

John Benjamins Publishing Company



This is a contribution from *The Mental Lexicon* 11:3
© 2016. John Benjamins Publishing Company

This electronic file may not be altered in any way.

The author(s) of this article is/are permitted to use this PDF file to generate printed copies to be used by way of offprints, for their personal use only.

Permission is granted by the publishers to post this file on a closed server which is accessible only to members (students and faculty) of the author's/s' institute. It is not permitted to post this PDF on the internet, or to share it on sites such as Mendeley, ResearchGate, Academia.edu.

Please see our rights policy on <https://benjamins.com/content/customers/rights>

For any other use of this material prior written permission should be obtained from the publishers or through the Copyright Clearance Center (for USA: www.copyright.com).

Please contact rights@benjamins.nl or consult our website: www.benjamins.com

Morphological schemas

Theoretical and psycholinguistic issues

Ray Jackendoff and Jenny Audring

Center for Cognitive Studies, Tufts University / Centre for Linguistics,
Leiden University

We propose a theory of the lexicon in which rules of grammar, encoded as declarative schemas, are lexical items containing variables. We develop a notation to encode precise relations among lexical items and show how this differs from the standard notion of inheritance. We also show how schemas can play both a generative role, acting as productive rules, and also a relational role, where they codify nonproductive but nevertheless prolific patterns within the lexicon. We then show how this theory of lexical relations can be embedded directly into a theory of lexical access and lexical processing, such that it can make direct contact with experimental findings.¹

Keywords: morphology, Construction Grammar, Construction Morphology, morphological processing, lexicon, lexical processing

Overall goals

Looking to the big picture, a cognitive science of language aspires to a unified account of the structure of language, how it is processed, and how it is instantiated in the brain. Roughly following Marr (1982), such an account involves (at least) three subtheories:

1. This article is largely excerpted from Jackendoff and Audring (forthcoming), which deals with these issues in much greater detail. For important suggestions, we are grateful to Geert Booij, Peter Culicover, Gina Kuperberg, Pim Levelt, Peter Hagoort, Ari Goldberg, three reviewers, and audiences at talks, workshops, and seminars at Tufts, Harvard, MIT, and the Max Planck Institute for Psycholinguistics. RJ is grateful to Peter Hagoort and the Max Planck Institute for Psycholinguistics for the opportunity to work on this project while a Visiting Fellow. JA is grateful to the Dutch national research organisation NWO for a Veni grant, #275-70-036.

- A theory of *representations*: the data structures stored in memory and those built online. This corresponds to Marr’s *computational theory* and Chomsky’s (1965) theory of *competence*.
- A theory of *processing*: the (virtual) architecture of the language processor, how representations are deployed in real time to create new structures, and how stored representations are acquired. Marr calls this an *algorithmic theory*, though strictly speaking, processing in the brain is most likely not in the nature of a step-by-step algorithm.
- A theory of *neural computation*: how the machinery of the processing theory is instantiated in the brain. Ultimately, this should include not only an account of brain localization, but also fine-scale details such as how neural encoding distinguishes the sound *b* from the sound *p* and the word *cat* from the word *dog*. Marr calls this an *implementational theory*.

Marr stresses (as Chomsky does not) that these theories must integrate with one another. The theories of processing and neural computation must be rich enough to support one’s postulated representations; conversely, the representational theory must lend itself to postulated processing mechanisms and neural implementation. If the theories conflict, there is no predetermined outcome as to which theory prevails. For instance, the linguist’s theory of competence should not be immune to evidence from performance. But neither should a theory of neural computation be immune to evidence from what an adequate theory of linguistic representation requires.

A theory of linguistic representations and processing should in addition seek integration between the theories of phonology, morphology, syntax, and semantics. And insofar as possible, a theory of the language faculty should seek integration with theories of other mental faculties. However, such integration must respect the details of the individual faculties: it must not result in ignoring much of what we know about language in order to fit it into a procrustean bed of, say, associationism.

The present article attempts some steps toward such an integrated theory. We focus on morphology, which highlights features of linguistic organization that are not so prominent in syntax (unless one thinks to look for them). Our approach, which we call Relational Morphology (RM: Jackendoff & Audring, forthcoming) grows out of the Parallel Architecture of Jackendoff (2002) and Culicover and Jackendoff (2005); we also draw freely on Construction Morphology (Booij, 2010). We will explore some aspects of a theory of morphological processing, raising many questions that we hope might encourage psycholinguists, neurolinguists, and computational modelers to try to fill out the picture, both theoretically and experimentally.

Representations in relational morphology

Rules, reconceived as schemas, are in the lexicon

A prominent construct within most linguistic theories has been the distinction between words and rules, or between the lexicon and the grammar, as though they are in different metaphorical “places” in the mind. In parallel, much of *psycholinguistics* has tended to treat storage of words as distinct from rules of grammar, a position explicitly advocated by e.g. Ullman 2015.² This split is rejected by the Parallel Architecture, by Construction Grammar (Goldberg, 1995; Hoffman & Trousdale, 2013), and by Construction Morphology. These approaches argue that grammatical rules are themselves lexical items – that is, the grammar is part of the lexicon. Here is how the reasoning goes.

In practically every linguistic theory, a word contains pieces of structure on three levels: its semantic structure, its syntactic features, and its phonology.³ In the Parallel Architecture framework, these levels are in principle independent, each with its own characteristic conditions of well-formedness. But each is also linked to the others: *this* phonological string can be linked to *this* syntax and *this* semantics. We call these connections **interface links**, and we notate them with indices that show which parts of structure on one level correspond with structure on another level (other frameworks use different notations with similar effect). The indices are to be thought of as marking the ends of association lines. In the easiest cases, the mapping is trivial, as in *sheep*.

- (1) Semantics: SHEEP₁
 Morphosyntax: N₁
 Phonology: /ʃi:p/₁

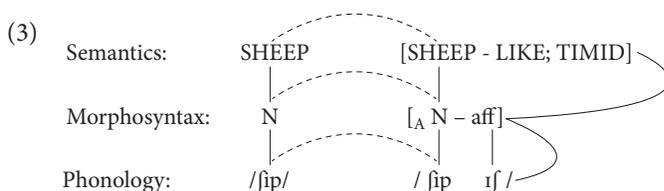
More interesting is a morphologically complex word, such as *sheepish* in (2).

- (2) Semantics: [SHEEP₁-LIKE; TIMID]₂
 Morphosyntax: [_A N₁ - aff₃]₂
 Phonology: /ʃi:p₁ iʃ₃/₂

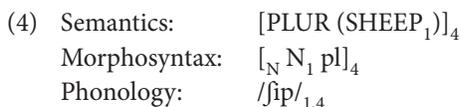
2. Other strains of psycholinguistics, e.g. connectionism, deny altogether that there are rules (Rumelhart and McClelland 1986).

3. – with limited exceptions. Words like *hello* and *ouch* have no syntactic category: they can constitute utterances on their own but only combine into utterances paratactically (*Hello, Joe*) or in quotes (“*Ouch!*” *she cried*.) A few other words lack semantics and serve only as “grammatical glue”, such as the underlined words in *I think that it’s raining* and *I do not like green eggs and ham*. Literate individuals will have a fourth level of structure, orthography.

Unpacking the notation in (2): The morphosyntax encodes the fact that this word is an adjective composed of a noun plus an affix. Coindex 2 links the semantics, morphosyntax, and phonology of the whole word, just as coindex 1 links the layers in (1). Coindex 1 links the syntactic category Noun with the meaning SHEEP and the phonology /ʃip/. Coindex 3 links the affix with the phonology /ɪʃ/.⁴ However, coindex 1 also links the relevant parts of *sheepish* to the word *sheep*. In this role, cutting across words, the coindexation serves as a *relational link*. If we notate the links within (1) and (2) as association lines rather than coindices, as in (3), we can see that the interface links (solid lines) connect levels within a lexical item, while the relational links (dashed lines) connect parts that are the same in different lexical items:⁵



The interface links in (2) establish a 1:1 relation between the parts of the semantics, the morphosyntax, and the phonology. However, such a transparent relation is not always present. For example, consider the plural word *sheep*, shown in (4).



(4) has a plural feature in morphosyntax, as is evident from determiner and verb agreement (*those sheep are...*); but there is no corresponding part in phonology. Rather, the full phonological string has two indices: one for its connection to the stem *sheep* (coindex 1), and one for its connection to the word as a whole (coindex 4). There is no “zero morpheme” in the phonology, as in many theories.

Given this format for representing words, the notation can easily be enriched so as to implement rules. Extracting the contribution of *sheep* (1) to *sheepish* (2), we arrive at (5) as the pattern whose instances include *sheepish*, *childish*, *sluggish*, and so forth.

4. One might wonder if coindex 3 should appear also in semantics, indexed on *LIKE*. Our reasons for not doing so are beyond the scope of this article.

5. The reader can also see why we have chosen coindexation as our notation rather than association lines: they’re a mess!

- (5) Semantics: $[X_y\text{-LIKE}]_z$
 Morphosyntax: $[{}_A N_y - \text{aff}_3]_z$
 Phonology: $/\dots_y \text{ɪ}_3 /_z$

Following the terminology of Construction Morphology, we call (5) a **schema** rather than a rule, to stress that it is declarative rather than procedural. (The term *schema* goes back in linguistics at least to Bybee and Slobin (1982), with more general antecedents cited in Rumelhart 1980 all the way back to Kant. Construction Grammarians would call (5) a **construction**.) (5) is not a procedure to convert an “input” into an “output,” by adding *-ish* to a noun to form an adjective; rather, it *licenses* or *motivates* its instances, a notion to which we return in a moment. This schema is formally like the words that instantiate it, with the same three layers of structure. However, instead of a lexical base, it has open slots or **variables**: *X* in semantics, an unspecified noun in morphosyntax, and an arbitrary phonological string in phonology.

Next consider the coindexation in (5). Coindex 3 links the affix in morphosyntax with the phonology *-ish*. It also serves as a relational link to the affix in *sheepish* and all other *-ish* adjectives. However, the layers of the schema as a whole are linked by the **variable coindex** *z*, and the variables within the schema are linked by the variable coindex *y*. These coindices link (5) to any lexical entry with parallel coindexation. For instance, index *y* links to index 1 in *sheepish* and to index 5 in *childish*, shown in (6); and index *z* links to index 2 in *sheepish* and index 6 in *childish*. In other words, the constant coindex 3 in (5) links to what is the same across the instances, and the variable coindices link to what is different but parallel in structure.

- (6) Semantics: $[\text{CHILD}_5\text{-LIKE; SILLY}]_6$
 Morphosyntax: $[{}_A N_5 - \text{aff}_3]_6$
 Phonology: $/\text{tʃaɪld}_5 \text{ɪ}_3 /_6$

Space prohibits us from discussing how coindexation applies in more complex morphological phenomena crosslinguistically; see Jackendoff and Audring (forthcoming). For the coindexation notation in phrasal syntax, see Jackendoff (2002) and Culicover and Jackendoff (2005).

Two functions of schemas

Replacing rules by schemas may seem like a small move, but there is an important difference. Schemas have the same format as words, differing only in that part of their structure consists of variables and variable coindices. Hence words and schemas can be in the same ‘place’ in the theory – or in the mind. In other words,

this theory does not need two independent constructs of “lexicon” and “grammar,” and in that respect it is more constrained.

Another difference between schemas and rules lies in their function in the grammar. Traditional rules build up novel composite expressions from smaller parts, for example “add *-ish* to a noun to form an adjective.” Schemas too can be used to build up novel composite expressions, by **unifying** their variables with other pieces of structure. So, for instance, if one encounters a new noun, say *wug*, one can unify it with the variable in the *-ish* schema to construct a new adjective *wuggish* with the meaning ‘like a wug.’ We will call this the **generative** function of schemas.

But schemas also have a second function: they can capture generalizations among existing lexical items. In this **relational** function of schemas, they encode what is the same among their instances. Consequently, the instances take less work to learn, store, and/or process. Well-known approaches to this function are **motivation** and **inheritance**, which we take up in Section 2.3.

In their relational function, schemas need not exhaustively specify the contents of the items they motivate. For instance, *sheepish* does not just mean ‘like a sheep,’ it means something like ‘timid’ with perhaps an overlay of ‘embarrassed,’ as in a *sheepish grin*. But it is still motivated by the *-ish* schema.

However, the relational use of schemas is not confined to lexical items with idiosyncratic properties. Regular complex forms must be stored when embedded in idioms (such as the plural in *raining cats and dogs*) and when there is no corresponding singular form (as in *scads of words*). Moreover, experimental research (e.g., Baayen, Dijkstra, & Schreuder, 1997) has shown that speakers store many regular forms, such as Dutch *dieren* ‘animals,’ which could in fact be created generatively.⁶ These facts are problematic for a rule-based theory, which decrees that items generated by rules cannot be in the lexicon – what Langacker (1987) has called the “rule-list fallacy.” In contrast, a schema for the regular past tense can apply to motivate these stored forms. It is the same schema, the same generalization, but instantiated here through relational rather than generative means.

The two functions of schemas are not confined to morphology. (7) is the phrase structure rule for English transitive VPs, formulated as a schema (or “treelet” in Janet Fodor’s (1998) sense).

$$(7) \quad [{}_{VP} V - NP]$$

6. Here is a point where psycholinguistic research has a crucial bearing on the theory of representations.

This schema obviously functions generatively in constructing novel instances like *swallow the beer*. But it can also function relationally to motivate the syntactic structure of idioms like *chew the fat* ('converse idly') which cannot be built online from their parts, because, although they may share some semantic features, overall they mean the wrong thing.

Crucially, not all schemas have both functions. The suffix *-en* in *harden* has many instances: *whiten, darken, soften*, and so on. However, unlike the *-ish* suffix, which can be used freely to generate new words such as *Trumpish*, the *-en* suffix is not productive: there is *soften* but not **louden, thicken* but not **slicken*, and so on. A great deal of morphology, especially derivational morphology, is like this: there is a generalization, but nevertheless one has to store the existing instances, and one is not free (or at least not *very* free) to make up new instances. The schema for [_V A *-en*] is what has been called a "semiproductive rule" or a "lexical redundancy rule"; it has only a relational function, not a generative one.

Purely relational schemas also turn up in phrasal syntax, although less prominently, in what Culicover (1999) calls 'syntactic nuts' – bits of syntax that only occur with a few listed words. For example, the construction exemplified by *day after day* occurs with only a limited class of prepositions (8a) (Jackendoff, 2010); and the little determiner formula in *what a job* occurs only with a restricted set of forms preceding the indefinite article (8b).

- (8) a. N-P-N: *day after day, week by week, face to face*
 but: **student before student, gun beside gun*
- b. [_{Det} X a]: *what a job, such a job, quite a change*
 but: **who a professor, *where a city*

One has to learn the instances of these patterns one by one. Hence these are syntactic patterns that cannot be encoded by productive generative rules. We conclude that syntax as well as morphology requires schemas functioning relationally.

In addition to schemas that have both the generative and relational functions, and schemas that have only the relational function, we might wonder if there could also be schemas that have only the generative function. The answer would be no. Such schemas cannot exist, because any instance of a schema that can be generated online can then be stored, where it falls under the relational function. One can store items of all sizes – catch phrases, song lyrics, even the entire Koran – without losing the internal structure that connects them to the grammatical patterns in the language.

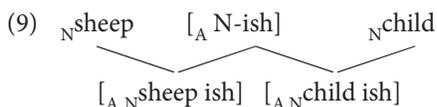
Hence there are basically two kinds of schemas: (a) productive schemas, which both serve the function of traditional generative rules and also capture generalizations over stored lexical items, and (b), nonproductive schemas, which have only the relational function. Thus one might consider productive schemas as

relational schemas that have “gone viral”; in effect, the generative function is the innovative add-on.

This conclusion about rules of grammar reorients the perspective of linguistic theory. Modern linguistics has been primarily concerned with the creativity of language, Humboldt’s “infinite use of finite means,” and therefore it has focused on generative rules that create novel utterances. For instance, the second paragraph of Berwick and Chomsky (2016: 1) says: “Generative grammar sought ... to ... explain what we will call the Basic Property of language: that a language is a finite computational system yielding an infinity of expressions...” However, once it is recognized that rules/schemas also serve in a relational function – and especially that many rules/schemas have *only* this function – it becomes apparent that we ought to focus at least as much on the relations within the lexicon, which cannot be described in terms of traditional generative rules. The relational function is the foundation on which the generative function builds.⁷

Inheritance: Spelling out the relational function

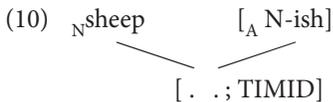
We now return to the question of what it means for a schema to “motivate” its instances, in the sense widely invoked in morphology (e.g., Daelemans, De Smet, & Gazdar, 1992; Booij, 2010, to appear; Lakoff, 1987; Goldberg, 1995; Radden & Panther, 2004; and indeed de Saussure, 1916/1977 (133)). A common position in these sources is that families of words like *sheep*, *sheepish*, and [_A N-*ish*] motivate each other through an **inheritance hierarchy**, of the sort illustrated in (9). The items lower in the hierarchy are taken to inherit structure from items they are connected to higher in the hierarchy. Thus *sheepish* and *childish* inherit from *sheep* and *child* respectively, and both inherit from the [_A N-*ish*] schema.



Inheritance is especially attractive because it has also been frequently invoked as a factor in the organization of concepts (Murphy, 2002): the concept *poodle* inherits from *dog*, which inherits from *animal*, and so on. Hence inheritance is a domain-general theoretical construct that requires no special machinery for morphology or even for language per se. But what does inheritance mean? What do the lines signify?

7. Distributed Morphology (Halle & Marantz, 1993; Siddiqi, to appear) can be seen as an attempt to treat nonproductive schemas in terms of generative rules. See Jackendoff and Audring (forthcoming) for a critique.

A common interpretation of inheritance (e.g., Collins & Quillian, 1969; Pollard & Sag, 1994; Riehemann, 1988) is that it serves to keep the lexicon maximally economical: any information present in a higher node of the hierarchy does not have to be specified in a lower node. For instance, *sheepish* inherits almost everything from the two higher nodes and therefore can be listed something like this:



Following Jackendoff (1975), we call this the **impoverished entry theory**: the idea is that lexical items contain only information that cannot be inherited from elsewhere.

Despite its intuitive appeal to elegance and minimum description length, there are good reasons to reject this position. Consider a case from conceptual inheritance: RJ's concept of his late cat Peanut can't specify the color of her paws without linking it to her paws. But by hypothesis, the fact that she has paws is inherited from his concept of cats in general, and is not a direct part of his concept of Peanut. In other words, according to this theory, the concept of Peanut per se contains nothing that the color can be the color of. As Bybee (2001:7) puts it (citing Langacker, 1987), "if predictable properties are taken away from objects, they become unrecognizable."

Another argument comes from the theory of acquisition. In order to construct a schema such as $[\text{A N-ish}]$, one must generalize over existing lexical items whose details are present in memory. The impoverished entry theory forces one to claim that once the schema is established, the redundant details of all the instances used to establish it are immediately erased from memory, thereby optimizing the lexicon. We find this implausible (though psycholinguistic evidence might prove us wrong).⁸ Similarly, in the course of acquiring a new complex word, one must first discover its details, and only then determine what schemas it falls under and what its base is. The impoverished entry theory suggests that as soon as one establishes the new word's relation to a base and to one or more schemas, all its redundant features are immediately expunged. Again, we find this implausible. (Similar arguments can be found in Langacker, 1987 and Booij, to appear.)

A third argument against impoverished entries is that inheritance is taken to be asymmetric: one lexical entry is the ancestor of the other. But consider pairs like *assassin/assassinate* or *linguist/linguistics*. On the basis of phonology, the second

8. This does not preclude the possibility that redundant information, like any stored information, may decay and be forgotten. The point is that until it decays, such information is redundant; the impoverished entry theory claims this is impossible.

member of each pair should inherit from the first, just like *sheep/sheepish*. However, the semantic dependency goes the other way: an assassin is ‘someone who assassinates people’ and a linguist is ‘someone who does linguistics.’ This mixed dependency cannot be expressed in terms of standard inheritance with impoverished entries: Which is the ancestor and which is the dependent? What gets omitted from which entry?

On the basis of these and other arguments (Jackendoff and Audring forthcoming), we conclude that economy is the wrong criterion for a theory of lexical storage in the brain. An alternative that we find plausible is that the brain *embraces* redundancy, at least up to a point. For instance, languages seem to have no problem marking thematic roles redundantly, through word order, case marking, and verb agreement. Further afield, the visual system has at least five partially redundant mechanisms for depth perception: lens accommodation, eye convergence, stereopsis, occlusion, and perspective (Marr, 1982). Redundancy appears to have the effect of making mental computation more robust (Libben, 2006: 6).

These precedents encourage us to adopt a **full-entry theory**, in which lexical items are encoded in their entirety, even where redundant.⁹ In this approach, the lexical entry of *sheepish* is like (2), with its full structure, rather than like (10), in which it has been evacuated of content. We explicate motivation in terms of shared structure: the *-ish* schema (3) and the word *sheep* (1) motivate (2) by virtue of the structure they have in common. The relational links, encoded by coindexation, show precisely what parts of their respective structures the items share.

This formulation avoids the difficulties for the impoverished entry theory. First, the fact that Peanut has paws is directly present within her conceptual structure, coindexed with the paws of the more general schema for cats. Hence the idiosyncratic color of Peanut’s paws has something to be attributed to. Second, there is no need to erase redundant information in instances of newly constructed schemas or in newly encountered instances of existing schemas. Third, unlike standard inheritance, coindexation is not inherently directional, and it can apply to the three levels of representation independently where necessary. This permits the mixed connections between *assassin* and *assassinate* to be expressed by coindexation as in (11): the semantics of *assassinate* is shared with part of the semantics of *assassin* (coindex 8), while the phonology of *assassin* is shared with part of the phonology of *assassinate* (coindex 7).

9. We acknowledge that there are important questions about what it means to code an item “in its entirety.” Does that include phonetic detail? Does it include semantic detail that might be termed “real-world knowledge”? We must leave these questions open. However, we resist the other extreme, the position that one stores all and only fully detailed exemplars.

- (11) a. Semantics: [PERSON WHO [MURDERS POLITICIAN]₈]₇
 Morphosyntax: N₇
 Phonology: /əsæsən/₇
- b. Semantics: [MURDER POLITICIAN]₈
 Morphosyntax: [_V N₇ – aff₉]₈
 Phonology: /əsæsən₇ eɪt₉/₈

We call such relations “sister” relations, in contrast to the mother-daughter relations of standard inheritance.

Summary so far and prospectus

To sum up the theory of representation:

- The lexicon contains words and schemas in the same format: pieces of semantic, morphosyntactic, and phonological structure, connected by interface links.
- Schemas differ from words only in that they contain variables.
- Words and schemas explicitly motivate each other through relational links, which designate pieces of structure shared among fully specified lexical entries, whether words or schemas.
- Both productive and nonproductive schemas can be used to motivate the structure of listed lexical items. Productive schemas can also be used to generate novel expressions on line. Both morphology and syntax have both kinds of schemas.

Acquisition of productive schemas relies on the relational function

In order to be able to generate an indefinitely large number of novel utterances, we obviously need productive schemas, functioning generatively. But the status of schemas in their *relational* function is less clear – especially nonproductive schemas, which function only relationally and never generate anything. One might legitimately wonder if the relational function is necessary at all. For instance, Pinker (1999) takes the position that there are rules for regular morphology – in our terms, productive schemas used generatively – but that there are only associations and analogies, not rules, for nonproductive patterns. (He does not reflect on productive schemas used relationally.)

There do exist some small families of words that do not fit any general pattern. For instance, the pairs *bomb/bombard*, *laugh/laughter*, *hate/hatred*, and *compare/*

comparison have apparent suffixes that are at best exceedingly rare in modern English. Such pairs certainly do not warrant an abstract schema with variables. Nor is there a general schema that connects pairs like *assassin* and *assassinate* in the precise way they are related in (11). It is reasonable to think such cases are connected only as sisters.

On the other hand, fairly general – though nonproductive – schemas are intuitively more appealing for the hundreds of adjectives that end in *-ous* (*joyous, glorious, adventurous, avaricious* and so on) and the thousands of nouns that end in *-ion*. Nevertheless, it is worthwhile to insist: Are schemas really needed for these? Wouldn't analogy and association be enough?

Here is an argument from acquisition. Everyone who believes in rules of grammar agrees that children are trying to construct the productive rules of their language – here, the productive schemas – on the basis of primary linguistic input. The input of course contains no rules: rather, children must be discovering (or inferring) patterns in the input. This requires appropriate evidence to be stored in memory, in the form of stored words and fixed phrases. In other words, roughly following Tomasello (2003), Culicover and Nowak (2003), Bybee (2010), and many others, the acquisition of a schema has to be item-based.

However, just storing items is not enough: there must be some process that ranges over long-term memory (LTM), seeking commonalities among items.¹⁰ In terms of the present framework, this process creates sister links among items that appear to share structure, for instance between *sheep* and *sheepish*, between *assassin* and *assassinate*, and between *chew* and *chew the fat*.

On the basis of such sister links, the learner must (unconsciously) create hypotheses for putative rules. In the present framework, such hypotheses take the form of tentative schemas, which are in the same format as words. The parts of the linked sisters that are the same form the constants in the schema (e.g. *ish*), and

10. We might conceptualize this process in various ways. First, there might be a global “web crawler” that (unconsciously) trolls LTM for similarities, and establishes links among similar items. Alternatively, individual lexical items might send out signals, looking for partners, and linking when they find a match. This might take place on first presentation of a word, or later, once the word is stored in memory. None of these possibilities really lends itself to neural implementation as it is currently understood (we think); one does not normally think of neural networks as adding connections.

Individuals undoubtedly differ in how thoroughly they search for shared structure. In particular, we should not exclude the possibility that children try out hypotheses that adults would find rather far-fetched. For instance, at about three years old, RJ's grandson Ezra responded to someone's remark about Saturday that “it makes you sad.” For another case, Lila Gleitman confesses (p.c.) that until well into adulthood, she was convinced that *robinson* meant an adventure involving a sea voyage, on the strength of *Swiss Family Robinson* and *Robinson Crusoe*.

the parts that differ (e.g. *sheep, child, slug*) are replaced by variables in the schema. The acquisition process does not need to shift from the format of words to an entirely different format for rules (a criticism mounted by Bybee, 1995 of Pinker's "words and rules" dichotomy) – a formal advantage of the schema approach.

Now suppose that a learner has posited a tentative schema on the basis of some collection of sister items in LTM. Its function vis-à-vis these items is by necessity relational from the outset: the instances are already in memory. Crucially, though, the learner has no way of knowing in advance whether or not this schema can be used generatively as well – that is, whether it is productive. If in the end it is not – as will be the case most of the time, especially in (derivational) morphology¹¹ – the learner is left with a nonproductive schema that can only motivate lexical items that have been heard. In other words, nonproductive schemas are a logical steppingstone to productive schemas, and the relational function of schemas is a *prerequisite* to their taking on a generative function.

A nonproductive schema can also aid in learning new words. If the child has the *-ish* schema and encounters a new instance, say *wolfish*, it is possible to decompose the word and at least partly interpret it, perhaps with the help of context. This differs from simply interpreting the new word on the basis of analogy with known words. By invoking a schema, the learner does not have to compare the new word unselectively with all potential sisters, whatever their degree and dimensions of similarity. The schema in effect offers a prestructured or "precompiled" basis for analogy.

Theory of processing: Basic assumptions

We now turn to the question of how nonproductive schemas might play a role in morphological processing. Our intuition is that schemas should facilitate the processing of morphologically complex words. This intuition seems to be confirmed at least for compounds: Fiorentino and Poeppel (2007) find that compounds are processed faster than comparable monomorphemic words. We wish to show how such a difference could come about in the case of nonproductive suffixation. In order to do so, we now briefly set out our thinking on language processing in

11. What criteria does the learner use to determine whether a schema is productive? Goldberg (2005) offers some possibilities, as does Yang (2016). Here we observe only that the obvious factor of frequency cannot be the only factor in play. For instance, English has a construction (*all*) *X-ed out* 'worn out from too much X', as in *I'm coffeed/Olympic'd/historied/knitted out*. This construction is completely productive in choice of noun or verb for X. Yet it is quite infrequent; one hears it perhaps three times per year.

general (building on Jackendoff, 2002, 2007), and in particular on the processing of spoken words. Our approach shares a great deal with other theories of the lexicon. However, following Marr's desideratum of integrating representational and processing theories, we build features of our linguistic theory directly into the account of processing.

As in many approaches, we treat knowledge of language as a network consisting of linked nodes. Like some versions of lexical networks (e.g., Levelt, Roelofs, & Meyer, 1999), its nodes fall into three layers, corresponding to semantic, morphosyntactic, and phonological structures. Such a network has no independent level of "lexical nodes." Rather, a "lexical node" is a complex of semantic, morphosyntactic, and phonological nodes connected by interface links.

A difference from most network models (including neural networks) is that we do not regard nodes in the network as simplex monads. Rather, they are *structures*, as in Booij (2010), Bybee (2010), and Kapatsinski (2007). For instance, the lexical entry for *sheepish* comprises semantic, morphosyntactic, and phonological nodes, connected by interface links, and each node has the internal structure shown in (2) above. In other words, the contents of the network mirror the representational theory's description of morphologically complex words.

Another important feature of our conception of the lexical network is that the relations among words are not instantiated by simple links among nodes. Rather, because the nodes have internal structure, their connections can be encoded by the relational links posited in Section 2, which pinpoint the regions of similarity among structures. So again, the representational theory determines details of the processing theory.

Yet another difference from traditional networks, again following the representational theory, is that the network includes not only all the words, but also all the rules, in the form of schemas; these too are lexical items, with all the privileges and responsibilities thereof.

We conceive of the lexicon as supporting spreading activation, in much the sense of standard approaches going back at least to Collins and Loftus (1975). However, again we propose a refinement arising from the representational theory: since nodes are now connected by interface and relational links, these are what form the paths for spreading activation. The strength of spreading activation can then be modulated in part by how extensive the linkage is.

We make some further assumptions about processing that are not linked directly to the representational theory. Departing from pure network models, but in concurrence with many other approaches, working memory (WM) is to be regarded as a functional component distinct from long-term memory (LTM); it does not consist simply of the parts of LTM that are active. LTM contains the lexical network of "knowledge of language" that we have just described; it corresponds roughly to

the “memory” component in Hagoort’s (2005) MUC (“memory, unification, control”) approach. It is in LTM that schemas fulfill their relational function, through their links to more fully specified items. In contrast, working memory (WM) is the functional component in which pieces of lexical items are assembled into larger structures, either to create an utterance (in production), or to analyze and parse an input (in comprehension). WM is where productive schemas perform their generative function; it corresponds roughly to Hagoort’s “unification” component. (For arguments for this division of labor between LTM and WM, including differences from Baddeley’s (1986) conception of working memory, see Jackendoff, 2002, 2007; Marcus, 1998, 2001; Gallistel & King, 2009.)

Following much contemporary thought, tracing back to Swinney (1979), Tanenhaus, Leiman, and Seidenberg (1979), and Woods (1980), as well as to cohort theory and its descendants (Marslen-Wilson & Tyler, 1980; Marslen-Wilson, 1987; Marslen-Wilson & Zwitserlood, 1989), we assume that processing is “promiscuous”: in language comprehension, everything in LTM sufficiently similar to the current input is activated and retrieved into WM, where it serves as a candidate for “what is being heard,” in competition with other candidates. The degree to which any particular candidate is activated in WM depends on a number of factors, including its current level of activation in LTM (see below), the relative strength of its competitors, and how well it fits the current context.

The competition among candidates need not be resolved immediately, as can be seen from minimal pairs such as (12) and (13), involving word identification and syntactic structure respectively. An ambiguity arises at the italicized words, but it cannot be resolved until the semantic processing of the underlined words. (We know of no experiments testing these particular situations.)

- (12) a. That’s not *a parent*; it’s actually a teacher.
 b. That’s not *apparent*; it’s actually quite obscure.
- (13) a. *The* more attractive and intuitively plausible theory isn’t always right.
 [*the* = determiner]
 b. *The* more attractive and intuitively plausible a theory is, the better chance it has of becoming popular. [*the* = mark of comparative correlative construction (Culicover & Jackendoff, 2005)]

A final assumption of our approach concerns the correlation between the speed of lexical access and the corpus frequency of the word being identified (Oldfield & Wingfield, 1965). This actually involves two correlations. First, reaction time correlates with resting activation: the higher an item’s level of resting activation, the faster it responds (Baayen, Dijkstra, & Schreuder, 1997; Bybee, 1985, 1995 uses the term “lexical strength”; see also Plag, 2003). In turn, concurring with practically every theory of lexical acquisition, an item’s level of resting

activation is incrementally driven up by repeated use, along the familiar lines of Hebbian learning. Thus there is no direct representation of frequency in the brain. Rather, frequency in a written corpus stands proxy for frequency of use, which affects resting activation – which in turn affects reaction time. (Notice here the interplay between representational and processing theories. The structure of an item is determined by the representational theory, but its resting activation belongs to the theory of processing.)

Activation of a node in the LTM network can be raised above resting level through a number of familiar mechanisms.

- After an item is called by WM, we find it reasonable to assume that it does not return to resting activation immediately, but takes a little while to settle down. This provides an account of *identity priming*: if the item recurs in the input soon enough, its level is still above resting activation, and thus it is easier to summon it back to WM, resulting in more rapid identification.
- A node's activation may also be raised by activation that has spread along a relational link from another activated lexical item in LTM. This again makes it quicker to respond to a call from WM. This is *neighborhood priming*.
- A third possible source of heightened activation is semantic priming from the understanding of context. Such priming need not be lexical: for instance (12a) might be favored over (12b) because the speaker accompanied the utterance of *that* with a point to a person in a school.
- Finally, among neighborhood primes are fixed expressions containing the word in question. For a trivial instance, the input *baa baa black...* activates the fixed expression whose continuation is *sheep*, which in turn primes *sheep*. Hence hearing /ʃ/ activates the candidate *sheep* more strongly than the candidate *shoe*.

The Parallel Architecture's continuity between words and rules has an interesting consequence here. Since schemas are lexical entries, they too have a resting activation and can be called into WM, with a strength based on their frequency. This immediately yields an account of *syntactic priming*: if, for instance, one has heard a ditransitive dative structure (e.g. *Kate handed Sam a banana*), one is more likely to favor producing a ditransitive (e.g. *Max sent Anne a message*) over a prepositional dative (*Max sent a message to Anne*). If words and rules are entirely different mental entities, syntactic priming requires a separate mechanism, as noted by Bock and Loebell (1990) in an early description of the phenomenon. On the present account, syntactic priming is just another form of identity priming. The ditransitive dative construction is a lexical item in LTM, and in this case it has not entirely returned to resting activation, so at the moment it is easier to reactivate than usual (Wittenberg & Snedeker, 2014).

This conception of WM processing offers a connection with probabilistic/predictive approaches to syntactic parsing (for syntax: Hale, 2003, 2011; Levy, 2008; Kuperberg & Jaeger, 2015; for morphology, O'Donnell, 2015). The basic idea of this approach is that at any particular moment, the processor is making predictions about what is to come in the input; these predictions are not all-or-nothing, but probabilistic. As new input comes in, its degree of surprisal – basically its deviation from prediction – affects processing time and electrophysiological measures of processing effort. Surprisal is measured via corpus frequency and cloze probability. Continuations with low surprisal (*baa baa black ... sheep*) are processed faster and elicit a lower processing cost than continuations with high surprisal (*baa baa black ... shoe*).

To translate this approach into present terms: A candidate interpretation of the input typically includes structure that continues beyond the input. This extra structure can be considered a prediction. For instance, an initial /ʃi.../ activates the cohort *she, sheep, sheen, sheet, sheath*, and so on, each of which can be considered a prediction about the continuation. The probability assigned to each prediction can be reframed as the relative strength of activation of each of the candidates, which, as we have seen, varies with frequency and priming. A low-surprisal continuation is one that conforms to a high-strength candidate, such as the continuation *...eep* after *baa baa black sh...* A high-surprisal continuation is one that conforms to a low-strength candidate, such as the continuation *...oe* in the same context. A high-surprisal continuation has the effect of dethroning a leading candidate and elevating a lowly one; this seems a reasonable cause for increased processing effort, in proportion to the disparity between the two competitors. We are not aware of any attempts to adapt the mathematics of the surprisal perspective to a framing in terms of competitive activation, but it should be possible.

On the other hand, the architecture we have proposed presents severe challenges to a theory of neural implementation. For one thing, it is unknown how to neurally implement any sort of structured representations such as the three levels of *sheepish* or *chew the fat*. Still more problematic is how material is retrieved from LTM into WM, and how WM unifies these retrieved pieces by instantiating variables. A third problem is how new words or phrases are learned: this requires adding new representations to the LTM network by copying them from WM, where they have been constructed in response to input. We are not aware of any proposals about neural computation that can fulfill these functions (with the possible exception of Smolensky & Legendre, 2006). We take it that one of the Big Questions for cognitive neuroscience is how to reconcile these apparently incompatible demands from the representational theory and the theory of neural implementation.

Schemas in processing

We now return to the question of how morphological schemas in the sense of Section 2 play a role in processing.

Much of the literature on morphological processing is concerned with the question of lookup versus computation: When processing morphologically complex words, does one process them as whole words, or does one decompose them into constituents (Nootheboom, Weerman, & Wijnen, 2002)? Various positions have been proposed:

- All decomposition, a.k.a. full parsing (Clahsen, 1999; Taft, 2004)
- All whole-word, a.k.a. full listing (Butterworth, 1983; and in a way Rumelhart & McClelland, 1986)
- Simultaneous decomposition and whole-word retrieval, in competition, a.k.a. Race model (Baayen, 1993)
- Simultaneous decomposition and whole-word retrieval, potentially reinforcing each other (Schreuder & Baayen, 1995; Kuperman et al., 2008))

Given the non-compositional semantics of many morphologically complex words, we can immediately reject the all-decomposition theory. Given speakers' ability to understand novel morphologically complex words – particularly in languages with exuberantly productive morphological systems, such as Turkish (Hankamer, 1989), we can immediately reject the all-whole-word theory. The choice thus comes down to the latter two theories, both in the spirit of promiscuous access. Of these two, we favor the latter: when the decompositional and whole-word strategies result in incompatible candidates, they compete. But when they result in compatible candidates, they reinforce each other, and their redundancy creates a more robust outcome and potentially a faster reaction time.

A crucial part of the analysis is Relational Morphology's hypothesis that, in the representational theory, schemas are lexical items right alongside words. Hence in processing they should behave like just like words. In particular, they should have a resting activation correlated with the frequency with which they are used; and they should be able to be primed by related lexical items or schemas, thereby raising their activation above resting level.

In general, a schema ought to be easier to access than its instances, for two reasons. First, the frequency of a schema is the sum of the frequencies of its instances, so (at least for a first approximation) its resting activation ought to be higher than any of them. Second, an instance of a schema (say *sheepish*) might be expected to activate the schema itself (here, [_A N-*ish*]) more strongly than it activates sister instances (say *childish*). The reason is that the schema's variable does

not conflict with the *sheep* in *sheepish*, whereas the *child* in *childish* does, reducing the intensity of spreading activation.

To see how schemas might play a role in processing, we look step by step at the activation of four types of words: (a) as a baseline, a monomorphemic word such as *hurricane*; (b) a novel bimorphemic word such as *purpleness*; (c) a stored word with a legitimate suffix but whose base is not a lexical item, such as *scrumptious*; (d) a stored bimorphemic word such as *sheepish*.

We idealize to the situation where the word is being heard in isolation. External factors such as priming and syntactic or semantic context will affect all four cases comparably. For example, if previous context has set up a syntactic expectation of an adjective, say *the very...*, this bias will equally boost the activation of *scrumptious* and *sheepish*, and inhibit the activation of *hurricane* and *purpleness*. We also note that visually presented words will have a different timecourse, because the word can be perceived all at once rather than sequentially.

The monomorphemic word *hurricane* sets the background for the other three cases.

- By the time /hərɪk/ is heard, the uniqueness point has been reached, and the only viable candidate remaining is indeed *hurricane*.
- Hence the continuation of the input is predicted to be /eyn/, and when that comes, the prediction is confirmed.

Next consider *purpleness*, a novel word with a lexical base. This is an altogether plausible word, given *blackness*, *whiteness*, and *redness*, but it is not a stored lexical item (at least for the authors, before they made it up as an example).

- By the time /pɪpl/ is heard, the cohort has been reduced to *purple*, whose semantics primes that of other color words to some extent.
- But this time there is no hint that *purpleness* is coming, since (by assumption) there is no such word in the lexicon that could be activated.
- When /nəs/ is heard, the lexical item *purple* is now incompatible with the input (“This isn’t a word I know”), and it may start to decay in WM.¹²
- However, /nəs/ activates the [_N A-ness] schema, which calls for a preceding adjective to instantiate its variable.
- *Purple*, still (somewhat) active in WM, satisfies the variable. It becomes reactivated and unifies with the [_N A-ness] schema, accounting for the input.
- Meanwhile, [_N A-ness] might prime all the listed *-ness* words to some degree.

12. If the word were not heard in isolation, the syllable /nəs/ might be part of a following word, e.g. *purple nasturtium*, leading to further hypotheses in competition.

Hence this is a pure case of what the literature calls decomposition or computation.

The third case is *scrumptious*. This has the common suffix *-ous*, but it is attached to a nonce base *scrum(p)(t)* (speakers differ on the exact pronunciation). The literature calls such bases “bound roots.” They are ubiquitous in English morphology, for instance *deleterious*, *generous*, *egregious*; *commotion*, *contraption*, *compassion*; *deprecate*, *insulate*, and *vindicate*. To our knowledge, neither morphological theory nor psycholinguistic research has had much to say about such words. However, they are no marginal phenomenon. For instance, at least a third of the hundreds of *-ous* words in English are of this character, and the proportion of verbs with *-ate* is if anything higher.¹³

Our intuition is that such items ought to take less effort to identify than equally long and comparably frequent monomorphemic words such as *orchestra* and *hurricane*. Walking through the processing of *scrumptious*:

- By the time /skrʌmptʃ/ is heard, the cohort in WM has been reduced to the single word *scrumptious*; unlike with *purpleness*, there is no word **scrumpt* in the lexicon that could be activated.
- The word *scrumptious* in WM activates the [_A X-ous] schema.
- It also primes all the other *-ous* words such as *joyous* and *glamorous*. But the schema is primed more than its instances, because its variable does not conflict with /skrʌmptʃ/, whereas /dʒɔj/ and /glæməɪ/ do conflict with it. (Non-adjectival words such as *artifice* may be primed too, but even less, as they are morphosyntactically unlike *scrumptious*.)
- The resting activation of [_A X-ous] is higher than that of all its instances, as discussed above. Hence already at the point where only /skrʌmptʃ/ has been heard, the [_A X-ous] schema is ready to go.
- When the whole word /skrʌmptʃəs/ has been heard, the lexical item *scrumptious* has already achieved a presence in WM and can satisfy the input.
- But in addition, now the [_A X-ous] schema is called to WM, again priming all the *-ous* words a little.
- [_A X-ous] then partially satisfies the input, redundantly.

We conjecture that this redundancy should add strength and/or robustness and/or speed to the identification of the input as the word *scrumptious*, compared to a monomorphemic word, which does not get the extra boost. We eagerly await the experiment. (Schreuder & Baayen, 1995 argue that redundancy speeds processing in a different set of cases.)

13. These estimates are based on an informal count by RJ of the words he recognizes in a reverse dictionary created by Mark Lindsay and brought to our attention by Mark Aronoff. (The exact numbers are of no consequence.)

Finally, we return to *sheepish*, which combines the conditions of the previous two cases: the whole word is a lexical item, and the base is also a lexical item.

- At the point where /ʃip/ has been heard, it activates the lexical items *sheep*, *sheepish*, *shepherding*, and perhaps others, in proportion to their resting activation.
- All these lexical items prime each other, along with other neighbors.
- In addition, once *sheepish* is primed, it primes all the other *-ish* words, but especially the [_A N-*ish*] schema (just as *scrumptious* primes the [_A X-*ous*] schema above).
- When the whole word /ʃipɪʃ/ has been heard, the lexical items *sheep* and *shepherding* are incompatible with the input and start to decay; *sheepish* becomes activated still more.
- But now [_A N-*ish*] is called to WM, and because it has been primed, it comes relatively quickly.
- It satisfies the input, redundantly supporting the affix of *sheepish* (parallel to *-ous* in *scrumptious*).
- In addition, its variable can be unified with *sheep*, creating a structure that is compatible with the input (parallel to the case of *purpleness*). This revives the activation of *sheep* (but does not revive *shepherding*). Hence the interpretation of the input as the word *sheepish* is supported by both the lexical item *sheepish* and the assembly of the lexical items *sheep* and [_A N-*ish*].

We therefore expect *sheepish* to be identified more robustly and faster than monomorphemic *hurricane*, purely compositional *purpleness*, and partially compositional *scrumptious*, all else being equal.

The compositional route naturally takes more steps than the whole-word route: the base has to be retrieved, the affix schema has to be retrieved, and the base has to be unified with the affix. Each step takes time. So the contribution of the whole word will tend to dominate in determining the speed of processing. However, if the whole word is infrequent, hence relatively slow, but the base and the affix are both frequent, hence relatively fast, the compositional route may be faster and/or stronger. This is in accord with Baayen and Lieber (1991), who show that the compositional route takes precedence if the affix in question has a large number of low-frequency instances, and particularly if the base is of relatively high frequency compared to the derived word. (There may in addition be individual differences in this balance. For instance, Reifegerste (2014) shows that older speakers tend to rely on storage to a greater extent than younger speakers, who make relatively greater use of composition.)

Hay and Baayen (2005) discuss such gradient results and conclude that the compositionality of morphologically complex items is not discrete but graded: words are divided into morphemes *to such-and-such a degree*. We suggest that the gradient they observe is a result of the relative strength and/or speed of whole-word access vs. compositional access. In turn, compositional access depends on (a) the resting activation of the affix (which depends on how numerous and how frequent its instances are), (b) how similar the remaining part of the word is to a lexical base (in *sheepish*, a lot; in *malicious*, less, because of the phonological differences from *malice*; in *scrumptious*, not at all), and (c) if there is a lexical base, its own resting activation.

Conclusions

We have sketched here the theory of Relational Morphology, based on many antecedents in the literature. It is important that this is not simply a theory of morphology: it serves as an additional component of the Parallel Architecture, helping to fulfill the aspiration for an integrated theory of linguistic description. In particular, we have taken care to show that many innovations we have proposed for morphology apply to syntax as well.

We have furthermore tried to show that RM can be embedded directly into a processing theory of some sophistication. Our reasoning rests on a large number of premises that we have tried to lay out here.

From the representational theory of Relational Morphology:

- Words and rules (now schemas) are stated in the same format, namely pieces of linguistic structure connected by interface links.
- Words and schemas are stored together in a network of lexical relations, whose connections are formalized as relational links.
- Productive rules/schemas are only one part of the grammar; lexical relations are at least equally as important.

From the processing theory in which Relational Morphology is embedded:

- Schemas are stored in memory with a resting activation related to their frequency, just like words.
- Therefore schemas can prime their instances and be primed by them.
- Schemas unify in WM with their bases, just like verbs unify with their arguments.

Admittedly, we have only scratched the surface here. For instance, similar analyses need to be constructed for visual word recognition (both in ordinary reading and in eye-tracking), for production in picture naming and continuous speech, for phoneme and word monitoring, for masked priming, and other natural and experimental tasks. Nor have we said anything about the many other complex effects (e.g. neighborhood density) cited by e.g. deJong (2004), Kuperman et al. (2008), Moscoso del Prado Martín et al. (2004), Bien, Baayen, and Levelt (2011), Kapatsinski (2007), and Amenta and Crepaldi (2012). Nevertheless, we hope our approach offers an incentive to revisit the experimental literature and to design new experiments. We also hope that it might invite computational modeling that allows one to fine-tune the parameters of the theory such as activation strength, the effects of spreading activation, and the relative timing of the many events taking place in lexical access, in order to best simulate the experimental results.

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*. doi:10.3389/fpsyg.2012.00232
- Baayen, H. (1993). On frequency, transparency and productivity. In G. Booij & J. van Marle (Eds.), *Yearbook of Morphology 1992*, 181–208. Dordrecht: Kluwer.
- Baayen, R. H., Dijkstra, T., & Schreuder, R. (1997). Singulars and plurals in Dutch: Evidence for a parallel dual-route model. *Journal of Memory and Language*, 37, 94–117. doi:10.1006/jmla.1997.2509
- Baayen, R. H., & Lieber, R. (1991). Productivity and English derivation: A corpus-based study. *Linguistics*, 29, 801–844. doi:10.1515/ling.1991.29.5.801
- Baddeley, A. (1986). *Working memory*. Oxford: Clarendon Press.
- Berwick, R., & Chomsky, N. (2016). *Why only us?* Cambridge, MA: MIT Press.
- Bien, H., Baayen, R. H., & Levelt, W. J. M. (2011). Frequency effects in the production of Dutch deverbal adjectives and inflected verbs. *Language and Cognitive Processes*, 26, 683–715. doi:10.1080/01690965.2010.511475
- Bock, K., & Loebell, H. (1990). Framing sentences. *Cognition*, 35, 1–39. doi:10.1016/0010-0277(90)90035-l
- Booij, G. (2010). *Construction morphology*. Oxford: Oxford University Press.
- Booij, G. (to appear). Inheritance and motivation in Construction Morphology. In N. Gisborne & A. Hippisley (Eds.), *Default inheritance*. Oxford: Oxford University Press.
- Butterworth, B. (1983). Lexical representation. In B. Butterworth (Ed.), *Language production* (vol. 2, pp. 257–294). San Diego, CA: Academic Press.
- Bybee, J. (1985). *Morphology: A study of the relation between meaning and form*. Amsterdam: John Benjamins. doi:10.1075/tsl.9
- Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes*, 10, 425–455. doi:10.1080/01690969508407111

- Bybee, J. (2001). *Phonology and language use*. Cambridge: Cambridge University Press.
doi:10.1017/CBO9780511612886
- Bybee, J. (2010). *Language, usage and cognition*. Cambridge: Cambridge University Press.
doi:10.1017/CBO9780511750526
- Bybee, J., & Slobin, D. (1982). Rules and schemas in the development and use of the English past tense. *Language* 58: 265–289. doi:10.1353/lan.1982.0021
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Clahsen, H. (1999). Lexical entries and rules of language: A multidisciplinary study of German inflection. *Behavioral and Brain Sciences*, 22, 991–1060.
- Collins, A., & Loftus, E. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, 82, 407–428. doi:10.1037/0033-295X.82.6.407
- Collins, A., & Quillian, M. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 9, 240–247. doi:10.1016/S0022-5371(69)80069-1
- Culicover, P. W. (1999). *Syntactic nuts*. Oxford: Oxford University Press.
- Culicover, P. W., & Jackendoff, R. (2005). *Simpler Syntax*. Oxford: Oxford University Press.
doi:10.1093/acprof:oso/9780199271092.001.0001
- Culicover, P. W., & Nowak, A. (2003). *Dynamical grammar*. Oxford: Oxford University Press.
- Daelemans, W., De Smet, K., & Gazdar, G. (1992). Inheritance in natural language processing. *Computational Linguistics*, 18, 205–218.
- deJong, N. (2002). *Morphological families in the mental Lexicon*. Nijmegen: Max Planck Institute for Psycholinguistics.
- de Saussure, F. (1916). *Cours de linguistique générale*, ed. by C. Bally & A. Sechehaye, with the collaboration of A. Riedlinger, Lausanne and Paris: Payot; trans. by W. Baskin, *Course in general linguistics*, Glasgow: Fontana/Collins, 1977.
- Fiorentino, R., & Poeppel, D. (2007). Compound words and structure in the lexicon. *Language and Cognitive Processes*, 22, 953–1000. doi:10.1080/01690960701190215
- Fodor, J. D. (1998). Learning to parse? In D. Swinney (Ed.), *Anniversary Issue of Journal of Psycholinguistic Research*, 27(2), 285–318.
- Gallistel, C. R., & King, P. A. (2009). *Memory and the computational brain*. Malden, MA: Wiley-Blackwell. doi:10.1002/9781444310498
- Goldberg, A. (1995). *Constructions: A Construction Grammar approach to argument structure*. Chicago: University of Chicago Press.
- Goldberg, A. (2005). *Constructions at work*. New York: Oxford University Press.
doi:10.1093/acprof:oso/9780199268511.001.0001
- Hagoort, P., 2005. On Broca, brain, and binding: A new framework. *Trends in Cognitive Sciences*, 9, 416–423. doi:10.1016/j.tics.2005.07.004
- Hale, J. (2003). The information conveyed by words in sentences. *Journal of Psycholinguistic Research*, 32, 101–123. doi:10.1023/A:1022492123056
- Hale, J. (2011). What a rational parser would do. *Cognitive Science*, 35, 399–443.
doi:10.1111/j.1551-6709.2010.01145.x
- Halle, M., & Marantz, A. (1993). Distributed morphology and the pieces of inflection. In K. Hale & S. J. Keyser (Eds.), *The view from building 20* (pp. 111–176). Cambridge, MA: MIT Press.
- Hankamer, J. (1989). Morphological parsing and the lexicon. In W. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 392–408). Cambridge, MA: MIT Press.
- Hay, J., & Baayen, H. (2005). Shifting paradigms: Gradient structure in morphology, *Trends in Cognitive Science*, 9(7), 342–348. doi:10.1016/j.tics.2005.04.002

- Hoffman, T., & Trousdale, G. (Eds.). (2013). *The Oxford handbook of Construction Grammar*. Oxford: Oxford University Press. doi:10.1093/oxfordhb/9780195396683.001.0001
- Jackendoff, R. (1975). Morphological and semantic regularities in the Lexicon, *Language*, 51(3), 639–671. doi:10.2307/412891
- Jackendoff, R. (2002). *Foundations of language*. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780198270126.001.0001
- Jackendoff, R. (2007). A Parallel Architecture perspective on language processing. *Brain Research*, 1146, 2–22. doi:10.1016/j.brainres.2006.08.111
- Jackendoff, R. (2010). *Meaning and the lexicon*. Oxford: Oxford University Press.
- Jackendoff, R., & Audring, J. (forthcoming). *The texture of the mental Lexicon*.
- Kapatsinski, V. (2007). Frequency, neighborhood density, age-of-acquisition, lexicon size, neighborhood density and speed of processing: Towards a domain-general, single-mechanism account. In S. Buescher, K. Holley, E. Ashworth, C. Beckner, B. Jones, & C. Shank (Eds.), *Proceedings of the 6th annual High Desert Linguistics Society Conference* (pp. 121–140). Albuquerque, NM: High Desert Linguistics Society.
- Kuperberg, G., & Jaeger, F. (2015). What do we mean by prediction in language processing? *Cognition and Neuroscience*, 31(1), 32–59.
- Kuperman, V., Bertram, R., & Baayen, R. H. (2008). Morphological dynamics in compound processing. *Language and Cognitive Processes*, 23, 1089–1132. doi:10.1080/01690960802193688
- Lakoff, G. (1987). *Women, fire, and dangerous things*. Chicago: University of Chicago Press. doi:10.7208/chicago/9780226471013.001.0001
- Langacker, R. (1987). *Foundations of Cognitive Grammar* (vol. 1). Stanford, CA: Stanford University Press.
- Levelt, W., Roelofs, A., & Meyer, A. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22, 1–75.
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3), 1126–1177. doi:10.1016/j.cognition.2007.05.006
- Libben, G. (2006). Why study compounds? An overview of the issues. In G. Libben & G. Jarema (Eds.), *The representation and processing of compound words* (pp. 1–21). Oxford: Oxford University Press.
- Marcus, G. (1998). Rethinking eliminative connectionism. *Cognitive Psychology*, 37, 243–282. doi:10.1006/cogp.1998.0694
- Marcus, G. (2001). *The algebraic mind*. Cambridge MA: MIT Press.
- Marr, D. (1982). *Vision*. San Francisco: Freeman.
- Marslen-Wilson, W., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition*, 8, 1–71. doi:10.1016/0010-0277(80)90015-3
- Marslen-Wilson, W. (1987). Functional parallelism in spoken language understanding. *Cognition*, 25, 71–102. doi:10.1016/0010-0277(87)90005-9
- Marslen-Wilson, W., & Zwitserlood, P. (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 576–585. doi:10.1037/0096-1523.15.3.576
- Moscato del Prado Martín, F., Bertram, R., Häikiö, T., Schreuder, R., & Baayen, R. H. (2004). Morphological family size in a morphologically rich language: The case of Finnish compared with Dutch and Hebrew. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(6), 1271–1278. doi:10.1037/0278-7393.30.6.1271
- Murphy, G. (2002). *The big book of concepts*. Cambridge, MA: MIT Press.

- Nooteboom, S., Weerman, F., & Wijnen, F. (Eds.). (2002). *Storage and computation in the language faculty*. Dordrecht: Kluwer. doi:10.1007/978-94-010-0355-1
- O'Donnell, T. J. (2015). *Productivity and reuse in language: A theory of linguistic computation and storage*. Cambridge, MA: MIT Press. doi:10.7551/mitpress/9780262028844.001.0001
- Oldfield, R., & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology*, 17, 273–281. doi:10.1080/17470216508416445
- Pinker, S. (1999). *Words and rules*. New York: Basic Books.
- Plag, I. (2003). *Word-formation in English*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511841323
- Pollard, C., & Sag, I. (1994). *Head-driven Phrase Structure Grammar*. Chicago: University of Chicago Press.
- Radden, G., & Panther, K.-U. (2004). Introduction: Reflections on motivation. In G. Radden & K.-U. Panther (Eds.), *Studies in linguistic motivation* (pp. 1–46). Berlin, New York: Mouton de Gruyter.
- Reifegerste, J. (2014). *Morphological processing in younger and older people: Evidence for flexible dual-route access*. Nijmegen: Max Planck Institute for Psycholinguistics.
- Riehemann, S. (1988). Type-based derivational morphology. *Journal of Comparative Germanic Linguistics*, 2, 49–77. doi:10.1023/A:1009746617055
- Rumelhart, D. (1980). Schemata: The building blocks of cognition. In R. Spiro, B. Bruce, & W. Brewer (Eds.), *Theoretical issues in reading comprehension* (pp. 33–58). Hillsdale, NJ: Erlbaum.
- Rumelhart, D., & McClelland, J. (1986). On learning the past tense of English verbs. In J. McClelland, D. Rumelhart, & the PDP Research Group (Eds.), *Parallel distributed processing* (vol. ii, pp. 216–271). Cambridge, MA: MIT Press.
- Schreuder, R., & Baayen, R. H. (1995). Modeling morphological processing. In L. Feldman (Ed.), *Morphological aspects of language processing* (pp. 131–154). Hove: Erlbaum.
- Siddiqi, D. (to appear). Distributed Morphology. In J. Audring & F. Masini (Eds.), *The Oxford handbook of morphological theory*. Oxford: Oxford University Press.
- Smolensky, P., & Legendre, G. (2006). *The harmonic mind*. Cambridge, MA: MIT Press.
- Swinney, D. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, 18, 645–659. doi:10.1016/S0022-5371(79)90355-4
- Taft, M. (2004). Morphological decomposition and the reverse base frequency effect. *Quarterly Journal of Experimental Psychology*, 57A, 745–765. doi:10.1080/02724980343000477
- Tanenhaus, M., Leiman, J. M., & Seidenberg, M. (1979). Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. *Journal of Verbal Learning and Verbal Behavior*, 18, 427–440. doi:10.1016/S0022-5371(79)90237-8
- Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge MA: Harvard University Press.
- Ullman, M. T. (2015). The declarative/procedural model: A neurobiologically motivated theory of first and second language. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (2nd ed., pp. 135–158). New York: Routledge.
- Wittenberg, E., & Snedeker, J. (2014). It takes two to kiss, but does it take three to give a kiss? Categorization based on thematic roles. *Language, Cognition and Neuroscience*, 29(5), 635–641. doi:10.1080/01690965.2013.831918

- Woods, W. (1980). Multiple theory formation in speech and reading. In R. Spiro, B. Bruce, & W. Brewer (Eds.), *Theoretical issues in reading comprehension* (pp. 59–82). Hillsdale, NJ: Erlbaum.
- Yang, Charles. (2016). *The price of linguistic productivity*. Cambridge, MA: MIT Press.

Corresponding address

Ray Jackendoff
Center for Cognitive Studies
Tufts University
Medford, MA 02155
USA
ray.jackendoff@tufts.edu