Understanding
& Measuring
Morphological
Complexity

EDITED BY MATTHEW BAERMAN,
DUNSTAN BROWN, & GREVILLE G. CORBETT

Understanding and Measuring Morphological Complexity

Edited by

MATTHEW BAERMAN, DUNSTAN BROWN,
AND GREVILLE G. CORBETT

OXFORD

UNIVERSITY PRESS

Contents

8

Computational complexity of abstractive morphology

VITO PIRRELLI, MARCELLO FERRO,
AND CLAUDIA MARZI

8.1 Introduction

In a constructive perspective on morphological theorizing, roots and affixes are the basic building blocks of morphological competence, on the assumption that the lexicon is largely redundancy-free. The speaker, having identified the parts of a word form, proceeds to discard the full form from the lexicon. Fully inflected forms are parsed (in word recognition) or generated (in word production) on line from their building blocks. This is in sharp contrast with a second type of perspective, named abstractive, which treats word forms as basic units and their recurrent parts as abstractions over full forms. According to this perspective, full forms are the founding units of morphological processing, with sub-lexical units resulting from the application of morphological processes to full forms. Learning the morphology of a language thus amounts to learning relations between fully stored word forms, which are concurrently available in the speaker's mental lexicon and jointly facilitate processing of morphologically related forms.

The essential distinction between constructive and abstractive approaches (due to Blevins 2006) is not coextensive with the opposition between morpheme-based and word-based views of morphology. There is an obvious sense in which constructive approaches are morph-based, as they rely on the combination of minimal sub-lexical units. On the other hand, abstractive approaches are typically word-based, since words are taken to be essential to morphological generalizations. However, one can endorse an abstractive view of morphologically complex words as being internally structured into recognizable constituent parts. According to this view, constituent parts are analysed as 'emergent' from independent principles of lexical organization, whereby full lexical forms are redundantly stored and mutually related through entailment

relations (Matthews 1991, Corbett and Fraser 1993, Pirrelli 2000, Burzio 2004, Booij 2010). Conversely, a constructive orientation can underlie a word-based view of morphological structure. This is the case of Stump's notion of 'paradigm function' (Stump 2001), mapping the root of a lexeme and a set of morphosyntactic properties onto the paradigm cell occupied by an inflected form of the lexeme.

More importantly for our present purposes, the two views are also orthogonal to the distinction between symbolic (or rule-based) and sub-symbolic (or neurally inspired) word processing. First, we observe that, in spite of their being heralded as champions of the associative view on morphological structure, the way connectionist neural networks have been used to model morphology learning adheres, in fact, to a strictly derivational view of morphological relations, according to which a fully inflected form is always produced (or analysed) on the basis of a unique, underlying lexical form. By modelling inflection as a phonological mapping function from a lexical base to its range of inflected forms, connectionist architectures are closer to a rule-free variant of the classical constructive view, than to associative models of the mental lexicon. Conversely, according to Albright and Hayes' Minimal Generalization algorithm (2003), speakers conservatively develop structure-based rules of mapping between fully inflected forms. These patterns are based on a cautious inductive generalization procedure: speakers are confident in extending a morphological pattern to other forms to the extent that (i) the pattern obtains for many existing word forms, (ii) there is a context-similarity between words that comply with the pattern, and (iii) there is a context-difference between those word forms and word forms that take other patterns. Accordingly, the speaker's knowledge of word structure is more akin to one dynamic set of relations among fully inflected forms, in line with an overall abstractive approach.[1]

Finally, the separation line between constructive and abstractive views has little to do with the controversy over the psychological reality of morpheme-like units. Constructivists tend to emphasize the role of morphs in both morphology acquisition and processing, whereas in the abstractive view morphs emerge as abstractions over full forms. In fact, we take it to be one of the most important contributions of emergentist approaches to language inquiry the observation that self-organization effects provide an explanatory basis for important aspects of language knowledge. We shall show that self-organization is a determinant of lexical competence and that emergent morphological patterns do play an important role in word processing and acquisition.

---

[1] In fact, the sheer number of leading forms which need to be learned for a speaker to be able to master an entire paradigm is not a defining feature of abstractive approaches. In simple inflectional systems, a single form may well be sufficient to produce all remaining forms of the same paradigm. The computational question of how many different forms are needed to predict the entire range of formal variation expressed by a specific paradigm or by a class of morphologically related paradigms has lately been addressed on either information-theoretic (conditional entropy of cells and paradigms, Moscoso del Prado Martín et al. 2004) or set-theoretic grounds (see Stump and Finkel's contribution to the present volume).

It could be tempting to suggest that the abstractive and constructive views just represent two complementary modes of word processing: (i) a production mode for the constructive perspective, assembling sub-lexical bits into lexical wholes, and (ii) a recognition mode for the abstractive perspective, splitting morphologically complex words into their most recurrent constituent morphs. Blevins (2006) shows that this is not true empirically, as the two approaches lead to very different implications (such as the 'abstractive' possibility for a word form to have multiple bases or 'split-base effect') and allow for a different range of representational statements (e.g. the sharp separation between actual data and abstract patterns arising from data in the constructive approach).

In this chapter, we intend to show that the two views are also computationally different. We contend that a number of problems arising in connection with a sub-symbolic implementation of the constructive view are tackled effectively, or disappear altogether, in a neurally inspired implementation of the associative framework, resting on key-notions such as self-organization and emergence. In particular, we will first go over some algebraic prerequisites of morphological processing and acquisition that, according to Marcus (2001), perceptron-like neural networks are known to meet to a limited extent only (section 8.2). A particular variant of Kohonen's Self-Organizing Maps (2001) is then introduced to explore and assess the implications of a radically abstractive approach in terms of its computational complexity (section 8.3). Experimental data are also shown to provide a quantitative assessment of these implications (section 8.4). Finally, some general conclusions are drawn (section 8.5).

8.2 The algebraic basis of constructive morphology

Marcus (2001) introduces a number of criterial properties defining any descriptively adequate morphological system.

First, the system must have a way to distinguish variables from instances, to be able to make such a simple statement as '*play* is an English verb stem', where *play* is a constant and stem is a variable which can be assigned any English verb stem. As we shall see, such a distinction is essential to capture the intuitive notion of repetition of the same symbol within a string, as when we say that the string *pop* contains two instances (tokens) of the same *p* (type). Second, a proper morphological system must avail itself of the formal means to represent abstract relationships between variables. Even the most straightforward default rule-like statement such as the following

(1) $PROGR(Y) = Y + ing,$

meaning that a progressive verb form in English is built by affixing *ing* after any verb stem Y, requires specification of two relations: (i) an identity relation between the

lexical root *Y* to the left of '=' and the stem *Y* in the right-hand side of the equation,[2] and (ii) the ordering relation between the stem and the progressive inflectional ending *-ing*. Third, there must be a way to bind a particular instance (a constant value) to a given variable. For example, *Y* in (1) can be assigned a specific value (e.g. the individual stem *play* to yield the form *playing*), as the output of the concatenation operation expressed by '+'. Fourth, the system must be able to apply operations to arbitrary instances of a variable. For example, a concatenation operation must be able to combine *ing* with any input value assigned to *Y*, unless *Y*'s domain is otherwise constrained. Finally, the system must have a way to extract relations between variables like (1), on the basis of training examples (*walking*, *sleeping*, *playing*, etc.). Marcus goes on to argue that any morphological system must be able to meet these prerequisites, irrespective of whether it is implemented symbolically as a rule-based processor, or sub-symbolically, as a neurally inspired artificial network. In fact, traditional connectionist models of morphological competence are taken by Marcus to fail to meet at least some of these prerequisites. What is called by Rosenblatt (1962: 73) 'a sort of exhaustive rote-learning procedure', in which every single case of a general relation must be learned individually, appears to be an essential feature of perceptron-like artificial neural networks. Such a characteristic behaviour extends to the family of multi-layered perceptrons classically used to simulate morphology learning. The consequences of this limitation are far-reaching, as detailed in the ensuing sections.

8.2.1 Binding

In a multi-layered perceptron, a variable (or type)—say the letter *p*—can be represented by a single node or, alternatively, by a distributed pattern of nodes, on the input layer. A single instantiation of a *p* in a string like *put* is encoded on the input layer by activating the corresponding node(s). We say that an instance of *p* is bound to its type node(s) through node's activation. However, for a string like *pop* to be properly encoded, two instances of *p* must be distinguished—say p_1 (or *p* in first position) and p_3 (or *p* in third position). Activating the same *p* node twice would not do, as the whole string is encoded as an integrated activation pattern, where all nodes representing input letters are activated concurrently. Logically, the problem amounts to binding the input layer node (the letter type) to a specific individual, a letter token. When it comes to representing words, a token can be identified uniquely by specifying its type together with the token's position in the word string. In this context, the solution to the binding problem amounts to assigning a specific order relationship to a given type. A traditional strategy for doing this with neural networks is known as conjunctive coding.

---

[2] Computationally, this can be implemented by introducing a copy operator, replicating the content of a given memory register into another register, irrespective of the particular nature of such content. As we shall see later on, this is not the only possible computational interpretation of equation (1).

8.2.2 Coding

In conjunctive coding (Coltheart et al. 2001, Harm and Seidenberg 1999, McClelland and Rumelhart 1981, Perry et al. 2007, Plaut et al. 1996), a word form like *pop* is represented through a set of context-sensitive nodes. Each such node ties a letter to a specific serial position (e.g. {P_1, O_2, P_3}), as in so-called positional coding or, alternatively, to a specific letter cluster (e.g. {_PO, POP, OP_}), as is customary in so-called Wickelcoding. Positional coding makes it difficult to generalize knowledge about phonemes or letters across positions (Plaut et al. 1996, Whitney 2001) and to align positions across word forms of differing lengths (Davis and Bowers 2004), as with *book* and *handbook*. The use of Wickelcoding, on the other hand, while avoiding some problems of positional coding, causes an acquisitional dead-lock. Speakers of different languages are known to exhibit differential sensitivity to symbol patterns. If such patterns are hard-wired in the input layer, the same processing architecture cannot be used to deal with languages exhibiting differential constraints on sounds or letters.

8.2.3 Relations and the problem of variables

Consider the problem of providing a neurally inspired version of the relation expressed in (1). The equation conveys an identity mapping between the lexical base in PROGR(Y) and the verb stem Y to the right of '='. A universally quantified relation of this kind is common to a variety of default mapping relations holding in the morphologies of virtually any language. Besides, this sort of free generalization seems to be a fundamental feature of human classificatory behaviour. Straightforward linguistic evidence that people freely generalize universally quantified one-to-one mappings comes from morphological reduplication (or immediate repetition), as found in Indonesian pluralization, whereby the plural of *buku* ('book') is *buku-buku*. Reduplication patterns like this have quite early roots in language development. Seven-month-old infants are demonstrably able to extract an underlying abstract structure from two minutes' exposure to made-up syllabic sequences cast into an *AAB* pattern (*ga-ga-na*, *li-li-ti*, etc.) and to tell this habituation structure from other test patterns containing repetitions in different positions (e.g. *ABB*) (Saffran et al. 1996, Marcus et al. 1999, Gerken 2006).

   A one-to-one mapping function can easily be implemented in a network that uses one node to represent each variable (Figure 8.1, left). The model can freely generalize an identity relation by setting the connection weight to 1. However, it is not at all clear how a one-node-per-variable solution can represent the highly non-linear types of morphological exponence abundantly documented in the literature (see Anderson, section 2.2, this volume for a thorough discussion), such as stem-internal vowel alternation in base/past-tense English pairs such as *ring–rang*, *bring–brought*, and *hang–hung* where different mapping functions are required depending
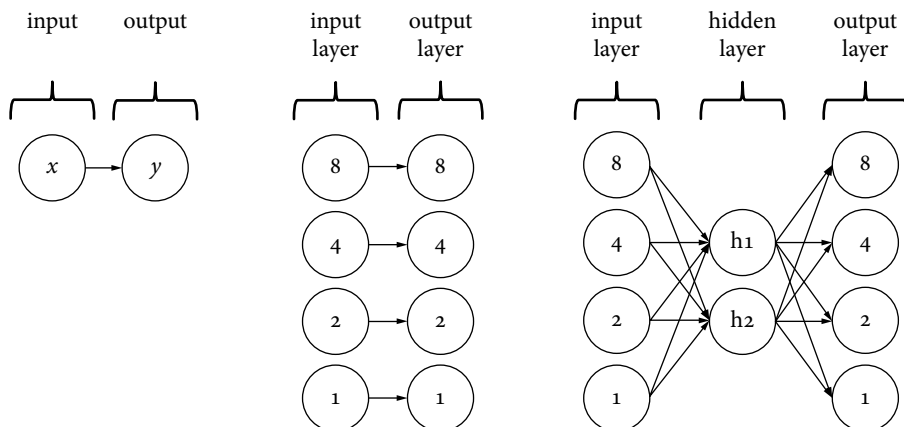
FIGURE 8.1 One-to-one and one-to-many relations in one-node-per-variable and many-nodes-per-variable neural networks.

on both the vowel and its surrounding context. A descriptively more adequate solution is provided by a many-nodes-per-variable perceptron, where each node represents a specific value assigned to each variable, and different connections emanating from the same node can either enforce an identity relation (Figure 8.1, centre) or make room for other mapping relations, with hidden layer units accounting for inter-node influences (Figure 8.1, right). However more flexible, a many-node-per-variable perceptron, when trained by back-propagation, can learn a universally quantified one-to-one mapping relation only if it sees this relation illustrated with respect to each possible input and output node (Marcus 2001). Back to Figure 8.1 (centre), a perceptron trained on '8', '4', and '2' nodes only, is not in a position to extend an identity relation to an untrained '1' node. An implication of this state of affairs is that a multi-layered perceptron can learn the relation expressed by equation (1) only for strings (in the domain of Y) that fall entirely in the perceptron's training space.

8.2.4 Training independence

Technically, the problem of dealing with relations over variables is a consequence of the back-propagation algorithm classically used to train a multi-layered perceptron. Back-propagation consists in altering the weights of connections emanating from an activated input node, for the level of activation of output nodes to be attuned to the expected output. According to the delta rule in equation (2), connections between the jth input node and the ith output node are in fact changed in proportion to the difference between the target activation value $\hat{h}_i$ of the ith output node and the actually observed output value h_i:

(2) $\triangle w_{i,j} = \gamma(\hat{h}_i - h_i)x_j$

where $w_{i,j}$ is the weight of the connection between the jth input node and the ith output node, γ is the learning rate, and x_j is the activation of the jth input node. If x_j is null, the resulting change $\triangle w_{i,j}$ is null. In other words, back-propagation never alters the weights of connections emanating from an input node in Figure 8.1 (centre), if the node is never activated in training. This is called by Marcus (2001) training independence. Interposition of a layer of hidden units mediating input–output mapping (Figure 8.1, right) does not remedy this. There is no direct way in which a change in the connections feeding the output node '1' can affect the connections feeding a different output node.

8.2.5 Preliminary discussion

In many respects, multilayer perceptrons are taken to provide a basic neurally inspired associationist mechanism for morphology processing and acquisition. Much of their success history is due to the methodological allure of their simplicity. It seems reasonable to investigate the inherent representational and procedural limitations of this class of models to understand more of the task of word learning and processing and its basic computational components. Among the main liabilities of multilayer perceptrons, the following ones strike us as particularly critical in the light of our previous discussion.

Training locality. Due to training independence, training one node does not transfer to another node. There is no way to propagate information about connections between two (input and output) nodes to another pair of such nodes. This is not only detrimental to the development of abstract open-ended relations holding over variables (as opposed to pair-wise memorized relationships between specific instances), but also to the modelling of word paradigms as complex morphological interfaces. Over the last fifteen years, considerable evidence has accrued on the critical role of paradigm-based relations as an order-principle imposing a non-local organizing structure on word forms memorized in the speaker's mental lexicon, facilitating their retention, accessibility, and use, while permitting the spontaneous production and analysis of novel words. A number of theoretical models of the mental lexicon have been put forward to deal with the role of these global constraints in (i) setting an upper bound on the number of possible forms a speaker is ready to produce (Carstairs and Stemberger 1988), (ii) accounting for reaction times in lexical decision and related tasks (Baayen et al. 1997, Orsolini and Marslen-Wilson 1997, and others), (iii) explaining production errors by both adults and children (Bybee and Slobin 1982, Bybee and Moder 1983, Orsolini et al. 1998), and (iv) accounting for human acceptability judgements and generalizations over nonce verb stems (Say and Clahsen 2002, Albright 2002). We believe that computational models of morphology acquisition and processing have something to say about all these issues. The basic challenge is how it is possible for such global paradigm-based constraints to emerge on the basis of local learning steps (Pirrelli et al. 2004).

Coding. Another important related point has to do with the nature of input representations. It has often been suggested that distributed representations may circumvent the connectionist failure to encode universally quantified variables (Elman 1998). If the same pool of nodes is used to encode different input patterns, then at least some of the mapping relations holding for one pattern will also obtain for the other ones, thus enforcing a kind of generalization. For example, if each input node represents not a word but a part of a word, then training one word has an impact on nodes that are part of the representation of other words. However, distributed representations can be of help only if the items to which one must generalize share that particular contrast (pool of nodes) on which the model was trained. If the model was trained on a different contrast than the one encoded in the test item, generalization to the test item will fail. More seriously, distributed representations can raise unnecessary coding ambiguity if each individual is encoded as an activation pattern over the pool of nodes corresponding to its variable. If two individuals are activated simultaneously, the resulting joint activation pattern can make it difficult to disentangle one individual from the other. In fact, the representational capacity of the network to uniquely bind a particular individual decreases with the extent of distribution. One extreme is a purely localist representation, with each input node coding a distinct individual. The other extreme is the completely distributed case, where each of the 2^N binary activation patterns over N nodes represents a distinct individual. In this case, no two patterns can be superimposed without spuriously creating a new pattern.

Alignment. Lack of generalization due to localist encoding is an important issue in acquisition of morphological structure, as it correlates with human perception of word similarity. The problem arises whenever known symbol patterns are presented in novel arrangements, as when we are able to recognize the English word *book* in *handbook*, or the shared root *mach* in German *machen* and *gemacht*. Conjunctive coding of letters is closer to the localist extreme, anchoring a letter either to its position in the string, or to its surrounding context. Since languages wildly differ in the way morphological information is sequentially encoded, ranging from suffixation to prefixation, circumfixation, apophony, reduplication, interdigitation, and combinations thereof, alignment of lexical roots in three diverse pairs of paradigmatically related forms like English *walk–walked*, Arabic *kataba–yaktubu* ('he wrote'–'he writes') and German *machen–gemacht* ('make'–'made' past participle), requires substantially different encoding strategies. If we wired any such strategy into lexical representations (e.g. through a fixed templatic structure separating the lexical root from other morphological markers) we would in fact slip morphological structure into the input, making input representations dependent on languages. This is not wrong *per se*, but it should be the outcome, not a prerequisite of lexical acquisition. A cognitively plausible solution is to let the processing system home in on the right sort of encoding strategy through repeated exposure to a range of language-specific families of morphologically related words. This is what conjunctive coding in classical connectionist architectures cannot do.

For example, in German *machen* and *gemacht* the shared substring *mach* would be indexed to different time positions. In Arabic, the set of Wickelcodes {_Kᴀ, ᴋAт, ᴀTᴀ, тAв, ᴀBᴀ, ʙA_} encoding the perfective form *kataba* turns out to have an empty intersection with the set {_Yᴀ, ʏAᴋ, ᴀKт, ᴋTu, тUʙ, uBu, ʙU_} for the imperfective form *yaktubu*.

Relations between variables and the copy operator. The apparent need for algebraic variables and open-ended relationships between variables in morphological rules such as (1) makes the computational problem of implementing copying mechanisms centre-stage in word processing. This is illustrated by the two-level transducer of Figure 8.2, implementing a classical constructive approach to word parsing and generation. The input tape represents an input memory register containing an abstract representation of the Italian form *vengono* ('they come'), with *ven* representing the verb root, '$V3$' its conjugation class, and '3p_pres_ind' the feature specification 'third person singular of present indicative' ('+' is a morpheme boundary). The string resulting from the application of the finite state transducer to the input string is shown on the output tape. The mapping relationship between characters on the two tapes is represented by the colon (':') operator in the centre box, with input characters appearing to the left of ':', and the corresponding output characters to its right. When no colon is specified, the input letter is intended to be left unchanged. Note that 'ε' stands for the empty symbol. Hence the mapping '[+ : ε]' reads 'delete a marker of morpheme boundary'. Most notably, transductive operations of this kind define declarative, parallel, and bi-directional relationships. We can easily swap the two tapes and use, for word parsing, the same transducer designed for word generation.[3]

Multi-layered perceptrons seem to have little to match such a rich weaponry of formal tools and computational operations. Nonetheless, it should be appreciated that multiple copying of unbounded strings is an operation that goes beyond the computational power of finite-state transducers such as the one in Figure 8.2. This is an immediate consequence of the fact that states represent the order of memory of a transductive automaton. For an automaton to be able to replicate an exact copy of any possible string, irrespective of its length, the number of available states needed to precompile all copies is potentially unbounded. As pointed out by Roark and Sproat (2007), if fully reduplicative morphological phenomena require potentially unbounded copying, they are probably among the very few morphological phenomena which are not in the purview of finite-state processing devices.

It seems natural to establish a parallelism between the two-level architecture in Figure 8.2 and the multi-layered perceptron of Figure 8.1 (right). In both cases, we

---

[3] Due to the hybrid status of lexical representations in two-level automata, where both surface and abstract morphological information is combined on the same level, abstract features (e.g. conjugation class) must be linearly arranged and interspersed with phonological information. This may strike the reader as rather awkward and other neater solutions have been entertained in the literature. However, as argued by Roark and Sproat (2006), they do not add up to the computational power of two-level finite state machines.
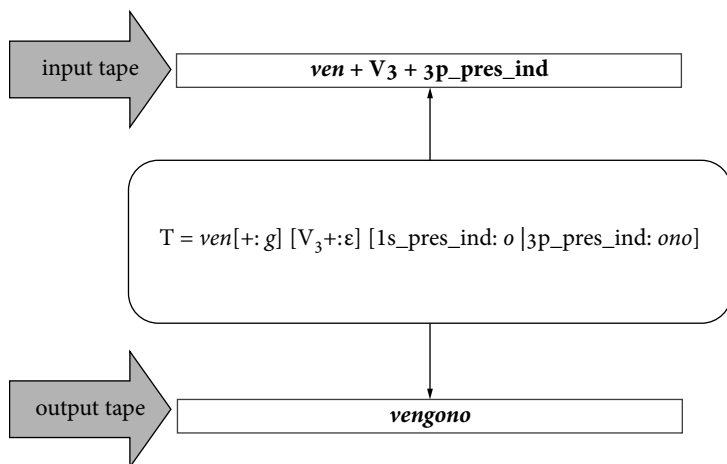
FIGURE 8.2   A two-level finite state transducer for the Italian irregular form *vengono* 'they come'.

have an input and an output representation layer, which could be likened to digital registers. In both cases, mapping relations enforce the requirement that at least part of the representation in either layer is also found in the other layer. It is also tempting to conceptualize such relation as a copy operation. However, although this is indeed the case for finite-state transducers, it is not literally true of multi-layered perceptrons, where data copying is implemented as weight adjustment over inter-layer connections. According to Marcus, this is a fundamental liability of multi-layered perceptrons. Since they cannot implement unbounded copying, he argues, they are not suitable computer models of morphology learning.

In fact, we observed that universal copying is not a problem for perceptrons only. Finite-state transducers are in principle capable of dealing with multiple copying, as long as the number of strings in the domain of the copy operation is bounded. This is somewhat reminiscent of the local constraint on weight adjustments in multi-layered perceptrons, where the domain of identity mapping must be defined with respect to each possible input and output node. Admittedly, the two limitations are of a different nature and are due to different reasons. Nonetheless, it is important to appreciate that universal copying is a thorny computational issue, and its necessity for word processing and learning should be argued for with care.

8.2.6 Interim summary

In this section, we have been through a number of criticisms levelled at classical connectionist networks dealing with issues of word processing with no symbolic manipulation of typed variables. In the remainder of this chapter, we will make the suggestion that many of the liabilities pointed out by Marcus in connection with

multi-layered perceptrons in fact apply to a specifically constructive approach to morphological processing. In particular, they are due to (i) the idea that morphological inflection and derivation require a formal mapping from lexical bases to full forms, on (ii) the assumption that underlying bases and inflected forms are represented on two distinct layers of connectivity (or, equivalently, in two distinct registers). Such a theoretically loaded conceptualization makes computational models of inflection and derivation unnecessarily complex. In the following sections we propose an alternative neurally inspired view of word processing and learning that is strongly committed to an abstractive perspective on these and related issues. The approach amounts to a re-interpretation of equation (1) in terms of different computational operations than copying and concatenation. It seems likely that an adequate account of morphology acquisition will have a place both for memory mechanisms that are sensitive to frequency and similarity, and for processing operations that apply to all possible strings. A fundamental assumption of the computational framework proposed here is that both aspects, lexical memory and processing operations, are in fact part and parcel of the same underlying computational mechanism. This mechanism is considerably simpler than advocates of algebraic symbol manipulation are ready to acknowledge, and points in the direction of a more parsimonious use of variables and universally quantified one-to-to mappings. This is not to imply that these notions play no role in morphology learning or language learning in general, but that their role in word learning and processing should be somewhat qualified.

8.3 A computational framework for abstractive morphology

Kohonen's Self-Organizing Maps (SOMs) (Kohonen 2001) define a class of unsupervised artificial neural networks that mimics the behaviour of small aggregations of neurons (pools) in the cortical areas involved in the classification of sensory data (brain maps). In such aggregations, processing consists in the activation of specific neurons upon presentation of a particular stimulus. A distinguishing feature of brain maps is their topological organization (Penfield and Roberts 1959): nearby neurons in the map are activated by similar stimuli. There is evidence that at least some aspects of their neural connectivity emerge through self-organization as a function of cumulated sensory experience (Kaas et al. 1983). Functionally, brain maps are thus dynamic memory stores, directly involved in input processing, exhibiting effects of dedicated long-term topological organization.

Topological Temporal Hebbian SOMs (hereafter TSOMs, Ferro et al. 2011, Marzi et al. 2012c) represent a variant of classical SOMs making room for memorizing time series of symbols as activation chains of map nodes. This is made possible by a level of temporal connectivity (or temporal connection layer, Figure 8.3, left), implemented as a pool of re-entrant connections providing the state of activation of the map at the immediately preceding time tick. Temporal connections encode the
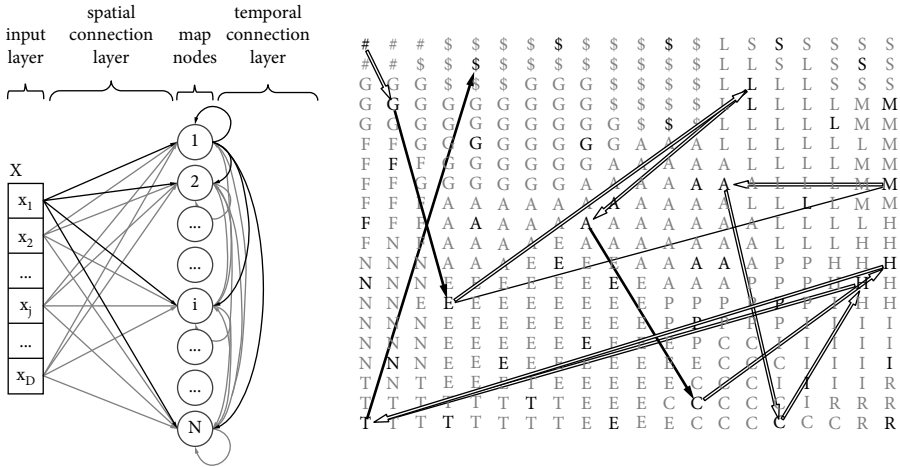
Figure 8.3 **Left**: Outline architecture of a TSOM. Each node in the map is connected with all nodes of the input layer. Input connections define a communication channel with no time delay, whose synaptic strength is modified through training. Connections on the temporal layer are updated with a fixed one-step time delay, based on activity synchronization between *BMU(t–1)* and *BMU(t)*. **Right**: A two-dimensional 20 × 20 TSOM, trained on German verb forms, showing activation chains for *gelacht* and *gemacht* represented as the input strings #,G,E,L,A,C,H,T,$ and #,G,E,M,A,C,H,T,$ respectively. See text for more details.

map's probabilistic expectations of upcoming events on the basis of past experience. In the literature, there have been several proposals of how this idea of re-entrant temporal connections can be implemented (see Voegtlin 2002 and references therein). In this chapter, we will make use of one particular such proposal (Koutnik 2007, Pirrelli et al. 2011).

In language, acoustic symbols are concatenated through time to reach the hearer as a time-bound input signal. With some qualifications, also written words can be defined as temporal patterns, conceptualized as strings of letters that are produced and input one letter at a time, under the assumption that time ticks are sampled discretely upon the event of letter production/presentation. In our view, this abstract conceptualization highlights some peculiar (and often neglected) aspects of the linguistic input and provides an ecological setting for dealing with issues of language production, recognition, and storage at an appropriate level of abstraction. TSOMs are designed to simulate and test models of language processing targeted at this level of abstraction.

8.3.1 *Input protocol: peripheral encoding*

In our idealized setting, each word form is represented by a time-series of symbols (be they phonological segments or printed letters) which are administered to a TSOM by encoding one symbol at a time on the input layer (Figure 8.3, left). The input

layer is implemented as a binary vector whose dimension is large enough for the entire set of symbols to be encoded, with one pattern of bits uniquely associated with each symbol. Patterns can partially overlap (distributed representations) or can be mutually orthogonal (localist representations). In all experiments presented here, we adopted orthogonal codes. A whole word is presented to a map starting with a start-of-word symbol ('#') and ending with an end-of-word symbol ('$'). Map's re-entrant connections are initialized upon presentation of a new word. The implication of this move is that the map's activation state upon seeing the currently input word form has no recollection of past word forms. Nonetheless, the map's overall state is affected by previously shown words through long-term learning, as detailed in what follows.

8.3.2 Parallel activation: recoding

Upon presentation of one symbol on the input layer all map nodes are activated simultaneously through their input (or spatial) connections. Their resulting level of activation at time t, $h_i(t)$, is also determined by the contribution of re-entrant connections, defined over the map's temporal layer, which convey information about the state of map's activation at time t-1, according to the following equation:

(3) $h_i(t) = \alpha \cdot h_{S,i}(t) + \beta \cdot h_{T,i}(t)$.

Equation (3) defines the level of activation of the ith node at time t as a weighted summation of the contribution $h_{S,i}$ of the spatial layer and the contribution $h_{T,i}$ flowing through the map's temporal layer. Given (3), we define the map's Best Matching Unit at time t (hereafter $BMU(t)$) as the node showing the highest activation level at time t.

$BMU(t)$ represents (a) the map's response to input, and (b) the way the current input stimulus is internally recoded by the map. As a result, after a string of letters is presented to the map one character at a time, a temporal chain of BMUs is activated. Figure 8.3 (right) illustrates two such temporal chains, triggered by the German verb forms *gelacht* ('laughed', past participle) and *gemacht* ('made', past participle), which are input as the series #,G,E,L,A,C,H,T,$ and #,G,E,M,A,C,H,T,$ respectively. In the figure, each node is labelled with the letter the node gets most sensitive to after training. Pointed arrows represent temporal connections linking two consecutively activated BMUs. They thus depict the temporal sequence of symbol exposure (and node activation), starting from the symbol '#' (anchored in the top left corner of the map) and ending with '$'. Temporal BMU chains of this kind represent how the map recodes input words.

8.3.3 Long-term dynamic: training

In the learning phase, at each time step t, $BMU(t)$ adjusts its connection weights over the spatial and temporal connection layers (Figure 8.3, left) and propagates adjustment to neighbouring nodes as an inverse function of their distance from $BMU(t)$.

Adjustment makes connection weights closer to values in the input vector. This amounts to a topological specialization of individual nodes which get increasingly sensitive to specific symbol types. Notably, specialization does not apply across the board. Due to the propagation function, adjustment is stronger for nodes that are closer to *BMU(t)*. Moreover, the topological range of propagation shrinks as learning progresses. This means that, over time, the propagation wave gets shorter and shorter, reaching fewer and fewer nodes.

The temporal layer presents a similar dynamic. Adjustment of Hebbian connections consists of two steps: (i) potentiate the strength of association from *BMU(t-1)* to *BMU(t)* (and its neighbouring nodes), and (ii) depress the strength of association from all other nodes to *BMU(t)* (and neighbouring nodes) (Figure 8.4, left). The two steps enforce the logical entailment *BMU(t)* → *BMU(t-1)* and the emergence of a context-sensitive recoding of symbols. This means that the same symbol type will recruit different *BMUs* depending on its preceding context in the input string. This is shown in Figure 8.3 (right) where the letter *a* in *gemacht* and *gelacht* activates two different nodes, the map keeping memory of their different left contexts. Interestingly,
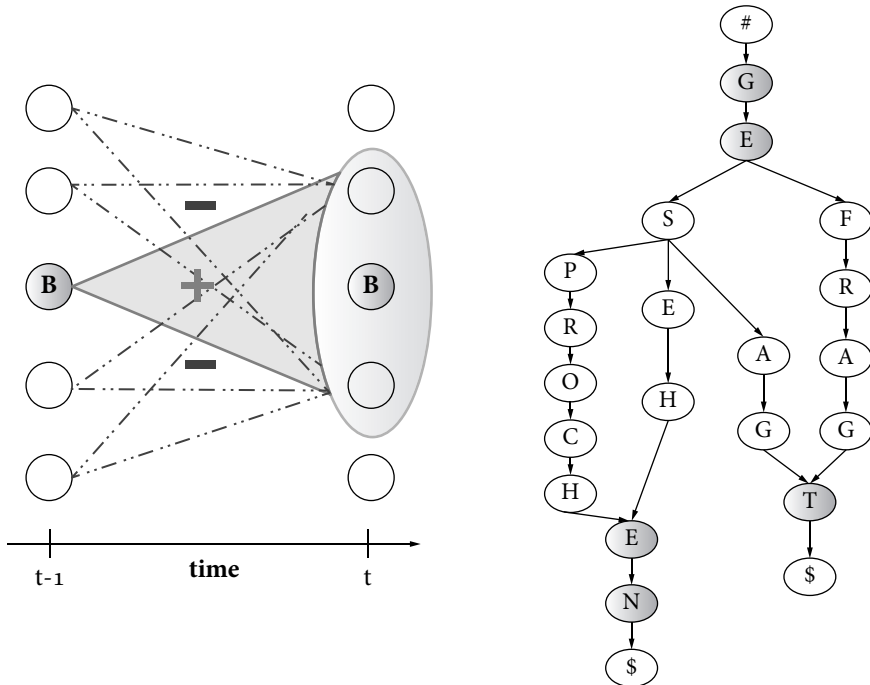

FIGURE 8.4 Left: Topological propagation of long-term potentiation (solid lines) and long-term depression (dotted lines) of temporal re-entrant connections over two successive time steps. B-nodes indicate *BMUs*. **Right**: Word-graph representation of German past participles.

the Markovian order of the map memory is not limited to one character behind but can rise as the result of dynamic recoding. If space on the map allows, the different *BMU*s associated with *a* in *gemacht* and *gelacht* will, in turn, select different *BMU*s for the ensuing *c*, as shown in Figure 8.3. Were nodes trained independently, the map would thus exhibit a tendency to dedicate distinct activation chains to distinct input strings (memory resources allowing). This happens if the topological range of the propagation function for temporal connections goes down to 0. In this case, the map can train temporal connections to each single node independently of connections to neighbouring nodes. However, if the topological range of the propagation function for temporal connections does not get to 0, adjustments over temporal connections will transfer from one node to its neighbouring nodes (Figure 8.4, left). In this condition, activation chains triggered by similar words will tend to share nodes, reflecting the extent to which the map perceives them as identical. The topological distance between chains of activated *BMU*s responding to similar input strings tells us how well the map is aligning the two strings.

8.4 Implications of TSOMs for word processing and storage

TSOMs are dynamic models of string storage. They can be seen as mimicking the self-organizing behaviour of a mental lexicon growing in size as a function of cumulated lexical experience. The more word forms are stored in a TSOM through lexical exposure, the more complex its overall internal structure becomes, with recurrent patterns of morphological structure being internally recoded as shared chains of node activation. This is shown in Figure 8.4 (right) (adapted from Marzi et al. 2012a), where the activation chains for the German past participles *gesprochen* ('spoken'), *gesehen* ('seen'), *gesagt* ('said'), and *gefragt* ('asked') are unfolded and vertically arranged in a word graph.

It is important to appreciate that this behaviour is the outcome of an internally generated dynamic, not a function of either direct or indirect supervision. The map is in no way taught what the expected morphological structure of input word forms is. Structure is an emergent property of self-organization and a function of exposure to input representations. In a strict sense, TSOMs offer no output representations. In a broader sense, input and output representations co-exist on the same layer of map circuitry. It is the level of node connectivity that changes through learning. Output morphological structure can thus be defined as an overall patterning of long-term connectivity. Such a radical change of perspective has important implications both theoretically and computationally.

8.4.1 *Coding*

A fundamental difference between TSOMs and traditional perceptrons is that in TSOMs we distinguish among levels of input coding. In particular, note that the

input vector, peripherally encoded on the input layer in Figure 8.3 (left), is eventually recoded on the map proper. It is the latter level that provides the long-term representation upon which morphological structure is eventually perceived.[4] Since recoding is the result of learning, different cumulated input patterns may lead to different recoded representations. By neglecting the important distinction between levels of input coding, perceptrons exhibit a principled difficulty in dealing with input representations. Conjunctive coding can naturally deal with issues of input specificity if two conditions are met: (i) there is a topological level of encoding where two nodes that are sensitive to the same symbol type share identical internal representations; and (ii) context-sensitive specialization on symbol tokens is acquired, not wired-in. In TSOMs, both conditions are met: (i) nodes that are selectively sensitive to the same symbol type have overlapping representations on the spatial layer, and (ii) nodes become gradually sensitive to time-bound representations through specialization of their temporal connections. Hence, different input instances of the same symbol receive differently recoded representations, thus effectively dealing with variable binding when the same symbol type occurs more than once in the same input string.

8.4.2 *Time vs Space: Differences in recoding strategies*

A TSOM recoding strategy may reflect either a bias towards the local context where a symbol is embedded (i.e. its immediately preceding symbols), or a tendency to conjunctively encode the symbol *and* its position in the string. Choice of either strategy is governed by the relative prominence of parameters α and β in equation (3), with α modulating the map's sensitivity to symbol encoding on the input layer, and β weighing up the map's expectation of upcoming symbols on the basis of past input. The two recoding variants (hereafter dubbed spatio-temporal and temporal respectively) have an impact on the overall topology of the map. In a spatio-temporal map, topologically neighbouring nodes tend to be sensitive to symbol identity, with sub-clusters of nodes being selectively sensitive to context-specific instances of the same symbol. Conversely, in a temporal map neighbouring nodes are sensitive to different symbols taking identical positions in the input string (Marzi et al. 2012a). This is shown in Figure 8.5, plotting the average topological dispersion of *BMU*s activated by symbols in the 1st, 2nd, …, 9th position in the input string on a spatio-temporal (upper plot) and a temporal map (lower plot). Dispersion is calculated as the average distance between nodes discharging upon seeing any symbol appearing in a certain position in the input string, and is given as a fraction of 100 over the map's diagonal. Dispersion values thus range between 0 per cent (when all symbols activate the same

---

[4] This is in fact confirmed by what we know about levels of letter representation in the brain, ranging from very concrete location-specific patterns of geometric lines in the occipito-temporal area of the left hemisphere, to more frontal representations abstracting away from physical features of letters such as position in the visual space, case, and font (Dehaene 2009).
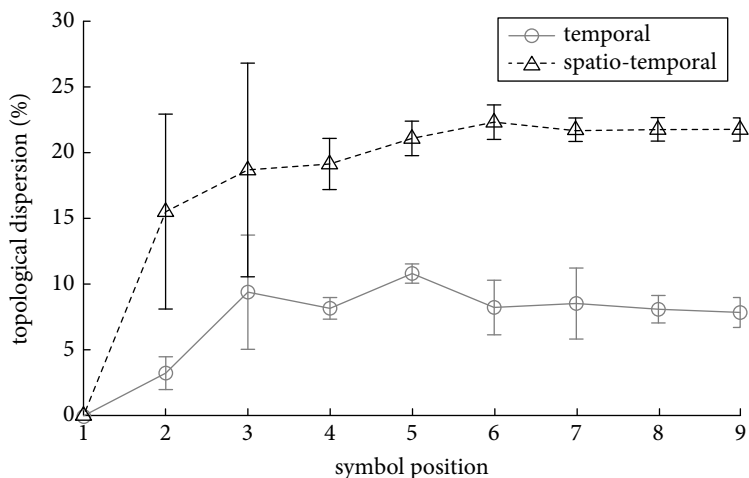
FIGURE 8.5 Topological dispersion of symbols on temporal and spatio-temporal maps, plotted by their position in input words. Dispersion values are given as a fraction of 100 over the map's diagonal.

node, this is true of the start-of-word symbol '#') to 100 per cent, when two nodes lie at the opposite ends of the map's diagonal.

Marzi et al. (2012b) show that the two recoding strategies exhibit a differential behaviour. Due to their greater sensitivity to temporal ordering, temporal maps are better at recalling an input string by reinstating the corresponding activation chain of *BMU*s in the appropriate order, when the input string is already familiar to the map. This is due to temporal maps being able to build up precise temporal expectations through learning, and avoid possible confusion arising from repetition of the same symbols at different positions in the input string. On the other hand, spatio-temporal maps are better at identifying substrings that are shared by two or more word forms, even when the substrings occur at different positions in the input string. This seems to be convenient for acquiring word structure in non-templatic morphologies. For a TSOM to be able to perceive the similarity between—say—*finden* and *gefunden*, the pieces of morphological structure shared by the two forms must activate identical or topologically neighbouring nodes. Clearly, due to the combinatorial properties of morphological constituents, this is done more easily when recoded symbols are abstracted away from their position in time.

We can measure the sensitivity of a map to a specific morphological constituent by calculating the topological distance between the chains of *BMU*s that are activated by the morphological constituent appearing in morphologically related words. For example, we can calculate how close on the map are nodes that discharge upon presentation of a verb stem shared by different forms of the same paradigm. The shorter
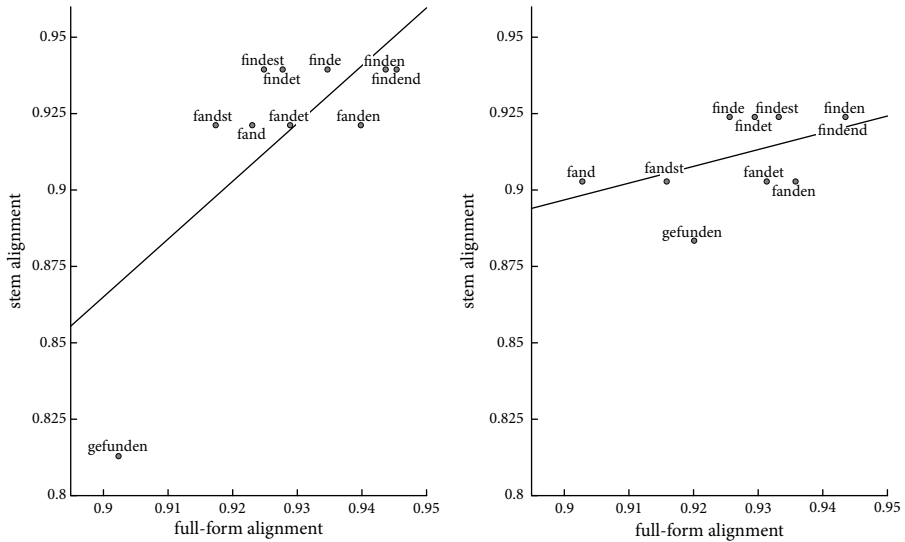
FIGURE 8.6 Alignment plots of the *finden* paradigm on a temporal (left) and a spatio-temporal (right) map.

the average distance is, the greater the map's sensitivity to the shared morphological structure. If *D* is such an average distance normalized as a fraction of the map's diagonal (or dispersion value), then 1-*D* measures the alignment between activation chains discharging on that morphological constituent.

Plots in Figure 8.6 show how well a temporal map (left) and a spatio-temporal map (right) perceive the alignment between the stem in each inflected form of German *finden* ('find') and the stems in all other forms of the same verb paradigm. Stem alignment scores (*y* axis) are plotted against the perceived alignment of each full form with all other forms in the same paradigm (full-form alignment, *x* axis). Intercept values of the linear interpolation of alignment scores in the two maps show that a spatio-temporal map can align intra-paradigmatic forms better than a temporal map can. In particular, spatio-temporal maps prove to be less sensitive to shifts in stem position, as shown by the different alignment values taken by the form *gefunden*, which is perceived as an outlier in the temporal map.

This is consistently replicated in the inter-paradigmatic perception of affixes. Box plots in Figure 8.7 show how scattered are activation chains that discharge upon seeing a few inflectional endings of Italian and German verbs (Marzi et al. 2012b). Activation chains are more widely scattered on temporal maps than they are on spatio-temporal maps. This is true of both German and Italian inflectional endings, confirming a systematic difficulty of temporal maps in perceiving the formal identity of inflectional endings that occur at different positions in word forms.
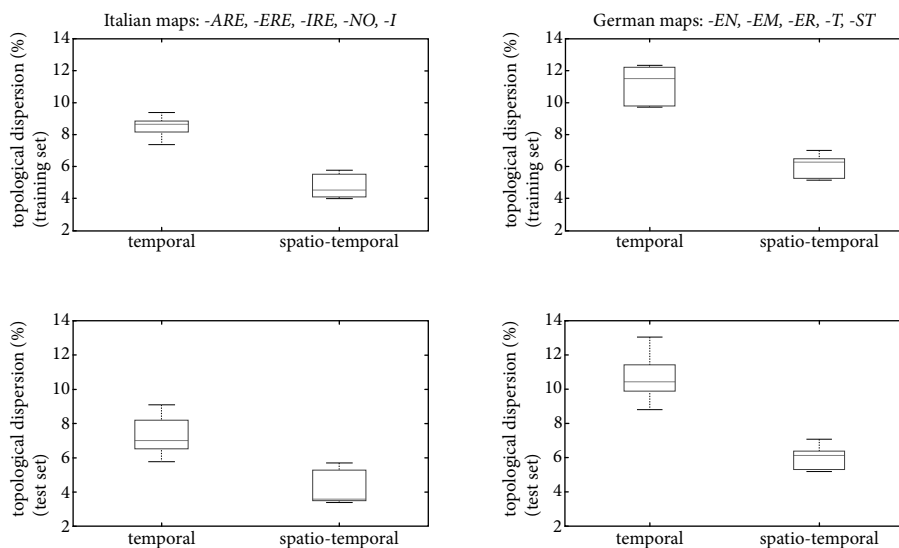
FIGURE 8.7 Top: (Topological) dispersion values across activation patterns triggered by selected inflectional endings on temporal and spatio-temporal maps for Italian (left) and German (right) known word forms. **Bottom**: (Topological) dispersion values for the same set of inflections calculated on unknown word forms. In all panels, dispersion values are given as a fraction of 100 over the map's diagonal.

It is worth emphasizing, in passing, that a TSOM identifies inflectional endings only on the basis of the formal contrast between fully inflected word forms. In German preterite forms, it is easier for a map to perceive a possible segmentation *te-st* by aligning *glaubte* with *glaubtest*, than the segmentation *te-s-t* on the basis of the contrast between *glaubtest* and *glaubtet*. The latter segmentation requires discontinuous alignment between the final *t*'s in the two forms, which is something the map does more reluctantly, and only under certain distributional conditions. This behaviour appears to be in keeping with Bank and Trommer's 'subanalysis complexity hypothesis' (this volume).

8.4.3 Training independence

Training independence is taken by Marcus to be responsible for perceptrons' inability to generalize connectivity patterns to untrained nodes. As nodes are trained independently, nothing learned in connection with one specific node can be transferred to another node. This is not true of TSOMs, where learning requires topological propagation of adjusted spatial and temporal connections, modulated by the neighbourhood distance between each map's node and the current *BMU(t)*. Topological propagation plays an important role in allowing the map to generalize over unseen activation
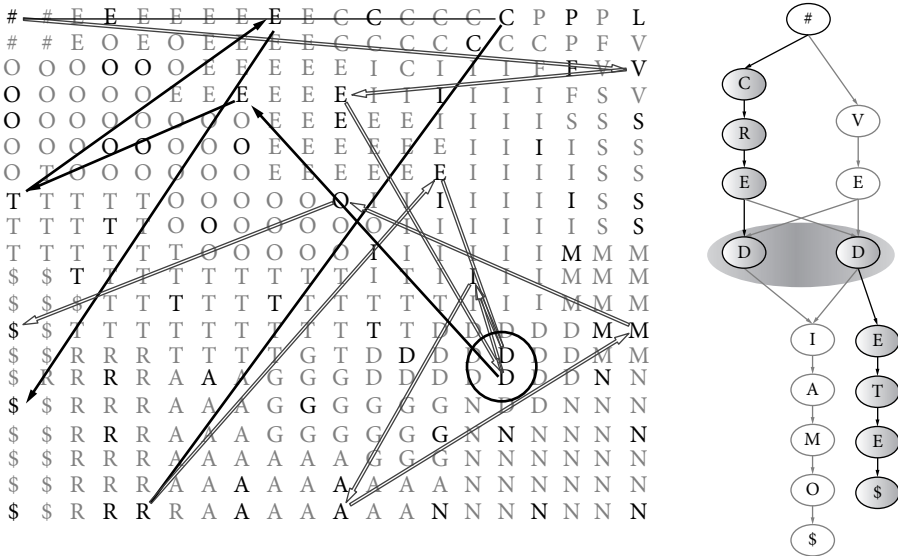
FIGURE 8.8 *BMU activation chains for vediamo-vedete-crediamo on a 20 × 20 map (left) and their word-graph representation (right). In the graph, nodes that are strongly activated by the novel form credete are highlighted in grey. Topological proximity of D-nodes allows the map to bridge the transitional gap between cred- and -ete.*

chains. Figure 8.8 (left) shows the activation chains triggered by the Italian verb forms *vediamo* 'we see', *vedete* 'you see' (second person plural), and *crediamo* 'we believe', on a 20×20 map trained on all these forms. The same map is eventually prompted with the unknown *credete* 'you believe' (second person plural), to test its capacity to anticipate a novel form. The crucial generalization step here involves the circled area including D-nodes that are activated upon seeing stem-final *d* in either *ved-* or *cred-*. When *credete* is shown to the map, spread of temporal activation over the circled area will raise an expectation for *-ete* to follow, based on the activation chain of *vedete*. The word graph in Figure 8.8 (right) illustrates this. Topological propagation of temporal connections over neighbouring map nodes enforces inter-node training dependence, allowing expectations of upcoming symbols to be transferred from one activation chain to another, based on topological contiguity.

An important consequence of this generalization strategy is that inferences are made cautiously, on the basis of a local, analogy-based, inter-paradigmatic extension. This seems to be a logically necessary step to take for a learner to be able to capture the wide range of stem variation and stem selection phenomena attested in inflectional systems of medium and high complexity. Another interesting implication of the generalization strategy enforced by a TSOM is that inflectional endings are better aligned if they are immediately preceded by the same left context. Hence, not any
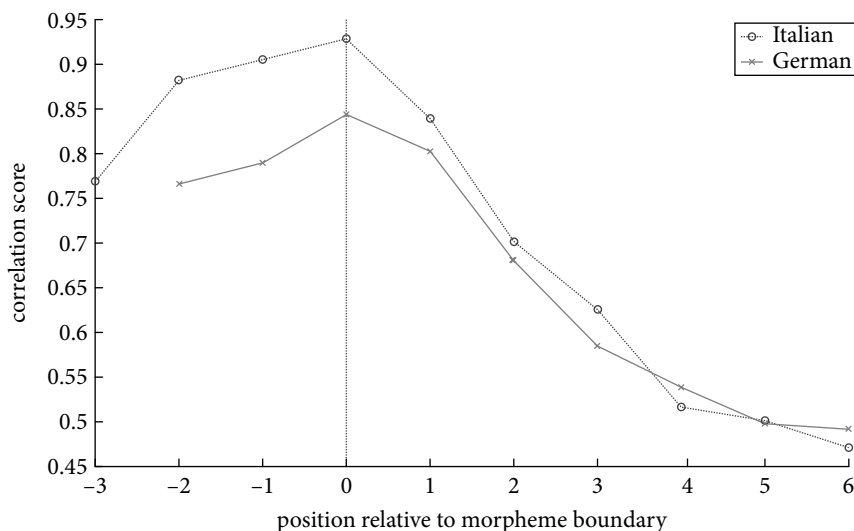
FIGURE 8.9 Correlation coefficients between alignment scores and recall scores for novel words, plotted by letter positions relative to the stem-ending boundary. Position order on the *x* axis is right to left: 0 corresponds to the first letter of the inflectional ending; negative values correspond to letters coming after the first letter of the inflectional ending, and positive values index letters preceding the inflectional ending. The plot shows that correlation rapidly decreases as an inverse function of the letter distance from the stem-ending boundary on either side.

analogy between two stems will work, but only a shared sub-string in the immediate context of the inflectional ending. This is shown by the graph of Figure 8.9, plotting the correlation between how accurately the map is recalling a novel input word, and how well the novel word's activation chain is aligned with the closest activation chain of a known word. As we move away from the stem-ending boundary of the novel word on either side, correlation of alignment scores with recall accuracy becomes increasingly weaker and statistically less significant.

This is exactly what Albright and Hayes' Minimal Generalization algorithm would expect: alignment between nodes activated by a novel word and nodes activated by known words matters less as nodes lie further away from the morphemic boundary. Unlike Minimal Generalization, however, which requires considerable *a priori* knowledge about word structure, mapping rules and the specific structural environment for their application, the generalization bias of TSOMs follows from general principles of topological memory and self-organization.

8.4.4 *Universally quantified variables*

According to Marcus, a default morphological relation like (1) (repeated in (4) for the reader's convenience) requires the notion of a copy operation over a universally

quantified variable. An abstractive computational model of morphological competence such as the one embodied by the TSOM of Figure 8.8 does not require a copy operation.

(4) $PROGR(Y) = Y + ing$,

In fact, equation (4) only imposes the requirement that the base form and progressive form of a regular English verb share the same stem. In TSOMs, sharing the same stem implies that the same morphological structure is used over again. There is no computational implication that this structure is copied from one level of circuitry to another such level. This is well illustrated by the activation chains plotted in Figure 8.8 for the Italian verb forms *vediamo* and *vedete*, which appear to include the same word initial activation chain.

   Another interesting related issue is whether a TSOM can entertain the notion of an abstract morphological process such as prefixation. If a map A, trained on prefixed word forms, is more prone to recognize a word form with an unfamiliar prefix than another map B never trained on prefixed words, we can say that map A developed an expectation for a word to undergo prefixation irrespectively of the specific prefix. To ascertain this, we trained two maps on two sets of partially instantiated paradigms: the first map was exposed to fifty Italian verb paradigms, and the second map to fifty German verb paradigms. Each paradigm contained all present indicative forms, all past tense forms, the infinitive, past participle, and gerund/present participle forms, for a total of fifteen paradigm cells. Whereas Italian past participles are mostly suffixed (with occasional stem alternation), German past participles are typically (but not always) prefixed with *ge-*. After training, we assessed to what extent the two maps were able to align stems in <PREF-X, X> pairs, where X is a known infinitive, and PREF an unknown prefix. Results are shown in Figure 8.10, where stem dispersion is tested on fifty made-up Italian and German pairs, instantiating a <*ri-X, X*> (e.g. *riprendere*, *prendere*) and a <*zu-X, X*> (e.g. *zumachen*, *machen*) pattern respectively. The German map is shown to consistently align prefixed stems with base stems better than the Italian map does, with over 50 per cent of German pairs activating exactly the same *BMU*s, thus proving to be more tolerant towards variability of stem position in the input string. This evidence cannot simply be interpreted as the result of a more flexible coding strategy (e.g. spatio-temporal recoding as opposed to temporal) since both maps were trained with the same parameter set. Rather, it is the map's acquired familiarity with prefixed past participles in German that makes it easier for similarly (but not identically) prefixed patterns to be recognized.

8.5 General discussion

Human lexical competence is known to require the fundamental ability to retain sequences of linguistic items (e.g. letters, syllables, morphemes, or words) in the
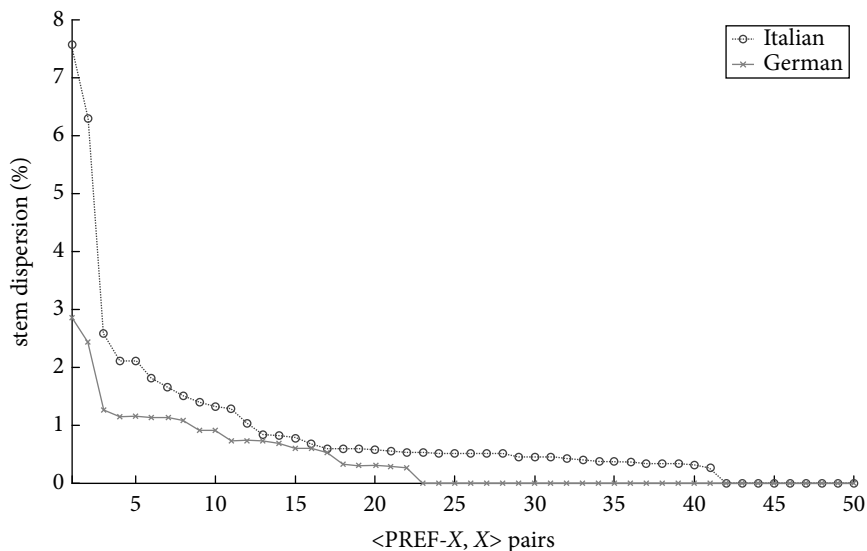
FIGURE 8.10 Stem dispersion over <PREF-*X*, *X*> pairs in Italian and German verb forms. Dispersion values are given as a fraction of 100 over the map's diagonal.

working memory (Gathercole and Baddeley 1989, Papagno et al. 1991). Speakers appear to be sensitive to frequency effects in the presentation of temporal sequences of verbal stimuli. Items that are frequently sequenced together are stored in the long-term memory as single chunks, and accessed and executed as though they had no internal structure. This increases fluency and eases comprehension. Moreover, it also explains the possibility to retain longer sequences in short-term memory when familiar chunks are presented. The short-term span is understood to consist of only a limited number (ranging from three to five according to recent estimates, e.g. Cowan 2001) of available store units. A memory chunk takes one store unit of the short-term span irrespective of length, thus leaving more room for longer sequences to be temporarily retained. Furthermore, chunking produces levels of hierarchical organization of the input stream: what is perceived as a temporal sequence of items at one level may be perceived as a single unit on a higher level, to become part of more complex sequences (Hay and Baayen 2003). Finally, parts belonging to high-frequency chunks tend to resist being perceived as autonomous elements in their own right and being used independently. As a further implication of this 'wholeness' effect, frequently used chunks do not participate in larger word families such as inflectional paradigms (Marzi et al. 2012b).

   We take this type of evidence to illustrate principles of lexical self-organization and to shed light on the intimate interplay between processing and storage in language acquisition. TSOMs provide a general framework for putting algorithmic hypotheses of the processing–storage interaction to the severe empirical test of a computer implementation. Furthermore, unlike classical perceptron-like neural architectures

trained on back-propagation, they allow scholars to entertain a purely abstractive view of morphological competence, based on a number of realistic assumptions concerning acquisition of word structure. We recapitulate them in what follows.

Ecology of training conditions. In modelling word acquisition, it is important to be clear about what is provided in input, and what is expected to be acquired. Choice of input representation may in fact prejudge morphological generalizations to a considerable extent (see also Stump and Finkel's consideration of representational issues, this volume). Machine learning algorithms tend to make fairly specific assumptions on word representations. Even memory-based approaches to morphology induction (Daelemans and van den Bosch 2005), which make 'lazy' use of abstraction-free exemplars, are no exception. To ensure that only features associated with aligned symbols (letters or phonological segments) are matched, Keuleers and Daelemans (2007) align exemplar representations to the right and cast them into a syllabic template. Indeed, for most European languages, we can construe a fixed-length vector representation by aligning input words to the right, since inflection in those languages typically involves suffixation and sensitivity to morpheme boundaries. However, this type of encoding presupposes considerable knowledge of morphology of the target language and does not possibly work with prefixation, circumfixation, and non-concatenative morphological processes in general. Likewise, most current unsupervised algorithms (see Hammarström and Borin 2011 for a recent survey) model morphology learning as a segmentation task, assuming a hard-wired linear correspondence between sublexical strings and morphological structure. However, both highly fusional and non-concatenative morphologies lend themselves grudgingly to being segmented into linearly concatenated morphemes. In assuming that word forms simply consist of a linear arrangement of time-bound symbols, TSOMs take a minimalist view on matters of input representation. Differences in self-organization and recoding are the by-product of acquired sensitivity of map nodes to recurrent morphological patterns in the training data. In our view, this represents a principled solution to the notorious difficulty of multi-layered perceptrons in dealing with issues of input binding and coding. Finally, it addresses the more general issue of avoiding input representations that surreptitiously convey structure or implicitly provide the morphological information to be acquired. Models of word acquisition should be more valued for adapting themselves to the morphological structure of a target language, than for their inductive or representational bias.

Storage and processing. In models of memory, mnemonic representations are acquired, not given. In turn, they affect acquisition, due to their being strongly implicated in memory access, recognition, and recall. TSOMs process words by activating chains of memory nodes. Parts of these chains are not activated by individual words only, but by classes of morphologically related words: for example, all forms sharing the same stem in regular paradigms, or all paradigmatically homologous forms sharing the same suffix. An important implication of this view is that both

lexical forms and sub-lexical constituents are concurrently represented on the same level of circuitry by Hebbian patterns of connectivity linking stored representations. This eliminates the need for a copy operator transferring information from one level to another, according to a mechanistic interpretation of the variable Y in a default generalization like $PROGR(Y) = Y + ing$. Moreover, robust recoding of time-bound information effectively minimizes the risk of creating spurious patterns due to the superposition of lexical and sub-lexical chains.

Another consequence of this view is that morphological productivity is no longer conceptualized as a mapping function between a (unique) lexical base (stored in the lexicon) and an inflected form (generated on-line). Any inflected form can act as a base for any other inflected form. It is important to appreciate, however, that a base is not defined in terms of structural factors such as economy of storage or simplicity of mapping operations, but rather in terms of usage-oriented factors such as time of acquisition, frequency-based entrenchment, probabilistic support of implicative relations, and gradient capability of discriminating inflectional class among alternative hypotheses. As long as we are ready to accept that lexical representations are not dichotomized from lexical entailments, the reconceptualization of word structure offered by TSOMs is in line with a genuinely abstractive view of morphology.

Generalization and training. Marcus (2001) correctly emphasizes that no generalization is possible with training independence. Generalization implies propagation, transferring of information from one node to another node, not just adaptation of connectivity between individual input and output nodes. Propagation is also independently needed to address another apparent paradox in incremental acquisition. Human speakers are very good at inferring generalizations on the basis of local analogies, involving the target word and its closest neighbours (Albright 2002). However, analogy is a pre-theoretical notion and any linguistically relevant analogy must take into account global information, such as the overall number of forms undergoing a particular change under structurally governed conditions. How can global analogies be inferred on the basis of local processing steps? Topological propagation offers a solution to this problem. Although a *BMU* is identified on the basis of a pairwise similarity between each node and the current input stimulus only, the connectivity pattern of a *BMU* is propagated topologically on the spatial and temporal layers of a map. This prompts information transfer but also competition between conflicting patterns of local analogy. It is an important result of our simulations that the neat effect of competition between local analogies is the emergence of a global, paradigm-sensitive notion of formal analogy (Marzi et al. 2012a).

All issues raised here appear to address the potential use of computational models of language acquisition for theoretical linguists and psycho-linguists focusing on the nature of grammar representations. We argue that there is wide room for cross-disciplinary inquiry and synergy in this area, and that the search for the most appropriate nature of mental representations for words and sub-word constituents,

and a better understanding of their interaction in a dynamic system, can provide such a common ground. At our current level of understanding, it is very difficult to establish a direct correspondence (Clahsen 2006) between language-related categories and macro-functions (rules vs exceptions, grammar vs lexicon) on the one hand, and neuro-physiological correlates on the other hand. As an alternative approach to the problem, in the present chapter we have focused on a bottom-up investigation of the computational complexity and dynamic of psycho-cognitive micro-functions (e.g. perception, storage, alignment, and recall of time series) to assess their involvement in language processing, according to an indirect correspondence hypothesis. Albeit indirect, such a perspective strikes us as conducive to non-trivial theoretical and computational findings.

To date, progress in cross-disciplinary approaches to language inquiry has been hindered by the enormity of the gap between our understanding of some low-level properties of the brain, on the one hand, and some very high-level properties of the language architecture on the other. The research perspective entertained in these pages highlights some virtues of modelling, on a computer, testable processing architectures as dynamic systems. Arguably, one of the main such virtues lies in the possibility that by implementing a few basic computational functions and their dynamic interrelations, we can eventually provide some intermediate-level building blocks that would make it easier to relate language to cognitive neuroscience.

References

Abbott, C. (1984). 'Two feminine genders in Oneida'. *Anthropological Linguistics* 26, 125–37.

Abrams, P. (2006). *Onondaga Pronominal Prefixes*. PhD thesis, University of Buffalo.

Ackerman, F., J. Blevins, and R. Malouf (2009). 'Parts and wholes: Patterns of relatedness in complex morphological systems and why they matter'. In J. Blevins and J. Blevins (eds), *Analogy in Grammar: Form and Acquisition*, pp. 54–82. Oxford: Oxford University Press.

Aikhenvald, Alexandra Y. (2003). *A grammar of Tariana*. Cambridge: Cambridge University Press.

Aikhenvald, Alexandra Y. (2008). *The Manambu language of East Sepik, Papua New Guinea*. Oxford: Oxford University Press.

Albright, Adam. (2002). 'Islands of reliability for regular morphology: Evidence from Italian'. *Language* 78, 684–709.

Albright, Adam, and Bruce Hayes (2003). 'Rules vs. Analogy in English past tenses: A computational/experimental study'. *Cognition* 90, 119–61.

Anderson, Gregory D. S., and K. David Harrison (2008). Remo (Bonda). In G. D. S. Anderson, (ed.), *The Munda languages*. London: Routledge, pp. 557–632.

Anderson, Stephen R. (1988). 'Morphological change'. In Frederick J. Newmeyer (ed.), *Linguistics: The Cambridge survey*, vol. I, 324–62. Cambridge: Cambridge University Press.

Anderson, Stephen R. (1992). *A-Morphous morphology*. Cambridge: Cambridge University Press.

Anderson, Stephen R. (2005). *Aspects of the theory of clitics*. Oxford: Oxford University Press.

Anderson, Stephen R. (2011). 'Stress-conditioned allomorphy in Surmiran (Rumantsch)'. In Martin Maiden, John Charles Smith, Maria Goldbach, and Marc-Olivier Hinzelin (eds), *Morphological autonomy: Perspectives from Romance inflectional morphology*, pp. 13–35. Oxford: Oxford University Press.

Anderson, Stephen R. (to appear). 'The morpheme: Its nature and use'. In Matthew Baerman (ed.), *Oxford handbook of inflection*. Oxford: Oxford University Press.

Andrews, James Richard (1975). *Introduction to Classical Nahuatl*. Austin: University of Texas Press.

Aronof, Mark (1994). *Morphology by itself*. Cambridge, MA: MIT Press.

Baayen, R. Harald, Rochelle Lieber, and Robert Schreuder (1997). 'The morphological complexity of simplex nouns'. *Linguistics*, 35, 861–77.

Baayen, R. Harald, Petar Milin, Dusica Filipović Đurđević, Peter Hendrix, and Marco Marelli (2011). 'An amorphous model for morphological processing in visual comprehension based on naive discriminative learning'. *Psychological Review* 118, 438–81.

Baerman, Matthew (2004). 'Directionality and (Un)Natural Classes in Syncretism'. *Language* 80(4), 807–27.

Baerman, Matthew, Dunstan Brown, and Greville G. Corbett (2005). *The syntax–morphology interface: A study of syncretism*. Cambridge: Cambridge University Press.

Baerman, Matthew, Greville G. Corbett, and Dunstan Brown (eds) (2010). *Defective paradigms: Missing forms and what they tell us*. Oxford: Oxford University Press.

Baerman, Matthew, Greville G. Corbett, Dunstan Brown, and Andrew Hippisley (eds) (2007). *Deponency and morphological mismatches*. Oxford: Oxford University Press.

Baeza-Yates, Ricardo, and Berthier Ribeiro-Neto (1999). *Modern Information Retrieval*. New York, NY: ACM Press, Addison-Wesley.

Baker, Mark C. (1995). *The polysynthesis parameter*. New York: Oxford University Press.

Bauer, Laurie (2004). *Morphological productivity*. Cambridge: Cambridge University Press.

Beard, Robert (1995). *Lexeme–morpheme base morphology*. Albany, NY: SUNY Press.

Bickel, Balthasar (2003). 'Belhare'. In G. Thurgood and R. J. LaPolla (eds), *The Sino-Tibetan languages*. London: Routledge, pp. 546–70.

Blake, Barry J. (1979). *A Kalkatungu grammar*. Canberra: Dept. of Linguistics, Research School of Pacific Studies, Australian National University.

Blevins, J. (2004). 'Inflection classes and economy'. In Z. G. Muller Gereon and Lutz Gunkel (eds), *Explorations in Nominal Inflection*, pp. 375–96. Berlin: Mouton de Gruyter.

Blevins, J. (2006). 'Word-based morphology'. *Journal of Linguistics* 42, 531–73.

Blevins, J. (2013). 'Word-based morphology from Aristotle to modern wp (word and paradigm models)'. In K. Allan (ed.), *The Oxford handbook of the history of linguistics*, pp. 41–85. Oxford: Oxford University Press.

Boas, Franz (1947). 'Kwakiutl grammar, with a glossary of the suffixes'. *Transactions of the American Philosophical Society* 37(3), 201–377.

Bobaljik, Jonathan David (2007). *On Comparative Suppletion*. Ms., University of Connecticut.

Bobaljik, Jonathan David (2012). *Universals in comparative morphology: Suppletion, superlatives, and the structure of words*. Cambridge: MIT Press.

Bochner, H. (1993). *Simplicity in Generative Morphology*. Berlin: Mouton de Gruyter.

Boelaars, J. H. M.C. (1950). *The linguistic position of south-western New Guinea*. Leiden: E. J. Brill.

Booij, Geert (2010). *Construction Morphology*. Oxford: Oxford University Press.

Broadwell, George Aaron (2006). *A Choctaw reference grammar*. Lincoln: University of Nebraska Press.

Brøndal, Viggo (1940). 'Compensation et variation, deux principes de linguistique générale'. *Scientia* 68, 101–9. Reprinted: Id. 1943. *Essais de linguistique générale*, pp. 105–16. København: Munskgaard.

Brown, Dunstan, Carole Tiberius, and Greville G. Corbett (2007). 'The alignment of form and function: Corpus-based evidence from Russian'. *International Journal of Corpus Linguistics* 12, 511–34.

Burzio, Luigi (2004). 'Paradigmatic and syntagmatic relations in Italian verbal inflection'. In J. Auger, J. C. Clements, and B. Vance (eds), *Contemporary Approaches to Romance Linguistics*. Amsterdam: John Benjamins.

Bybee, Joan L., and Carol L. Moder (1983). 'Morphological classes as natural categories'. *Language* 59, 251–70.

Bybee, Joan L., and Dan Isaac Slobin (1982). 'Rules and Schemas in the Development and Use of the English Past Tense'. *Language* 58, 265–89.

Bybee, Joan L., R. D. Perkins, and W. Pagliuca (1994). *The evolution of grammar: Tense, aspect and modality in the languages of the world*. Chicago: University of Chicago Press.

Cairns, Charles E. (1986). Word structure, markedness, and applied linguistics. *Markedness. Proceedings of the twelfth annual linguistics symposium of the University of Wisconsin-Milwaukee, March 11–12, 1983*, ed. by Fred R. Eckman, Edith A. Moravcsik, and Jessica R. Wirth, pp. 13–38. New York: Plenum.

Carstairs, Andrew (1983). 'Paradigm economy'. *Journal of Linguistics* 19, 115–28.

Carstairs, Andrew (1984). 'Outlines of a constraint on syncretism'. *Folia linguistica* 18, 73–85.

Carstairs, Andrew (1987). *Allomorphy in inflection*. London: Croom Helm.

Carstairs, Andrew, and Paul Stemberger (1988). 'A processing constraint on inflectional homonymy'. *Linguistics* 26, 601–18.

Carstairs-McCarthy, Andrew (1994). 'Inflection classes, gender, and the principle of contrast'. *Language* 70, 737–88.

Carstairs-McCarthy, Andrew. (2010). *The evolution of morphology*. Oxford: Oxford University Press.

Chafe, W. L. (1977). The evolution of third person verb agreement in the Iroquoian languages. In C. Li (ed.), *Mechanisms of Syntactic Change*, pp. 493–524. Austin, Texas: University of Texas Press.

Clahsen, Harald (2006). 'Linguistic perspectives on morphological processing'. In D. Wunderlich, (ed.), *Advances in the Theory of the Lexicon*, pp. 355–88. Berlin: Mouton de Gruyter.

Chomsky, Noam, and Morris Halle (1968). *The sound pattern of English*. New York: Harper & Row.

Chumakina, Marina (2011). 'Morphological complexity of Archi verbs'. In Gilles Authier and Timur Maisak (eds), *Tense, aspect, modality and finiteness in East Caucasian languages* (Diversitas Linguarum 30), pp. 1–24. Bochum: Brockmeyer.

Chumakina, Marina, Dunstan Brown, Harley Quilliam, and Greville G. Corbett (2007). *Archi: A Dictionary of the Archi Villages, Southern Daghestan, Caucasus.* <http://www.smg.surrey.ac.uk/archi/linguists/>.

Chumakina, Marina, and Greville G. Corbett (2008). 'Archi: The challenge of an extreme agreement system'. In Aleksandr V. Arxipov (ed.), *Fonetika i nefonetika. K 70-letiju Sandro V. Kodzasova* [Festschrift for S. V. Kodzasov], pp. 184–94. Moskva: Jazyki slavjanskix kul'tur.

Clackson, James (2007). *Indo-European linguistics: An Introduction*. Cambridge: Cambridge University Press.

Coltheart, Max, Kathleen Rastle, Conrad Perry, Robyn Langdon, and Johannes Ziegler (2001). 'DRC: A dual route cascaded model of visual word recognition and reading aloud'. *Psychological Review* 108, 204–56.

Corbett, Greville G. (2007). 'Deponency, syncretism and what lies between'. *Proceedings of the British Academy* 145, 21–43.

Corbett, Greville G. (2012). *Features*. Cambridge: Cambridge University Press.

Corbett, Greville G., and Norman M. Fraser. (1993). 'Network morphology: A DATR account of Russian nominal inflection'. *Journal of Linguistics* 29, 113–42.

Cowan, Nelson (2001). 'The magical number 4 in short-term memory: A Reconsideration of mental storage capacity'. *Behavioral and Brain Sciences* 24, 87–114.

Croft, William (2003). *Typology and universals*. 2nd edition. Cambridge: Cambridge University Press.

Csúcs, Sándor (1998).'Udmurt'. In D. Abondolo (ed.), *The Uralic languages*. London and New York: Routledge, pp. 276–304.

Culicover, Peter W. (2013). *Grammar and Complexity: Language at the Intersection of Competence and Performance*. Oxford: Oxford University Press.

Culy, Christopher (1985). 'The complexity of the vocabulary of Bambara'. *Linguistics and Philosophy* 8, 345–51.

Daelemans, Walter, and Antal van den Bosch (2005). *Memory-based language processing*. Cambridge: Cambridge University Press.

Dahl, Östen (2004). *The growth and maintenance of linguistic complexity*. Amsterdam: John Benjamins.

Davies, John (1989). *Kobon*. London: Routledge.

Davis, J. Colin, and S. Jeffrey Bowers (2004). 'What do letter migration errors reveal about letter position coding in visual word recognition?' *Journal of Experimental Psychology: Human Perception and Performance* 30, 923–41.

Dehaene, Stanislas (2009). *Reading the brain.* New York: Penguin.

Dimmendaal, Gerrit Jan (1983). *The Turkana language*. Dordrecht: Foris.

Donohue, Mark (1999). *Warembori*. Muenchen: Lincom Europa.

Donohue, Mark (2008). 'Complex predicates and bipartite stems in Skou'. *Studies in Language* 32(2), 279–335.

Donohue, Mark (2011). 'Case and configurationality: Scrambling or mapping?' *Morphology* 21, 499–513.

Drabbe, P. (1947). *Nota's over de Jénggalntjoer-taal*. Merauke: Archief.

Drabbe, P. (1950). 'Talen en Dialecten van Zuid-West Nieuw-Guinea'. *Anthropos* 45, 545–74.

Driem, George van (1987). *A grammar of Limbu*. Berlin: Mouton de Gruyter.

Ehala, Martin (2009). 'Linguistic strategies and markedness in Estonian morphology'. *Sprachtypologie und Universalienforschung* 1/2, 29–48.

Elman, Jeff (1998). 'Generalization, simple recurrent networks, and the emergence of structure'. *Proceedings of the 20th Annual Conference of the Cognitive Science Society.* Mahwah. NJ: Lawrence Erlbaum Associates.

El-Solami-Mewis, Catherine (1987). *Lehrbuch des Somali*. Leipzig: VEB Verlag Enzyklopädie.

Evans, Nicholas (1995a). *A grammar of Kayardild: With historical-comparative notes on Tangkic*. Berlin: Mouton de Gruyter.

Evans, Nicholas (1995b). 'Multiple case in Kayardild: Anti-iconic suffix order and the diachronic filter'. In Frans Plank (ed.), *Double case: Agreement by suffixaufnahme*, pp. 396–428. New York: Oxford University Press.

Ferro, Marcello, Claudia Marzi, and Vito Pirrelli (2011). 'A Self-Organizing Model of Word Storage and Processing: Implications for Morphology Learning'. *Lingue e Linguaggio* (10)2, 209–26. Bologna: il Mulino.

Finkel, Raphael, and Gregory Stump (2007). 'Principal parts and morphological typology'. *Morphology* 17, 39–75.

Finkel, Raphael, and Gregory Stump (2009). 'Principal parts and degrees of paradigmatic transparency'. In James P. Blevins and Juliette Blevins (eds), *Analogy in grammar: Form and acquisition*, pp. 13–53. Oxford: Oxford University Press.

Fortescue, M. (1980). 'Affix ordering in West Greenlandic derivational processes', *International Journal of American Linguistics* 46(4), 259–78.

Gathercole, E. Susan, and Alan D. Baddeley (1989). 'Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study'. *Journal of Memory and Language* 28, 200–13.

Gazdar, G., G. Pullum, B. Carpenter, E. Klein, T. Hukari, and R. Levine (1988). 'Category structures'. *Computational Linguistics* 14, 1–19.

Gerken, Lou Ann (2006). 'Decisions, decisions: Infant language learning when multiple generalizations are possible'. *Cognition* 98, B67–B74.

Gershenson, Carlos, and Nelson Fernández (2012). 'Complexity and information: Measuring emergence, self-organization, and homeostasis at multiple scales'. *Complexity* 18, 29–44.

Givón, Talmy (1971). 'Historical syntax and synchronic morphology: An archaeologist's field-trip'. *Proceedings of the Chicago Linguistic Society* 7, 394–415.

Goldsmith, John A. (2010). 'Segmentation and morphology'. In A. Clark, C. Fox, and S. Lappin (eds), *Handbook of Computational Linguistics and Natural Language Processing*. Oxford: Blackwell.

Greenberg, Joseph (1966). *Language universals, with special reference to the feature hierarchies*. The Hague: Mouton. Reprinted 2005. Berlin: Mouton de Gruyter.

Halle, Morris (1997). 'Distributed Morphology: Impoverishment and Fission'. In Y. K. Benjamin Bruening and M. McGinnis (eds), *Papers at the Interface*. Vol. 30 of *MIT Working Papers in Linguistics*, pp. 425–49. Cambridge MA: MITWPL.

Halle, Morris, and Alec Marantz (1993). 'Distributed Morphology and the Pieces of Inflection'. In K. Hale and S. J. Keyser (eds), *The View from Building 20*, pp. 111–76. Cambridge MA: MIT Press.

Hammarström, Harald, and Lars Borin (2011). 'Unsupervised learning of morphology'. *Computational Linguistics* 37(2), 309–50.

Hankamer, J. (1989). 'Morphological parsing and the lexicon'. In W. Marslen Wilson (ed.), *Lexical Representation and Process*, pp. 392–408. Cambridge, MA: MIT Press.

Hargus, Sharon (1997). 'The Athabaskan disjunct prefixes: Clitics or affixes?' In Jane Hill, P. J. Mistry, and Lyle Campbell (eds), *The life of language: Papers in linguistics in honor of William Bright.* Berlin: Mouton de Gruyter.

Harm, W. Michael, and Mark S. Seidenberg (1999). 'Phonology, reading acquisition, and dyslexia: Insights from connectionist models'. *Psychological Review* 106, 491–528.

Harris, Zellig Sabbatai (1955). 'From phoneme to morpheme'. *Language* 31, 190–222.

Haspelmath, Martin (2006). 'Against markedness (and what to replace it with)'. *Journal of Linguistics* 42, 25–70.

Hawkins, John A. (2004). *Efficiency and complexity in grammars*. Oxford: Oxford University Press.

Hay, Jennifer, and R. Harald Baayen (2003). 'Phonotactics, parsing and productivity'. *Italian Journal of Linguistics* 15, 99–130.

Hippisley, Andrew (2010). 'Lexical analysis'. In Nitin Indurkhya and Fred J. Damerau (eds) *Handbook of natural language processing, second edition* pp. 31–58. Boca Raton, FL: CRC Press, Taylor and Francis Group.

Hjelmslev, Louis (1935). *La catégorie des cas*. Copenhagen: Munskgaard.

Hockett, Charles F. (1947). 'Problems of morphemic analysis'. *Language* 23, 321–43.

Hocket, Charles F. (1958). *A course in modern linguistics*. New York: Macmillan Company.

Huntley, David (1993). 'Old Church Slavonic'. *The Slavonic languages*, ed. by Bernard Comrie and Greville G. Corbett, pp. 125–87. London: Routledge.

Jacobi, Hermann (1886). *Ausgewählte Erzählungen in Mâhârâshtrî*. Leipzig: Hirzel.

Jakobson, Roman (1939). 'Signe zéro'. In *Mélanges de linguistique, offerts à Charles Bally*, pp. 143–52. Genève: Georg.

Jurafsky, Daniel, and James H. Martin (2009). *Speech and language processing*. London: Pearson Education Ltd.

Kaas, Jon H., Michael M. Merzenich, and Herbert P. Killackey (1983). 'The reorganization of somatosensory cortex following peripheral nerve damage in adults and developing mammals'. *Annual Review of Neuroscience* 6, 325–56.

Keuleers, Emmanuel, and Walter Daelemans (2007). 'Morphology-based learning models of inflectional morphology: A Methodological Case Study'. In V. Pirrelli (ed.), *Psychocomputational issues in morphology and processing, Lingue e Linguaggio* 2, 151–74.

Kibrik, Aleksandr E. (1977a). *Opyt strukturnogo opisanija arčinskogo jazyka.* Volume 2: *Taksonomičeskaja grammatika*. Moskva: Izdateľstvo Moskovskogo Universiteta.

Kibrik, Aleksandr E. (1977b). *Opyt strukturnogo opisanija arčinskogo jazyka.* Volume 3: *Dinamičeskaja grammatika*. Moskva: Izdateľstvo Moskovskogo Universiteta.

Kibrik, Aleksandr, Alexandre Arkhipov, Mikhail Daniel, and Sandro Kodzasov (2007). *Archi text corpus.* <http://www.philol.msu.ru/~languedoc/eng/archi/corpus.php>.

Kibrik, Aleksandr E., Sandro Kodzasov, Irina Olovjannikova, and Džalil Samedov (1977). *Opyt strukturnogo opisanija arčinskogo jazyka.* Volume 1: *Leksika, fonetika*. Moskva: Izdateľstvo Moskovskogo Universiteta.

Koenig, Jean-Pierre, and Karin Michelson (2012). 'The (non)universality of syntactic selection and functional application'. In C. Piñon (ed.), *Empirical studies in syntax and semantics*, Volume 9, pp. 185–205. Paris: Centre National de la Recherche Scientifique.

Koenig, Jean-Pierre, and Karin Michelson (in press). 'Invariance in argument realization: The case of Iroquoian'. *Language*, 90(4).

Kohonen, Teuvo (2001). *Self-organizing maps*. Berlin Heidelberg: Springer-Verlag.

Kossmann, Maarten G. (2007). 'Berber morphology'. In A. Kaye (ed.), *Morphologies of Asia and Africa*, pp. 429–46. Winona Lake, Indiana: Eisenbrauns.

Kostić, Aleksandar, and Milena Božić (2007). 'Constraints on probability distributions of grammatical forms'. *Psihologija* 40, 5–35.

Kostić, Aleksandar, Tania Markovic, and Aleksandar Baucal (2003). 'Inflectional morphology and word meaning: Orthogonal or co-implicative domains'. *Morphological Structure in Language Processing*, ed. by R. H. Baayen and R. Schreuder, pp. 1–44. Berlin: Mouton de Gruyter.

Koutnik, Jan (2007). 'Inductive modelling of temporal sequences by means of self-organization'. *Proceedings of International Workshop on Inductive Modelling* (IWIM 2007), pp. 269–77. Prague: Czech Technical University.

Langer, Hagen (1991). *Ein automatisches Morphemsegmentierungsverfahren für das Deutsche*. PhD thesis, Georg-August-Universität zu Göttingen.

Lanman, Charles R. (1880). 'A statistical account of noun-inflection in the Veda'. *Journal of the American Oriental Society* 10, 325–601.

Leumann, Manu (1977). *Lateinische Laut- und Formenlehre*. München: Beck.

Lounsbury, F. (1953). *Oneida verb morphology. Yale University Publications in Anthropology* 48. New Haven, CT: Yale University Press.

Lovins, Julie Beth (1968). 'Development of a stemming algorithm'. *Mechanical Translation and Computational Linguistics* 11(2), 22–31.

Lynch, John (1978). *A grammar of Lenakel*. Vol. 55 of *Series B*, Pacific Linguistics B.

Mace, John (2003). *Persian grammar: For reference and revision*. London: Routledge.

Mallon, Mick (1991). *Introductory Inuktitut: Reference grammar*. Montreal: Arctic College—McGill University Inuktitut Text Project.

Maiden, Martin (2005). 'Morphological autonomy and diachrony'. Yearbook of Morphology 2004, 137–75.

Maiden, Martin, and Paul O'Neill (2010). 'On morphomic defectiveness'. *Proceedings of the British Academy* 163, 103–24.

Marchand, Hans (1969). *The categories and types of present-day English word formation*. 2nd edition. Minchen: Beck.

Marcus, Gary F. (2001). *The algebraic mind: Integrating connectionism and cognitive science*. Cambridge, MA: MIT Press.

Marcus, Gary F., S. Vijayan, Shoba Bandi Rao, and Peter M. Vishton (1999). 'Rule learning in 7-month-old infants'. *Science* 283, 77–80.

Marzi, Claudia, Marcello Ferro, Claudia Caudai, and Vito Pirrelli (2012a). 'Evaluating Hebbian self-organizing memories for lexical representation and access'. *Proceedings of the 8th International Conference on Language Resources and Evaluation* (LREC 2012), 886–93, Istanbul.

Marzi, Claudia, Marcello Ferro, and Vito Pirrelli (2012b). 'Prediction and generalisation in word processing and storage'. *Proceedings of the 8th Mediterranean Morphology Meeting* (8th MMM 2011), 114–31. University of Patras, Greece.

Marzi, Claudia, Marcello Ferro, and Vito Pirrelli (2012c). 'Word alignment and paradigm Induction'. *Lingue e Linguaggio* (11)2, 251–74.

Matthews, P. H. (1974). *Morphology: An introduction to the theory of word-structure*. Cambridge: Cambridge University Press.

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

McCarthy, John (2010). *An introduction to Harmonic Serialism*. Ms., University of Massachusetts, Amherst.

McClelland, James L., and David E. Rumelhart (1981). 'An interactive activation model of context effects in letter perception: Part 1. An account of basic findings'. *Psychological Review* 88, 375–407.

Meir, Irit, Wendy Sandler, Carol Padden, and Mark Aronof (2010). 'Emerging sign languages'. In Marc Marschark and Patricia Elizabeth Spencer (eds), *Oxford handbook of deaf studies, language, and education*, vol. 2, pp. 267–80. New York: Oxford University Press.

Miestamo, Matti, Kaius Sinnemäki, and Fred Karlsson (2008). *Language complexity: Typology, contact, change*. Amsterdam: John Benjamins.

Milin, Petar, Victor Kuperman, Aleksandar Kostić, and R. Harald Baayen (2009). 'Words and paradigms bit by bit: An information-theoretic approach to the processing of paradigmatic

structure in inflection and derivation'. In James P. Blevins and Juliette Blevins (eds), *Analogy in grammar: Form and acquisition*, pp. 214–52. Oxford: Oxford University Press.

Milizia, Paolo (2013). *L'equilibrio nella codifica morfologica*. Roma: Carocci.

Milizia, Paolo (2014). 'Semi-separate exponence in cumulative paradigms: Information-theoretic properties exemplified by Ancient Greek verb endings'. *Linguistic Issues in Language Technology* 11(4), 95–123.

Mohri, Mehryar, and Richard Sproat (2006). 'On a common fallacy in computational linguistics'. In M. Suomen, A. Arppe, A. Airola, O. Heinämäki, M. Miestamo, U. Määttä, J. Niemi, K. K. Pitkänen, and K. Sinnemäki (eds), *A Man of Measure*: *Festschrift in Honour of Fred Karlsson on his 60th Birthday*. *SKY Journal of Linguistics* 19, 432–9.

Moscoso del Prado, Martín, Fermín Aleksandar Kostic, and R. Harald Baayen (2004). 'Putting the bits together: An information-theoretical perspective on morphological processing'. *Cognition* 94(1), 1–18.

Müller, Gereon (2003). 'On decomposing inflection class features: Syncretism in Russian noun inflection'. In L. Gunkel, G. Müller, and G. Zifonun (eds), *Explorations in nominal inflection*, pp. 189–228. Berlin: Mouton de Gruyter.

Müller, Gereon (2005). *Subanalyse verbaler Flexionsmarker*. Ms., Universität Leipzig.

Noreen, Adolf (1923). *Altisländische und altnorwegische Grammatik.* Halle: Niemeyer.

Oberlies, Thomas (2001). *Pāli: A grammar of the language of the Theravāda* Tipiṭaka. Berlin: Mouton de Gruyter.

O'Neill, Paul (2011). 'The notion of the morphome'. In M. Goldbach, M. Maiden, and J.-C. Smith (eds), *Morphological autonomy: Perspectives from Romance inflectional morphology*, pp. 70–94. Oxford: Oxford University Press.

Orsolini, Margherita, Rachele Fanari, and Hugo Bowles (1998). 'Acquiring regular and irregular inflection in a language with verb classes'. *Language and Cognitive Processes* 13, 425–64.

Orsolini, Margherita, and William Marslen-Wilson (1997). 'Universals in morphological representation: Evidence from Italian'. *Language and Cognitive Processes* 12, 1–47.

Ortmann, Albert (1999). 'Affix repetition and non-redundancy in inflectional morphology'. *Zeitschrift für Sprachwissenschaft* 1, 76–120.

Packard, Jerome L. (2000). *The morphology of Chinese: A linguistic and cognitive approach*. Cambridge: Cambridge University Press.

Papagno Costanza, Tim Valentine, and Alan Baddeley (1991). 'Phonological short-term memory and foreign-language vocabulary learning'. *Journal of Memory and Language* 30, 331–47.

Parks, Douglas Richard (1976). *A grammar of Pawnee*. New York: Garland.

Penfield, Wilder, and Lamar Roberts (1959). *Speech and brain mechanisms.* Princeton, NJ: Princeton University Press.

Perry, Conrad, Johannes C. Ziegler, and Marco Zorzi (2007). 'Nested incremental modeling in the development of computational theories: The CDP+ model of reading aloud'. *Psychological Review*, 114(2), 273–315.

Pirrelli, Vito (2000). *Paradigmi in Morfologia. Un approccio interdisciplinare alla flessione verbale dell'italiano*. Pisa-Roma: Istituti Editoriali e Poligrafici Italiani.

Pirrelli, Vito, Basilio Calderone, Ivan Herreros, and Michele Virgilio (2004). 'Non-locality all the way through: Emergent global constraints in the Italian morphological lexicon'. *Proceedings of 7th Meeting of the ACL SIGPHON*.

Pirrelli, Vito, Marcello Ferro, and Basilio Calderone (2011). 'Learning paradigms in time and space. Computational evidence from Romance languages'. In M. Maiden, John C. Smith, Maria Goldbach, and Marc-Olivier Hinzelin (eds), *Morphological autonomy: Perspectives from Romance inflectional morphology*, pp. 135–57. Oxford: Oxford University Press.

Pischel, Richard (1900). *Grammatik der Prakrit-Sprachen*. Strassburg: Trübner.

Plank, Frans, and Wolfgang Schellinger (1997). 'The uneven distribution of genders over numbers: Greenberg Nos. 37 and 45'. *Linguistic Typology* 1, 53–101.

Plaut, David, James McClelland, Mark Seidenberg, and Karalyn Patterson (1996). 'Understanding normal and impaired word reading: Computational principles in quasi-regular domains'. *Psychological Review* 103, 56–115.

Polinsky, Maria, and Eric Potsdam (2001). 'Long-distance agreement and topic in Tsez'. *Natural Language and Linguistic Theory* 19, 583–646. Kluwer Academic Publishers.

Porter, Martin (1980). 'An algorithm for suffix stripping'. *Program* 3, 130–7.

de Reuse, Willem Joseph (1994). *Siberian Yupik Eskimo: The language and its contacts with Chukchi*. Salt Lake City: University of Utah Press.

Rice, Keren (2000). *Morpheme order and semantic scope: Word formation in the Athabaskan verb*. Cambridge: Cambridge University Press.

Rice, Keren (2011). 'Principles of affix ordering: An overview'. *Word Structure* 4(2), 169–200.

Rice, Keren (2012). 'Morphological complexity in Athabaskan languages: A focus on discontinuities'. Presented at SSILA workshop on Morphological Complexity in the Languages of the Americas, Annual Meeting of the Linguistic Society of America, Portland, Oregon, January 2012.

Rissanen, Jorma (2007). *Information and complexity in statistical modeling*. New York: Springer.

Roark, Brian, and Richard Sproat (2007). *Computational approaches to morphology and syntax*. Oxford: Oxford University Press.

Rosenblatt, Frank (1962). *Principles of neurodynamics*. New York: Spartan.

Round, Erich R. (2009). *Kayardild morphology, phonology and morphosyntax*. Yale University PhD dissertation.

Round, Erich R. (2010). 'Autonomous morphological complexity in Kayardild'. Paper presented at the Workshop on Morphological Complexity, Harvard University, 12 January 2010.

Round, Erich R. (2011). 'Morphemes as a level of representation capture unity of exponence across the inflection–derivation divide'. *Linguistica* 51, 217–30.

Round, Erich R. (2013). *Kayardild morphology and syntax*. Oxford: Oxford University Press.

Round, Erich R. (forthcoming). 'Kayardild inflectional morphotactics is morphomic'. In A. Luis and R. Bermudez-Otero (eds), *The morphome debate*. Oxford: Oxford University Press.

Round, Erich R. (in prep). *Paradigmatic evidence for morphomic organisation in Kayardild inflection*.

Ruhlen, Merritt (1987). *A guide to the world's languages: Classification*. Vol. 1. Stanford: Stanford University Press.

Russell, S. and P. Norvig (2009). *Artificial intelligence: A modern approach*. 3rd edition. Upper Saddle River, NJ: Prentice Hall.

Sagot, Benoît (2013). 'Comparing complexity measures'. Paper presented at the Workshop 'Computational Approaches to Morphological Complexity', 22 February, 2013, Paris.

Sagot, Benoît, and Géraldine Walther (2011). 'Non-canonical inflection: data, formalisation and complexity measures'. In *Actes de SFCM 2011 (2nd workshop on Systems and Frameworks for Computational Morphology)*, Zürich, Switzerland.

Saffran, Jenny, Richard N. Aslin, and Elissa L. Newport (1996). 'Statistical Learning by 8-Month-Old Infants'. *Science* 274, 1926–8.

Saffran, Jenny, Elissa L. Newport, and Richard N. Aslin (1996).'Word segmentation: The role of distributional cues'. *Journal of Memory and Language* 35, 606–21.

Salanova, Andres Pablo (2012). 'Reduplication and verbal number in Mebengokre'. Presented at SSILA workshop on Morphological Complexity in the Languages of the Americas, Annual Meeting of the Linguistic Society of America, Portland, Oregon, January 2012.

Sampson, Geoffrey, David Gil, and Peter Trudgill (2009). *Language Complexity as an evolving variable* (Studies in the evolution of language 13). Oxford: Oxford University Press.

Sapir, Edward (1921). *Language*. New York: Harcourt, Brace & World.

Say, Tessa, and Harald Clahsen (2002). 'Words, rules and stems in the Italian mental lexicon'. In S. Nooteboom, Fred Weerman, and Frank Wijnen (eds), *Storage and computation in the language faculty*, pp. 96–122. Dordrecht: Kluwer Academic Publishers.

Schmitt, Rüdiger (1981). *Grammatik des Klassisch-Armenischen*. Vol. 32 of *Innsbrucker Beiträge zur Sprachwissenschaft*. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.

Schönig, Claus (1998). 'Azerbaidjanian'. In L. Johanson and E. A. Csató (eds), *The Turkish Languages*, pp. 248–60. London: Routledge.

Segel, Esben (2008). 'Re-evaluating zero: When nothing makes sense'. *Skase Journal of Theoretical Linguistics* 5(2), 1–20.

Seidel, August (1900). *Swahili Konversations grammatik*. Heidelberg: Julius Groos.

Shannon, Claude E. (1948). 'A mathematical theory of communication'. *Bell System Technical Journal* 27, 379–423.

Shannon, Claude E. (1951). 'Prediction and entropy of printed English'. *The Bell System Technical Journal* 30, 50–64.

Spencer, Andrew (2007). 'Extending deponency: implications for morphological mismatches'. *Proceedings of the British Academy* 145, 45–70.

Steever, Sanford B. (1998).'Kannada'. In S. B. Steever (ed.), *The Dravidian languages*, pp. 129–57. London: Routledge.

Stump, Gregory T. (1993). 'On rules of referral'. *Language* 69, 449–79.

Stump, Gregory T. (2001). *Inflectional morphology*. Cambridge: Cambridge University Press.

Stump, Gregory T. (2002). 'Morphological and syntactic paradigms: Arguments for a theory of paradigm linkage'. *Yearbook of Morphology* 2001, 147–80.

Stump, Gregory T. (2012). 'The formal and functional architecture of inflectional morphology'. In A. Ralli, G. Booij, S. Scalise, and A. Karasimos (eds), *Morphology and the Architecture of Grammar: Online Proceedings of the Eighth Mediterranean Morphology Meeting*, pp. 245–70.

Stump, Gregory, and Raphael A. Finkel (2013). *Morphological typology: From word to paradigm*. Cambridge: Cambridge University Press.

Thumb, Albert, and Richard Hauschild (1959). *Handbuch des Sanskrit: Eine Einführung in das sprachwissenschaftliche Studium des Altindischen*. II Teil. Heidelberg: Winter.

Trommer, Jochen, and Gereon Müller (eds) (2006). *Subanalysis of argument encoding in distributed morphology*. Vol. 84 of *Linguistische Arbeits Berichte*. Institut für Linguistik: Universität Leipzig.

Vaillant, André (1958). *Grammaire comparée des langues slaves*. Tome II. *Morphologie*. Deuxième partie: *Flexion pronominale*. Lyon-Paris: IAC.

Vajda, Edward J. (2010). 'A Siberian link with Na-Dene languages'. In James Kari and Ben A. Potter (eds), The Dene–Yeniseian connection (Anthropological Papers of the University of Alaska New Series, Vol. 5(1–2)), pp. 33–99. Fairbanks: Department of Anthropology, University of Alaska Fairbanks.

Voegtlin, Thomas (2002). 'Recursive self-organizing maps'. *Neural Networks* 15, 979–91.

Von Hinüber, Oskar (1968). *Studien zur Kasussyntax des Pāli, besonders des Vinaya-piṭaka*. München: Kitzinger.

Weiers, Michael (2011). 'Moghol'. In J. Janhunen (ed.), *The Mongolic Languages*, pp. 248–64. Routledge.

Whitney, Carol (2001). 'Position-specific effects within the SERIOL framework of letter-position coding'. *Connection Science* 13, 235–55.

Wiese, Bernd (2004). 'Categories and paradigms: On underspecification in Russian declension'. In Gereon Müller, Lutz Gunkel, and Gisela Zifonun (eds), *Explorations in nominal inflection*, pp. 321–72. Berlin: Mouton de Gruyter.

Wunderlich, Dieter, and Ray Fabri (1994). 'Minimalist morphology: An approach to inflection'. *Zeitschrift fü Sprachwissenschaft* 20, 236–94.

Yu, Alan C. L. (2007). *A natural history of infixation.* Oxford: Oxford University Press.

Zúñiga, Fernando (2000). *Mapudungun*. München: Lincom Europa.

Zwicky, A. (1985). 'How to describe inflection'. In M. Niepokuj, M. van Clay, V. Nikiforidou, and D. Feder (eds), *Proceedings of the Eleventh Annual Meeting of the Berkeley Linguists Society*, pp. 372–86. Berkeley, CA: Berkeley Linguistics Society.