Title
A network model of English derivational morphology

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 46(0)

Authors
Monakhov, Sergei
Diessel, Holger

Publication Date
2024

A Network Model of English Derivational Morphology

Sergei Monakhov (sergei.monakhov@uni-jena.de)
Friedrich-Schiller-Universität Jena,
Fürstengraben 1, 07743 Jena, Germany

Holger Diessel (holger.diessel@uni-jena.de)
Friedrich-Schiller-Universität Jena,
Fürstengraben 1, 07743 Jena, Germany

Abstract

Models of word recognition and production diverge on the question of how to represent complex words. Under the morpheme-based approach, each morpheme is represented as a separate unit, while under the word-based approach, morphemes are represented in lexical networks. The word-based approach is consistent with construction morphology and recent research on the grammar network. However, while the network view of constructions has become popular in recent years, there is little computational and experimental research on this topic. In the current study, we used a computational network model (based on graph theory) and an experiment to investigate the Romance component of English morphology. Specifically, we provide evidence that complex words can be conceptualised as paths in a weighted directed network of morphemes.

Keywords: grammar network; complex words; construction morphology; lexical knowledge; English morphology

Introduction

There are two general approaches to the study of morphology. The morpheme-based approach, in which morphemes are the basic units of morphological structure (Halle, 1973; Kiparsky, 1982), and the word-based approach, in which morphological structure is derived from complex words (Matthews, 1972; Bybee 1985; Blevins, 2016). The two approaches make very different assumptions about how complex words are formed and processed (Butterworth, 1983; Forster and Taft, 1994; Smolka, Preller, and Eulitz, 2014; Milin, Smolka, and Feldman, 2017; Smolka, Libben, and Dressler, 2019; Plag and Winther Balling, 2020). In the morpheme-based approach, complex words are generated by rules, similar to rules in syntax, whereas in the word-based approach, they are structured by paradigmatic relations.

The morpheme-based approach provides an intuitive explanation for regular patterns of morphology, but many complex words are irregular or idiosyncratic. Morpheme boundaries are often gradient and fuzzy (Hay, 2002), and the meaning of complex words does not always reflect the meaning of their morphological parts (Plag, 2018). Moreover, there is a large body of research showing that frequency of use affects the processing of complex words.

The word-based approach is more flexible in dealing with morphological idiosyncrasies and frequency effects. Some word-based models use correspondence rules that are independent of usage and experience (Aronoff, 1994), but most word-based models rely on mechanisms such as pattern recognition and entrenchment (which are sensitive to frequency of use) in order to explain how complex words are formed and processed. In these usage-based models, morphological structure emerges from recurrent patterns of language use.

A pioneer of usage-based morphology is Bybee (1985). In Bybee's network model, each word has a unique representation that reflects the language users' experience with particular lexical units. Bybee's model is consistent with various other network models of morphology (Hay and Baayen, 2005; Booij, 2010) and with recent research on the grammar network (Diessel, 2019, 2023; Schmid, 2020). However, while the network view of grammar has become popular in recent years, there is little computational and experimental research on this topic.

In this paper, we present a computational network model and an experiment to investigate the mechanisms that govern the formation and processing of complex words. Our model is based on network science, a theoretical and computational approach that employs the mathematical tools of graph theory to study complex systems in the real world (Barabási, 2016). The network science approach offers an alternative to connectionism and related computational approaches. Forty years ago, Rumelhart and McClelland (1986) presented an artificial neural network (ANN) of the English past tense, inspired by Bybee's early work on morphology (Bybee and Slobin, 1982), which sparked an intense debate about the nature of grammar (Pinker and Prince, 1988; Elman et al., 1996).

ANNs are subsymbolic systems that represent words by a set of nodes with different activation values. A typical ANN consists of a layer of input nodes, one or more layers of hidden nodes, and a layer of output nodes. ANNs can learn to map a given input pattern representing a linguistic symbol (such as the base form of an English verb) to a particular output pattern representing a related symbol (such as the past tense form of that verb).

ANNs are powerful pattern recognition engines that can learn all kinds of associations from statistical regularities in the training data, but they suffer from a common weakness: ANNs are often difficult to interpret. Some researchers argue that the global activation patterns of the network represent linguistic generalizations (Elman et al., 1996), but the relationship between the activation patterns of an ANN and linguistic generalizations is very abstract and subject to an ongoing debate (Seidenberg and Plaut, 2014).

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

More recently, Baayen and colleagues have developed a related approach to morphology that uses more concrete representations than ANNs and a different learning mechanism, called discriminative learning (Baayen et al., 2011; Milin et al., 2017). In this approach, word forms are represented by a set of n-grams of different sizes. The main task of a discriminative learning network is to map a given pattern of n-grams onto a particular pattern of semantic features. In a series of experiments, Baayen and colleagues have shown that their model makes excellent predictions about a variety of factors that influence morphological processing.

Discriminative learning models are less abstract than traditional ANNs, but they also eliminate the notion of morpheme. To be sure, the traditional notion of morpheme is problematic. Morphological structure is gradient, and the idea that complex words can always be divided into discrete morphemes is inconsistent with the linguistic data (Hay, 2002; Hay and Baayen, 2005). This does not mean, however, that morphological analysis can dispense with the notion of morpheme. There is ample evidence that people recognize (meaningful) parts in complex words and that the forms of complex words are governed by morphophonemic processes that presuppose some notion of morpheme (see Amenta and Crepaldi (2012) for a review).

Consistent with this evidence, we devised a network model of morphology in which morphemes are analysed as gradient symbolic units. Our network model, as well as connectionist and discriminative learning models, is consistent with the usage-based thesis that grammar is derived from speakers' experience with concrete linguistic tokens. However, in our model, each morpheme is represented as a separate node in the network. Importantly, the morphological nodes of our network are not basic units of language structure, as in the traditional morpheme-based approach, but emergent entities that are derived from repeated patterns of language use.

Our model accounts for both the emergence of morphological structure and the processing and formation of complex words. With regard to the emergence of morphological structure, we argue that a morphological network is constructed from individual words by means of distributional analysis. Distributional analysis is driven by two general factors: similarity and frequency. When two or more words include overlapping parts, these parts are singled out as linguistic units. Most of these units are classical morphemes, combining a particular form with meaning, but our model also recognises empty and semantically vague morphemes. Since the emerging units are derived from statistical patterns of linguistic data, they are gradient and idiosyncratic.

With regard to the processing and formation of complex words, we argue that they are represented as paths in the emergent, self-organising morphological network. In the current study, we use a key concept of network science, the notion of shortest path, which refers to the task of finding all shortest or most optimal paths connecting two non-adjacent nodes in a network. We show that words derived along short (optimal) paths are more easily formed and recognised than words derived along long (suboptimal) paths.

Data

Our study focuses on the Romance component of English. Since English has borrowed heavily from Latin and French, the English lexicon contains a very large number of Romance-based words and an extensive set of derivational affixes that are derived from these words. We decided to focus on the Romance component of the English lexicon because this component is particularly interesting for investigating how the statistical properties of an emergent morphological network affect the processing and formation of complex words.

The network model we use draws on data from a multi-billion internet corpus of modern English (*enTenTen21*, Sketch Engine; Jakubíček et al., 2013). The network contains the 315 most frequent Romance bases included in this corpus. All possible allomorphs and graphical representations of each base were taken into account. For each base, we selected the 1,000 most frequent lemmas (with a minimal frequency threshold of five tokens). The obtained lists were checked, using the Oxford English Dictionary, to exclude words with non-Romance bases that were incorrectly suggested by our search algorithm. Overall, this resulted in 35,166 unique English words including a Romance base and at least one other morpheme. All words were segmented into the strings of sequentially related morphemes and then turned into a weighted directed network.

Formally, we created a graph that comprised two sets of nodes, one for the Romance bases and another for all encountered derivational affixes. As for the links, a link between any two morpheme nodes was added only if this linear sequence of morphemes was attested in at least one lexeme in the dataset. The links were weighted by type frequency, indicating the number of unique words that include a particular sequence of adjacent morphemes.

Hypotheses

In the network approach to modelling lexical knowledge, a complex word can be conceived of as a path connecting two or more nodes. For almost any pair of nodes in the network, there exist multiple possible paths, but some paths are more 'beaten' or 'well-trodden' than the others. We hypothesize that the majority of complex English words are created along most heavily weighted paths (that is, paths with high type frequencies).

There are various algorithms to calculate the weight or length of a path in a network (Sniedovich, 2006; Bauer et al., 2010). Regardless of which algorithm one uses, the goal is to find the shortest path between nodes such that the sum of the weights of its constituent links is minimised (Schrijver, 2012). In order to align our measure of path weight with the metaphor of shortest path, we use the multiplicative inverses of the weights, which can be conceptualised as distances between adjacent nodes. Since the sequences of morphemes that appear in many lexemes are heavily weighted, they appear closer to each other (when we use the transformed weights) than the sequences of morphemes that appear in just a few words.

Our assumptions may be formulated as follows. First, Romance bases and derivational affixes can be represented in a directed network. Second, this network emerges from the processing of individual words by speakers learning English. Third, different links of this network have different weights, depending on the number of times a particular link has been used in deriving words. Fourth, new derivations are more likely to arise along the paths comprised of heavily weighted links.

If this is true and the constructed network is an accurate representation of lexical knowledge, we expect to find support for the following research hypotheses: (i) new bases and affixes can be learnt through an analysis of shortest paths in the network, without any knowledge of the syntactic categories of words and their semantics; (ii) in an experimental setting, possible words derived along shorter paths will get higher acceptability ratings than possible words formed along longer paths; and (iii) the path length of complex words is a helpful factor in explaining the variance in people's accuracy of visual word recognition and their reaction times.

Computational modelling

The complex words that entered our network were manually segmented into morpheme sequences based on information from the Oxford English Dictionary. This makes the network a purely theoretical construct, and yet our first hypothesis is that this network can emerge from the processing of individual words by speakers learning English. Of course, the best way to test this hypothesis would be to build the network entirely from the bottom up, starting with free lexical bases. The main obstacle to achieving this goal is the small number of words in our dataset. With only 315 lexical bases, our network includes only a small fraction of the English lexicon.

A less ambitious step in the same direction will be to show, without loss of generality, that our model can segment new complex words into allomorphs and morphemes. If our network can in principle acquire new morpheme nodes by shortest path analysis of new input, then it logically follows that, with a more representative dataset, the same learning procedure can be used, first, to acquire new affixes from two-morpheme words including known free bases, and second, to acquire new bound bases from two-morpheme words including known affixes.

The computational algorithm we designed to segment new words had to solve two related tasks. First, it had to segment new complex words into allomorphs (or allographs), and second, it had to map the segmented allomorphs to morphemes. In the simplest case, an allomorph node is connected to one morpheme node, but many Romance morphemes have multiple allomorphs that vary with the context. For example, the words *impolite*, *incorrect*, *irregular*, and *illegal* include four different allomorphs (*im*, *in*, *ir*, and *il*) that are represented by four separate nodes linked to the same morpheme node IN in our model. Moreover, some allomorphs have homophones connected to different morpheme nodes. For example, the segment *lect* is a homophone that is linked to two separate morpheme nodes, one representing the lexical base of words such as *select* (Latin *legĕre* 'to collect, choose') and another one representing the lexical base of words such as *delectable* (Latin *lacere* 'to entice'). All links are weighted using the multiplicative inverses of the established mappings (based on type frequency).

The algorithm works as follows. First, we selected a particular morpheme (base or affix) and deleted all words including this morpheme from the dataset. Then, the initial network was constructed from the remaining (manually parsed) words. And finally, the deleted words were fed back into the network and analyzed by the model. These words were first ordered from most to least frequent allomorph and then, within each group, from most to least frequent word. By feeding words into the model in this order, it is possible to proceed from structures with fewer elements to more complicated ones (cf. Sigurd, Eeg-Olofsson, and Van Weijer, 2004).

For each new word, the model generates all possible segmentations and then assigns the segmented units to the morphemes using the shortest path analysis. For example, the word *concept* can be segmented in many different ways, e.g., *con-cept*, *co-nc-ep-t*, *c-o-n-c-e-p-t*. Considering all possible segmentations, the model seeks to map the parsed-out units to existing morpheme nodes. For example, when the word *concept* is segmented into *con* and *cept*, these segments will be mapped to the morphemes COM and CEPT. If the mapping is successful, existing links are re-evaluated (strengthened). If a parsed-out unit does not match any existing allomorph, there are two possibilities. If the parsed-out unit is formally similar to an existing morpheme node, the model creates a new link to that node, with an initial weight determined by a metric similar to the Levenshtein distance between two sequences (minimal number of insertions, deletions, or substitutions needed to transform one string into another). If there is no existing morpheme node similar to a parsed-out unit, the model creates a new morpheme node. In this way, the model produces all possible segmentations and morphemic derivations for a new word fed into the network.

When all possible derivations are available, the model selects the one that results in the shortest possible path and deletes all others. If the selected derivation uses an established path, the used links are strengthened. If the selected derivation involves a new path, the model uses new links and, if necessary, also creates new morpheme nodes. Crucially, the computational algorithm we use gives priority to established paths. Creating new links and nodes is costly and only chosen if an analysis along an existing path is not available. If there are several existing paths, the model selects the one with the shortest distance.

For example, when the model sees the word *contraception* for the first time and does not yet know the base *cept*, it selects a derivation that connects *contra* and *ion* to existing morpheme nodes (i.e., CONTRA and ION) and creates a new morpheme node CEPT for *cept* because this is the most efficient derivation. Interestingly, when the model encounters *contraception* and is not yet familiar with the morpheme CONTRA but already knows the morphemes CON and TRA[NS] (in addition to CEPT and ION), it misanalyses *contra* as a composite form and links the segments *con* and *tra* to existing morpheme nodes (because this is more efficient than creating a new morpheme node for CONTRA). However, once the model encounters words such as *contravene* and *contravention*, it recognizes that *contra* is a prefix that requires a separate morpheme node, and corrects the initial analysis of *contraception*, as the derivation of *contra→cept→ion* is more efficient than the derivation *con→tra→cept→ion*, provided that *contra* can be linked to an existing morpheme node.

We tested the model's performance on five bases, both free and bound, two prefixes, and two suffixes (the prefix *re* was learnt from the words where it appears in the second position to make things a bit more complicated for the model). All words with respective morphemes were deleted from the dataset and then analysed by the model as described above. It took from one to three passes, depending on the morpheme, to obtain the results presented in Table 1.

Table 1: Results of the computational modelling.

Bases	Number of words	Accuracy
place / *plac*	79	1.0
norm	149	0.96
tort	117	0.97
ceive / *ceiv* / *cept*	342	0.99
clude / *clud* / *clus*	206	0.96
Prefixes	Number of words	Accuracy
anti	687	0.98
re	1,225	0.97
Suffixes	Number of words	Accuracy
ness	729	0.96
able / *abil* / *ible*	2,143	0.95

Segmenting complex words into morphemes is a challenging task. Many English words of Romance origin have unclear morpheme boundaries, some morphemes are homographic, and many morphemes have several allomorphs that vary with the context. Nevertheless, our model segmented new words into allomorphs and analyzed the segmented parts as morphemes with a very high degree of accuracy (97%, on average). We take this as evidence for our first hypothesis that our network model can acquire new allomorphs and morphemes through distributional analysis, based on an algorithm of shortest path. We showed that our model not only learns to segment words into allomorphs but also learns to categorize allomorphs such as *ceive* and *cept* or *able*, *abil*, and *ible* as instances of the same morpheme.

Experimental evidence

The creation of new words is usually motivated by semantic and pragmatic factors; but in this study, we concentrated on the influence of people's linguistic experience on word formation. To test our second hypothesis, namely, to see whether possible words derived along shorter paths will get higher acceptability ratings than possible words formed along longer paths, we conducted an experiment where the subjects were presented with pairs of possible words and asked to indicate which member of the pair was more likely to be a word of English.

The pairs of possible words were constructed as follows. First, we selected five base nodes with different weights in the network: *fact* (512 types), *cept* (484), *tract* (474), *join* (333), and *cert* (96). For each base, we chose two prefix and two suffix nodes and constructed four possible words in such a way as to ensure that (i) each link was attested in the network, and (ii) the sums of reciprocal path weights for each word were different, so that the words could be easily ranked. To these four words, one more word with the same base, one different prefix and one different suffix was added, such that two sequences of morphemes therein were unattested in the network. So overall, for each base, we obtained five possible words (e.g., *in-ex-cept-or-al*, *in-ex-cept-al-or*, *ex-in-cept-or-al*, *ex-in-cept-al-or*, and *ad-in-cept-ing-al*). Each possible word was paired in the experiment with four of its counterparts of the same base, which gave us ten combinations, 50 pairs in total.

Each word pair in our data was evaluated by 24 different people (1,200 participants overall), and so any word in any pair could theoretically win from zero to 24 of these contests. While analysing the results, we treated each stimulus in the data as a Bernoulli trial in which each word might win or lose, depending on the probability of success associated with this word's path length. Thus, each word, when tested against a word consisting of the same morphemes but derived along a different path, was a part of 24 independent Bernoulli trials with equal probability of success, and the outcome followed the Binomial distribution $Y \sim \mathrm{Binomial}(n, \theta)$.

We used the Markov chain Monte Carlo sampling approach to construct the following posterior distributions: θ_{probable}, $\theta_{\text{less probable}}$, and $\theta_{\text{improbable}}$, where the words are labelled (i) 'probable' if they are formed along shortest paths; (ii) 'less probable' if they are formed along longer paths; and (iii) 'improbable' if they contain non-existing links.

For inference, we used a hierarchical beta-binomial model where the individual words' probabilities of success θ_{ij} were pooled and treated as coming from each group's specific Beta distribution $\theta_i \sim \mathrm{Beta}(\alpha_i, \beta_i)$, the shape parameters of which were, in turn, obtained from a uniform distribution: $\alpha_i \sim \mathrm{Uniform}(0, 10)$ and $\beta_i \sim \mathrm{Uniform}(0, 10)$. We sampled 45,000 θs from the three posterior distributions of interest. The obtained credible intervals for the three groups' probabilities of success are given in Table 2.

Table 2: Probabilities of success.

group of words	θ		
	0.05	0.50	0.95
probable	.56	.59	.62
less probable	.48	.51	.53
improbable	.29	.32	.36

The results suggest that our hypothesis is borne out. Although there is a great deal of variation in the data, the words derived along the shortest paths were judged by most speakers to be probable while the words with unattested links were generally considered improbable. Interestingly, the posterior distribution of 'less probable' words' θ includes values both above and below 0.5, which indicates that these words' probability of success is no better than random guessing.

Lexical decision tasks

To investigate whether the path length of complex words in our network is a helpful factor in explaining the variance in people's accuracy of visual word recognition and their reaction times, we used the data provided by the English Lexicon Project (Balota et al., 2007) and the MorphoLex database (Sánchez-Gutiérrez et al., 2018).

The English Lexicon Project affords access to a large set of lexical characteristics, along with behavioral data from visual lexical decision and naming studies of 40,481 words. Of interest to us were the following variables: (i) the mean lexical decision latency (in msec) for a particular word across participants; (ii) the proportion of accurate responses for a particular word; (iii) frequency norms, based on the HAL corpus; and (iv) length of words in letters.

From the MorphoLex database, we obtained the data on the following additional morphological variables: (i) percentage of more frequent words in the morphological family of a particular morpheme; (ii) family size (the number of word types with this morpheme); (iii) cumulative morpheme token frequency; and (iv) probability of the morpheme being encountered in a hapax legomenon (affix productivity measure).

Table 3: Reaction times model coefficients.

	estimate	SE	p
constant	752.87	7.603	< 0.001
path weight	45.768	4.985	< 0.001
length	20.753	0.566	< 0.001
frequency	-24.108	0.696	< 0.001
MorphoLex	…	…	…

Table 4: Response accuracy model coefficients.

	estimate	SE	p
constant	0.555	0.011	< 0.001
path weight	-0.124	0.007	< 0.001
length	0.013	0.0008	< 0.001
frequency	0.032	0.001	< 0.001
MorphoLex	…	…	…

Overall, there are 6,301 complex words in our dataset that are also present in the English Lexicon Project and MorphoLex databases. We fitted two linear regression models to the data: one with the mean lexical decision latency as the dependent variable, another with the proportion of accurate responses as the dependent variable. Both variables were regressed on the same set of predictors which included length of words, their frequency norms, and their path weights in our network. While fitting each model, we also controlled for all the additional morphological variables from the MorphoLex database (respective coefficients are omitted in Tables 3 and 4 to avoid clutter).

The path length of complex words was found to be a reliable predictor of both mean lexical decision latency and the proportion of accurate responses. The longer the path along which a certain word is derived, the more time it takes to visually recognise the word and the lower the accuracy of such recognition. It is clear that these results are in line with our expectations. What makes them even more prominent is that the words in the MorphoLex database contain, on average, much fewer morphemes (M = 1.75) than the words in our dataset (M = 3.62). It stands to reason that the importance of path weight measure should only increase with the increasing word length.

Conclusion

In the usage-based approach, morphological structure is emergent from lexical relations connecting words, or lexical strings, with overlapping properties. One can think of lexical relations as associations that arise from distributional analysis in language acquisition and change. Since distributional analysis is driven by similarity and type frequency, morphological structure is gradient and not limited to classical morphemes. Yet, while the emergent units of morphological structure are not always separable from their lexical hosts, they enter into new sequential relations once they have been parsed out of a lexical sequence. If the emergent sequential links are instantiated in a large number of word types, they become productive, that is, extendable to new lexemes.

In this study, we have used a computational model and an experiment to investigate whether complex English words of Romance origin can be represented as a weighted directed network of morphemes. Our computational model showed that a morphological network like the one we analyse can in fact emerge from the processing of individual words by speakers learning English. We were able to verify that such a network can be constructed via the distributional analysis of word forms only, without any knowledge of the syntactic categories of words and their semantics. Crucially, we provided evidence that unfamiliar bases and affixes can be learnt given the rest of the network and that it is possible, as a result of such analysis, to account for allomorphy of the type that is built into the network, for example, to recognise that *ceive* and *cept* are the same unit and should be stored as one node.

Of major importance for the model was the emergent network connectivity as measured by shortest paths' distances. Since type frequency is known to be an important determinant of productivity, we hypothesized that the network path length plays an important role in forming new English words. We conducted an experiment in which participants were asked to rate possible new English words that our network derived along optimal and suboptimal paths. The experiment confirmed that speakers judge new possible words more easily acceptable if they are formed along optimal paths (that is, paths with high type frequencies). The path length of complex words was also found to be a good predictor of both reaction time and response accuracy in the visual word recognition tasks of the English Lexicon Project.

In actual language use, new words are always meaningful and motivated by a particular communicative intention; but the word forms we used in our experiments were semantically opaque and unmotivated. Nevertheless, though participants could not easily map these forms onto a specific meaning, their acceptability judgements varied with the probability levels of different word groups created by our network, indicating that path weight affects the formation and processing of new words as an independent factor.

In conclusion, network science provides a powerful framework for analyzing language use and language emergence. If we think of language as an encompassing network, the approach can also be applied to many linguistic phenomena including phenomena in the domain of syntax (Diessel, 2020, 2023). The network approach is consistent with the way psychologists and neuroscientists analyze the human mind and brain (Sporns, 2012) and accords with the emergentist view of grammar (MacWhinney et al., 2020). If language is a complex dynamic network, linguists and cognitive scientists need new instruments for analyzing language structure. We believe that network science provides an excellent toolkit for linguistic research on morphology and all other aspects of the language system.

References

Amenta, S., & Crepaldi, D. (2012). Morphological processing as we know it: An analytical review of morphological effects in visual word identification. *Frontiers in Psychology 3*: 1–12.

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

Baayen, R. H., Milin, P., Durdevic, D. F., Hendrix, P., & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological review 118*: 438–481.

Balota, D. A., Yap, M. J., Hutchison, K. A. et al. (2007). The English Lexicon Project. *Behavior Research Methods 39*: 445–459.

Barabási, A.-L. (2016). *Network science*. Cambridge: Cambridge University Press.

Bauer, R., Delling, D., Sanders, P., Schieferdecker, D., Schultes, D., & Wagner, D. (2010). Combining hierarchical and goal-directed speed-up techniques for Dijkstra's algorithm. *ACM Journal of Experimental Algorithmics 15*: 1–31.

Blevins, J. P. (2016). *Words and paradigms*. Oxford: Oxford University Press.

Booij, G. (2010). Construction morphology. *Language and Linguistics Compass 4(7)*: 543–555.

Butterworth, B. (1983). Lexical representation. In B. Butterworth (Ed.), *Language Production*. London: Academic Press, pp. 257–294.

Bybee, J. (1985). *Morphology. A study of the relation between meaning and form*. Amsterdam: John Benjamins.

Bybee, J., & Slobin, D. I. (1982). Rules and schemas in the development of the English past tense. *Language 58*: 265–289.

Diessel, H. (2019). *The grammar network. How linguistic structure is shaped by language use*. Cambridge: Cambridge University Press.

Diessel, H. (2020). A dynamic network approach to the study of syntax. *Frontiers in Psychology 11*.

Diessel, H. (2023). *The Constructicon. Taxonomies and networks*. Cambridge: Cambridge University Press.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunckett, K. (1996). *Rethinking innateness. A connectionist perspective on development*. Cambridge: Bradford Books/MIT Press.

Forster, K. I., & Taft, M. (1994). Bodies, antibodies, and neighborhood density effects in masked form-priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition 20*: 844–863.

Halle, M. (1973). Prolegomena to a theory of word formation. *Linguistic Inquiry 4*: 3–16.

Hay, J. (2002). From speech perception to morphology: Affix-ordering revisited. *Language 78(3)*: 527–555.

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

Jakubíček, M., Kilgarriff, A., Kovář, V., Rychlý, P., & Suchomel, V. (2013). The TenTen corpus family. In A. Hardie & R. Love (Eds.), *7th international corpus linguistics conference CL. Abstract book*. UCREL, pp. 125–127.

Kiparsky, P. (1982). From cyclic phonology to lexical phonology. In H. van der Hulst & N. Smith (Eds.), *The Structure of Phonological Representations*. Dordrecht: Foris, pp 131–176.

MacWhinney, B., Kempe, V., Brooks, P. J., & Li, P. (Eds.) (2022). *Emergentist approaches to language*. Lausanne: Frontiers Media SA.

Matthews, P. (1991). *Morphology* (2nd ed.). Cambridge: Cambridge University Press.

Milin, P., Smolka, E., & Feldman, L. B. (2017). Models of lexical access and morphological processing. In E. M. Fernández and H. S. Cairns (Eds.), *The Handbook of Psycholinguistics*. Wiley-Blackwell, pp. 240–268.

Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition 28*: 73–193.

Plag, I. (2018). *Word formation processes in English*. Cambridge: CUP.

Plag, I., & Winther Balling, L. (2020). Derivational morphology: An integrative perspective on some fundamental questions. In V. Pirrelli, I. Plag & W. Dressler (Eds.), *Word Knowledge and Word Usage: A Cross-Disciplinary Guide to the Mental Lexicon*. Berlin: De Gruyter, pp. 295–335.

Rumelhart, D. E., & McClelland, J. L. (1986). On learning the past tenses of English verbs. In D. E. Rumelhart, J. L. McClelland, & The PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition*. Vol. 2. Cambridge, MA: MIT Press, pp. 216–271.

Sánchez-Gutiérrez, C. H., Mailhot, H., Deacon, S. H. et al. (2018). MorphoLex: A derivational morphological database for 70,000 English words. *Behavior Research Methods 50*: 1568–1580.

Smolka, E., Libben, G., & Dressler, W. U. (2019). When morphological structure overrides meaning: evidence from German prefix and particle verbs. *Language, Cognition and Neuroscience 34(5)*: 599–614.

Smolka, E., Preller, K., & Eulitz, C. (2014). 'Verstehen' ')understand') primes 'stehen') 'stand'): Morphological structure overrides semantic compositionality in the lexical representation of German complex verbs. *Journal of Memory and Language 72(1)*: 16–36.

Schmid, H.-J. (2020). *The dynamics of the linguistic system. Usage, conventionalization, and entrenchment.* Oxford: Oxford University Press.

Schrijver, A. (2012). On the history of the shortest path problem. In M. Grötschel (Ed.), *Documenta Mathematica. Extra volume: Optimization stories. 21st International symposium on mathematical programming (Berlin, August 19–24, 2012)*, pp. 155–167.

Seidenberg, M. S., & Plaut, D. C. (2014). Quasiregularity and its discontents: The legacy of the past tense debate. *Cognitive Science 38*: 1190–1228.

Sigurd, B., Eeg-Olofsson, M., & Van Weijer, J. (2004). Word length, sentence length and frequency — Zipf revisited. *Studia Linguistica 58*: 37–52.

Sniedovich, M. (2006). Dijkstra's algorithm revisited: the dynamic programming connexion. *Control and Cybernetics 35(3)*: 599–620.

Sporns, O. (2012). *Networks of the brain*. Cambridge: MIT Press.