Available online at www.sciencedirect.com

SciVerse ScienceDirect

Lingua 133 (2013) 360–373

Lingua

www.elsevier.com/locate/lingua

On the analogical modelling of the English past-tense: A critical assessment

Clive A. Matthews *

School of Language and Communication Studies, University of East Anglia, Norwich NR4 7TJ, United Kingdom

Received 30 October 2009; received in revised form 28 March 2013; accepted 3 April 2013

Abstract

The formation of English past tenses has provided the arena in which alternative computational models of psycholinguistic processes have been most intensely tested. The debate has mainly revolved around accounts based on a single, analogising mechanism as exemplified by connectionist models and the dual mechanism theory of Pinker (1999) which combines elements of a pattern associator memory with a rule-based system. Prasada and Pinker (1993) provide evidence against the connectionist approach by showing that their networks failed to reproduce the behaviour of their English speaking subjects when forming the past tense of a subset of nonce verbs. Eddington (2000a) provides an alternative single mechanism account of this data using exemplar-based, analogising models which are claimed to produce behaviour comparable to Prasada and Pinker's subjects. On the basis of these results he argues that not only is a general rejection of single mechanism, analogising accounts premature but also that exemplar-based models ''may prove to have advantages over their connectionist counterparts''.

This paper tests these theoretically important claims. It is shown that the contrast between the particular exemplar-based theory, analogical modelling (AM), and connectionist model which forms the basis of Eddington's arguments is, in fact, a function of differences in the mappings the two simulations compute and that once these are removed, the AM advantage disappears. A detailed analysis of the predictions of the particular AM simulation presented further shows that, although the output bears superficial similarities with those produced by Prasada and Pinker's subjects, other properties are more problematic.

Keywords: Exemplar-based computation; Analogical modelling; Single-mechanism account; English past tense

1. Introduction

Since its inception, the defining assumption of the cognitive sciences has been the *computational* or *information processing* hypothesis, namely that the mind/brain is ''a device capable of automatically inputting, storing, manipulating and outputting information in virtue of inputting, storing, manipulating and outputting representations of that information'' (van Eckardt, 1993:50). Since the nature and form of the implicated representations and the form of their manipulation are left unspecified, this proposal represents no more than a ''schema for a set of *versions* of a research framework that specifies the unrevisable elements and either leaves the revisable elements blank or specifies them in terms of a range of options'' (van Eckardt, 1993:139, emphasis in original). Accordingly, a wide variety of different models are compatible with the computational hypothesis including systems as simple as look-up tables (Churchland and Sejnowski, 1992:69–70).

DOI of original article: http://dx.doi.org/10.1016/S0024-3841(99)00043-1.
* Tel.: +44 1603 593430.
E-mail address: clive.matthews@uea.ac.uk.

0024-3841/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.lingua.2013.04.002

Within this space of possibilities the literature has tended to concentrate on ''classical'' rule-based theories using symbols with a combinatorial syntax (Pylyshyn, 1984) and connectionist (or neural) networks based upon principles of analogy using abstract representations organised around the ''statistical central tendency'' of the data (Bechtel and Abrahamsen, 2002).

Which type of computational system best describes human cognition is, naturally, an empirical question. Within the domain of language the arena in which symbolic and connectionist models have been most intensely tested has been the formation of the English past tense. That this has turned out to be the proving ground is, apart from an element of historical happenstance, largely driven by the contrasting properties of regular and irregular forms, differences which appear to favour one type of account over the other. On the one hand, regular inflection (e.g. *smile*/*smiled*) appears a paradigmatic candidate for treatment via symbolic rule, in this case concatenating the suffix -*ed* to the verb's stem. Certainly children's production of over-regularised forms such as *goed* and *hitted* were taken as prime evidence of rule-governed linguistic processes in the early years of the ''cognitive revolution'' (Berko, 1958; Kuczaj, 1977). On the other hand, irregularly inflected verbs (e.g. *go*/*went*, *hit*/*hit*, *fling*/*flung*), due to their unpredictable nature, seem to entail storage in memory but, because of various ''islands of reliability'' (e.g. *fling*/*flung*, *swing*/*swung*, *wring*/*wrung*, etc.), a memory which has the properties of a pattern associator.

The majority of researchers (although see Chomsky and Halle, 1968; Halle and Mohanan, 1985; Albright and Hayes, 2003, for dissenting voices) are in agreement that the irregulars are acquired and stored via some form of pattern association process, typically assumed to be realised by some form of connectionist model (Rumelhart and McClelland, 1986). The disagreement mainly concerns the regulars. Notwithstanding the rule-like appearance of these verbs, proponents of connectionist models claim that not only are they capable of generating regular forms but that the use of this single computational mechanism provides an explanation for the fact that the exceptions also share properties with the regular forms: for example, nearly two thirds of the irregulars have forms ending in /d/ or /t/ (McClelland and Patterson, 2002a). In contrast, others argue that the differences between the two classes are qualitative rather than graded and that the regular class is best accounted for in terms of a rule-based system. Some have taken this as showing that past-tense formation is therefore the result of two different computational processes; this dual-mechanism approach assumes an -*ed* concatenation rule for the regulars and an associative memory for the irregulars (Pinker, 1998, 1999).

As the debate over single or dual model accounts has developed, a wide range of phenomena have been brought to bear on the issue including data on acquisition, the distributions of verbs in phonological space, genetic impairment, aphasia and so. It is impossible to review this large, and often contradictory, literature in this paper (as a starting point see McClelland and Patterson, 2002a,b; Pinker and Ullman, 2002a,b; Pinker, 2006). Rather the focus here is on a single but oft discussed piece of data: that presented in Prasada and Pinker (1993, henceforth P&P). P&P compare the outputs of a neural net with those produced by English speaking informants on a set of nonce items. Ling and Marinov (1993) point out that this is a more realistic test of a computer simulation than predicting the forms of actual verbs since native speakers benefit from having memorised the exceptions. Following work initiated by Bybee and Moder (1983), P&P constructed a test set of 60 pseudo-verbs, divided equally into regular- and irregular-like items. These two subsets in turn consisted of three groups of 10 pseudo-verbs designated as *prototypical*, *intermediate* and *distant*. The prototypical pseudo-irregulars were constructed on the basis of rhymes with existing irregular verbs; the intermediate pseudo-irregulars were derived from the prototypical items by changing either the initial or final consonant cluster; the distant pseudo-irregulars were similarly derived from the prototypical items but by changing both initial and final consonant clusters. The pseudo-regulars were created in the same way but in this case with the prototypical nonce verbs rhyming with several existing high-frequency verbs which included vowels occurring in the pseudo-irregular set.[1] P&P found, as did Bybee and Moder (1983), that their English speaking subjects were more reluctant to supply irregular inflections to the pseudo-irregulars the greater their distance from existing (irregular) verbs (see Fig. 1). In other words, although many suggested *splung* as the past tense of *spling* (cf. *cling*/*clung*, *fling*/*flung*, *sling*/*slung*, *string*/*strung*, *swing*/*swung*, etc.), few supplied *trusp* as the equivalent form of *trisp*. Such results are consistent with a pattern associator memory model. On the other hand, phonological distance had virtually no effect on the willingness of their subjects to regularly inflect the pseudo-regulars

[1] It has been noted that P&P's test set is problematic on various grounds. For example, Albright and Hayes (2003:134) note that ''Prasada and Pinker designed their novel verbs using informal methods, such as finding verbs that rhymed with many regulars/irregulars, or changing just one phoneme vs. multiple phonemes to obtain greater distance. One problem with this approach is that it provides no quantitative control for how many existing rhymes a novel verb has, how similar they are, and so on. In addition, as Prasada and Pinker themselves note, this procedure introduces a confound: the only way for a novel verb to be dissimilar to all existing regulars is for it to be dissimilar to *all* English words. As a result, the verbs in Prasada and Pinker's ''distant from existing regulars'' condition were phonologically deviant as English words, e.g. *ploamph* and *smairg*.'' In line with these criticisms Albright and Hayes constructed their own test set of nonce items. Although this set could have been used with the simulations reported here, it was decided to retain P&P's test items since these were the focus of Eddington's paper. Although Albright and Hayes' items provide a more appropriate test set, P&P's own series have still remained something of a benchmark in subsequent work on analogical modelling (e.g. Keuleers, 2008).

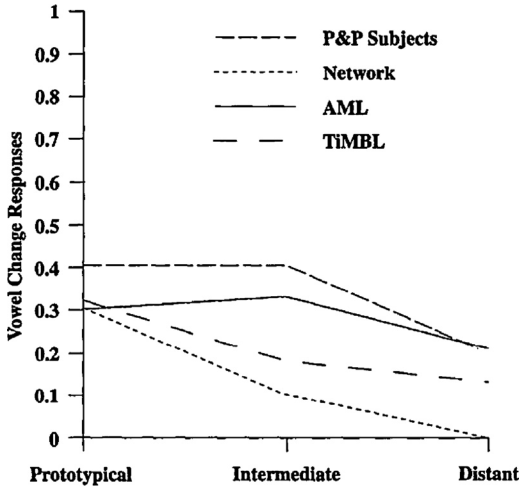

Fig. 1. Mean number/probability of vowel change past tense forms for pseudo-irregulars (Eddington's Fig. 3). Note that the legend keying the data plots inverts TiMBL's results with those of P&P's subjects. In other words, the second series up represents P&P's subjects' results and the top series TiMBL's.

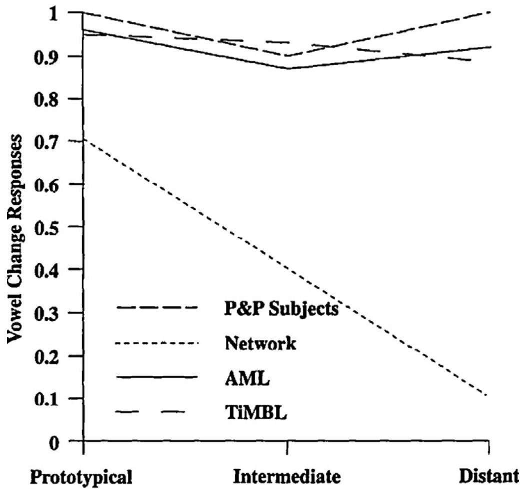

Fig. 2. Mean number/probability of regularly suffixed past tense forms for pseudo-regulars (Eddington's Fig. 4). The series labelling errors identified in the previous footnote also occur in this graph. In addition note that the y-axis should be labelled ''Suffixed Responses''.

(Prasada and Pinker, 1993:28; Pinker, 1999:143); they were as happy with *smeeltht* for *smeelth* as *plipt* for *plip* (see Fig. 2).[2] This pattern of behaviour is consistent with a rule-based model since the item's phonology is irrelevant to the regular rule's application.

P&P also fed the dataset of invented verbs to a trained neural network of the same type as that used by Rumelhart and McClelland (1986). As expected, given their analogising abilities, the net produced very similar outcomes to P&P's subjects with respect to the pseudo-irregulars. With the pseudo-regulars, however, there was a marked divergence of behaviour: the network treated them more like irregulars and decreasingly applied the regular suffix as the verbs became more unusual (see Figs. 1 and 2). Subsequent work using more sophisticated connectionist

[2] As Albright and Hayes (2003:134) point out this claim is slightly misleading. They note that, in fact, the distant regular verbs *did* receive lower participant ratings by P&P's informants. P&P's own suspicion was that these results might have been caused by the phonological deviance of items such as *ploamphed* and *smairged*. To correct for this possible confound P&P subtracted stem phonological well-formedness ratings from the past tense ratings; it was only then that the similarity effects for regulars appeared to disappear.

architectures (e.g. Plunkett and Marchman, 1993; Daugherty and Hare, 1993) has also failed to convincingly produce results matching those of P&P's subjects (for some discussion see Pinker, 1999; Marcus, 2001).

Eddington (2000a; see also Eddingtom, 2004), however, presents a series of results which cast doubt on the idea that P&P's results cannot be replicated within an analogy-based account. The models used by Eddington are also organised around a single analogising mechanism but in this case they are exemplar-based rather than connectionist (see Chandler, 2009, for a general overview). Such models rather than using a summary, abstract representation of the dataset as with neural nets, store (memorise) each instance of the learning set independently as an individual feature-vector paired with a category label. Generalisation to new examples is achieved by identifying similar instances within this set on the basis of which an outcome may be extrapolated. For example, assuming the unknown verb *spling* matched with *cling*, *fling* and *wring*, each of which forms its past tense through an ɪ/ʌ vowel change, its past tense form would be predicted to be *splung*. Because the items in the ''analogical set'' may be associated with different category labels (for example, *spling* may also match with *ring* from the ɪ/æ class), the model may identify a range of possible outcomes, each of which can be associated with a probability depending on how many instances in the set belong to the particular classification.[3] Although there is a large body of material in the literature on concept formation which explores (and supports) exemplar-based models (see Shanks, 1995; Murphy, 2002, for reviews) and a number of recent extensions into the field of Natural Language Processing (see Daelemans and van den Bosch, 2005; Skousen et al., 2002 for reviews), such models are still relatively unfamiliar in the mainstream linguistic literature.

Eddington's study made use of two particular exemplar-based models: Analogical Modeling of Language (these days simply known as Analogical Modeling (AM)) (Skousen, 1989; Skousen et al., 2002) and the Tilburg Memory Based Learner (TiMBL) (Daelemans and van den Bosch, 2005). It is beyond the scope of this paper to describe the technicalities of AM (the reader is referred to the references just cited as well as Eddington's paper for details) save to say that membership of the analogical set, determined by (unweighted) feature similarity, consists of those exemplars from the training set which share the greatest number of features with the test item (the nearest neighbours) as well as more distant matches (but only if they do not increase the uncertainty about the outcome compared with the closer matches). This latter class allows for more common patterns (e.g. the regular past-tense forms) to apply even if the instances are only marginally similar. Procedurally, the model does not distinguish between the regular and irregular instances in the training set.

Eddington presented P&P's test set to simulations of both AM and TiMBL. Figs. 1 and 2 (reproduced from Eddington, 2000a:294–295) plot the results. Fig. 1 records the results for the pseudo-irregulars in terms of the average probability that a vowel changing irregular form for each sub-group was predicted. Fig. 2 similarly shows the results for the pseudo-regulars but this time in terms of the probability that the simulations predicted a regular suffixed response.

As may be seen the results of the exemplar-based models appear to have more in common with P&P's informants than the output of the neural net; indeed Eddington claimed that the models ''mirror . . .[P&P's] . . . choices quite closely'' (Eddington, 2000a:294) whilst Skousen (2002a:23) is even more positive saying that the model ''correctly predicts'' the experimental findings. Eddington drew two conclusions from this: first that single mechanism, analogically-based models *are* capable of replicating P&P's data contrary to the claims previously rehearsed and, second, that exemplar-based theories are superior analogically-based cognitive models in comparison with the particular connectionist model adopted by P&P.

If Eddington's conclusions are well-founded, they are, clearly, of great significance to the general debate regarding the nature of cognitive processing and as such merit close scrutiny. That is the intention of this paper. Since the assessment is of necessity detailed, only AM will be considered in what follows. In part this choice is driven by Daelemans et al.'s (1997) finding that AM tends to outperform TiMBL on certain tasks and so possibly represents the best-case model. Further, to a broad approximation, proponents of AM adopt a more cognitive stance with respect to their simulations compared with those working with TiMBL (Skousen et al., 2002). The structure of the paper that follows divides into three parts. Section 2 argues that are a number of significant differences between Eddington's two models which cast doubt on the idea that they are truly comparable. Section 3 reports on the replication of the original study with the connectionist and analogical models more carefully aligned; the results show that with these adjustments both models perform equally well, contrary to Eddington's original conclusion. Section 4 raises a number of issues relating to the output of the AM model.

2. An initial assessment of Eddington's AM/connectionist comparison

Both the connectionist and AM models utilised by Eddington produce a mapping from root verbs to their associated past tense. The core issue is whether the reported difference in the output of the two functions is a consequence of the

[3] The probabilities quoted throughout this paper are based on the number of ''pointers'' associated with each item in the analogical set (as exemplified in Tables 2 and 3). For details see Eddington (2000a).

differing *computational processes* producing these mappings. For this question to be properly addressed, it is necessary that various features of the models – such as training sets, input and output representations, etc. – be identical. This section highlights a number of differences between the two models which cast doubt on their suitability for comparison. [Some of the following points were first noted in Matthews, 2005:288–289.]

Consider the input representations. P&P's connectionist simulation, which furnishes the results for Eddington's point of comparison with AM, is a replication of Rumelhart and McClelland's (1986) work and, hence, uses input and output representations based on Wickelphones. A Wickelphone is an ordered triple of phonemes representing the preceding phoneme, the phoneme of interest and the following phoneme. A word is represented as an unordered set of such forms: *hit*, for example, is associated with the set {#hi, hit, it#} (where '#' represents a word boundary).[4] The use of Wickelphones has been criticised on a variety of theoretical and technical grounds (Pinker and Prince, 1988:96–101; Lachter and Bever, 1988:208–215) and is clearly not an ideal form of representation. Since even powerful learning algorithms are of little avail if the representations are inappropriate, most subsequent connectionist research has tended to use a slot/feature form of representation (e.g. MacWhinney and Leinbach, 1991; Plunkett and Marchman, 1991, 1993; Ling and Marinov, 1993; Hare and Elman, 1995). The AM input is also naturally encoded in slot/feature template based upon a standard phonemic representation. That said, the AM inputs go beyond the mere phonemic since they also include information regarding the word's syllabic structure, whether its final syllable is stressed as well as the identity of the final phoneme (see Eddington, 2000a:292 for details). All of this information is included because, following earlier research by Derwing and Skousen (1994:210–214), it proves ''to be optimal'' for predicting past tense forms (Eddington, 2000a:292). If this is true, the fact that this information is not available to the connectionist model puts it at a potential disadvantage in any act of comparison.

The output representations of the two models are also significantly different. The connectionist simulation maps onto a phonological representation (in terms of Wickelfeatures) of the past tense of the input. The AM model, however, produces a radically different output, namely a *category label* indicating the class of past tense verb to which the item belongs.[5] Some connectionist models have also adopted this ''classifier'' approach (e.g. Hare and Elman, 1992; Plunkett and Nakisa, 1997; Hahn and Nakisa, 2000) with a corresponding improvement in outcomes when compared with networks mapping onto phonological outputs. Pinker and Ullman (2002a:458), however, describe this type of approach as turning ''inflection into a multiple-choice test among a few innate possibilities''. This somewhat disparaging remark partially reflects what is perceived to be a reduction in the complexity of the computational task. So, for example, the production of the particular inflected phonological form is factored out of the task and left to some subsequent, unspecified computational process. In addition, as Hahn and Nakisa (2000:323) note, the classifier approach obviates the need to overcome the alignment problem (Bullinaria, 1997:238); a problem which arises in the phonology-to-phonology approach since the length of the output strings do not match those of the input in any consistent way.[6] With different inputs and outputs it is difficult to see the point of comparing Eddington's two models.

There is further mismatch which potentially confounds the comparison. It is well-known in the machine learning literature (e.g. Banko and Brill, 2001) that as the size of the database increases, so does the accuracy of generalisation; in fact, as Daelemans and van den Bosch (2005:54) note, the improvement tends to approximate to a log-linear curve. Although dataset size may have an effect on a system's output, Eddington does not control for this: his AM dataset consists of 848 exemplars which is double the 422 items P&P used to train their neural net. It is not impossible that this difference skewed the results in AM's favour.

The reporting of the results in Figs. 1 and 2 also raises a potential problem of comparability. The probability scores reported by P&P for the connectionist simulation represent the *actual* number of forms the network produced for the relevant subset of test items: the value of 0.3 for vowel changes within the prototypical set, for example, records that 3 of the 10 outputs produced by the network belonged to this category. The corresponding AM probability of 0.32, however, represents the average probability assigned to the vowel-change category for each word. This average, however, does not entail that AM assigned 3 of the inputs to the vowel changing category; the same average could be recorded even if each item were actually predicted to be regular.

[4] The actual input to (and output from) the system is via a set of lower level representations called Wickelfeatures which include a distinctive feature for each of the phonemes contained in the Wickelphone.

[5] More recent work using exemplar-based models has focussed on producing fully-specified phonological forms as output (for some discussion see Chandler, 2010; Keuleers, 2008; see also Albright and Hayes, 2003).

[6] In an earlier version of this paper it was also argued that a classifier approach would remove the need to find a means of preventing the regular suffix from being applied to an inflected form of the stem – resulting in, say, *broked* – rather than to the stem itself (Marcus, 2001:77–79) on the grounds that such blended forms are rarely produced by children (Marcus et al., 1992) and adults (Prasada and Pinker, 1993). A reviewer, however, has pointed out that children do produce such forms and possibly more commonly than the previous authors suggest (see, for example, Kuczaj (1977:594) as do some adults with language impairment (see, for example, Ullman et al., 1997:271).

Table 1
Sample encodings of dataset.

reg	= 0rI0 = = 0siv01v	‘‘receive’’ – regular
irrv	= = =0 = = = 0du0=1u	‘‘do’’ – irregular with vowel change
irreg	= = = 0 = = sprEd01d	‘‘spread’’ – irregular with no vowel change

As a final area of difficulty we might note that although Eddington asserts that the instance-based results ‘‘mirror [P&P’s] subjects’ choices quite closely’’ (p. 294), no statistical analysis is presented to support this claim. If one wanted to dispute the claim, it is unclear how to decide the issue.

As this section has shown, there are a variety of reasons which prevent a proper assessment of Eddington’s claims for the superiority of AM over connectionist models on the basis of the supplied data. The following section reports on simulations which attempt to rectify these problems and so allow for a more appropriate comparison.

3. A re-comparison of connectionism with analogical modelling

The simulations reported in this section are intended to use the two models to compute the same input/output function on the basis of the same training set. Since it offered the simplest solution, the decision was taken to use a neural net which maps onto class labels rather than phonological forms.

The AM encoding of the input data essentially replicated Eddington’s original study. Each item was coded as a feature vector consisting of 14 variables representing the phonemic content of the penultimate and final syllables of the verb, the stress value of the final syllable and the verb’s final phoneme. Both syllables were encoded in terms of six variables representing three onset consonants, the vocalic peak and two coda consonants.[7] To align the nuclei of each syllable, assignment proceeded from the vowel outwards, with 0 (zero) used to mark the edges of the syllable. Any unassigned variables were filled with the null variable (=). For monosyllabic verbs the penultimate syllable position was represented as = = = 0 = =. Variables 13 and 14 recorded whether the final syllable was stressed (1) or not (0) and the verb’s final phoneme, respectively. This latter information is included since, being centred on the vowel, the final phoneme may occur in different positions within the syllabic representation. Finally, each verb was classified as forming its past tense either through the addition of the regular suffix (‘‘reg’’),[8] some kind of vowel change (‘‘irrv’’) or some other irregular form not involving a vowel change (‘‘irreg’’). Some sample representations are shown in Table 1.

A set of 4254 verbs was initially coded based on the training set of Albright and Hayes (2003).[9] The phonemic transcriptions of these verbs reflected American English pronunciations. From this set a subset of the 848 most frequent verbs were chosen to form the basic training set. This set comprised 726 regular verbs, 99 irregulars with a vowel change and 23 other irregulars.[10] The test set comprised the 60 nonce items from P&P’s paper.

The connectionist simulation used a standard feedforward network with a hidden layer of 20 units. Although the output of a network is, in general, sensitive to the number of hidden units, no attempt was made to optimise performance through experimentation with the hidden layer; it is quite possible, therefore, that the results reported here do not represent the best-case. The input layer consisted of 201 units notionally divided into 14 pools of units. Each pool corresponded to one of the variables in the AM vectors and each position within the pool to a particular character. The size of the pool depended upon the number of characters found to occur in that position (given the database). For example, since only two characters, <0> or <s>, occurred in the third onset position of the final syllable (variable 7), the seventh pool consisted of two units; the 14th pool, on the other hand, had to represent 34 different characters and so consisted of 34 units. It is

[7] This differs slightly from Eddington’s encoding. First, the onset to the final syllable only included the last two consonants. Secondly, the penult syllable did not include the onset at all.

[8] It may be thought that this slips rules into the story by the backdoor; closer consideration suggests not. [I thank a reviewer for the following point.] Linguistic rules, being generalisations, involve statements over variables; the regular past-tense rule below may be read as ‘‘to form the past tense of X, where X is a verb, concatenate ‘‘ed’’ to it’’.

$$[X]_V \rightarrow [X + ed]_{Vpast}$$

It is because the variable abstracts away from the verb’s meaning, phonology, frequency of occurrence and so on that *-ed*-concatenation can, in principle, apply to any verb. AM’s behavioural class ‘‘regular verb’’, however, does not directly make reference to such an abstract statement; rather, it simply keys the operation ‘‘concatenate ‘‘ed’’ to the particular item being processed.

[9] Data downloaded from http://www.linguistics.ucla.edu/people/hayes/rulesvsanalogy/.

[10] Unlike Eddington this set was taken from the English portion of the CELEX database (Baayen et al., 1995) rather than the Brown corpus (Francis and Kučera, 1982). As with the minor coding differences mentioned in fn 9, any discrepancies with Eddington’s study are not relevant here since the issue is how the two classifiers perform *relative to one another given this dataset and coding*.

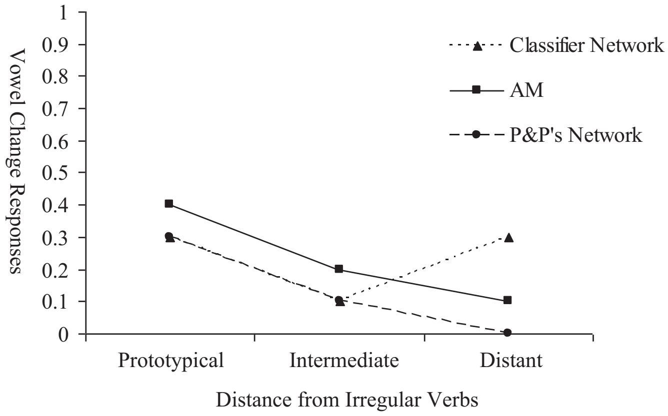

Fig. 3. Number of vowel change past-tense forms for pseudo-irregulars (expressed as a probability).

unlikely that this crudest of approaches represents the optimal means of encoding the data and the results reported below should be interpreted in this light. The null variable in any position was represented as a block of 0s. The output layer consisted of three units, one for each of the three categories of past tense.

The AM simulations were run using AM::Parallel,[11] version 2.3 of an implementation of the AM algorithm written in a combination of Perl and C (Stanford, 2002; Parkinson, 2002). The connectionist simulations used the neural network development system NeuralWorks Professional II/Plus (© NeuralWare, Inc.). The network was initially trained using the back propagation algorithm (with a learning rate of 0.1 and sigmoid transfer function) until all 848 items in the dataset were correctly assigned to one of the three output categories. The test set was applied to this network.

The outputs produced by the two simulations are presented in Appendix 1. Figs. 3 and 4 represent these results in the same way as Figs. 1 and 2 from Eddington's paper. In order to be consistent with P&P's network results, the probability scores represent the number of *actual* forms classified by the two systems for the relevant category.[12] Although there are slight differences, the pattern of results for the pseudo-irregulars for both models as shown in Fig. 3 appears to be similar to that reported by Eddington (Fig. 1).

The network results for the pseudo-regulars presented in Fig. 4, however, are very different. The contrast between the output of the classifier network and that produced by P&P's connectionist simulation is striking. In particular, although P&P's model failed to add *-ed* to the distant regulars, the classifier model showed no such reluctance and produced results which closely approximate to the AM model.

In order to more properly quantify the degree of correspondence between the two classifiers, each output was classified as 0 if categorised as regular and 1 if categorised as irregular (either with or without a vowel change – although

[11] Available at http://humanities.byu.edu/am/amdownloads.html.

[12] The following table gives the AM average predicted probability in the same way as recorded by Eddington.

	Prototypical	Intermediate	Distant
Vowel change irregulars	0.3964	0.3151	0.1693
Suffixed regulars	0.9744	0.8616	0.7519

These figures are not precisely those reported in Eddington (2000a: 293) which were: vowel change irregulars – 0.32, 0.33, 0.21; suffixed regulars – 0.96, 0.87, 0.92.

Because the network's unit outputs typically sum to 1.00 for a particular input, it is also possible to treat these outputs as rough approximations to probability values. The following table records the average for these outputs.

	Prototypical	Intermediate	Distant
Vowel change irregulars	0.2632	0.1879	0.3057
Suffixed regulars	0.9508	0.6700	0.7237

Although not identical with the scores recorded in Figs. 3 and 4, the trends are very similar. Assuming Eddington's original models produced similar results, the potential discrepancy regarding the different methods of recording probability for the net and AM noted in section 2 in Eddington's graphs (Figs. 1 and 2) is unlikely to be serious.

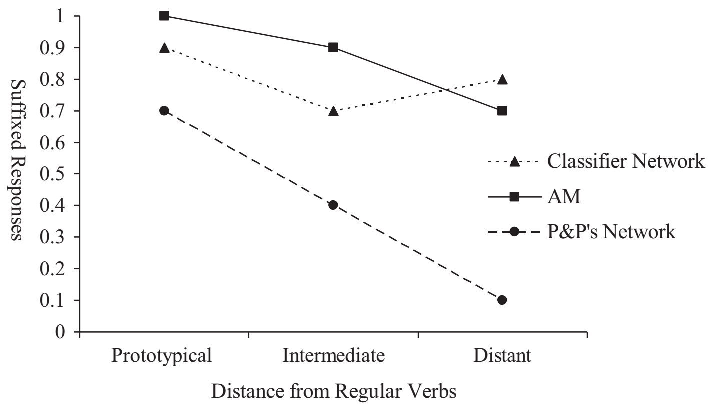

Fig. 4. Number of suffixed past-tense forms for pseudo-regulars (expressed as a probability).

spling was the only example of the latter). On the basis of this McNemar's χ^2 test (Gillick and Cox, 1989) shows that the differences between the two classifiers was insignificant ($p < 0.424$).[13] The McNemar tests applied to the two subsets similarly show no significant differences between the models – regulars ($p < 0.250$) and irregulars ($p = 1.000$).

The results of this section show clearly that once a neural network is constituted as a classifier, and uses the same input representations and training set, AM's advantage as reported by Eddington disappears. This conclusion is consistent with previously reported comparisons of the two models (Daelemans et al., 1993; Matthews, 2005). The next section considers the nature of AM's performance in a little more detail.

4. AM's classificatory behaviour

Eddington (2000a:281) claims that his AM model (and, presumably, the simulations reported here) ''successfully mirrored [P&P's] subject's responses''. In terms of the general trend of the curves in Figs. 1–4 this appears to be roughly true (although statistically unproven). On closer inspection, however, there are various aspects where this interpretation is more questionable.

Considering the data presented by P&P in their Appendix (1993:52–54) two facts are apparent: (i) their subjects show a strong preference for regular suffixation in all cases, including the prototypical pseudo-irregulars, but even so (ii) some irregular forms are produced for each example.[14] This pattern of results is not replicated by the particular AM model and dataset used in this study. So, for example, since no regular items are members of the relevant analogical sets, both *spling* and *meep* are assigned a zero probability of being regular – a somewhat unfortunate result for this simulation if this is taken as implying that a regular form is impossible. In contrast, the connectionist classifier produced a very low level of activation for the same examples (*spling* = 0.01, *meep* = 0.06) which might be argued to suggest some slight probability that these items permit a regular form. That said, the network produced negative outputs for two instances (*smeeg* = −0.09, *smeej* = −0.07) which could also be interpreted as meaning no regular form. Surprisingly these results occur even though 86% of the training set is composed of regular verbs. From this perspective the irregular items appear to be producing a disproportionate effect.

From the opposite perspective, those instances assigned an AM probability of 1.0 of being regular – *cleef*, *nist*, *glip*, *gloke*, *pleem*, *nace*, *ploab*, *ploag*, *smaib*, *smaig* – are predicted as having no possible irregular form. Again an unfortunate conclusion since, as P&P's subjects clearly appreciated, any verb could potentially be irregular even if only a member of the no-change class. The connectionist classifier produced an even greater number of regular only outputs – as may be seen in Appendix 1.

[13] Cross-validation is often used as a means of evaluating and comparing learning algorithms (Weiss and Kulikowski, 1991). The process works by dividing the overall set of items into *k*-groups and testing the algorithm *k*-times leaving out one of the subsets from the training set each time and using it as the test set. In this way each data item has a chance of being validated against. This is not appropriate in the present case since (a) the dataset is relatively small and (b) the nonce set cannot be included as part of a training set since by definition they do not have a past tense.

[14] The same roughly applies to the results reported by Albright and Hayes (2003:155–158) although there are a number of examples where their subjects failed to sanction any irregular forms. AM simulations with the Albright and Hayes' test set (not discussed here) allowed irregular forms for all but a small number of these examples. Since no comparable connectionist model was tested on the same test set, it is not possible to compare results across these two models.

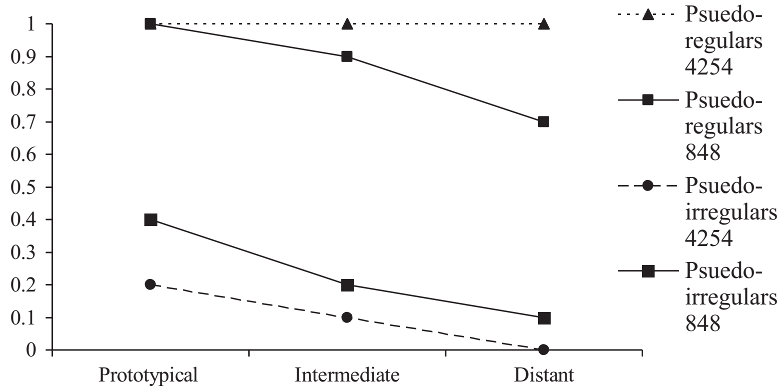

Fig. 5. Comparison of AM probabilities in relation to training set size of 4254 vs. 848. (Pseudo-regular = probability of *-ed* suffix; pseudo-irregular = probability of vowel change.).

Part of the explanation for these results lies in the size of the training set: the set simply does not contain enough regular verbs for the algorithm to find analogical matches for *spling* and *meep*. Increasing the number of items in the dataset – and so more closely approximating the lexicon of an adult native speaker – overcomes this. So, using the full 4254 Albright and Hayes (2003) word set results in *meep* being assigned a 0.63 probability of being regular on the basis of matches to verbs such as *heap*, *seep*, *reap* and *peep*, verbs which were not in the original, smaller training set.[15] In general the larger training set with its increase in both the number and proportion of regulars[16] results in the probability of an item being classified as irregular dropping across the board as shown in Fig. 5.[17] Intuitively it is less clear from these results whether AM's predictions still ''mirror'' P&P's subjects' responses but McNemar's statistic shows that the difference in classification is not significant ($p < 0.008$).

Given that the nature of the training set has a significant effect on the outcomes of the model as previously mentioned, which results (those on the basis of the 848 or 4254 training sets) should form the basis of comparison with P&P's subjects? Eddington justifies the constitution of his small set partially on the grounds of robust frequency effects in lexical access reported in the psycholinguistic literature (examples cited are Allen et al., 1992; Scarborough et al., 1977; MacKay, 1982). From this it is assumed that ''frequent forms are more readily available, and therefore, more likely to be selected as analogs'' (p. 292). This is, perhaps, not an unreasonable assumption[18] but note that the cited evidence only identifies the *type* of verb (i.e. high frequency) which should be included in the training set, not the *number*. The statement that ''*therefore*, the 848 most frequent English verbs were extracted'' (p. 292, emphasis added) clearly does not follow; certainly no psycholinguistic evidence on lexical access is presented to justify this kind of number. In a later paper (Eddington, 2002a:297) it is noted that the choice of 848 items was based on the fact they were ''sufficient to simulate native speaker intuitions'' but whether this is meant to represent the linguistic knowledge available to the speaker/hearer is another matter. Interestingly, more recent AM simulations experiment with much larger datasets: for example Chandler (2002) on English past tenses uses 1617 monosyllabic English verbs; Eddington (2002b) on Spanish gender assignment uses 2416 nouns whilst Eddington (2000b) on Spanish stress assignment utilises a training set of 4970 words.

Because the three-way categorisation is so blunt in this particular simulation, the classifications produced are rather uninformative; although AM classifies *spling* as a vowel-changing irregular (with a probability of 0.93), this does not inform us whether the predicted form is *splung* (by analogy with *fling-flung*) or *splang* (*sing-sang*). By examining *spling*'s analogical set as shown in Table 2, however, it can be seen that in effect *splung* is the preferred option with a probability of 0.51 compared with *splang*'s 0.28.

[15] The effect on *spling* is not as dramatic but it also is given a probability of 0.04 of being regular.

[16] Regulars (4040) = 94.97% (compared with 85.61% in small training set); irregular with vowel change (172) = 4.04% (11.67%); irregular with no vowel change (42) = 0.99% (2.71%).

[17] The results based on the larger training set showed an increase in the number of items which received a probability of 1.0 for regular: up from 10 to 23 (out of the 60). As the number of regulars increase in the dataset it is not surprising that the proportion of irregulars decrease in the analogical sets; what is unexpected is that the irregulars disappear altogether in 13 of the sets. This surprising result is in need of an explanation.

[18] Although not unreasonable, it is still contentious. Balota and Chumbley (1984), for example, argue that frequency effects are the result of post-access mechanisms, although they are probably in the minority in adopting this view.

Table 2
Analogical set for *spling* on basis of 848 word training set. [Column 1: classification category (irrv = irregular with vowel change, irreg = irregular without vowel change); column 2: instance from dataset matching input; column 3: number of pointers to instance; column 4: percentage of pointers relative to total frequency.].

irrv	bring	120	13.97%
irrv	ring	120	13.97%
irreg	spread	8	0.93%
irrv	sing	120	13.97%
irrv	swing	120	13.97%
irreg	split	51	5.94%
irrv	cling	160	18.63%
irrv	fling	160	18.63%
Statistical Summary			
irrv	vowel-change	800	93.13%
irreg	no-vowel-change	59	6.87%

Table 3
Analogical set for *cleef* on basis of 848 word training set. [Column 1: classification category (reg = regular verb); column 2: instance from dataset matching input; column 3: number of pointers to instance; column 4: percentage of pointers relative to total frequency.].

reg	laugh	88	32.35%
reg	lift	28	10.29%
reg	clean	16	5.88%
reg	shift	24	8.82%
reg	drift	28	10.29%
reg	stuff	88	32.35%
Statistical Summary			
reg	regular	272	100.00%

Considering Table 2 it is worth noting that the set also predicts that *splought* (by analogy with *bring*/*brought*) and *spling* (*split*/*split*) are possibilities (but, not of course, the regular *splinged*).[19] Although these last two are ‘‘logical’’ possibilities, the data from Prasada and Pinker (1993) (and supported by the results of Albright and Hayes, 2003:157) suggests that they are not contemplated by their human subjects as possibilities. The analogical sets are full of such marginal examples – *prect* for *preek* (by analogy to *creep*/*crept*); *frownk* for *frink* (*freeze*/*froze*), etc. – forms which even the most creative punster is unlikely to contemplate (Prasada and Pinker, 1993; fn 4). It may be assumed that the difficulty here resides, at least partly, in the fact that the matching algorithm is not constrained in its notion of similarity. For example, although phonological rhyme may be important to speakers (Marchman, 1997), this particular AM model does not place any special emphasis on this property so allowing *preek* to match *creep* in the above.[20]

Further consideration of the analogical sets reveals other anomalies. For example, as previously noted, on the basis of the particular dataset used *cleef* is assigned to the class of regulars with a probability of 1.0 on the basis of the analogical set shown in Table 3.

The problem is that this set includes examples from each of the allomorphic variants for the regular past tense; is the model, therefore, predicting that both *cleefd* and *cleefəd* are possible forms along with (the correct) *cleeft*?[21] Derwing and Skousen (1994:213–214) and Skousen (2002b:43–44) have previously noted this problem with respect to regulars, but it

[19] Although the connectionist output does not produce an analogical set, a potential point of comparison are the results obtained with a net whose output units represented the fine-grained 27-way classification described by Pinker and Prince (1988). For each test item the majority of these output units showed a degree of activation. That said, there was typically a sharp division between those outputting <0.04 and those producing an output >0.1. Taking the latter for *spling*, the predicted past-tense forms on this simulation were *spling* (0.43), *splang* (0.12), *splung* (0.16), *splingt* (0.26) and *splinged* (0.27).

[20] I thank one of the reviewers for this point. The connectionist model used also predicts *prect* as a possible past-tense form (although with a very low probability) so this is not a problem unique to the AM model. As one of the reviewers noted, the phonological rhyme effect is one that an empirically adequate model will need to encompass in some way.

[21] The neural network mentioned in fn 19 similarly predicted *cleefd* and *cleeft* (but not *cleefəd*) as possible past-tense forms although its preferred form was *claught*. These various examples show that using a more fine-grained form of categorization compared with Eddington's rather blunt three-way classification does not lead to an improvement in either model's performance nor to distinguishing one model from the other (in terms of these simulations). No significant improvements are seen as the size of the training set increases given this level of granularity.

cleef	0	k	l	i	f	0	1	f
laugh	=	0	l	@	f	0	1	f
lift	=	0	l	I	f	t	1	t
clean	0	k	l	i	n	0	1	n
shift	=	0	S	I	f	t	1	t
drift	0	d	r	I	f	t	1	t
stuff	0	s	t	V	f	0	1	f

Fig. 6. ''Variegated'' feature similarities for *cleef*'s analogical set.

should be equally borne in mind that the problem extends to the irregular categories: for example, on the basis of the *fling* occurring in its analogical set, *flape* is predicted to belong to the ɪ → ʌ class even though it does not contain the requisite vowel, whilst *goav* is predicted to have a possible suppleted form, whatever that might be, due to *go* being included in its analogical set.[22]

These unfortunate analogues arise as a consequence of the matching algorithm relying on what Albright and Hayes (2003:121) call ''variegated similarity''. In other words, because each variable carries equal weight, exemplars may be similar to the target in different ways. Fig. 6 illustrates the variable nature of the matching with respect to the analogical set of *cleef*. This clearly shows that although the representations have been deliberately coded so as to isolate both the vocalic core of each syllable as well as the final phoneme as the determinant of the allophonic class, the matching process does not ensure that either variable is accorded special emphasis.

Various solutions have been suggested in the AM literature to block predicting ill-formed input such as *cleefd* and *cleefəd*. One approach has been to assume a single regular past tense category, ''add /d/'', and then rely on phonological rules (modulo the comments in fn 8) to either insert a schwa or devoice the /d/ depending upon the environment (Derwing and Skousen, 1994:213–214; Skousen, 2002b:44). Derwing and Skousen felt unhappy with this solution partially on the grounds that it represents ''an ad hoc and unparsimonious addendum to the analogical model'' (p. 214). It should also be noted that their approach would not obviously rescue the *flape*/*goav*-type of examples noted here. More sophisticated approaches to phonological specifications have recently been developed for analogical models (see Keuleers, 2008; Chandler, 2010 for some discussion) which represent more compelling solutions to these problems.

5. Conclusions

Eddington (2000a) argues that, contrary to previous claims, homogeneous, analogising models are capable of producing the kinds of differential behaviour shown by native English speaking informants with respect to regular and irregular inflections first identified by Prasada and Pinker (1993) as long as those models are exemplar-based. From this he further argues that these models prove superior computational models to certain connectionist accounts of cognition (at least in this domain). The results and arguments presented in this paper cast doubt on both these conclusions.

The empirical contrast between exemplar-based and connectionist models which Eddington identifies has been shown to have been a function of the differences in the datasets and the mappings which the simulations were computing; once these differences are removed, the AM advantage disappears.

With respect to the detail of the AM simulation, although there are certain superficial similarities in output with P&P's subjects, closer inspection shows a lack of match with respect to certain important properties; for example, in some instances the model's predictions are too restricted – failing to predict possible regular or irregular outcomes – whilst in

[22] The fine-grained connectionist simulation did not categorise *flape* as belonging to the ɪ → ʌ class but did similarly classify *goav* as having a potential suppleted form (0.08).

others the suggested forms are phonologically impossible. It is far from clear, then, whether the particular model of AM presented by Eddington had ''successfully predicted the [P&P] subjects' choices'' (p. 295) is truly justified. Similar problems equally apply to the connectionist simulations used.

It should be borne in mind that the discussion throughout the paper has revolved around two particular implementations of the computational models in question. As such it could be argued that recent developments, particularly within analogical modelling, have reduced the force of the analysis relating, as it does, to somewhat dated implementations. This may or may not be so. There are, however, a couple of important lessons to be learnt from the considerations of this paper. The first is that the success or not of a simulation can only be judged by considering the detail of the model's performance and not through some broad metric of achievement. This leads to the second, and more important, lesson. As has been demonstrated throughout, the performance of a particular implementation (both AM and connectionist) can vary widely depending on how the data is coded, the nature of the output categories, the size of the training set and so on. As a consequence, the researcher has to be careful to distinguish between implementational decisions which produce a desired outcome from those which can be justified on independent grounds. It is this latter constraint on theoretical adequacy which places the greatest demand on the modeller and it is hoped that this paper contributes in some small way to the call for greater circumspection when choosing how to implement a particular model.

Acknowledgements

This paper has taken a considerable length of time to reach this published form and I thank the journal's editorial staff for their patience during this process. I am especially indebted to the very detailed comments received from the external reviewers who have helped to considerably improve the content of the paper. Naturally any errors or omissions remain the author's responsibility.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.lingua.2013.04.002.

References

Albright, A., Hayes, B., 2003. Rules vs. analogy in English past tenses: a computational/experimental study. Cognition 90, 119–161.

Allen, P., McNeal, M., Kvak, D., 1992. Perhaps the lexicon is coded as a function of word frequency. Journal of Memory and Language 31, 826–844.

Baayen, H., Piepenbrock, R., Gulikers, L., 1995. The CELEX Lexical Database (Release 2). Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA.

Balota, D., Chumbley, J., 1984. Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. Journal of Experimental Psychology: Human Perception and Performance 10, 340–357.

Banko, M., Brill, E., 2001. Mitigating the paucity-of-data problem: exploring the effect of training corpus size on classifier performance for natural language processing. In: Proceedings of the Conference on Human Language Technology, San Diego, CA.

Bechtel, W., Abrahamsen, A., 2002. Connectionism and the Mind: Parallel Processing, Dynamics, and Evolution in Networks. Basil Blackwell, Oxford.

Berko, J., 1958. The child's learning of English morphology. Word 14, 150–177.

Bullinaria, J., 1997. Modeling reading, spelling, and past tense learning with artificial neural networks. Brain and Language 59, 236–266.

Bybee, J., Moder, C., 1983. Morphological classes as natural categories. Language 59, 251–270.

Chandler, S., 2002. Skousen's analogical approach as an exemplar-based model of categorization. In: Skousen, R., Lonsdale, D., Parkinson, D. (Eds.), Analogical Modeling: An Exemplar-based Approach to Language. John Benjamins, Amsterdam, pp. 51–105.

Chandler, S., 2009. Exemplar-based models. In: Eddington, D. (Ed.), Quantitative and Experimental Linguistics. Lincom Europa, Munich, pp. 100–158.

Chandler, S., 2010. The English past tense: analogy redux. Cognitive Linguistics 21, 371–417.

Chomsky, N., Halle, M., 1968. The Sound Pattern of English. MIT Press, Cambridge, MA.

Churchland, P.S., Sejnowski, T., 1992. The Computational Brain. Bradford Books. MIT Press, Cambridge, MA.

Daelemans, W., van den Bosch, A., 2005. Memory-Based Language Processing. Cambridge University Press, Cambridge.

Daelemans, W., Gillis, S., Durieux, G., van den Bosch, A., 1993. Learnability and markedness in data-driven acquisition of stress. In: Ellison, T., Scobbie, J. (Eds.), Computational Phonology, Edinburgh Working Papers in Cognitive Science, vol. 8. pp. 157–178.

Daelemans, W., Gillis, S., Durieux, G., 1997. Skousen's analogical modeling algorithm: a comparison with lazy learning. In: Jones, D., Summers, H. (Eds.), New Methods in Language Processing. University College Press, London, pp. 3–15.

Daugherty, K., Hare, M., 1993. What's in a rule? The past tense by some other name might be called a connectionist net. In: Mozer, M., Smolensky, P., Touretzky, D., Elman, J., Wiegend, A. (Eds.), Proceedings of the 1993 Connectionist Models Summer School. Erlbaum, Hillsdale, NJ, pp. 149–156.

Derwing, B., Skousen, R., 1994. Productivity and the English past tense: testing Skousen's Analogical Model. In: Lima, S., Corrigan, R., Iverson, G. (Eds.), The Reality of Linguistic Rules. John Benjamins, Amsterdam, pp. 193–218.
Eddingtom, D., 2004. Issues in modeling language processing analogically. Lingua 114, 849–871.
Eddington, D., 2000a. Analogy and the dual-route model of morphology. Lingua 110, 281–298.
Eddington, D., 2000b. Spanish stress assignment within the analogical modeling of language. Language 76, 92–109.
Eddington, D., 2002a. Dissociation in Italian conjugations: a single-route account. Brain and Language 81, 291–302.
Eddington, D., 2002b. Spanish gender assignment in an analogical framework. Journal of Quantitative Linguistics 9, 49–75.
Francis, N., Kučera, H., 1982. Frequency Analysis of English Usage: Lexicon and Grammar. Houghton Mifflin, Boston, MA.
Gillick, L., Cox, S., 1989. Some statistical issues in the comparison of speech recognition algorithms. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, vo1. 1. pp. 532–535.
Hahn, U., Nakisa, R., 2000. German inflection: single route or dual route. Cognitive Psychology 41, 313–360.
Halle, M., Mohanan, K., 1985. Segmental phonology of modern English. Linguistic Inquiry 16, 57–116.
Hare, M., Elman, J., 1992. A connectionist account of English inflectional morphology. In: Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society. Erlbaum, Mahwah, NJ, pp. 265–270.
Hare, M., Elman, J., 1995. Learning and morphological change. Cognition 56, 61–98.
Keuleers, E., 2008. Memory-based learning of inflectional morphology. Unpublished doctoral dissertation. Universiteit Antwerpen, Antwerp.
Kuczaj, S., 1977. The acquisition of regular and irregular past tense forms. Journal of Verbal Learning and Verbal Behavior 16, 589–600.
Lachter, J., Bever, T., 1988. The relation between linguistic structure and associative theories of language learning: a constructive critique of some connectionist learning models. Cognition 28, 195–247.
Ling, C., Marinov, M., 1993. Answering the connectionist challenge: a symbolic model of learning the past tenses of English verbs. Cognition 49, 235–290.
MacKay, D., 1982. The problems of flexibility, fluency, and speed-accuracy trade-off in skilled behaviour. Psychology Review 89, 60–94.
MacWhinney, B., Leinbach, J., 1991. Implementations are not conceptualizations: revising the verb learning model. Cognition 40, 121–157.
Marchman, V., 1997. Children's productivity in the English past tense: the role of frequency, phonology, and neighborhood structure. Cognitive Science 21, 283–304.
Marcus, G., 2001. The Algebraic Mind: Integrating Connectionism and Cognitive Science. MIT Press, Cambridge, MA.
Marcus, G., Pinker, S., Ullman, M., Hollander, J., Rosen, T., Xu, F., 1992. Over-generalization in language acquisition. Monographs of the Society for Research in Child Language 57.
Matthews, C., 2005. French gender attribution on the basis of similarity: a comparison between AM and connectionist models. Journal of Quantitative Linguistics 12, 262–296.
McClelland, J., Patterson, K., 2002a. 'Words or Rules' cannot exploit the regularity in exceptions: reply to Pinker and Ullman. Trends in Cognitive Sciences 11, 464–465.
McClelland, J., Patterson, K., 2002b. Rules or connections in past-tense inflections: what does the evidence rule out? Trends in Cognitive Sciences 11, 465–472.
Murphy, G., 2002. The Big Book of Concepts. Bradford Books. MIT Press, Cambridge, MA.
Parkinson, D., 2002. Running the Perl/C version of the Analogical Modeling program. In: Skousen, R., Lonsdale, D., Parkinson, D. (Eds.), Analogical Modeling: An Exemplar-based Approach to Language. John Benjamins, Amsterdam, pp. 365–383.
Pinker, S., 1998. Words and rules. Lingua 106, 219–242.
Pinker, S., 1999. Words and Rules: The Ingredients of Language. Weidenfeld and Nicolson, London.
Pinker, S., 2006. Whatever happened to the past tense debate? In: Bakovic, E., Ito, J., McCarthy, J. (Eds.), Wondering at the Natural Fecundity of Things: Essays in Honor of Alan Prince. Lingusitics Research Center, University of California, Santa Cruise. Available online: http://repositories.cdlib.org/lrc/prince/11.
Pinker, S., Prince, A., 1988. On language and connectionism: analysis of a parallel distributed processing model of language acquisition. Cognition 28, 73–193.
Pinker, S., Ullman, M., 2002a. The past and future of the past tense. Trends in Cognitive Sciences 6, 456–463.
Pinker, S., Ullman, M., 2002b. Combination and structure, not gradedness, is the issue. Trends in Cognitive Sciences 6, 472–474.
Plunkett, K., Marchman, V., 1991. U-shaped learning and frequency effects in a multi-layered perceptron: implications for child language acquisition. Cognition 38, 43–102.
Plunkett, K., Marchman, V., 1993. From rote learning to system building: acquiring verb morphology in children and connectionist nets. Cognition 48, 21–69.
Plunkett, K., Nakisa, R., 1997. A connectionist model of the Arabic plural system. Language and Cognitive Processes 12, 807–836.
Prasada, S., Pinker, S., 1993. Generalisation of regular and irregular morphological patterns. Language and Cognitive Processes 8, 1–56.
Pylyshyn, Z., 1984. Computation and Cognition: Toward a Foundation for Cognitive Science. Bradford Books. MIT Press, Cambridge, MA.
Rumelhart, D., McClelland, J., 1986. On learning the past tenses of English verbs. In: Rumelhart, D., McClelland, J. (Eds.), Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Vol. 2 Psychological and Biological Models. Bradford Books. MIT Press, Cambridge, MA, pp. 216–271.
Scarborough, D., Cortese, C., Scarborough, H., 1977. Frequency and repetition effects in lexical memory. Journal of Experimental Psychology 3, 1–17.
Shanks, D., 1995. The Psychology of Associative Learning. Cambridge University Press, Cambridge.
Skousen, R., 1989. Analogical Modeling of Language. Kluwer, Dordrecht.
Skousen, R., 2002a. An overview of Analogical Modeling. In: Skousen, R., Lonsdale, D., Parkinson, D. (Eds.), Analogical Modeling: An Exemplar-based Approach to Language. John Benjamins, Amsterdam, pp. 11–26.
Skousen, R., 2002b. Issues in Analogical Modeling. In: Skousen, R., Lonsdale, D., Parkinson, D. (Eds.), Analogical Modeling: An Exemplar-based Approach to Language. John Benjamins, Amsterdam, pp. 27–48.

Skousen, R., Lonsdale, D., Parkinson, D. (Eds.), 2002. Analogical Modeling: An Exemplar-based Approach to Language. John Benjamins, Amsterdam.
Stanford, T., 2002. Implementing the Analogical Modeling algorithm. In: Skousen, R., Lonsdale, D., Parkinson, D. (Eds.), Analogical Modeling: An Exemplar-based Approach to Language. John Benjamins, Amsterdam, pp. 385–409.
Ullman, M., Corkin, S., Coppola, M., Hickok, G., Growden, J., Koroshetz, W., Pinker, S., 1997. A neural dissociation within language: evidence that the mental dictionary is part of declarative memory, and that grammatical rules are processed by the procedural system. Journal of Cognitive Neuroscience 9, 266–276.
van Eckardt, B., 1993. What is Cognitive Science? Bradford Books. MIT Press, Cambridge, MA.
Weiss, S., Kulikowski, C., 1991. Computer Systems that Learn. Morgan Kaufmann, San Mateo, CA.