language users to infer event structure, instead of morphological case marking or word order. If we are willing to accept the abstraction of disregarding word order as a way of expressing argument structure, the Naming Game models suggest a way for general cognitive principles to do so.

Regier, Kemp & Kay (this volume) address semantic variation as shaped by simple functional principles. Their model is built on notions of efficient communication. They look at various existing semantic configurations in different languages in terms of informativeness and simplicity, showing that there are many near-optimal solutions to the same problem (such as coding color terms). Regier, Kemp & Kay's approach to the evolution of linguistic system is more detailed than the Naming Game paradigm. An implementation in an iterated model would be of great interest, so that its predictions on language evolution can be better understood.

The experimental paradigm of iterated learning across generations (Griffiths & Kalish, 2007; Kirby et al., 2008; Reali & Griffiths, 2009) has recently arisen as a way to investigate word transmission. This paradigm is inspired by the traditional game 'Telephone' or 'Chinese Whispers', in which one person whispers a message to another, who repeats it to the next person, and so on until it reaches the end of the chain. The transmission errors accumulate, so the final message may be quite different from the initial message. In the iterated learning paradigm, each person learns a mini-language rather than a single message. Their productions on a final test serve as input to the next person in the chain. The paradigm provides a model system for language transmission across generations. Using this paradigm, Kirby et al. (2008) show that a language with random pairings of word forms to meanings tends to become regularized over generations into a language with a morpho-syntactic structure. The accidental similarities across amongst forms that come about through imperfect learning in the transmission chain can serve to define combinatorial units. These in turn become available to refer to previously unseen events. Reali & Griffiths (2009) explore the role of word frequency in the transmission process. When there are two competing words for the same object, learning by adults generally tracks the frequencies in the input. However, the iterated learning paradigm reveals slight biases that may be too subtle to detect reliably within a single generation. Griffiths & Kalish (2007) present a more general analysis of how the iterated learning paradigm reveals the prior assumptions that learners share as they approach the task.

Very little research on learning of artificial languages has included child or adolescent learners. One research strand on children's language learning has focused on how children adapt variable input to reduce learning demands (Hudson Kam & Newport, 2005). In learning the grammatical morphemes of an artificial language, children in Hudson Kam & Newport (2005) reduced variability in the input language to create a language that was regularized in a single generation, unlike the adult participants who produced inconsistent output from inconsistent input. An important finding is that in the process of reducing variability, children created new forms that did not exist in the input. Hudson et al. concluded that children and adults learned differently, with children imposing statistical properties on inconsistent input to form a consistent, regular language. This latter suggestion fits with K. Smith's (2011) suggestion that the child drive to reduce variability of the input in a word learning task could be attributed to the need to increase predictability in the learning task. If so, then children could be a driving force in innovation in language, in morphological creativity and word creation.

The finding that children create new forms goes back at least to Berko (1958). The relevant point here is that children do so more extensively than adult language users. (Pierrehumbert 2006 provides an example of adults generalizing morpho-phonological patterns to novel forms. In this study, young adults extent the pattern of velar softening – as in *electric-electricity* – to pseudo-Latin nonce words such as *hovac-hovacity*.) Ramscar et al. (2012) link this difference between children and adults to the effect of pre-existing categories on adult discriminative learning. This finding is important as it explains why children and adult learners behave differently, lending more explanatory power to theories that emphasize the role of children in morphological change.

A further limitation of the initial iterated learning results is that each learner receives input from only one member of the previous generation. In fact, children learn a language from several adults. As the mathematical analysis of Niyogi & Berwick (2009) shows, the variety of different language systems that can ultimately form is significantly affected by the number of distinct input streams to each learner. This is just one of the social variables that have the potential to affect the detailed dynamics of the lexicon. Social biases that learners bring to the task are not necessarily shared, and the space of possible lexicons resulting from an iterated learning task could exhibit systematic dependencies on the social perceptions of the learners.

Iterated learning paradigms do not lead to the emergence of a structure that is comparable in complexity to the morphology of natural languages – the emergent patterns reported by Kirby et al. (2008) are relatively simple. This is due to the limits in time or population scale imposed by laboratory experiments. Relying on longer time scales and larger populations we can study linguistic processes on the scale that they operate. Laboratory experiments, however, highlight that general cognitive principles can directly influence the way linguistic data are perceived and transmitted at an individual level, as well as the extent to which the effects of these can accumulate through a longer time period.

Experimental and computational work on language evolution offers important insights into how communicative principles – like striving towards communicative efficiency – and general cognitive principles – such as principles of human categorization – can shape a language system. Iterated learning paradigms, both within and across generations, are illuminating because they show how relatively small errors or biases can, by accumulating over time, lead to qualitatively distinct morphological outcomes. The interaction of random variation with systematic functional pressures also helps to explain typological diversity.

Morphology in use: The role of word frequency in change

The above section describes iterated learning experiments which demonstrate how structure, including morphological structure, change over multiple generations. It is also clear from studying trajectories of change in morphology that repeated use-in-context affects the course of evolution of morphological patterns in other ways. This is most clearly seen by considering the role of frequency of usage, and its apparent effect on the evolution of morphological patterns and morphologically complex forms.

Frequency can act as both a conservative and innovative force upon trajectories of diachronic change, leading to two seemingly contradictory effects. On one hand, token frequency can be a conservative force protecting high-frequency structures from analogical leveling. As Bybee & Thompson (1997) have shown, frequently used expressions are often resistant to analogical change. For instance, in English there has been continuous pressure to regularize irregular verb forms. Since the time of Old English, nearly 200 verbs have lost the stem vowel alternation and have adopted the regular past tense form. Synchronically, we find that most of the verbs that are still irregular are very frequent. Bybee & Thomson's hypothesis is that the frequent use has strengthened their representation in memory, which is why they have resisted the pressure from analogical change. An example is given by Anshen & Aronoff (1988), who hypothesize that in surviving singular-plural ablaut pairs in English (such as *foot* ~ *feet*) the plural form has either higher frequency than the singular, the two have about the same frequency, or the plural has absolute high frequency. All lower frequency forms have succumbed to the analogical force of regularization.

The mechanisms for this finding have already been discussed above. Frequent forms are reliably encountered in everyone's experience, and robustly encoded though many instances (in instance-based models) or reliable statistics (in other approaches). Infrequent forms may never be encountered, in which case more regular forms are productively created by recombining parts of other words. Or else, they may be encountered but poorly learned, placing them at a disadvantage in competing with productively formed combinations.

Berg (2011) also claims that morphemes and words showing allomorphy should always be more frequent than invariant units within the same domain. The diachronic counterpart of this hypothesis is that allomorphy should emerge in high-frequency items earlier, and be lost in high-frequency items later, than in low-frequency ones. In other words, allomorphy should be diachronically stable in high-frequency units but more ephemeral in less frequent ones. These observations follow from the fact that larger statistical samples are available for frequent words or morphemes, making the learning of sub-cases possible. This does not mean, however, that low frequency items are necessarily invariant; in many morphologically languages, some allophonic processes apply to all words. Vowel harmony in Finnish, Estonian, or Hungarian provides an example. Such systems arise under Berg's account when alternations exhibited for high-frequency units are generalized and applied to low frequency words as well, as in the case of (fully productive) Hungarian vowel harmony.

On the other hand, grammaticalization theory claims that high token frequency forms are very prone to phonetic reduction and loss of semantic features. They are therefore held to be prone to phonetic and phonological processes which can erode morphological boundaries, and to processes causing semantic drift. Diessel (2007) argues that frequency is a driving force of phonetic reduction and grammaticalization (cf. Bybee 2006). Particularly high frequency of use of lexical items in certain contexts can also lead to movement along the so-called 'cline' of grammaticalization posited by Hopper & Traugott (2003). A classic example is the development of French particles. Latin expresses negation with the particle *nōn,* which was then reinforced in French by the word *pas*, meaning `step'. *Pas* later lost its meaning entirely in negative constructions and, after the disappearance of *non*, started to serve as the exclusive negative marker.

Over long periods of time, frequent forms can move along the cline from content word - grammatical word - clitic - inflectional affix. A paucity of longitudinal data means, however, that it is difficult to distinguish between the effects of high frequency and the effects of increasing frequency. If a word form was less frequent before, and becomes more frequent, then the functional pressure towards minimum length encoding (described above in section 4) would also shorten it. It is less clear why a form that is high frequent and very well-learned should become shortened over time even if its frequency is stable. One possibility is that frequencies are not generally stable, and that high-frequency words tend to become even more frequent through a "rich-get-richer" process (contrary to the assumptions of Reali & Griffiths, 2009). Another possibility is that the phonological and semantic dynamics studied in grammaticalization theory primarily concerns situations in which the frequency of a form increases over time due to exogenous factors.

Frequency distributions are thus crucial to the evolving dynamic system because they affect aspects of morphological production, processing, and productivity. Shifts in frequency will change the dynamics of the system. Concomitantly, shifts in inferred morphological structure, of course, also lead to shifts in frequency distributions. Thus these effects all constantly evolve together as part of a tightly interconnected system.

Conclusion

Morphology is set of associations between meaning and form. These associations can be partial, overlapping, messy and gradient. The associations are learned as generalizations over lexical knowledge. Much of this learning takes place as part of early language acquisition, but generalizations continue to form and evolve throughout adulthood. There can be considerable individual and social variation in morphological knowledge and use. Individuals differ in the lexical inventories over which generalizations can be formed. They also differ in terms of their social identities and alliances, and their cognitive strategies, all of which may impact morphological learning and usage.

The morphological system is transmitted between individuals and generations. The transmission process is imperfect, and, over time, this imperfect transmission can create or destroy morphological patterns. The production and perception of morphology is highly sensitive to frequency and predictability in context, in a way that shapes trajectories of language change.

Even though emergentist approaches, as we argue above, allow for a better approximation of the inherent gradience of morphological patterns, there are several remaining challenges for morphological theory. These include the organization of morphological paradigms (Baayen et al., 2011; Albright, 2002; Rebrus & Törkenczy, 2000; Fabri, 2009), paradigmatic gaps (Rice, 2003; Lukács et al., 2009; Törkenczy, 2002; Daland et al., 2007), allomorph selection (Bye, 2008), the way the structural properties of a language affect morphological patterns, as in the case of grammaticalization (Bybee, 2010), as well as instances where observed productivity is either much larger or much smaller than what linguistic models would lead us to expect – this interesting problem is scrutinized by Pierrehumbert (2010).

As much as the literature discussed in this chapter falls short of providing a complete, comprehensive view of natural language morphology, we believe that the inherent properties of morphological patterns warrant an emergentist approach. Word structure ’emerges’ through a complex interplay of learnability, generalizability, frequency and social dynamics, which together shape the formation, evolution, and production of morphological patterns.

References

Albright, Adam C. 2002. *The identification of bases in morphological paradigms*: University of California, Los Angeles dissertation.

Anshen, Frank & Mark Aronoff. 1988. Producing morphologically complex words. *Linguistics* 26. 641–655.

Arbib, Michael. 2014. Language evolution – an emergentist view. In Brian MacWhinney & William O'Grady (eds.), *The handbook of language emergence*, Wiley-Blackwell.

Baayen, R Harald. 1994. Derivational productivity and text typology. *Journal of quantitative linguistics* 1(1). 16–34.

Baayen, R Harald. 2010. Demythologizing the word frequency effect: A discriminative learning perspective. *The Mental Lexicon* 5(3). 436–461.

Baayen, R. Harald, Ton Dijkstra & Robert Schreuder. 1997. Singulars and plurals in Dutch: Evidence for a parallel dual-route model. *Journal of Memory and Language* 37(1). 94–117.

Baayen, R Harald, Petar Milin, Dusica Filipović DJurdjević, Peter Hendrix & Marco Marelli. 2011. An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological review* 118(3). 438.

Baayen, R Harald & Anneke Neijt. 1997. Productivity in context: a case study of a Dutch suffix. *Linguistics* 35(3). 565–588.

Beckner, Clayton. 2013. Quantitative determinants of prefabs: a corpus-based, experimental study of multiword units in the lexicon. University of New Mexico PhD Thesis.

Beckner, Clayton & Joan Bybee. 2014. Emergence at the crosslinguistic level: Attractor dynamics in language change. In Brian MacWhinney & William O'Grady (eds.), *The handbook of language emergence*, Wiley-Blackwell.

Berg, Thomas. 2011. A diachronic frequency account of the allomorphy of some grammatical markers. *Journal of Linguistics* 47(1). 31–64.

Berko, Jean. 1958. The child's learning of English morphology. Radcliffe College PhD Thesis.

Beuls, Katrien, Luc Steels & Sebastian Höfer. 2012. The emergence of internal agreement systems. In Luc Steels (ed.), *Experiments in language evolution*, Amsterdam: John Benjamins.

Blanchard, Daniel, Jeffrey Heinz & Roberta Golinkoff. 2010. Modeling the contribution of phonotactic cues to the problem of word segmentation. Journal of Child Language. 37:03. 487-511.

Blake, Barry J. 1994. *Case.* Cambridge Textbooks in Linguistics. Cambridge University Press.

Blumenthal-Dramé, Alice. 2013. *Entrenchment in usage-based theories. what corpus data do and do not reveal about the mind* Topics in English linguistics 83. Berlin: De Gruyter Mouton.

Bybee, Joan L. 1985. *Morphology: A study of the relation between meaning and form*, vol. 9. Amsterdam: John Benjamins.

Bybee, Joan L. 2006. *Frequency of use and the organization of language.* Oxford University Press.

Bybee, Joan L. 2010. *Language, usage and cognition*, vol. 98. Cambridge University Press.

Bybee, Joan L. & Sandra Thompson. 1997. Three frequency effects in syntax. In *Proceedings of the 23rd annual meeting of the Berkeley Linguistics Society: General session and parasession on pragmatics and grammatical structure*, 378–388. Berkeley: Berkeley Linguistics Society.

Bye, Patrik. 2008. Allomorphy - selection, not optimization. In Sylvia Blaho, Martin Krämer & Patrik Bye (eds.), *Freedom of analysis*, Berlin & New York: Mouton de Gruyter.

Carstairs-McCarthy, Andrew, Miriam Butt & Tracy Holloway King. 2004. What morphology can tell us about grammar. In *Proceedings of the LFG'04 Conference. Stanford: CSLI Publications*, .

Corbett, Greville G. 2006. *Agreement*, vol. 109. Cambridge University Press.

Creutz, Mathias & Krista Lagus. 2007. Unsupervised models for morpheme segmentation and morphology learning. *ACM Transactions on Speech and Language Processing (TSLP)* 4(1). 3.

Daelemans, Walter, Jakub Zavrel, Ko Van der Sloot & Antal Van den Bosch. 2007. TiMBL: Tilburg Memory Based Learner, version 6.1, Reference Guide. Tech. rep. ILK Research Group Technical Report Series no. 07-07.

Daland, Robert & Janet B Pierrehumbert. 2011. Learning diphone-based segmentation. *Cognitive Science* 35(1). 119–155.

Daland, Robert, Andrea D. Sims & Janet B. Pierrehumbert. 2007. Much ado about nothing: A social network model of russian paradigmatic gaps. In *Annual meeting-association for computational linguistics*, vol. 45 1, 936.

Dell, Gary S, Myrna F Schwartz, Nadine Martin, Eleanor M Saffran & Deborah A Gagnon. 1997. Lexical access in aphasic and nonaphasic speakers. *Psychological review* 104(4). 801.

Devitt, Amy J. 1989. *Standardizing written English: Diffusion in the case of Scotland 1520-1629*. Cambridge University Press.

Diessel, Holger. 2007. Frequency effects in language acquisition, language use, and diachronic change. *New Ideas in Psychology* 25. 108–127.

É Kiss, Katalin. 2002. *The syntax of Hungarian*. Cambridge University Press.

Ellis, Nick C. 1996. Working memory in the acquisition of vocabulary and syntax: Putting language in good order. *The Quarterly Journal of Experimental Psychology: Section A* 49(1). 234–250.

Fabri, Ray. 2009. Stem allomorphy in the Maltese verb. *Ilsienna – Our Language* 1. 1–20.

Frisch, Stefan A, Nathan R Large, Bushra Zawaydeh & David B Pisoni. 2001. Emergent phonotactic generalizations in English and Arabic. *Typological studies in Language* 45. 159–180.

Goldinger, Stephen D. 1996. Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology-Learning Memory and Cognition* 22(5). 1166–1182.

Goldsmith, John. 2001. Unsupervised learning of the morphology of a natural language. *Computational linguistics* 27(2). 153–198.

Goldwater, Sharon, Thomas L Griffiths & Mark Johnson. 2011. Producing power-law distributions and damping word frequencies with two-stage language models. *Journal of Machine Learning Research* 12. 2335–2382.

Gries, Stefan Th. & Martin Hilpert. 2008. The identification of stages in diachronic data: Variability-based neighbour clustering. *Corpora* 3(1). 59–81.

Gries, Stefan Th. & Martin Hilpert. 2012. Variability-based neighbor clustering: A bottom-up approach to periodization in historical linguistics. In Terttu Nevalainen & Elizabeth Closs Traugott (eds.), *The Oxford Handbook of the History of English*, Oxford University Press.

Griffiths, Thomas L & Michael L Kalish. 2007. Language evolution by iterated learning with bayesian agents. *Cognitive Science* 31(3). 441–480.

Hakkani-Tür, Dilek Z, Kemal Oflazer & Gökhan Tür. 2002. Statistical morphological disambiguation for agglutinative languages. *Computers and the Humanities* 36(4). 381–410.

Haspelmath, Martin & Andrea D. Sims. 2010. *Understanding morphology*. London: Arnold. 2nd edn.

Hay, Jennifer B. 2002. From Speech Perception to Morphology: Affix Ordering Revisited. *Language* 78(3): 527-555.

Hay, Jennifer B. 2003. *Causes and consequences of word structure*. New York and London: Routledge.

Hay, Jennifer B & R Harald Baayen. 2005. Shifting paradigms: gradient structure in morphology. *Trends in Cognitive Sciences* 9(7). 342–348.

Hilpert, Martin. 2011. Diachronic collostructional analysis: how to use it and how to deal with confounding factors. In Kathryn Allan & Justyna A. Robinson (eds.), *Current methods in historical semantics* (Topics in English linguistics 73), 133–160. Berlin: De Gruyter Mouton.

Hilpert, Martin & Stefan Th. Gries. 2009. Assessing frequency changes in multistage diachronic corpora: Applications for historical corpus linguistics and the study of language acquistition. *Literary and Linguistic Computing* 24(4). 385–401.

Hirsimäki, Ts, Js Pylkkönen & M. Kurimo. 2009. Importance of higher-order n-gram models in morph-based speech recognition. *IEEE Transactions on Audio, Speech, and Language Processing* 17(4). 724–732.

Hopper, P. C. & E. J. Traugott. 2003. *Grammaticalization* Cambridge Textbooks in Linguistics. Cambridge University Press.

Hudson Kam, Carla L & Elissa L Newport. 2005. Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development* 1(2). 151–195.

Johnson, Keith. 2005. Speaker normalization in speech perception. In R. Remez D. B. Pisoni (ed.), *The handbook of speech perception*, 363–389. Oxford: Blackwell.

Kan, Pui Fong & Jennifer Windsor. 2010. Word learning in children with primary language impairment: A meta-analysis. *Journal of Speech, Language and Hearing Research* 53(3). 739.

Keune, Karen, Roeland Van Hout & R. Harald Baayen. 2006. Socio-geographical variation in morphological productivity in spoken Dutch: a comparison of statistical techniques. In *Proceedings of JADT 2006*, .

Kirby, Simon & Morten H Christiansen. 2003. From language learning to language evolution. *Studies in the evolution of language* 3. 272–294.

Kirby, Simon, Hannah Cornish & Kenny Smith. 2008. Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences* 105(31). 10681–10686.

Komarova, Natalia L & Martin A Nowak. 2001. The evolutionary dynamics of the lexical matrix. *Bulletin of Mathematical Biology* 63(3). 451–484.

Kornai, András. 1994. *On Hungarian morphology.* Linguistica. Budapest: Hungarian Academy of Sciences.

Krott, Andrea, R Harald Baayen & Robert Schreuder. 2001. Analogy in morphology: modeling the choice of linking morphemes in Dutch. *Linguistics* 39(1; ISSU 371). 51–94.

Kytö, Merja. 1993. Third-person present singular verb inflection in early British and American English. *Language Variation and Change* 5. 113–139.

Li, Ping, Igor Farkas & Brian MacWhinney. 2004. Early lexical development in a self-organizing neural network. Neural Networks 17:8–9. 1345-1362.

Lukács, Ágnes, Péter Rebrus & Miklós Törkenczy. 2009. Defective verbal paradigms in Hungarian – description and experimental study. In *Defective paradigms: Missing forms and what they tell us*, Oxford University Press.

MacWhinney, Brian. 1978. The acquisition of morpho-phonology. *Monographs of the Society for Research in Child Development.* 1-123.

MacWhinney, Brian. 1983. Miniature linguistic systems as tests of universal operating principles. *Journal of Psycholinguistic Research* 12. 467-478.

MacWhinney, B. J., J. Leinbach, R. Taraban & J. L. McDonald. 1989. Language learning: Cues or rules? *Journal of Memory and Language* 28(3). 255–277.

Magnuson, James S, James A Dixon, Michael K Tanenhaus & Richard N Aslin. 2007. The dynamics of lexical competition during spoken word recognition. *Cognitive Science* 31(1). 133–156.

Magnuson, James S, Michael K Tanenhaus, Richard N Aslin & Delphine Dahan. 2003. The time course of spoken word learning and recognition: studies with artificial lexicons. *Journal of Experimental Psychology: General* 132(2). 202.

Medin, Douglas L & Marguerite M Schaffer. 1978. Context theory of classification learning. *Psychological review* 85(3). 207.

Miller, George A. 1995. Wordnet: a lexical database for English. *Communications of the ACM* 38(11). 39–41.

Monaghan, Padraic & Morten H. Christiansen. 2010. Words in puddles of sound: modeling psycholinguistic effects in speech segmentation. Journal of Child Language. 37:03. 545-564.

Nevalainen, Terttu & Helena Raumolin-Brunberg. 2003. *Historical sociolinguistics: Language change in Tudor and Stuart England*. London: Longman.

Niyogi, Partha & Robert C Berwick. 2009. The proper treatment of language acquisition and change in a population setting. *Proceedings of the National Academy of Sciences* 106(25). 10124–10129.

Nosofsky, Robert M. 1988. Similarity, frequency, and category representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 14(1). 54–65.

Pierrehumbert, Janet B. 2003. Probabilistic phonology: Discrimination and robustness. In Rens Bod Jennifer B. Hay, Stefanie Jannedy (ed.), *Probabilistic linguistics*, Cambridge: MIT Press.

Pierrehumbert, Janet B. 2006. The statistical basis of an unnatural alternation. *Laboratory phonology* 8. 81-107.

Pierrehumbert, Janet B. 2010. The dynamic lexicon. In A. Cohen, M. Huffmann & C Fougeron (eds.), *Handbook of laboratory phonology*, 173–183. Oxford University Press.

Plag, Ingo, Christiane Dalton-Puffer & R Harald Baayen. 1999. Productivity and register. *Journal of English Language and Linguistics* 3. 209–288.

Rácz, Péter, Viktória Papp & Jennifer B. Hay. forthcoming. Frequency and corpora. In Andrew R Hippisley (ed.), *Cambridge handbook of morphology*, Cambridge University Press.

Ramscar, Michael, Melody Dye & Malte Hubner. Forthcoming. When the fly flied and when the fly flew: How semantics can make sense of inflection. *Language and Cognitive Processes*.

Ramscar, Michael, Melody Dye & Joseph Klein. 2013. Children value informativity over logic in word learning. *Psychological science* 24(6). 1017–1023.

Ramscar, Michael, Melody Dye & Stewart M McCauley. 2012. Expectation and error distribution in language learning: The curious absence of "mouses" in adult speech. *Language, accepted* .

Ramscar, Michael, Daniel Yarlett, Melody Dye, Katie Denny & Kirsten Thorpe. 2010. The effects of feature-label-order and their implications for symbolic learning. *Cognitive Science* 34(6). 909–957.

Raumolin-Brunberg, Helena. 2005. The diffusion of subject *you*: A case in historical sociolinguistics. *Language Variation and Change* 17(1). 55–73.

Raumolin-Brunberg, Helena. 2006. Leaders of linguistic change in Early Modern England. In Roberta Facchinetti & Matti Rissanen (eds.), *Corpus-based studies of diachronic English* (Linguistic Insights 31), Peter Lang.

Reali, Florencia & Thomas L Griffiths. 2009. The evolution of frequency distributions: Relating regularization to inductive biases through iterated learning. *Cognition* 111(3). 317–328.

Rebrus, Péter & Miklós Törkenczy. 2000. Morfofonológiai jelenségek és a Lexikon [Morphophonological processes and the Lexicon]. In Ferenc Kiefer (ed.), *Strukturális magyar nyelvtan 3. morfológia [structural hungarian grammar 3. morphology]*, 763–947. Budapest: Akadémiai Kiadó.

Regier, Terry, Charles Kemp & Paul Kay. 2014. Word meanings across languages support efficient communication. In Brian MacWhinney & William O'Grady (eds.), *The handbook of language emergence*, Wiley-Blackwell.

Rice, Curt. 2003. When nothing is good enough: Dialectal variation in norwegian imperatives. *Nordlyd* 31/2.

Roberts, Gareth. 2008. Language and the free-rider problem: An experimental paradigm. *Biological Theory* 3(2). 174–183.

Rosch, Eleanor H. 1973. Natural categories. *Cognitive psychology* 4(3). 328–350.

Rumelhart, David E & James L McClelland. 1987. Learning the past tenses of English verbs: Implicit rules or parallel distributed processing. *Mechanisms of language acquisition* 195–248.

Säily, Tanja. 2008. *Productivity of the suffixes -ness and -ity in 17th century English letters: A sociolinguistic approach*. Helsinki University of Helsinki Masters Thesis.

Säily, Tanja & Jukka Suomela. 2009. Comparing type counts: the case of women, men and *-ity* in early English letters. In Antoinette Renouf & A. Kehoe (eds.), *Corpus linguistics: Refinements and reassessments*, 87–109. Amsterdam: Rodopi.

Skousen, Royal. 2002. *Analogical modeling: An exemplar-based approach to language*. Amsterdam: John Benjamins.

Smith, Kenny. 2011. Learning bias, cultural evolution of language, and the biological evolution of the language faculty. *Human Biology* 83(2). 261–278.

Smith, Linda B, Chen Yu & Alfredo F Pereira. 2011. Not your mother's view: The dynamics of toddler visual experience. *Developmental science* 14(1). 9–17.

Steels, Luc. 2005. What triggers the emergence of grammar. In *Proceedings of the Second International Symposium on the Emergence and Evolution of Linguistic Communication (AISB 2005 Convention, University of Hertfordshire, Hatfield, UK). AISB*, .

Stokes, Stephanie F. 2010. Neighborhood density and word frequency predict vocabulary size in toddlers. *Journal of Speech, Language and Hearing Research* 53(3). 670.

Stokes, Stephanie F., D. Bleses, H. Basbøll & C. Lambertsen. in press. Crosslinguistic evidence of statistical learning in emerging lexicons. *Journal of Speech, Language, and Hearing Research* .

Stokes, Stephanie F, Sophie Kern & Christophe dos Santos. 2012. Extended statistical learning as an account for slow vocabulary growth. *Journal of child language* 39(1). 105.

Tily, Harry, Susanne Gahl, Inbal Arnon, Neal Snider, Anubha Kothari & Joan Bresnan. 2009. Syntactic probabilities affect pronunciation variation in spontaneous speech. *Language and Cognition* 1(2). 147–165.

Törkenczy, Miklós. 2002. Absolute phonological ungrammaticality in output-biased phonology. *Approaches to Hungarian* 8. 311–324.

Tremblay, Antoine & R Harald Baayen. 2010. Holistic processing of regular four-word sequences: A behavioral and erp study of the effects of structure, frequency, and probability on immediate free recall. *Perspectives on formulaic language: Acquisition and communication* 151–173.

van Trijp, Remi. 2012. The evolution of case systems for marking event structure. In Luc Steels (ed.), *Experiments in cultural language evolution*, vol. 3, 169. Amsterdam: John Benjamins.

Biographical note

Péter Rácz is a post-doctoral research fellow at the New Zealand Institute of Language Brain and Behaviour (NZILBB), working on the Wordovators project. The project is aimed at gaining a better understanding of how words are created and used, in a close collaboration between Northwestern University and NZILBB, funded by the John Templeton Foundation. He is very excited about language modeling and socially motivated linguistic variation.

Janet B. Pierrehumbert is Professor of Linguistics at Northwestern University. Her research uses experimental and computational methods to study the sound structure of language. She is a member of the American Academy of Arts and Sciences, a Fellow of the Linguistic Society of America, and a Fellow of the Cognitive Science Society. She is the leader of the Wordovators project.

Jennifer B. Hay is Professor in Linguistics at the University of Canterbury and the founder and director of the New Zealand Institute of Language Brain and Behaviour. She has published articles on morphology, laboratory phonology and sociophonetics. She is one of the principal investigators of the Wordovators project.

Viktória Papp is a lecturer at the Department of Linguistics at the University of Canterbury. Her main research interests include gender, sociolinguistic variation, corpus linguistics, and forensic phonetics.