# A Constructivist Dual-Representation Model of Verb Inflection

**Gert Westermann** (gert@cogsci.ed.ac.uk)
Institute for Adaptive and Neural Computation, Division of Informatics
University of Edinburgh, 2 Buccleuch Place
Edinburgh EH8 9LW, Scotland UK

## Abstract

A constructivist neural network is presented that models impaired inflectional processing in German agrammatic aphasia. The model is based on a single mechanism and develops two types of representation through a constructivist learning process. The model accounts for data that has been taken as evidence for a dual mechanism theory of inflection, and it suggests an inflectional processing system that is based not on a distinction between regular and irregular cases, but between inflections that are easy and hard to learn. The model represents a succesful single-mechanism neural network account of verb inflections.

## Introduction

The debate between rule-based and association-based theories of inflection has been continuing for many years and has moved from the initial focus on the English past tense to other languages such as the German participle (e.g. Clahsen, 1999; Marcus *et al.*, 1995). The reason for this shift is that in English, the issues of "regularity" and "high frequency" are confounded which makes it difficult to distinguish between the different theories. By contrast, in the German participle the regular case does not apply to the majority of all verbs, making it a so-called "minority default" (Marcus *et al.*, 1995).

A popular recent theory of how inflections are formed is the *Dual Mechanism Theory* (DMT) that postulates two qualitatively distinct mechanisms for the production of regular and irregular cases (e.g. Clahsen, 1999; Pinker, 1991, 1997; Marcus *et al.*, 1995). According to the DMT, regular inflections are produced by a mental symbolic rule, whereas irregulars are stored in an associative lexicon. Based on these mechanisms, the DMT claims to account for differences in the processing of regular and irregular inflections: whereas regular forms are applied productively to novel forms independently of their similarity to existing forms (e.g., *faxed*), irregular inflections show similarity effects both in existing "families" (*read* → *read*, *lead* → *led*, *breed* → *bred*) and in the extension to novel forms (*cleed* → *cled*).

However, while considerable empirical research has established processing differences between regular and irregular forms on many different levels from acquisition over psycholinguistic and ERP studies to impaired adult processing (see Clahsen, 1999, for an overview), little progress has been made in the specification of the DMT. Particularly problematic is the question in which way the two mechanisms interact to produce the inflected form. Marcus *et al.* (1995) proposed the *Blocking Principle* which states that a lexical entry (indicating an irregular verb) blocks the application of the rule, but an implementation of this principle (Nakisa *et al.*, 1997) showed that in practice it involves parameters for which a useful setting cannot be found. Therefore, the DMT remains highly underspecified and thus hard to falsify. However, even in its underspecified form, the DMT is contradicted by some empirical data, e.g., frequency effects for regular English past tense (Stemberger and MacWhinney, 1986) and regular Dutch plural (Baayen *et al.*, 1997) forms, and similarity effects for regular German participles in agrammatic aphasia (Penke *et al.*, 1999).

In this paper I present a neural network model of inflectional processing in German agrammatic aphasia that accounts for dissociations between regular and irregular forms without postulating two qualitatively distinct mechanisms. Instead, the model develops two types of representations in a constructivist process, driven by the structure of the training data, and it displays emerging areas of functional specialization that correspond largely, but not completely, to the distinction between regular and irregular forms. The trained model is lesioned in different ways and it accounts for empirical data better than the DMT. Based on these results I propose a new theory of inflectional processing that is based on a distinction not between regular and irregular, but between "easy to learn" and "hard to learn" forms.

The rest of this paper is organized as follows: first, the structure of the German participle and the impairment profiles observed in agrammatic aphaisa are reviewed. Then, the network model, the data, and the training regime are described, followed by a detailed analysis of the performance of the model in comparison with agrammatic aphasics. Finally, the resulting new theory of inflectional processing is presented and related to the DMT.

## 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: *Weak* participles are formed by a (prosodically determined) prefix *ge*-, the verb stem, and the ending -*t*, e.g., *sagen* (say) → *gesagt* (said). *Strong* 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). A few strong verbs have idiosyncratic participle forms, e.g., *sein* (be) → *gewesen* (been). The third group are *mixed* verbs that take the weak ending *-t* but change their stems like strong verbs, e.g., *wissen* (know) → *gewusst* (known). It is generally claimed that the weak verbs form the regular class, while strong verbs are irregular, and the terms regular and irregular will here be used in this sense.

In contrast to English, German does not have a majority of regular tokens (each verb counted according to how often it occurs in a corpus), and the majority of types (each verb counted just once) is less pronounced than in English.

The CELEX database (Baayen *et al.*, 1993) lists 3015 German participles. After cleaning out some obvious errors and homophones and choosing the more frequent of different participle forms of one stem, 2992 participles remain. However, German verbs are often formed by modifying other existing verbs with a prefix or separable particle, e.g., the simplex verb *fahren* (drive) occurs in CELEX 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, all composite forms were combined into one simplex form.

For the simulation experiments described below, 20,000 verb tokens were randomly extracted from this corpus 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 applied to 18 verbs).

The structure of the resulting training corpus is shown in table 1.

| | type | | token | |
|---|---|---|---|---|
| Regular | 518 | (78.01%) | 9306 | (46.49%) |
| Irregular | 134 | (20.18%) | 9717 | (48.54%) |
| Mixed | 12 | (1.81%) | 995 | (4.97%) |
| Sum | 664 | (100.00%) | 20018 | (100.00%) |

Table 1: The structure of the training corpus.

## 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. Investigating the precise nature of these deficits can therefore lead to insights into the internal representation of inflectional morphology. Penke *et al.* (1999) 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. They found irregular inflections to be selectively imapaired in six of the subjects, and three showed no significant difference between regular and irregular participles (the remaining two made more irregular errors but their total number of errors was too small to establish a significant difference between regulars and irregulars). Penke *et al.* (1999) concluded that irregular inflection can be selectively imapired in agrammatic aphasia.

## The Network Model

For the simulations described in this paper, a constructivist neural network (CNN) model was developed that 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* for an area of the input space. The problem in building RBF networks is to decide on the number and positions of these receptive fields. The CNN 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 quickprop), and when learning no longer improves the 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 receptive field as similar: the receptive field 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. A similar network has already been successfully used to model the acquisition of the English past tense (Westermann, 1998).

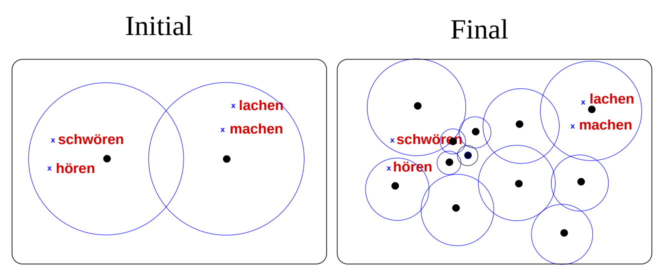


Figure 1: Receptive fields covering the input space at the beginning (left) and the end (right) of learning.

Figure 1 shows a hypothetical start and end state in a two-dimensional input space. While initially only two units 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.

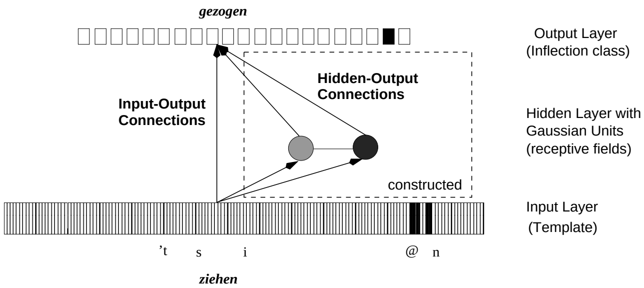


Figure 2: The initial architecture of the network.

## Data

The 664 German verbs were classified according to the way in which their participles are formed, resulting in a total of 22 classes, one of which was the "stem+-*t*" (regular) class, 6 were for mixed verbs, and 15 for irregular verbs.

The verbs were represented phonologically, and each phoneme was encoded 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 so that the endings occurred in the same slots.

The resulting network had 150 input units (three syllables with seven phonemes each represented by seven features, plus one stress-bit per syllable), and 22 output units for the 22 inflection classes.

## 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 (e.g., *holen* → *geholt* vs. *landen* → *gelandet*).

Five CNN models were trained on this corpus with different random initial weight settings. The networks were tested before the insertion of a new hidden unit. An output class was counted as correct when the corresponding unit, but no other unit, had an activation value over 0.7.

## Results

In order to model agrammatic aphasia, the CNN was lesioned in different ways. It was assumed that the removal of weights in the model corresponds to the destruction of neural tissue in the brain by a stroke.

### Localized Lesioning

The output in the CNN model is produced through two sets of connections: the direct connections between the input and the output layer that the network started out with, and the connections from the growing hidden to the output layer. A localized lesioning of these pathways in the CNN resulted in a double dissociation between regular and irregular verbs for four out of the five runs. The further analyses were conducted with these four networks.

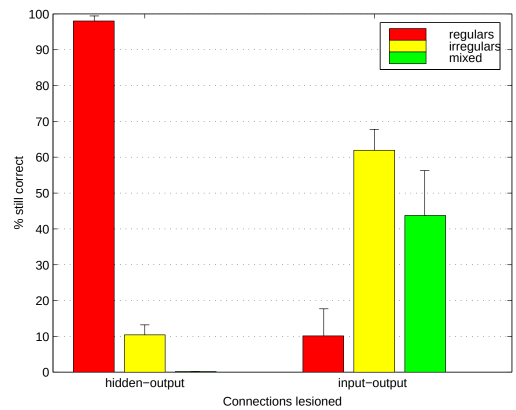


Figure 3: Double dissociation between regular and irregular (and mixed) verbs after lesioning the two pathways in the networks.

Figure 3 shows the results of lesioning the hidden-output (HO) and the direct input-output (IO) connections. Lesioning the HO connections resulted in a marked decrease of the performance of irregular and mixed verbs, with regular inflections remaining nearly fully intact. By contrast, lesioning the IO connections resulted in the opposite profile: performance of regulars was significantly more impaired than that of irregular and mixed verbs. It is important to note that this double dissociation emerged as a result of the structure of the training data together with the constructivist development of the model and was in no way prespecified.

Removing the HO connections in the network thus modeled the basic deficit in the inflection of agrammatic aphasics, namely, the breakdown of irregular and selective sparing of regular participles. Based on this result, the performance of the HO-lesioned CNN models was investigated with respect to the more detailed results reported by Penke *et al.* (1999).

Penke *et al.* (1999) found that 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 (i.e., their regular errors consisted mainly in using a wrong suffix or none at all). Testing the four corresponding CNN models for this behavior showed a good match of the aphasic profiles: the networks over-applied the regular class to 73.7% of all wrong irregulars (aphasics: 63.3%), but only 6.5% of all regular errors were irregularizations (aphasics: 14.3%). The other errors that can be made by the CNN models are no output, or ambiguous output when two (or more) output units are simultaneously activated.

Based on the assumption of two qualitatively distinct processing mechanisms for regular and irregular inflections, Penke *et al.* (1999) predicted and found a frequency effect in the aphasic production of irregulars, but not of regulars: there were significantly more errors for infrequent irregulars than for frequent ones, but no such effect occurred for regulars. When tested on the same verbs as the aphasic subjects, the CNN models equally showed a small frequency effect for irregulars but not for regulars: the error rate for low frequency irregulars (93.3%) was significantly higher than for high frequency irregulars (89.0%) (Wilcoxon, $p = 0.068$), but error rates for regulars did not differ statistically (1.7% for low frequency and 2.4% for high frequency regulars, $p = 0.273$).

Alternatively to a qualitative distinction, regulars and irregulars might represent 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 is "less regular" if it is dissimilar to other regulars but similar to irregulars. The reverse is true for irregulars (see also Daugherty and Seidenberg, 1992).

This assumption is attractive because it integrates mixed verbs which fall between regulars and irregulars in that they combine an irregular stem with the regular ending. Mixed verbs are generally ignored in the DMT because they are hard to consolidate with the proposed qualitative distinction between regulars and irregulars.

A regularity continuum would predicted that "less regular" regulars, being more similar to irregulars, should be more error prone than "very regular" regulars in agrammatic aphasics. Penke *et al.* (1999) analyzed the distribution of verbs with respect to stem vowels and found that for the stem vowel <e>, irregulars outnumber regulars, making regulars with this stem vowel less regular. Therefore, regular verbs with <e> should have a higher error rate because they are similar to irregulars.

This prediction was confirmed in the analysis of the aphasic data: all regular suffixation errors occurred with <e>-stems. While Penke *et al.* (1999) interpreted their results within the framework of a qualitative distinction between regulars and irregulars (allowing grading effects for both mechanisms with the qualitatively distinct verb groups influencing each other), a more plausible interpretation is that of a regularity continuum where a single mechanism underlies the production of both forms.

Testing the CNN model, which is based on such a single mechanism, for this effect yielded the same pattern of results as in the aphasic subjects: when tested on the same verbs, 4 out of 5 of the regular errors were for the stem vowel <e>, indicating that these verbs are treated more like irregulars.

In summary, by lesioning the HO connections in the CNN model, detailed aspects of the performance of agrammatic aphasics on German participle inflections could be modeled. These results comprise both those that have been claimed to be evidence for the dual mechanism theory (double dissociations; frequency effects only for irregulars) and those that contradict the predictions of the dual mechanism theory (regularity continuum effect).

## Global Lesioning

As shown in the previous section, the lesioning of the HO pathway in the CNN model can account for a selective impairment in the inflection of irregular verbs and thus model the performance of agrammatic aphasic subjects. This selective and total lesioning of one pathway might suggest that the processing of regular and irregular verbs is subserved by locally different brain structures that can be selectively affected by a stroke. To establish whether the observed profile could be modeled without this assumption, the effects of globally lesioning the network to different degrees was investigated, without making a distinction between the IO and the HO connections. Over 200 trials, the network was lesioned in 5%-steps by randomly removing weights from both sets of connections.

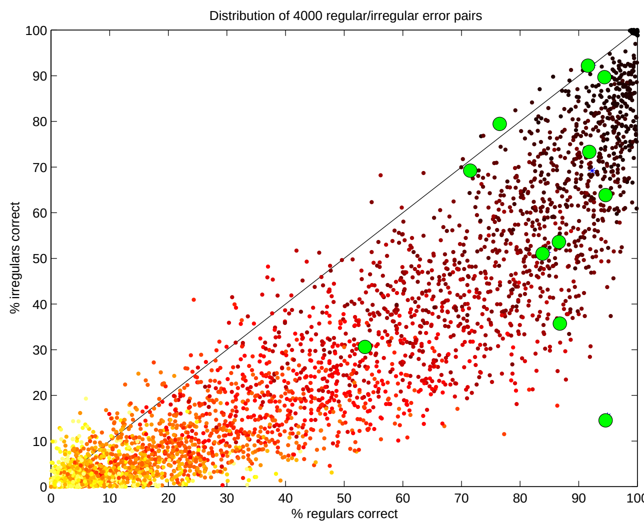


Figure 4: Performance on regulars vs. irregulars for 200 lesioning trials at 20 lesioning steps each (in 5%-steps). Greyscale indicates degree of lesioning (from dark to light). Data for the aphasic subjects are marked by circles.

The result of this global lesioning is shown in figure 4. The 4,000 lesioned networks showed some variety of regular vs. irregular errors, but, like with the aphasic subjects, there was never a selective sparing of irregulars with a breakdown on regular participles (top left of the plot). Instead, in most cases impairment of irregulars was stronger than of regulars (below the diagonal).

The data for the eleven aphasic subjects from (Penke *et al.*, 1999) are also displayed in figure 4. All aphasic data are within the range of performance predicted by the simulations, showing that although there is variability in the performance of agrammatic aphasics, differently lesioned CNNs can 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 inflections did not occur in any of the lesioning trials.

Why does global lesioning in the CNN lead to a profile in which irregular participles are more impaired than regulars? An answer to this question can be found by analyzing the connections in the model. Many of the IO connections are inhibitory, suppressing the activation of the wrong inflection class by other IO connections. This profile is due to the distributed representation of the input: overlapping representations between classes make the inhibition of wrongly activated classes necessary, and with increased lesioning this inhibition is lost, resulting in the activation of wrong output classes for regular and irregular verbs equally. By contrast, the HO connections from one receptive field usually contain only one strongly excitatory weight to the correct output class. Therefore, the HO weights do not tend to activate a wrong output class. This different weight structure can be explained by the localist nature of the receptive fields: due to the constructivist growth process, receptive fields tend to cover only verbs from one class. Therefore, representations for different classes do not overlap and inhibition is not required. An analysis of the distribution of the receptive fields over the verbs showed that they had been preferentially allocated for the difficult-to-learn irregular verbs. Therefore, a partial lesioning of the HO connections affected predominantly irregulars. Taken together, irregulars were impaired by the removal of weights in both the IO and the HO connections, while regulars were affected only by lesioned IO connections. Together, a global lesioning therefore led to a more pronounced breakdown for irregulars than for regulars.

A global lesioning profile in which regular inflections are selectively impaired could only arise from a total lesioning of the IO connections together with no or weak lesioning of the HO connections. Based on the CNN model therefore the prediction is made that a selective impairment of regular inflections in aphasics would be evidence for a locally separate processing of regular and irregular inflections in the brain, whereas the selective impairment of irregulars cannot be taken as evidence for such a separation.

## A Dual-Representation Theory of Verb Inflection

The results described in this paper show that the CNN can account for detailed empirical results from agrammatic aphasic inflectional processing. At the same time, the CNN avoids the problems of the DMT, namely, underspecification and contradiction to some empirical data.

Whereas the DMT proposes two mechanisms operating on a single representation of a verb stem, the CNN develops so that a single mechanism operates on two representations of the verb. Initially, the direct phonological input is used in the IO pathway to produce the output class. For verbs for which the output cannot be learned based on this structural representation alone, the CNN develops through a constructivist process additional representations in the hidden layer. In contrast to the structure-based input representations, these new representations are identity-based and localist: the activation of a hidden unit receptive field only indicates the presence of a certain input, without information about its structure. The CNN is therefore a single mechanism, but dual representation model. This dual representation view sheds a different light on the dissociations between regular and irregular forms. The DMT does not assume that any regular verbs are produced by the irregular mechanism, or vice versa. The common aphasic profile where both regular and irregular cases are partially impaired (albeit to different degrees) is therefore often attributed to performance errors or the unpredictability of aphasic impairment.

A more compelling explanation is offered by the CNN: here, the dissociations that become visible in the lesioning trials do not run clearly along the lines of regulars vs. irregulars. Instead, all verbs for which the inflection class cannot be learned in the direct IO pathway are shifted to the developing hidden layer and the HO pathway. This shift concerns regular, irregular, and mixed verbs, to different degrees. The dissociation between verbs is thus better described as *easy to learn* vs. *difficult to learn*, with the difficult forms relying on the hidden layer, whereas easy forms are produced in the IO pathway alone. This distinction can account better for the data such as mixed verbs, a regularity continuum, or the different aphasic profiles.

But what factors determine whether a form is easy or difficult to learn? The degree of difficulty is determined by several interacting distributional factors that can be derived from the principles of associative learning:

1. Frequency: a frequent transformation is easier to learn than an infrequent one. Therefore, inflection classes with a high summed token frequency will be easier to learn than those that only apply to rare verbs.

2. Class size: a transformation that applies to many different verbs is easier to learn than one that just applies to one verb. Therefore, inflection classes with many members (counted in types) are easier to learn than those confined to only a small group of verbs.

3. Similarity of class members to members of other classes: the inflection class of a verb is easier to learn if other similar verbs share the same class.

4. Ambiguity of inflectional morpheme: an inflection is easier to learn if it applies uniquely to members of its class, i.e., if it does not exist in other context as well. For example, the *-ed* suffix in English is highly indicative of the past tense: an analysis of the CELEX corpus showed that 99.6% of all word types in English that end in *-ed* are past tense forms. By contrast, the German irregular participle ending *-en* is much more ambiguous: it also occurs in verb infinitives (*gehen*, to go), noun plurals (*Wiesen*, meadows), and as part of noun singulars (*Drachen*, kite).

These factors influence each other, and further research will be needed to establish in detail how they interact. Nevertheless they show that the *regular—irregular* distinction is a good first approximation of the *easy—difficult* distinction: the regular inflection, although it does not apply to the most frequent individual verbs, is the single most frequent inflection in both English and German: 57.2% of English past tense tokens and 46.89% of German participle tokens are regular. At the same time, these classes are also the biggest in type size (88.4% and 64.7%, respectively). However, the third point, similarity of class members to members of other classes, does not separate along the lines of regular and irregular verbs: many regular verbs are similar to irregulars which should make them harder to learn in this view. And in fact the regularity continuum that has been shown for aphasics indicates that regulars that are similar to irregulars are more prone to impairment than others, that is, they rely more on storage in the lexicon.

A similar analysis of factors influencing errors in past tense formation has been conducted with school children (Marchman, 1997), where their errors on an elicited past tense production task were determined by frequency, the number of similar sounding stems in the same and in different inflection classes, and the phonological characteristics of the stem and past tense forms.

Taken together, although the dissociations of verbs into easy and difficult corresponds largely to the regular-irregular dissociation, it nevertheless suggests that the regular case is a post-hoc extraction and idealization of the developed structure of the inflectional processing system.

## Discussion

The results presented in this paper suggest a novel account of inflection learning and processing: it is a single mechanism system in which dual representations emerge from a constructivist learning process together with the structure of the environment. The system separates verbs along the lines of easy vs. hard to learn and can thus better explain empirical results that have so far been taken as evidence for the Dual Mechanism Theory. The qualitative distinction between regular and irregular inflections that lies at the core of the DMT, is a projection of formal linguistic analysis onto the human data. Because according to formal linguistics, human language data does not correspond to the abstract "competence" but is instead corrupted as "performance", any data that does not correspond to the predictions of the formal theory (i.e., regulars that behave like irregulars and vice versa) can therefore be attributed to performance. This method makes the DMT hard to falsify based on such data. By contrast, the CNN model is fully specified, and it shows how the actual human data can be modeled without recourse to a competence-performance distinction. Whereas the abstract category of "regularity" remains a good formal description of language structure, the fallacy is in drafting it into service as a *processing* category, as done in the DMT.

A way to test the validity of the CNN model empirically is to abandon the regular/irregular distinction in favour of an easy/hard distinction, by identifying "hard" regulars and "easy" irregulars. Such a distinction should then better predict impairment profiles in agrammatic aphasics and other aspects of dissociations in inflectional systems. More research along these lines will be needed to empirically verify the dual-representation model of verb inflection.

While connectionist, single-mechanism models of inflections have been rejected by proponents of the DMT (e.g. Clahsen, 1999; Pinker, 1997; Marcus *et al.*, 1995), the CNN model presents evidence that such models can account for inflectional processing more successfully than theories that rely on qualitatively distinct processing mechanisms.

**Acknowledgements** The author is now at Sony Computer Science Lab, 6 rue Amyot, 75005 Paris, France. (gert@csl.sony.fr).

## References

Baayen, H., Piepenbrock, R., and van Rijn, H. (1993). *The CELEX Lexical Database*. CD-ROM. Linguistic Data Consortium. University of Pennsylvania, PA.

Baayen, R. H., Dijkstra, T., and Schreuder, R. (1997). Singulars and plurals in Dutch: Evidence for a parallel dual-route model. *Journal of Memory and Language*, **37**(1), 94–117.

Clahsen, H. (1999). Lexical entries and rules of language: A multidisciplinary study of German inflection. *Behavioral and Brain Sciences*, **22**(6), 991–1013.

Daugherty, K. and Seidenberg, M. S. (1992). Rules or connections? The past tense revisited. In *Proceedings of the 14th Annual Conference of the Cognitive Science Society*, pages 259–264, Hillsdale, NJ. Erlbaum.

Marchman, V. A. (1997). Children's productivity in the English past tense: The role of frequency, phonology, and neighborhood structure. *Cognitive Science*, **21**(3), 283–304.

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.

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, editors, *Cognitive Models of Language Acquisiton*. MIT Press, Cambridge, MA.

Penke, M., Janssen, U., and Krause, M. (1999). The representation of inflectional morphology: Evidence from Broca's aphasia. *Brain and Language*, **68**, 225–232.

Pinker, S. (1991). Rules of language. *Science*, **253**, 530–535.

Pinker, S. (1997). Words and rules in the human brain. *Nature*, **387**, 547–548.

Stemberger, J. P. and MacWhinney, B. (1986). Frequency and the lexical storage of regularly inflected forms. *Memory & Cognition*, **14**, 17–26.

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