In: *Proceedings of the Ninth International Conference on Artificial Neural Networks (ICANN), 1999*

A Constructivist Neural Network Model of German Verb Inflection in Agrammatic Aphasia

Gert Westermann and **David Willshaw**
Institute for Adaptive and Neural Computation
University of Edinburgh, 2 Buccleuch Place
Edinburgh EH8 9LW, Scotland UK
gert@anc.ed.ac.uk, david@anc.ed.ac.uk

Martina Penke
Department of Linguistics
Heinrich-Heine Universität Düsseldorf
40225 Düsseldorf, Germany
penke@phil-fak.uni-duesseldorf.de

Abstract

We present a constructivist neural network that closely models the performance of agrammatic aphasics on German participle inflection. The network constructs a modular architecture leading to a double dissociation between regular and irregular verbs, and lesioning the trained network accounts for data obtained from aphasic subjects [6]. We analyze the internal structure of the network with respect to the representation of regular and irregular verbs and argue that the constructivist neural network presents a more plausible account of verb inflection than a recently proposed dual-mechanism theory.

1 Introduction

Recently there has been much research and controversy about the representation of inflectional morphology in the mind which has mainly addressed acquisition of the English past tense. It has become clear, however, that the English past tense confounds the issues of regularity, frequency and affixation: regular past tense forms have a higher frequency than irregulars, and they are the only forms that have an inflectional ending (`-ed`). Interest has therefore shifted to the German participle in which all verbs show separable endings and where the regular case is not the most frequent one.

In this paper we present a constructivist neural network that models the performance of agrammatic aphasics with the German participle, and we compare this model with a dual-mechanism theory of verb inflection. We show that the model can account for detailed human data, and that it presents a more successful and plausible account of verb inflection than the dual-mechanism theory.

Below, we first review the structure of the German participle, the dual-mechanism theory of inflection, and agrammatic aphasia. We then describe the constructivist neural network, the data used for the simulations, and the training regime. We analyze the network in terms of the internal representation it develops, and then report detailed results from experiments with lesioning of the trained network. Finally, we discuss the implications of the modelling results for the dual-mechanism theory of inflection and for the suitability of constructivist neural networks in the modelling of verb inflection.

1.1 The German Participle

German participles are comparable in usage to the English past tense in describing an event in the past. There are three groups of participles: *Regular* participles are formed by a (prosodically determined) prefix `ge-`, the verb stem, and the ending `-t`, e.g., *sagen* (say) →*gesagt* (said), *lachen* (laugh) →*gelacht* (laughed). *Irregular* participles take the ending `-en`, e.g., *geben* (give) → *gegeben* (given) and they may also change the verb stem, e.g., *gehen* (go) →*gegangen* (gone), *nehmen* (take) →*genommen* (taken). The third group are *mixed* verbs that take the regular ending `-t` but change their stems like irregulars: *wissen* (know) →*gewusst* (known), *denken* (think) →*gedacht* (thought).

The distribution of these verb groups in the German language is given in table 1.

	type		token	
Regular	1936	(64.71%)	40196	(46.89%)
Irregular	956	(31.95%)	41276	(48.16%)
Mixed	100	(3.34%)	4243	(4.95%)
Sum	2992	(100.00%)	85715	(100.00%)

Table 1: Distribution of the different verb groups in German (analyzed from the CELEX database). The numbers are participle frequencies.

In contrast to English, German does not have a majority of regular tokens[1], and the majority of types[2] is less pronounced than in English.

German verbs are often formed by modifying other existing verbs with a prefix or separable particle, e.g., in the CELEX database the simplex verb *fahren* (drive) occurs in 28 composite forms such as *hinausfahren*, *losfahren*, *fortfahren* etc. (drive out, drive off, continue). Since a prefix or particle do not alter the way in which the participle of a simplex verb is formed, we combined all composite forms into one simplex form, e.g., the 28 types made from *fahren* were combined into the single type *fahren*. This simplification increased the proportion of regular types in the corpus, indicating that most of the composite types that were lost in the simplex version were irregulars. Since there are many rare verbs in the CELEX database, we further removed all verbs

[1] each verb counted according to how often it occurs
[2] each verb counted just once

with a token frequency of less than ten, resulting in a corpus of 484 simplex types.

1.2 The Dual-Mechanism Theory of Inflection

Recently, a dual-mechanism theory of inflection has been proposed in response to homogeneous-architecture connectionist accounts [1, 3, 7].

The dual-mechanism theory of participle inflection postulates two qualitatively distinct pathways for the production of regular and irregular participles. While irregular forms are stored in an associative memory, regular forms are constructed each time they are produced by the application of a mental rule (for German: "add `-t` to the verb stem"). There are different theories of how these two pathways interact, the *Blocking Principle* [see e.g., 3] being the most prominent one: when a participle is to be produced, the associative memory is first searched for a corresponding entry and the retrieval of such an entry blocks activation of the rule-path. If no entry is found, however, the (default) rule is applied.

Westermann [9] argued for a middle position between homogeneous connectionist and dual-mechanism theories of verb inflection, describing a constructivist neural network that grows a modular architecture based on a single processing mechanism. The network successfully modelled human data and it developed specialized pathways for different participles, but in contrast to the dual-mechanism theory there was no qualitative distinction in the processing of regular and irregular verbs.

In this paper we use a similar constructivist neural network for modelling the performance of agrammatic aphasic subjects described in [6]. Comparing the results of our model with the human aphasic performance and investigating the internal structure of the network model can lead to insights into the human processing mechanisms for inflectional morphology.

1.3 Agrammatic Aphasia

Agrammatic Broca's aphasia is a language disorder that is generally caused by a stroke predominantly affecting anterior parts of the left hemisphere. One of the characteristic symptoms of Broca's aphasia is the tendency to omit or confuse inflections [4]. Investigating the precise nature of these deficits can therefore lead to insights into the internal representation of inflectional morphology. Penke et al. [6] analyzed data from eleven aphasic subjects who each produced 39 regular and 39 irregular participles in a sentence completion task with respect to regular and irregular errors, overregularizations and irregularizations, frequency effects, and effects of ablaut-patterns on error rates.

2 Network Model

For the simulations described in this paper, a modified version of the constructivist Supervised Growing Neural Gas (SGNG) network [2] was used. This model has previously been successfully applied to modelling the acquisition of the English past tense [9]. The SGNG algorithm constructively builds the hidden layer of a radial basis function (RBF) network. Each hidden unit has a Gaussian activation function and thus acts as a *receptive field* (rf) for an area of the input space. The problem in building RBF networks is to decide on the number and positions of the hidden units. The SGNG algorithm solves this problem by constructing the hidden layer during learning, adding units when and where they are needed. The network starts with just two units in the hidden layer, each covering roughly half of the input space (see figure 1). The network tries to learn the task with this architecture (by adjusting the weights with e.g., quickprop), and when learning no longer improves performance, a new unit is inserted. The place where the new unit is inserted is determined by the classification error resulting from treating inputs within one rf as similar: the rf that previously caused the highest error is shrunk and the new unit is inserted next to it. The idea here is that a unit which produces a high output error is inadequate, and therefore more structural resources are needed in that area.

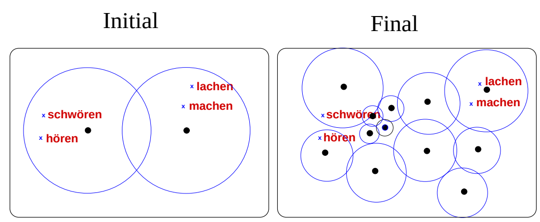

Figure 1: Receptive fields covering the input space at the beginning (left) and the end (right) of learning.

Figure 1 shows a hypothetical start and end state in a two-dimensional input space. While initially only two rfs cover the whole of the space, later hidden units have been inserted with different densities across the space to account for the specific learning task.

Figure 2 shows the network architecture. The input layer takes a phonological representation of the verb infinitive, and the output layer has one unit for each possible output class (see below). The hidden layer initially consists of only two units but is grown during learning. There are direct connections from the input to the output layer, and each hidden unit is fully connected to the output layer.

3 Data

The 484 verbs (see section 1.1) were classified according to the way in which their participles are formed (i.e., type of stem change and inflectional ending), resulting in a total of 21 classes, one of which was the regular class, four were for mixed verbs, and 16 for irregular verbs.

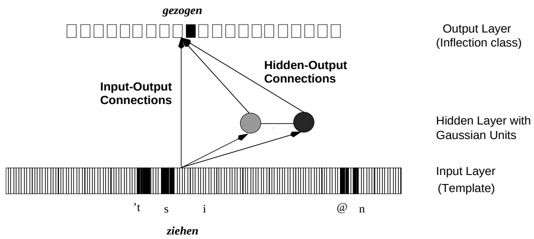

Figure 2: The initial architecture of the network.

Each phoneme was represented by a 7-bit feature vector with features such as *fricative*, *plosive*, *voiced* etc. for consonants, and *front*, *high*, *open* etc. for vowels. Presence of a feature was encoded with 1 and absence with -1.

For the training of the network, the phonological representation of the infinitive of each verb was then inserted into a template consisting of three syllables: XCCCVVCC-XCCCVVCC-XCCCVVCC; C stands for consonant, V for vowel, and X for whether the syllable is stressed or not. Since the endings of verbs are significant for the determination of the participle class, the verbs were right-aligned in this template. Table 2 shows examples of the template representation.

empfinden	`0---E-m-1pf-I-n-0d--@-n-`
schreiben	`0-------1Sr-W---0b--@-n-`
fliessen	`0-------1fl-i---0s--@-n-`
sein	`0-------0-------1z--W-n-`
`Template`	`XCCCVVCCXCCCVVCCXCCCVVCC`

Table 2: Template representation of the verb infinitives.

The resulting network had 150 input units (three syllables with seven phonemes each represented by seven features, plus one stress-bit per syllable), and 21 output units for the 21 inflection classes.

4 Training

The task to be learned by the network was the mapping from the phonological representation of the verb infinitive to the class of its participle. Viewing the learning of the participle as a classification task avoids confounding it with phonological details such as different pronunciation of regular forms depending on the last stem phoneme (*holen* →*geholt* vs. *landen* →*gelandet*).

From the corpus of 484 simplex verbs 20,000 verb tokens were randomly extracted according to their frequency. To ensure that each verb occurred at least once, all verb types which had not been randomly selected were added onto the resulting corpus with a token frequency of one (this concerned just one verb).

The structure of the resulting training corpus is shown in table 3.

The constructivist network was trained on this corpus as outlined in section 2. After 5411 epochs, performance on the training data reached 100% and all verbs were classified correctly for their participle class.

	type		token	
Regular	358	(73.97%)	9128	(45.64%)
Irregular	117	(24.17%)	9104	(49.18%)
Mixed	9	(1.86%)	1769	(5.18%)
Sum	484	(100.00%)	20001	(100.00%)

Table 3: Structure of the training corpus.

4.1 Developed Network Structure

The final structure of the network consisted of 318 hidden nodes (receptive fields, *rfs*), which corresponds to an average of 1.52 verbs per node. However, a closer investigation of how receptive fields responded to individual verbs revealed a significant distinction between regular and irregular verbs: while 220 rfs responded to regular verbs, 109 responded to irregulars (some rfs responded to both regulars and irregulars). Given the ratio of regulars and irregulars in the corpus, this means that on average, each regular verb used 60% of an rf while each irregular used 93%. In other words, regular verbs shared a receptive field with other verbs, while each irregular used almost a whole receptive field. For example, the four regular verbs *malen*, *baden*, *bahnen*, and *bannen* were all covered by the same rf, whereas there were never more than two irregulars covered by the same rf.

Numerical evidence for the uneven distribution of the rfs in input space comes from computing the mean and standard deviation of the average distance from each unit to its nearest neighbour, and comparing the result with a random distribution of units in input space. For the trained network the mean distance from a hidden unit to its nearest neighbour was $d = 8.70$ with a standard deviation $\sigma = 7.22$. This result was compared with averaging over 100 trials with 318 randomly distributed units which yielded $d = 75.08$ and $\sigma = 4.22$. d being significantly smaller in the network than in the random distribution indicates that the units populate only a subspace of the input space, while the higher variance shows a non-homogeneous distribution in this subspace.

The uneven distribution of resources indicates one advantage of constructivist learning, namely, the allocation of structure where it is needed to enhance learning, in this case, for the more difficult irregular verbs.

5 Lesioning Experiments

In order to model the results obtained from agrammatic aphasics [6] the network model was lesioned in different ways. It was assumed that the removal of weights in the network model corresponds to the destruction of neural tissue in the brain by a stroke.

The output in the network was produced through two routes or pathways: the direct connections between the input and the output layer that existed prior to the training of the network, and the connections from the growing receptive field layer to the output layer. We investigated the role of these two pathways by lesioning them individually and by randomly lesioning the whole network to different degrees.

5.1 Lesioning of pathways

Lesioning the individual pathways resulted in a double dissociation between regular and irregular verbs in the network, with mixed verbs behaving like irregulars. Figure 3 shows that when the direct input-output connections were removed, performance for the regular verbs was significantly worse than for irregulars. By contrast, lesioning of the hidden-output pathway lead to the opposite effect: performance for the regular verbs was hardly affected at all, while there was a marked decrease in performance for irregulars. This result corresponds to the findings reported in [6]: eight of the eleven tested aphasic subjects made significantly more errors with irregular verbs (41%) than with regulars (9%), the other three showed no significant differences between regular and irregular errors. Lesioning the connections from the receptive fields to the output layer in the network thus modelled the basic deficit in the inflection of agrammatic aphasics.

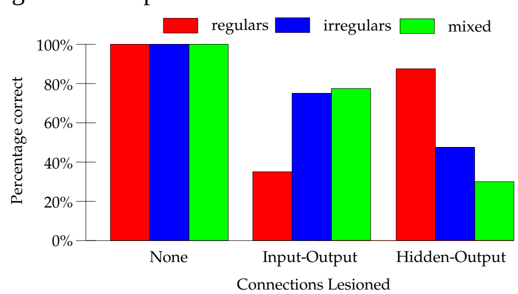

Figure 3: Double dissociation between regular and irregular (and mixed) verbs when lesioning the two pathways in the network.

Based on this result we investigated the performance of the lesioned network with respect to the more detailed results described in [6].

One of these results concerned the types of errors made by the aphasics: all subjects who made more errors on irregulars than on regulars overgeneralized the regular ending `-t` to irregular verbs, but they only rarely irregularized regular verbs. The same result was found in the network model: 65.2% of the irregular/mixed errors were overregularizations, but only 16.4% of the errors with regulars were irregularizations.

Another result obtained from the aphasic subjects was a frequency effect: based on the dual-mechanism hypothesis that regular forms are produced on the fly while irregular forms are fully stored in a lexicon, Penke et al. [6] predicted that the token frequency of irregular verbs would have an effect on the error rate with the more frequent irregulars being more robust, while no such effect should exist for regular verbs (since both frequent and infrequent regular participles are produced on the fly and thus not affected by a memory deficit). This prediction was confirmed in their experiments: the error rate for infrequent irregular participles was significantly higher than for frequent ones, while no such effect occurred for regulars. Again, we were able to model this result in the lesioned network (see figure 4): when tested on the same verbs as the aphasic subjects, frequent irregulars showed fewer suffixation-errors than infrequent ones, whereas error rates for frequent and infrequent regulars were nearly the same.

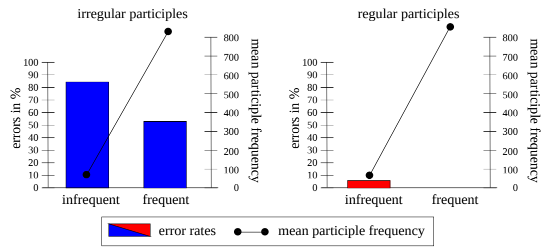

Figure 4: Frequency distribution of suffixation errors with regular and irregular participles in the neural network model. The lines shown are the mean participle frequencies, based on the CELEX-database. The columns show the distribution of errors for infrequent and frequent irregular and regular participles.

Based on a neural network model of the acquisition of the German participle, Westermann [8] had argued that regularity and irregularity are two ends of a continuum: a regular verb can be said to be “very regular” if it is similar to other regulars and dissimilar to irregulars. It can be said to be “less regular” if it is dissimilar to other regulars but similar to irregulars. The reverse is true for irregulars.

A regularity continuum would predict that “less regular” regulars, being more similar to irregulars, should be more error prone than “very regular” regulars. Penke et al. [6] analyzed the distribution of verbs with respect to stem vowels and found that for the stem vowel <e>, irregulars outnumber regulars, whereas for the stem vowels <au>, <ö>, <ä> and <ü>, regulars outnumber irregulars. Therefore, regular verbs with <e> should have a higher error rate because they are similar to irregulars. For irregular verbs, more overregularizations should occur for verbs with <au>, <ö>, <ä> and <ü> than for those with <e>, because they are “less irregular”.

This prediction was confirmed in the analysis of the aphasic data: all regular suffixation errors occurred with <e>-regulars. For infrequent irregulars, significantly more errors were made for verbs with <au>, <ö>, <ä> and <ü> compared with <e>.

We tested our lesioned network model for the same effect and found the same results: the error rate for <e> regulars was 78.9% compared with an average error rate for the four other stem types of 1.38%, indicating that <e>-regulars are more irregular than others. For irregulars the results were opposite: <e> forms had an error rate of 75% whereas the other forms were all wrong (100%).

5.2 Global Lesioning

The results reported so far indicate that the specific lesioning of the receptive field pathway in the network model can account for details of the morphological impairment in agrammatic aphasia. However, such selective lesioning would suggest that the pathways are locally distinct and that a stroke would always affect one but not the other pathway. In this case it is hard to explain why often in agrammatic aphasia regular inflection is selectively spared while irregular inflection is impaired, whereas it is not clear that the reverse case, i.e. the selective sparing of irregular inflection, occurs at all.

We therefore investigated the effects of globally lesioning the network to different degrees, without making a distinction between the input-output and the hidden-output pathways. Over 100 trials, the network was lesioned in 5%-steps by randomly removing weights from both pathways.

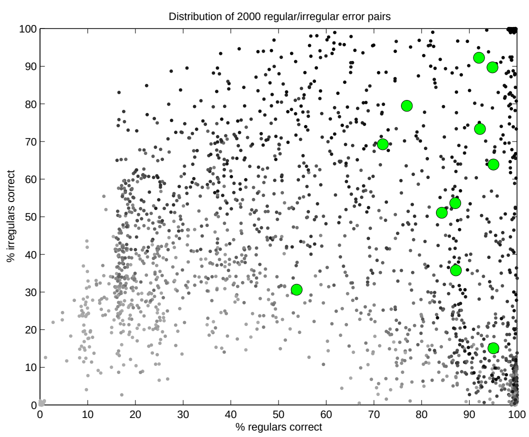

Figure 5: Performance on regulars vs. irregulars for 100 lesioning trials at 20 lesioning steps each. Greyscale indicates degree of lesioning (from dark to light). Data for the aphasic subjects are marked by circles.

Although the average tendency of the lesioned networks was to perform better on regulars than on irregulars, there were significant differences between individual trials and degrees of lesioning. Figure 5 shows the performance on regular vs. irregular participles for 100 random lesioning trials with 20 lesioning steps each (0%–100% in 5%-intervals). The data for the eleven aphasic subjects from [6] are also displayed. The figure shows that in the model the degree of lesioning (indicated by the greyscale) is only a rough predictor of performance. For example, selective sparing of regulars with almost total breakdown for irregulars (bottom right corner of the plot) occurs for a wide range of degrees of lesioning.

However, all aphasic data are well within the range of performance predicted by the simulations, and at the same time the model predicts that a selective sparing of irregulars (top left corner of the plot) should occur only very rarely or never at all. This result shows that although there is considerable variability in the performance of agrammatic aphasics, different lesioned networks can yet model the performance of each of them. The model is not over-general, however: like in aphasic subjects, a selective sparing of irregulars with a breakdown of regular inflection did not occur in any of the lesioning trials.

Why does global lesioning lead towards a selective impairment of irregulars while regulars are more or less spared? The answer could lie in the fact that each verb, regular and irregular, activates a receptive field in the hidden layer which contributes to the production of the correct output class. However, only the regular verbs are produced correctly even without the hidden layer pathway (section 5.1). Therefore, when connections in both pathways are lesioned, there is a higher probability for regular verbs to retain one intact pathway, leading to a correct production of the participle in more cases than for irregular verbs.

Connections from the receptive fields to the output layer are highly specific: there are often strong connections to the one output unit which represents the inflection class of the verb covered by a particular rf, and most of the other connections have very small weights. The loss of such a specialized connection will lead to the production of a wrong participle for that verb.

6 Discussion

The processing in the two pathways of the network model could be viewed as distinct: the direct input-output pathway operates on the structure, that is, the phonological representation of the input verb. By contrast, a receptive field constitutes a localist representation where one activated unit represents a whole verb. While this type of network might at first sight correspond to the dual-mechanism theory proposed in [1, 3, 7], a closer look reveals that there are significant differences: although the network develops two pathways to produce the correct participle of all verbs, it does not employ two distinct mechanisms in the production of regular and irregular forms. Instead, through the single mechanism of activation propagation and weight adaptation, each verb produces activation in each of the pathways, and the pathways collaborate to produce the correct participle class.

Furthermore, the constructivist network does not correspond to the principle of the dual-mechanism model which maintains that the participles of irregular verbs are stored as full forms whereas regular participles are produced by a rule mechanism: in the model, no full participle forms are stored; instead, they are produced through transformations of two different representations (distributed phonological and localist receptive field) of the verb infinitive.

The neural network model avoids many of the problem associated with the dual-mechanism theory of inflection:

- The collaboration between both pathways for every verb makes unnecessary an explanation of which pathway becomes active under which circumstances such as the Blocking Principle (see section 1.2). An implementation of this principle [5] showed that it does not improve the performance of a single-mechanism model.
- Mixed verbs which combine an (irregular) stem change with a (regular) `-t` ending do not fit into a dual-mechanism model that postulates encapsulated pathways without interaction, and they are generally ignored in dual-mechanism treatments of verb inflection [1, 3]. Further, the dual-mechanism theory cannot account for a regularity continuum which implies associative effects even for regulars. In the network model, however, the regularity continuum is natural and the integration of mixed verbs is straightforward: the degree of (ir)regularity is determined by the activation of each of the two paths.
- A dual-mechanism theory cannot easily account for the fact that agrammatic aphasia profiles in which irregular verbs are selectively spared but regulars are impaired seem to occur only very rarely. Such a profile would arise from a stroke affecting only the rule-path. On the other hand, the observed selective impairment of irregulars can in the dual-mechanism theory only be explained by a stroke affecting specifically the associative memory, which implies a local distinction between the two pathways. These strong assumptions, that the pathways are locally distinct but only one of them is ever selectively affected by a stroke, are unnecessary in the constructivist network model: the profiles observed in aphasics arise from the global lesioning of the whole system without a local distinction between pathways based on the fact that regulars can employ two representations of the system input and are therefore more robust than irregulars which rely on one highly specialized representation alone.
- In the network model, different lesioning trials lead to different performance profiles which nevertheless fall within the range observed in human aphasic data. The network model can therefore account simultaneously for the considerable variability between aphasic subjects and for the general absence of certain profiles (selective irregular sparing).

7 Conclusions

In this paper we have presented a constructivist neural network model that accounts for detailed results from the study of agrammatic aphasic processing of the German participle. Such results have often been used as arguments against "connectionist" accounts of morphological processing [1, 3], but these criticisms have confounded the issues of a single processing mechanism and a homogeneous network architecture. We have argued elsewhere [9] that constructivist neural networks growing a modular architecture avoid many of the shortcomings of fixed-architecture, homogeneous networks, and the results presented in this paper give further evidence for the claim that constructivist modular neural networks constitute plausible models of human morphological processing.

Acknowledgements The third author was supported by German Science Foundation grant SFB 282/C8. Thanks to Dieter Wunderlich for helpful comments on a draft of this paper and to David Sterratt for proof-reading.

References

[1] Clahsen, H. [in press]. Lexical entries and rules of language: A multidisciplinary study of German inflection, *Behavioral and Brain Sciences* .

[2] Fritzke, B. [1994]. Fast learning with incremental RBF networks, *Neural Processing Letters* **1**: 2–5.

[3] Marcus, G., Brinkmann, U., Clahsen, H., Wiese, R. and Pinker, S. [1995]. German inflection: The exception that proves the rule, *Cognitive Psychology* **29**: 189–256.

[4] Menn, L. and Obler, L. K. (eds) [1990]. *Agrammatic Aphasia: A Cross-Language Narrative Sourcebook*, Vol. I, Benjamins, Amsterdam.

[5] Nakisa, R. C., Plunkett, K. and Hahn, U. [1997]. A cross-linguistic comparison of single and dual-route models of inflectional morphology, *in* P. Broeder and J. Murre (eds), *Cognitive Models of Language Acquisiton*, MIT Press, Cambridge, MA.

[6] Penke, M., Janssen, U. and Krause, M. [in press]. The representation of inflectional morphology: Evidence from Broca's aphasia, *Brain and Language* .

[7] Pinker, S. [1991]. Rules of language, *Science* **253**: 530–535.

[8] Westermann, G. [1995]. *Konnektionistische Regelverarbeitung: Ein Neuronales Netzwerk zum Erlernen der Partizipienbildung im Deutschen*, Master's thesis, Technische Universität Braunschweig, Braunschweig, Germany.

[9] Westermann, G. [1998]. Emergent modularity and U-shaped learning in a constructivist neural network learning the English past tense, *Proceedings of the 20th Annual Conference of the Cognitive Science Society*, Erlbaum, Hillsdale, NJ, pp. 1130–1135.