Unifying syntactic theory and sentence processing difficulty through a connectionist minimalist parser

Sabrina Gerth · Peter beim Graben

Received: 26 May 2009 / Accepted: 31 August 2009 / Published online: 1 October 2009
© Springer Science+Business Media B.V. 2009

Abstract Syntactic theory provides a rich array of representational assumptions about linguistic knowledge and processes. Such detailed and independently motivated constraints on grammatical knowledge ought to play a role in sentence comprehension. However most grammar-based explanations of processing difficulty in the literature have attempted to use grammatical representations and processes per se to explain processing difficulty. They did not take into account that the description of higher cognition in mind and brain encompasses two levels: on the one hand, at the macrolevel, symbolic computation is performed, and on the other hand, at the microlevel, computation is achieved through processes within a dynamical system. One critical question is therefore how linguistic theory and dynamical systems can be unified to provide an explanation for processing effects. Here, we present such a unification for a particular account to syntactic theory: namely a parser for Stabler's Minimalist Grammars, in the framework of Smolensky's Integrated Connectionist/Symbolic architectures. In simulations we demonstrate that the connectionist minimalist parser produces predictions which mirror global empirical findings from psycholinguistic research.

Keywords Computational psycholinguistics · Human sentence processing · Minimalist Grammars · Integrated Connectionist/Symbolic architecture · Fractal tensor product representation

Introduction

Psycholinguistics assesses difficulties in sentence processing by means of several *quantitative measures*. There are *global measures* such as reading times of whole sentences or accuracies in grammaticality judgement tasks which provide metrics for overall language complexity on the one hand (Traxler and Gernsbacher 2006; Gibson 1998), and *online measures* such as fixation durations in eye-tracking experiments or voltage deflections in the event-related brain potential (ERP) paradigm on the other hand (Traxler and Gernsbacher 2006; Osterhout et al. 1994; Frisch et al. 2002). To explain the cognitive computations during sentence comprehension, theoretical and computational linguistics have developed *qualitative symbolic descriptions* for grammatical representations using methods from formal language and automata theory (Hopcroft and Ullman 1979). Over the last decades, Government and Binding theory (GB) has been one of the dominant theoretical tools in this field of research (Chomsky 1981; Haegeman 1994; Staudacher 1990). More recently, alternative approaches such as Optimality Theory (OT) (Prince and Smolensky 1997; Fanselow et al. 1999; Smolensky and Legendre 2006a, b) and the Minimalist Program (MP) have been suggested (Chomsky 1995). In particular Stabler's derivational minimalism (Stabler 1997; Stabler and Keenan 2003) provides a precise and rigorous formal codification of the basic ideas and principles of both, GB and MP. Such Minimalist Grammars have been proven to be *mildly context-sensitive* (Michaelis 2001; Stabler 2004) which makes them appropriate for the symbolic description of natural languages and also for psycholinguistic applications (Stabler 1997; Harkema 2001; Hale 2003a; Hale 2006; Niyogi and Berwick 2005; Gerth 2006). Other well-established formal accounts are e.g. Tree-Adjoining Grammar (TAG) (Joshi et al. 1975;

S. Gerth (✉)
Department of Linguistics, University of Potsdam,
Karl-Liebknecht-Str. 24-25, 14476 Potsdam/Golm, Germany
e-mail: sabrina.gerth@uni-potsdam.de

P. beim Graben
School of Psychology and Clinical Language Sciences,
University of Reading, Reading, Berkshire, UK

Joshi and Schabes 1997) and Head-Driven Phrase Structure Grammar (HPSG) (Pollard and Sag 1994).

The crucial task for computational psycholinguistics is to bridge the gap between qualitative symbolic descriptions at the theoretical side and quantitative results at the experimental side. One attempt to solve this problem is connectionist models for sentence processing. For instance Elman (1995) suggested *simple recurrent neural network* (SRN) architectures for predicting word categories of an input string. Such models have also been studied by Berg (1992), Christiansen and Chater (1999), Tabor et al. (1997), Tabor and Tanenhaus (1999), and more recently by Lawrence et al. (2000) and Farkas and Crocker (2008). However most previous connectionist language processors rely on context-free descriptions which are psycholinguistically rather implausible. Remarkable progress in this respect has been achieved by the *Unification Space Model* of Vosse and Kempen (2000) (see also Hagoort (2003, 2005)) and its most recent successor, *SINUS* (Vosse and Kempen this issue), deploying the TAG approach (Joshi et al. 1975; Joshi and Schabes 1997).

A universal framework for *Dynamic Cognitive Modeling* (beim Graben and Potthast 2009) is offered by Smolensky's *Integrated Connectionist/Symbolic* architectures (ICS) (Smolensky and Legendre 2006a, b; Smolensky 2006). It allows the explicit construction of neural realizations for highly structured mental representations by means of *filler/role decompositions* and *tensor product representations* (cf. Mizraji (1989, 1992) for a related approach). Moreover ICS suggests a dual aspect interpretation: at the macroscopic, symbolic level, cognitive computations are performed by the complex dynamics of distributed activation patterns; at the microscopic, connectionist level, these patterns are generated by deterministic evolution laws governing neural network dynamics (Smolensky and Legendre 2006a, b; Smolensky 2006; beim Graben and Atmanspacher 2009).

It is the purpose of this paper to present a unified global account for syntactic theory and sentence processing difficulty in terms of Minimalist Grammars and Integrated Connectionist/Symbolic architectures. We construct Minimalist Grammars for the lexical material studied in the psycholinguistic literature: (1) for the processing of verbs that are temporarily ambiguous with respect to a direct-object analysis versus a complement clause attachment in English (Frazier 1979); (2) for the processing of case-ambiguous noun phrases in scrambled German main clauses (Frisch et al. 2002). Then we describe a bottom-up parser for Minimalist Grammars that is able to process these sentences yet in a non-incremental way. The state descriptions of the parser are mapped onto ICS neural network architectures by employing filler/role decomposition and a new, hierarchical tensor product representation, which we

shall refer to as the *fractal tensor product representation*. The networks are trained using generalized Hebbian learning (beim Graben and Potthast 2009; Potthast and beim Graben 2009). For visualizing network dynamics through activation state space, an appropriate *observable model* in terms of principal component analysis is constructed (beim Graben et al. 2008a) that allows for comparison of different parsing processes. Finally, we suggest a global complexity measure in terms of temporally integrated principal components for quantitative assessment of processing difficulty.

Our work is a first step towards bridging the gap between symbolic computation using a psycholinguistically motivated formal account and dynamical representation in neural networks, combining the functionalities of established linguistic theories for simulating global sentence processing difficulties.

Method

In this section we construct ICS realizations for a minimalist bottom-up parser that processes sentence examples discussed in the psycholinguistic literature. First the materials will be outlined which reflect two different ambiguities: (1) direct-object versus complement clause attachment in English and (2) case-ambiguous nominal phrases in German.

Materials

English examples

It is well known that the following English sentences from Frazier (1979) elicit a mild garden path effect comparing the words printed in **bold** font:

- (1) The girl knew the answer **immediately**.
- (2) The girl knew the answer **was wrong**.

According to Frazier's minimal attachment principle (Frazier 1979), readers are being garden-pathed in sentence (2) because they interpret the ambiguous noun phrase "the answer" initially as the direct object of the verb "knew", which is the simplest structure. This processing strategy leads to a garden-path effect because "was wrong" cannot be attached to the computed structure and reanalysis becomes inevitable (Bader and Meng 1999; Ferreira and Henderson 1990; Frazier and Rayner 1982; Osterhout et al. 1994). Attaching the complement clause "the answer was wrong" in (2) to the phrase structure tree leads then to larger processing difficulty. In the event-related brain potential, a P600 has been observed for this kind of direct-object versus complement clause attachment ambiguity (Osterhout et al. 1994).

German examples

In contrast to English, the word order in German is relatively free, which offers the opportunity to vary syntactic processing difficulties for the same lexical items by changing their morphological case. The samples consist of subject-object versus object-subject sentences which are well-known in the literature for eliciting a mild garden path effect (Bader 1996; Bader and Meng 1999; Hemforth 2000). They were constructed similar to an event-related brain potentials study by Frisch et al. (2002) as follows:

(3) *Der Detektiv hat die Kommissarin gesehen.*
The detective_{MASCINOM} has the investigator_{FEMIAcc} seen.

‘The detective has seen the investigator.’

(4) *Die Detektivin hat den Kommissar gesehen.*
The detective_{FEMIAmbig} has the investigator_{MASCIACC} seen.

‘The detective has seen the investigator.’

(5) *Den Detektiv hat die Kommissarin gesehen.*
The detective_{MASCIACC} has the investigator_{FEMINOM} seen.

‘The investigator has seen the detective.’

(6) *Die Detektivin hat der Kommissar gesehen.*
The detective_{FEMIAmbig} has the investigator_{MASCINOM} seen.

‘The investigator has seen the detective.’

The sentences (3) and (4) have subject-object order whereas (5) and (6) have object-subject order. Previous work (Weyerts et al. 2002) has shown, that sentence (5) is harder to process than sentence (3) due to the scrambling operation which has to be applied to the object of sentence (5) and leads to higher processing load. A second effect for these syntactic constructions in German is that sentences (4) and (6) contain a case-ambiguous nominal phrase (NP). The disambiguation between subject and object takes place not before the second argument. Bader (1996) and Bader and Meng (1999) found that readers assume that the first NP is a subject which leads to processing difficulties at the second NP. In an event-related brain potentials study Frisch et al. (2002) showed that sentences like (6) lead to a mild garden path effect indicated by a P600. Additionally, Bader and Meng (1999) found that the garden path effect was strongest for sentences involving the scrambling operation which might be due to the fact that both processing difficulties add up in this case. We were able to model both effects—the scrambling operation as well as the disambiguation effect—on a global scale.

Minimalist Grammars

The following section will provide a short introduction into the formalism of Minimalist Grammars of Stabler (1997).

At first the formal definition and the tree building operations will be outlined, followed by an application to the English and German sentences of section “Materials”.

Definition of Minimalist Grammars

Following Stabler (1997), Minimalist Grammars (from here on referred to as MG) consist of a lexicon and structure building operations that are applied to lexical items and trees resulting from such applications. The items in the lexicon consist of syntactic and non-syntactic features (e.g. phonological and semantic features). Syntactic features are basic *categories*, namely *d*: determiner, *v*: verb, *n*: noun, *p*: preposition, *t*: tense (i.e. inflection in GB terminology), and *c*: complementizer. Categories are syntactic *heads* that select other categories as their complements or adjuncts. This is encoded by other syntactic features, called *selectors*: =*d* means “select determiner”, =*n* “select noun”, and so on. Furthermore, there are *licensors* (+*CASE*, +*WH*, +*FOCUS* etc.) and *licensees* (−*case*, −*wh*, −*focus* etc.). The licensor +*CASE*, e.g., assigns case to another lexical item that bears its counterpart −*case*. The tree structure building operations of MG are *merge* and *move*. They use the syntactic features to generate well-formed phrase structure trees.

Minimalist trees are either simple or complex. A simple tree is any lexical item. A complex tree is a binary tree consisting of lexical items as leaves and *projection indicators* “>” or “<” as further node labels. Each tree has a unique *head*: a simple tree is its own head; whereas the head of a complex tree is found by following the path indicated by “>” or “<” through the tree, beginning from the root. The head of the tree projects over all other leaves. However, every other leaf is always the head of a particular subtree, which is the *maximal projection* of that leaf. In this way, MG define a precise formalization of the basic ideas of Chomsky’s Minimalist Program, revising and augmenting Government and Binding theory. In order to simplify the notation, we can also speak about the (unique) feature of a tree, which is the first syntactic feature in the list of the tree’s head.

The merge operation appends simple trees (lexical items) or complex trees by using categories and selectors as depicted in Fig. 1.

The verb “know” in Fig. 1 (category *v*) has the selector =*d* as its feature while the determiner phrase (DP) “the answer” has as feature the category *d*. Thus, the verb selects the DP as its complement, yielding the merged tree for the phrase “know the answer”. The symbol “<” now indicates the projection relation: the verb immediately projects over the DP. Hence, the verb is the head of the complex tree describing a verb phrase (VP in GB notation). After accomplishing merge, the features of both trees are

Fig. 1 Merge operation for tree concatenation

Fig. 2 Move operation for tree transformation

deleted. The merge operation corresponds to the X-bar module in GB theory.

Figure 2 illustrates the move operation which transforms a tree into another tree for rearranging sentence arguments (Bader and Meng 1999). This operation is triggered by a licenser and a corresponding licensee which determines the maximal projection to be moved. The subtree possessing the licensee as its feature is extracted from the tree and merged with the remaining tree to its left, the former head also becomes the head of the transformed tree. After accomplishing the operation, licenser and licensee are deleted. Furthermore, the moved leaf and the trace λ of the extracted maximal projection can be co-indexed for illustrative purpose (but note, that indices and traces are not part of MG, they belong to syntactic metalanguage). The move operation corresponds to the modules for government (in case of case assignment) and “move α ” in GB.

Minimalist Grammars for the English sentences

Next, we explicitly construct Minimalist Grammars for the English sentence material of section “English examples”. Figure 3 shows the minimalist lexicon for sentence (1).

The first entry is a phonetically empty categorizer (category **c**) which takes a tense phrase (selector $=t$, i.e. an IP) as its complement. The second entry is a determiner phrase “the girl” (category **d**) which requires nominative case (licensee $-\text{nom}$).¹ The third entry in the lexicon describes the past tense inflection “-ed” which has category **t**. The feature of this item is $V=$, indicating that it firstly takes a verb as its complement and secondly that the phonetic features of the selected verb are prefixed to the phonetic features of the selecting head. In this way, MG describes *head movement with left-adjunction* (Stabler 1997). After that, a determiner phrase, here “the girl”, is selected ($=d$) and attached to its specifier position by a mechanism called shell formation (Stabler 1997) which occurs in our case when the verb phrase combines with further arguments on its left. As inflection governs nominative case in GB, it has licenser $+\text{NOM}$. As the fourth item, we have a determiner phrase “the answer” again which requires accusative case by its licensee $-\text{acc}$. This is achieved by the licenser $+\text{ACC}$ of the fifth item (**v**) “know” in accordance with GB. The verb takes a direct object ($=d$). Finally, the adverb “immediately” (**d**) serves as the verb’s modifier.

Only two movements take place to construct the well-formed grammatical tree: (1) the object “the answer” has to be moved into its object position (indexed with “ i ”); (2) “the girl” is shifted into the subject position of the tree (indexed with “ k ”). The overall syntactic process is easy to accomplish and leads to no processing difficulties (see appendix).

Figure 4 illustrates the minimalist lexicon of the English sentence (2) containing the complement clause.

There are two essential differences in comparison to Fig. 3. First, the verb “know” is encoded as taking a clause ($=c$) as its complement, instead of a direct object. Second, this clausal complement is introduced by a phonetically empty complementizer (indicated by the empty word ε) which merges with an IP ($=t$). Note, that an unambiguous reading can be easily obtained by replacing ε with “that”.

The same move operations like in sentence (1) have to be applied. In contrast to Fig. 3 the complement clause “the answer was wrong” is appended to the matrix clause by a complement phrase represented by the lexical item [$=t$; **c**; ε] which becomes empty after all its features ($=t$ and **c**) are checked (indicated by ε). This accounts for higher parsing costs for constructing the syntactic structure of a sentence containing a complement clause compared to a sentence with a nominal phrase (see appendix).

¹ Note that “the girl” is already a phrase that could have been obtained by merging “the” (**d**) and “girl” (**n**) together. We have omitted this step for the sake of simplicity.

Fig. 3 Minimalist lexicon of sentence (1) **Fig. 4** Minimalist lexicon of sentence (2)

Minimalist Grammars for the German sentences

Until now, the present work is one of the first studies which uses the MG formalism for German, so far it has been mostly applied to English (Stabler 1997; Harkema 2001; Hale 2003b; Niyogi and Berwick 2005; Hale 2006). In order to use MG for a language with relatively free word order, we introduce a new pair of features, +SCRAMBLE as a licenser and –scramble as a licensee into the formalism. These scrambling features expand the movement operation, thereby accounting for the possibility to rearrange arguments of the sentence signaled by the morphological case. Our approach is closely related to another suggestion by Frey and Gärtner (2002) who argue that scrambling (and also adjunction) have to be incorporated as *asymmetric operations* into MG. They distinguish a *scramble-licsee*, $\sim x$, from the conventional *move-licsee*, $-x$, in such a way that the corresponding licenser is not canceled upon feature-checking during scrambling, hence allowing for recursive scrambling. However, recursion is not at issue in our Minimalist Grammars modeling the German example sentences (3)–(6). Therefore we refrain from this complication by treating scrambling like conventional movement.

Figure 5 shows the lexicon of the subject-object sentence (3). Each lexical item contains syntactic features to trigger the merge and move operation. We adopt the classical Government and Binding theory here which states that subject and object have to be moved into their appropriate specifier positions (Haegeman 1994). The

lexicon for sentence (4) is obtained by exchanging *der Detektiv* “the detective_{MASCINOM}” with *die Detektivin* “the detective_{MASCIAMBIG}” and *die Kommissarin* “the investigator_{FEMIAMBIG}” with *den Kommissar* “the detective_{MASCIACC}” in the phonetic features, respectively.

Figure 6 illustrates the lexicon for the object-subject sentence (5). For this structure the scrambling features had to be introduced to move the object argument *die Detektivin* “the detective_{FEMIAMBIG}” from the lower right position upwards to the left-hand side of the verb *gesehen* “seen” and further upwards to the front of the sentence to assure the correct word for the object-subject sentence while maintaining the functional position. Furthermore another selector $T=$ indicates head movement with left adjunction, again. It is responsible for the movement of phonetic material to prefix the phonetic material of the selecting head (Stabler 1997), illustrated by the parenthesis around *hat* “has”: “/hat/” represents the phonetic material; “(hat)” indicates the interpreted semantic features.

Correspondingly, the lexicon for sentence (6) is obtained by exchanging *den Detektiv* “the detective_{MASCIACC}” with *die Detektivin* “the detective_{FEMIAMBIG}” and *die Kommissarin* “the investigator_{FEMIAMBIG}” with *der Kommissar* “the detective_{MASCINOM}” in the phonetic features from Fig. 6, respectively.

Minimalist parsing

This section outlines the algorithm of the minimalist parser developed by Gerth (2006). The parser takes as input the whole sentence divided into a sequence of tokens like:

the girl, know, –ed, the answer, immediately.

These tokens are used to retrieve the items from the minimalist lexicon Fig. 3, yielding the enumeration

$$S_0 = \{L_c, L_{\text{the girl}}, L_{\text{know}}, L_{-ed}, L_{\text{the answer}}, L_{\text{immediately}}\},$$

where the term $L_{\text{the girl}}$ denotes the MG feature array for the item “the girl”. The list S_0 is the initial *state description* of the MG parser showing all necessary lexical entries for the syntactic structure to be built. The parser operates on its state descriptions non-incrementally pursuing a bottom-up strategy within two nested loops: one for the domain of merge and the other for the domain of move. In a first loop each tree in the state description is compared with every other tree in order to check, whether they can be merged. In

Fig. 5 Minimalist lexicon of subject-object sentence (3) **Fig. 6** Minimalist lexicon of object-subject sentence (5)

this case, the merge operation is performed, whereupon the original trees are deleted from the state description and the merged tree is put onto its last position. In a second loop, every single tree is checked, whether it belongs to the domain of the move operation, in which case, move is applied, after that, the original tree is replaced by the transformed tree in the state description.

Thus, the initial state description S_0 is extended to S_1 where two (simple) trees in S_0 have been merged. Note that the rest of the lexical entries are just passed to the next state description without any change. Correspondingly, S_1 is succeeded by S_2 when either merge or move could have been successfully applied. The resulting sequence of state descriptions $S_0, S_1, S_2, \dots$ describes the parsing process. Since every state description is an enumeration of (simple or complex) minimalist trees, it could also be regarded as a *forest* in graph theoretic terms. Therefore, the minimalist operations merge and move can be extended to functions that map forests onto forests. This yields an equivalent description of the minimalist bottom-up parser.

Finally, the parser terminates when no further operation can be applied and only one tree remains in the state description.

An example parse of sentence (1) The lexicon for sentence (1) comprising the initial state S_0 was shown in Fig. 3.

At first, the two lexical items of “know” and “the answer” are merged, triggered by $=d$ and d which are deleted after merge has been applied (Fig. 7).

In the second step the lexical item of “immediately” is merged with the current tree (Fig. 8).

The move operation triggered by $-acc$ and $+ACC$ is accomplished in the third step which leads to a movement of “the answer” upwards in the tree leaving a trace indicated by λ behind. The involved sentence items are indexed with i (Fig. 9).

In step 4 the lexical item “-ed” is merged with the tree triggered by $V=$ and v . The head movement with left-adjunction results in a combination of “know” and “-ed”

Fig. 7 Step 1: merge **Fig. 8** Step 2: merge

leading to the inflected form “/knew/” prefixing its semantic features “(know)” (Fig. 10).

The item entry “the girl” is merged with the current tree (Fig. 11).

In step 6 move is applied to the lexical item “the girl” which leaves a trace indicated by λ_k behind (Fig. 12).

In the last parse step the lexical item for c is merged to the tree which leads to a grammatical minimalist tree with the only unchecked feature c as the head of the tree (indicating a CP) (Fig. 13).

Fractal tensor product representation

In order to unify symbolic and connectionist approaches in a global language processing model, the outputs of the minimalist parser have to be represented by trajectories in

Fig. 9 Step 3: move **Fig. 10** Step 4: merge **Fig. 11** Step 5: merge

suitable activation vector spaces. In particular this means that the state description S_t of the minimalist parser at processing time step t is mapped onto a training vector $\mathbf{u}(t) \in \mathbb{R}^n$ for an implementation in a neural network of n units. For achieving this we employ a hierarchy of tensor product representations which rely on a filler/role decomposition beforehand (Dolan and Smolensky 1989; Smolensky 1990; Smolensky and Legendre 2006a; Smolensky 2006; beim Graben et al. 2008a, b; beim Graben and Potthast 2009).

Fig. 12 Step 6: move

As we are interested only in syntax processing here, we firstly discard all non-syntactic features (i.e. phonological and semantic features) of the minimalist lexicons for the sake of simplicity. Then, we regard all remaining syntactic features and in addition the “empty feature” ε and both “head pointers” “>” and “<” as fillers f_i . We use two scalar Gödel encodings (beim Graben and Potthast 2009) for these fillers by integer numbers $g(f_i)$ for the English and German material respectively. The particular Gödel encoding used for the English sentences (1) and (2) in the present study is shown in Table 1.

Complementary the Gödel encoding used for the German sentences (3)–(6) is shown in Table 2.

Given the encodings of the fillers, the roles, which are the positions of the fillers in the feature list of a lexical entry, are encoded by fractional powers N^{-p} of the total number of fillers, which is $N_{eng} = 15$ for English and $N_{ger} = 17$ for German, when p denotes the p th list position. A complete lexical entry L is then represented by the sum of (tensor) products of Gödel numbers for the fillers and fractions for the list roles. Thus, the lexical entry for “-ed” in Fig. 3.

$$L_{-ed} = \begin{bmatrix} v = \\ = d \\ +NOM \\ t \\ -ed \end{bmatrix},$$

becomes represented by 15-adic rational number

$$g(L_{-ed}) = 7 \times 15^{-1} + 4 \times 15^{-2} + 11 \times 15^{-3} + 8 \times 15^{-4} = 0.4879. \quad (7)$$

In a second step, for minimalist trees, which are labeled binary trees with root labels that are either “>” or “<” and leaf labels that are lexical entries, we introduce three role positions

Fig. 13 Step 7: merge **Table 1** Gödel encoding for English minimalist lexicons

Filler f_i	Gödel code $g(f_i)$
ε	0
$>$	1
$<$	2
d	3
=d	4
v	5
=v	6
V=	7
t	8
=t	9
c	10
+NOM	11
-nom	12
+ACC	13
-acc	14

Consider the tree

Its tensor product representation is given through

$$\begin{aligned}
 &g(<) \otimes \mathbf{r}_1 \oplus g(L_l) \otimes \mathbf{r}_2 \oplus g(L_r) \otimes \mathbf{r}_3 \\
 &= 2 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + (4 \times 15^{-1} + 13 \times 15^{-2} + 5 \times 15^{-3}) \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \\
 &\quad + 14 \times 15^{-1} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 0.3259 \\ 0.9333 \end{pmatrix}, \quad (10)
 \end{aligned}$$

where L_l and L_r denote the feature arrays at the left and right leaf, respectively.

Moreover, complex trees are represented by Kronecker tensor products as outlined by beim Graben et al. (2008a) and beim Graben and Potthast (2009). Shortly a Kronecker product is an outer vector product of two vectors $\mathbf{f}_i \otimes \mathbf{r}_k$ (in our case a filler vector $\mathbf{f}_i$ with dimension n and a role vector $\mathbf{r}_k$ with dimension m) which results in an

$$r_1 \equiv \text{PARENT}, \quad r_2 \equiv \text{LEFTCHILD}, \quad r_3 \equiv \text{RIGHTCHILD}, \quad (8)$$

following Gerth (2006), beim Graben et al. (2008a) and beim Graben and Potthast (2009). By representing these roles as the canonical basis vectors in three-dimensional space,

$$r_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad r_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad r_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad (9)$$

tensor product representations for filler/role bindings of trees are obtained in the following way.

Table 2 Gödel encoding for German minimalist lexicons

Filler f_i	Gödel code $g(f_i)$
ε	0
$>$	1
$<$	2
d	3
=d	4
v	5
=v	6
t	7
=t	8
T=	9
c	10
+NOM	11
–nom	12
+ACC	13
–acc	14
+SCRAMBLE	15
–scramble	16

$n \times m$ -dimensional vector. The manner in which Gödel encoding and vectorial representation are combined in this construction implies a fractal structure in vector space (Siegelmann and Sontag 1995; Tabor 2000, 2003; beim Graben and Potthast 2009). Therefore, we refer to this combination as to the *fractal tensor product representation*.

In a final step, we have to construct the tensor product representations for the state descriptions S_t of the minimalist parser. Symbolically, this is an enumeration, or likewise a forest, of minimalist trees, on which the extended merge and move operations act according to a bottom-up strategy. In a first attempt, we tried to introduce one role for each position that a tree could occupy in this enumeration. Then a tensor product representation of a state description would be obtained by recursively binding minimalist trees as complex fillers to those roles. Unfortunately this leads to an explosion of vector space dimensions as a result of recursive vector multiplication. Therefore we refrained from that venture by employing an alternative encoding technique: The tensor product representations of all trees in a current state description are linearly superimposed (element-wise addition of vector entries) in a suitable embedding space (beim Graben et al. 2008a) which will be further outlined in section “Results”.

Neural network simulations

The minimalist parser described in section “Minimalist parsing” is an algorithm that takes one state description S_t at processing step t as input and generates an output S_{t+1} at time $t + 1$ by applying either merge or move to the

elements of S_t . Correspondingly, the fractal tensor product representation $\mathbf{u}(t) \in \mathbb{R}^n$ of the state description S_t constructed in section “Fractal tensor product representation” has to be mapped onto the representation $\mathbf{u}(t + 1) \in \mathbb{R}^n$ of its successor S_{t+1} by the state space representation of the parser. Hence, the parser becomes represented by a function $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}^n$ such that

$$\Phi(\mathbf{u}(t)) = \mathbf{u}(t + 1) \quad (11)$$

for every admissible time t . The desired map Φ can be straightforwardly implemented by an autoassociative neural network.

Tikhonov–Hebbian learning

Thus, we use the fractal tensor product representations of the state descriptions of the minimalist parser as training patterns for our neural network simulation. We employ a fully recurrent autoassociative Hopfield network (see Fig. 14) with continuous n -dimensional state space (Hopfield 1982; Hertz et al. 1991), and described by a synchronous updating rule resulting into the time-discrete dynamical evolution law

$$u_i(t + 1) = \sum_{j=1}^n w_{ij} f(u_j(t)). \quad (12)$$

Here, $u_i(t)$ denotes the activation of the i th neuron (out of n neurons) at time t , w_{ij} the weight of the synaptic connection between neuron j and neuron i and

$$f(u) = \frac{1}{1 + \exp[-\beta(u - \theta)]} \quad (13)$$

the logistic activation function with gain $\beta > 0$ and threshold θ .

Equation 12 can be written as a compact matrix equation

$$\mathbf{U} = \mathbf{W} \cdot \mathbf{V}, \quad (14)$$

where

$$\mathbf{W} = (w_{ij})_{\substack{1 \leq i \leq n \\ 1 \leq j \leq n}}$$

is the synaptic weight matrix and

$$\mathbf{V} = (f(u_i(t)))_{\substack{1 \leq i \leq n \\ 1 \leq t \leq T}}$$

denotes the nonlinearly transformed activation vectors

$$\mathbf{U} = (u_i(t))_{\substack{1 \leq i \leq n \\ 1 \leq t \leq T}}$$

for all times $1 \leq t \leq T$. The columns of the matrix $\mathbf{U}$ are given by the successive parsing states $\mathbf{u}(t)$ where t denotes the parse step and T is the actual duration of the overall parse.

Training the neural network corresponds then to solving an inverse problem described by Eq. 14, where the

Fig. 14 Sketch of a fully recurrent autoassociative neural network. Circles denote neurons and arrows their synaptic connections. Every neuron is connected to every other neuron. (Self-connections are omitted in this Figure)

unknown weight matrix $\mathbf{W}$ has to be determined from the given training patterns in $\mathbf{U}$. Equation 5 is strictly solvable only if matrix $\mathbf{V}$ is not singular, i.e. if $\mathbf{V}$ has an inverse $\mathbf{V}^{-1}$. In general, this is not to be expected. However, if the columns of $\mathbf{V}$ are linearly independent, $\mathbf{V}$ possesses a Moore–Penrose pseudoinverse

$$\mathbf{V}^\dagger = (\mathbf{V}^T \cdot \mathbf{V})^{-1} \cdot \mathbf{V}^T \quad (15)$$

that is often employed in Hebbian learning algorithms (Hertz et al. 1991).

Beim Graben and Potthast (2009) and Potthast and beim Graben (2009) have recently delivered an even more general solution for training neural networks in terms of Tikhonov regularization theory. They observed that a *regularized weight matrix* $\mathbf{W}_\alpha$ is given by a generalized *Tikhonov–Hebbian learning rule*

$$\mathbf{W}_\alpha = \mathbf{U} \cdot \mathbf{R}_\alpha \quad (16)$$

with Tikhonov-regularized pseudoinverse

$$\mathbf{R}_\alpha = (\alpha \mathbf{1} + \mathbf{V}^T \cdot \mathbf{V})^{-1} \cdot \mathbf{V}^T. \quad (17)$$

Here, the *regularization parameter* $\alpha \geq 0$ stabilizes the ill-posedness of the inverse problem by cushioning the singularities in $\mathbf{V}$ (beim Graben and Potthast 2009; Potthast and beim Graben 2009).

For training the connectionist minimalist parser via Tikhonov–Hebbian learning, we attempt the following scenario: every single parse p ($p = 1, \dots, 6$) of the sentences (1)–(6) is trained by one Hopfield network separately. Thus, we generated five connectionist minimalist parsers for the sentences (1) and (3)–(6), but not for (2) where the dimension of the required embedding space was too large (see Table 3 for details). For training we used training parameters $\beta = 10$ and $\theta = 0.3$. Interestingly, regularization was not necessary for that task; setting

Table 3 Embedding space dimensions for sentences (1) and (2)

Sentence	Embedding dimension n
(1) The girl knew the answer immediately.	6,561
(2) The girl knew the answer was wrong.	59,049

$\alpha = 0$ leads to the standard Hebb rule with Moore–Penrose pseudoinverse Eq. 15 (Hertz et al. 1991).

Observable models

The neural activation spaces obtained as embedding spaces from the fractal tensor product representation are very high dimensional (see Tables 3, 4). In order to visualize network dynamics, a method for data compression is required. This can be achieved by so-called *observable models* for the network’s dynamics. An observable is a number associated to a particular state of a dynamical system that can be measured by an appropriate measuring device. Examples for observables in cognitive neurodynamics are electroencephalogram (EEG), magnetoencephalogram (MEG) or functional magnetic resonance imaging (fMRI) (Freeman 2007; beim Graben et al. 2009). Formally, an observable is defined as a real-valued function

$$\varphi : \begin{cases} X & \rightarrow \mathbb{R} \\ \mathbf{u} & \mapsto s = \varphi(\mathbf{u}) \end{cases} \quad (18)$$

from a state space $X \in \mathbb{R}^n$ onto the real numbers such that $s = \varphi(\mathbf{u})$ is the measurement value of φ when the system is in state $\mathbf{u} \in X$. Taking a few number of observables, $\varphi_k, k = 1, \dots, K$ yields again a vectorial representation of the system in one *observable space* $Y = \varphi(X)$. The index k could be identified, e.g., with the k th recording electrode of the EEG or with the k th voxel in the fMRI.

In our connectionist minimalist parser, state space trajectories are the column vectors of the six particular training patterns $\mathbf{U}_p$ for $p = 1, \dots, 6$, or their dynamically simulated replicas, respectively. A common choice for data compression in multivariate statistics is the principal component analysis (PCA), which has been used as an observable model by beim Graben et al. (2008a)

Table 4 Embedding space dimensions for sentences (3)–(6)

Sentence	Embedding dimension n
(3) Der Detektiv hat die Kommissarin gesehen.	2,187
(4) Die Detektivin hat den Kommissar gesehen.	2,187
(5) Den Detektiv hat die Kommissarin gesehen.	6,561
(6) Die Detektivin hat der Kommissar gesehen.	2,187

previously. Therefore, here we pursue this approach further in the following way: for each minimalist parse p , the distribution of points belonging to the state space trajectory U_p is firstly standardized, resulting into a transformed distribution U_p^z (where the superscript z indicates z -transformation) with zero mean and unit variance. Then the columns of U_p^z are subjected to PCA, such that the greatest variance in U_p^z has the direction of the first principal component, the second greatest variance has the direction of the second principal component and so on. Our observable spaces Y_p are then spanned by the first, $y_1 = \text{PC\#1}$, and second, $y_2 = \text{PC\#2}$, principal components, respectively. For visualization in section “Phase portraits”, we overlap observable spaces Y_1 and Y_2 for the parses of the English sentences (1) and (2) and observable spaces Y_3 to Y_6 for the parses of the German sentences (3)–(6) in order to get *phase portraits* of these parses.

In a last step, we propose an observable for global processing difficulty. As the first principal component y_1 accounts for most of the variance in the data, thereby reflecting the volume of state space that is explored by the trajectories of the connectionist minimalist parser during syntactic language processing, we integrate y_1 over the temporal evolution of the system,

$$G = \sum_{t=1}^T |y_1(t)| \quad (19)$$

where the processing time t assumes values between $t = 1$ for the initial condition and $t = T$ for the final state of the parse.

Note that other useful observables could be for example Harmony (Smolensky 1986; Legendre et al. 1990a; Smolensky and Legendre 2006a; Smolensky 2006) or the change in global activation.

Results

In this section, we present the results from the fractal tensor product representation and subsequent neural network simulations, where firstly trajectories in neural activation space are visualized by phase portraits of the first two principal components (section “Phase portraits”). Secondly, we present the results for our global processing measure in section “Global analysis”, the first principal component integrated over time, as explained in section “Tikhonov–Hebbian learning”. Thirdly, we provide a correlation analysis between the first principal component and the number of tree nodes in the respective minimalist state descriptions to face a possible objection: that the first principal component of the state space representation might result into mere “node counting”.

Phase portraits

English sentences

In order to generate training patterns for neural network simulation, minimalist parses of the two English sentences (1) and (2) have been represented in embedding space by the fractal tensor product representation. In Table 3 we display the resulting dimensions of those vector spaces.

Comparing state space dimensions from Table 3 with the corresponding ones obtained from a more localist tensor product representation for context-free GB X-bar trees (beim Graben et al. 2008a), which lead to a total of 229,376 dimensions, it becomes obvious that fractal tensor product representations yield a significant reduction of embedding space dimensionality. However this reduction was not sufficient for training the Hopfield network with the clausal complement data (2). We are therefore only able to present results from the fractal tensor product representation for this example and no results of the neural network simulation. There were no differences between training patterns and simulated network dynamics for sentence (1).

Figure 15 shows the phase portraits spanned by the first two principal components for the English sentences. The start of the trajectories is indicated by the last words of the sentence: sentence (1) starts with coordinates (−53.54, 0.0415) and sentence (2) with coordinates (−160.376, 0.0140). The parses are initialized with different conditions in the state space representation according to the different minimalist lexicons. The nonlinear temporal evolution of trajectories through neural activation space is clearly visible.

Fig. 15 Phase portrait spanned by the first and second principal components, PC#1 and PC#2, of the fractal tensor product representation for minimalist grammar processing of English. Sentence (1): solid, sentence (2): dashed

German sentences

Minimalist parses of the four German sentences (3)–(6) have also been represented in embedding space. In Table 4 resulting dimensions of those vector spaces are presented.

We successfully trained four different Hopfield networks with the parses of the four German sentences (3)–(6) separately with the same parameter settings as for the English example (1). Again, regularization was not necessary for training.

Figure 16 shows phase portraits for the sentences (3) and (5), as well as (4) and (6). The trajectories for the subject-object sentences are exactly the same due to the same syntactic operations that have been applied.

Fig. 16 Phase portraits spanned by the first and second principal component, PC#1 and PC#2, of the fractal tensor product representation for minimalist grammar processing of German. **(a)** Scrambled construction. Subject-object sentence (3): solid, object-subject sentence (5): dashed. **(b)** Garden path construction. Subject-object sentence (4): solid, object-subject sentence (6): dashed

In Figure 16a processing of sentence (5) starts at coordinate (23.3727, -0.0386) while that of the subject-object sentence (3) begins at coordinate (-24.0083, 0.00746) representing the different initial conditions of syntactic structures. Both trajectories explore the phase space non-linearly and settle down in the parser's final states. As Fig. 16b shows, the sentences (4) and (6) start with nearly equal initial conditions in coordinates (-24.0083, 0.0746) and (-23.9778, 0.0762) respectively. They proceed equally for the first step and diverge afterwards. Finally they settle down in the final states, again.

Global analysis

The results of our global processing measure G between the sentences of the two languages as defined in Eq. 19 are shown in Table 5 and in Fig. 17 as a bar plot. Differences in the heights of the bars can be interpreted as reflecting the difference in syntactic operations that are inevitable to build the grammatical well-formed structure of the corresponding sentence.

As expected, the causal complement continuation in (2) is more difficult to process than the direct object continuation in (1). The sentences (3) and (4) exhibit exactly the same global processing costs G as there is no difference in building the syntactic structures of the subject-object sentences. The low G values can be attributed to the canonical

Table 5 Global processing measure G for the sentences (1)–(6)

Legend	Sentence	G
(1) do	The girl knew the answer immediately.	171.70
(2) cc	The girl knew the answer was wrong.	598.32
(3) svo	<i>Der Detektiv hat die Kommissarin gesehen.</i>	101.23
(4) svo	<i>Die Detektivin hat den Kommissar gesehen.</i>	101.23
(5) ovs	<i>Den Detektiv hat die Kommissarin gesehen.</i>	170.01
(6) ovs	<i>Die Detektivin hat der Kommissar gesehen.</i>	108.04

Fig. 17 Bar plots of the global processing measure G for the sentences (1)–(6)

Fig. 18 First principal component, PC#1, plotted against number of tree nodes. (1): blue square, (2): magenta cross, (3) and (4): black circle, (5): red plus, (6): green diamond (Color figure online)

subject preference strategy. Interesting, but not surprising is the fact that the scrambled sentence (5) results in a remarkably higher G value than the garden-path sentence (6). Furthermore, there is a slightly higher value of G for the garden-path sentence (6) in comparison to its control condition (4).

To visualize the fact that the fractal tensor product representation is not correlating with the overall number of nodes in the trees of the parser's state description, we show in Fig. 18 the values of the first principal component plotted against the corresponding number of nodes in the trees. In particular, we calculated the number of tree nodes of each parse step separately.²

As can be seen the points are not situated on a straight line which argues against a correlation of principal component and tree nodes. The correlation coefficient for PC#1 against the tree nodes is $r = -0.15$ which reflects a very weak correlation and argues further against a mere complexity measure of increasing nodes in the syntax tree.

Discussion

We have suggested a unifying Integrated Connectionist/Symbolic (ICS) account for global effects in minimalist

sentence processing. Our approach is able to capture the well-known psycholinguistic differences between ambiguous verbal subcategorization frames in English as well as case ambiguity in object before subject sentences in German. Symbolic computations are represented by trajectories in high-dimensional activation spaces of neural networks through fractal tensor product representations that can be visualized by appropriately chosen observable models. We have pointed out a global processing measure based on temporally integrated observables that successfully accounts for sentence processing difficulties. Modeling sentence processing through the combination of the minimalist grammar formalism and a dynamical system combines the functionalities of established linguistic theories and further accounts for the two levels of description of higher cognition in the brain and takes a step into a new perspective of achieving an abstract representation of language processes.

Though one crucial problem of Minimalist Grammars is that they do not simulate human language processing incrementally. Previous work by Stabler (2000) describes a Cocke–Younger–Kasami-like (CYK) algorithm for parsing Minimalist Grammars by defining a set of operations on strings of features that are arranged as chains (instead of phrase structure trees). Work by Harkema (2001) defines a Minimalist Grammars recognizer that works like an Early parser. So far none of the approaches could meet the claim of incrementality in Minimalist Grammars. Therefore, incremental minimalist parsers pursuing either left-corner processing strategies or employing parallel processing techniques would be psycholinguistically more plausible.

Representing such architectures in the ICS framework could also better account for limited cognitive resources, e.g. by restricting the available state space dimensionality as a model of working memory. ICS also provides suitable mechanisms such as graceful saturation that could be realized by state contraction in a neural network (Smolensky 1990; Smolensky 2006; Smolensky and Legendre 2006a). Finally, other observable models such as energy functions or network harmony (Smolensky 1986; Legendre et al. 1990a, b; Smolensky and Legendre 2006a; Smolensky 2006) could be related to both quantitatively measured processing difficulty in experimental paradigms on the one hand and to harmonic grammar (Legendre et al. 1990a, b; Smolensky and Legendre 2006a; Smolensky 2006) for the qualitative symbolic account of cognitive theory on the other hand. Although a lot of things have been looked into, this paper only claims to provide a proof of concept of how to integrate a particular grammar formalism within a dynamical system to model empirical phenomena known in psycholinguistic theory.

² Each lexical item corresponds to one node, further a root node with two daughters consists of three nodes in total (parent, left daughter, right daughter). A merge operation adds one node, while move increases the node count by two.

Acknowledgements We would like to thank Shravan Vasishth, Whitney Tabor, Titus von der Malsburg, Hans-Martin Gärtner and Antje Sauer mann for helpful and inspiring discussions concerning this work.

Appendix

In this appendix we present the minimalist parses of all example sentences from section “Materials”

English examples

The girl knew the answer immediately

$$\begin{aligned} & \begin{bmatrix} =t \\ c \end{bmatrix}, \begin{bmatrix} d \\ -nom \\ the\ girl \end{bmatrix}, \begin{bmatrix} V= \\ =d \\ +NOM \\ t \\ -ed \end{bmatrix}, \\ & \begin{bmatrix} d \\ immediately \end{bmatrix}, \begin{bmatrix} =d \\ =d \\ +ACC \\ v \\ know \end{bmatrix}, \begin{bmatrix} d \\ -acc \\ the\ answer \end{bmatrix} \end{aligned}$$

This example is outlined in section “Minimalist parsing”.

The girl knew the answer was wrong. (complement clause)

$$\begin{aligned} & \begin{bmatrix} =t \\ c \end{bmatrix}, \begin{bmatrix} d \\ -nom \\ the\ girl \end{bmatrix}, \begin{bmatrix} V= \\ =d \\ +NOM \\ t \\ -ed \end{bmatrix}, \begin{bmatrix} =c \\ v \\ know \end{bmatrix}, \begin{bmatrix} =t \\ c \\ \epsilon \end{bmatrix}, \\ & \begin{bmatrix} d \\ -nom \\ the\ answer \end{bmatrix}, \begin{bmatrix} =d \\ =d \\ +NOM \\ t \\ was \end{bmatrix}, \begin{bmatrix} d \\ wrong \end{bmatrix} \end{aligned}$$

1. step: Merge

2. step: Merge

3. step: Move

4. step: Merge

5. step: Merge

6. step: Merge

7. step: Merge

8. step: Move

9. step: Merge

German examples

Der Detektiv hat die Kommissarin gesehen

1. step: merge

2. step: move

3. step: merge

4. step: merge

5. step: move

6. step: merge

Die Detektivin hat den Kommissar gesehen

The sentence is parsed like the first sentence “Der Detektiv hat die Kommissarin gesehen.”

Den Detektiv hat die Kommissarin gesehen

1. step: merge

2. step: move

3. step: merge

4. step: merge

5. step: move

6. step: merge (head movement)

7. step: move (scrambling)

Die Detektivin hat der Kommissar gesehen

1. step: merge

2. step: move

3. step: merge

4. step: merge

5. step: move

6. step: merge (head movement)

At this point the derivation of the sentence terminates because there are no more features that could be checked. As there is still the licenser for the scrambling operation left the sentence is grammatically not well-formed and is not accepted by the grammar formalism.

References

- Bader M (1996) Sprachverstehen: Syntax und Prosodie beim Lesen. Westdeutscher Verlag, Opladen
- Bader M, Meng M (1999) Subject-object ambiguities in German embedded clauses: an across-the-board comparison. *J Psycholinguist Res* 28(2):121–143
- beim Graben P, Gerth S, Vasisht S (2008a) Towards dynamical system models of language-related brain potentials. *Cogn Neurodyn* 2(3):229–255
- beim Graben P, Pinotsis D, Saddy D, Potthast R (2008b) Language processing with dynamic fields. *Cogn Neurodyn* 2(2):79–88
- beim Graben P, Atmanspacher H (2009) Extending the philosophical significance of the idea of complementarity. In: Atmanspacher H, Primas H (eds) *Recasting reality*. Springer, Berlin, pp 99–113

- beim Graben P, Potthast R (2009) Inverse problems in dynamic cognitive modeling. *Chaos* 19(1):015103
- beim Graben P, Barrett A, Atmanspacher H (2009) Stability criteria for the contextual emergence of macrostates in neural networks. *Netw Comput Neural Syst* 20(3):177–195
- Berg G (1992) A connectionist parser with recursive sentence structure and lexical disambiguation. In: *Proceedings of the 10th national conference on artificial intelligence*, pp 32–37
- Chomsky N (1981) *Lectures on government and binding*. Foris
- Chomsky N (1995) *The minimalist program*. MIT Press, Cambridge
- Christiansen MH, Chater N (1999) Toward a connectionist model of recursion in human linguistic performance. *Cogn Sci* 23(4):157–205
- Dolan CP, Smolensky P (1989) Tensor product production system: A modular architecture and representation. *Connect Sci* 1(1):53–68
- Elman JL (1995) Language as a dynamical system. In: Port RF, van Gelder T (eds), *Mind as motion: explorations in the dynamics of cognition*. MIT Press, Cambridge, pp 195–223
- Fanselow G, Schlesewsky M, Cavar D, Kliegl R (1999) Optimal parsing: syntactic parsing preferences and optimality theory. *Rutgers Optim Arch* pp 367–1299
- Farkas I, Crocker MW (2008) Syntactic systematicity in sentence processing with a recurrent self-organizing network. *Neurocomputing* 71:1172–1179
- Ferreira F, Henderson JM (1990) Use of verb information in syntactic parsing: evidence from eye movements and word-by-word self-paced reading. *J Exp Psychol Learn Mem Cogn* 16:555–568
- Frazier L (1979) On comprehending sentences: syntactic parsing strategies. PhD thesis, University of Connecticut, Storrs
- Frazier L (1985) Syntactic complexity. In: Dowty D, Karttunen L, Zwicky A (eds), *Natural language parsing*. Cambridge University Press
- Frazier L, Rayner K (1982) Making and correcting errors during sentence comprehension: eye movements in the analysis of structurally ambiguous sentences. *Cogn Psychol* 14:178–210
- Freeman WJ (2007) Definitions of state variables and state space for brain-computer interface. Part 1. Multiple hierarchical levels of brain function. *Cogn Neurodyn* 1:3–14
- Frisch S, Schlesewsky M, Saddy D, Alpermann A (2002) The P600 as an indicator of syntactic ambiguity. *Cognition* 85:B83–B92
- Gerth S (2006) Parsing mit minimalistischen, gewichteten Grammatiken und deren Zustandsraumdarstellung. Unpublished Master's thesis, Universität Potsdam
- Gibson E (1998) Linguistic complexity: locality of syntactic dependencies. *Cognition* 68:1–76
- Frey W, Gärtner HM (2002) On the treatment of scrambling and adjunction in Minimalist Grammars. In: Jäger G, Monachesi P, Penn G, Wintner S (eds) *Proceedings of formal grammars*, pp 41–52
- Haegeman L (1994) *Introduction to government & binding theory*. Blackwell Publishers, Oxford
- Hagoort P (2003) How the brain solves the binding problem for language: a neurocomputational model of syntactic processing. *NeuroImage* 20:S18–S29
- Hagoort P (2005) On Broca, brain, and binding: a new framework. *Trends Cogn Sci* 9(9):416–423
- Hale JT (2003a) Grammar, uncertainty and sentence processing. PhD thesis, The Johns Hopkins University
- Hale JT (2003b) The information conveyed by words in sentences. *J Psycholinguist Res* 32(2):101–123
- Hale JT (2006) Uncertainty about the rest of the sentence. *Cogn Sci* 30(4):643–672
- Harkema H (2001) Parsing minimalist languages. PhD thesis, University of California, Los Angeles
- Hemforth B (1993) *Kognitives Parsing: Repräsentation und Verarbeitung kognitiven Wissens*. Infix, Sankt Augustin
- Hemforth B (2000) *German sentence processing*. Kluwer Academic Publishers, Dordrecht
- Hertz J, Krogh A, Palmer RG (1991) *Introduction to the theory of neural computation*. Perseus Books, Cambridge
- Hopcroft JE, Ullman JD (1979) *Introduction to automata theory, languages, and computation*. Addison-Wesley, Menlo Park
- Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational abilities. *Proc Natl Acad Sci USA* 79(8):2554–2558
- Joshi AK, Schabes Y (1997) Tree-adjointing grammars. In: Salomaa A, Rosenberg G (eds) *Handbook of formal languages and automata*, vol 3. Springer, Berlin, pp 69–124
- Joshi AK, Levy L, Takahashi M (1975) Tree adjunct grammars. *J Comput Syst Sci* 10(1):136–163
- Lawrence S, Giles CL, Fong S (2000) Natural language grammatical inference with recurrent neural networks. *IEEE Trans Knowl Data Eng* 12(1):126–140
- Legendre G, Miyata Y, Smolensky P (1990a) Harmonic grammar—a formal multi-level connectionist theory of linguistic well-formedness: theoretical foundations. In: *Proceedings of the 12th annual conference cognitive science society*. Cognitive Science Society, Cambridge, pp 388–395
- Legendre G, Miyata Y, Smolensky P (1990b) Harmonic grammar—a formal multi-level connectionist theory of linguistic well-formedness: an application. In: *Proceedings 12th annual conference cognitive science society*. Cognitive Science Society, Cambridge, pp 884–891
- Michaelis J (2001) Derivational minimalism is mildly context-sensitive. In: Moortgat M (ed) *Logical aspects of computational linguistics*, vol 2014. Springer, Berlin, pp 179–198 (Lecture notes in artificial intelligence)
- Mizraji E (1989) Context-dependent associations in linear distributed memories. *Bull Math Biol* 51(2):195–205
- Mizraji E (1992) Vector logics: The matrix-vector representation of logical calculus. *Fuzzy Sets Syst* 50:179–185
- Niyogi S, Berwick RC (2005) A minimalist implementation of Hale-Keyser incorporation theory. In: Sciallo AMD (ed) *UG and external systems language, brain and computation, linguistik aktuell/linguistics today*, vol 75. John Benjamins, Amsterdam, pp 269–288
- Osterhout L, Holcomb PJ, Swinney DA (1994) Brain potentials elicited by garden-path sentences: evidence of the application of verb information during parsing. *J Exp Psychol Learn Mem Cogn* 20(4):786–803
- Pollard C, Sag IA (1994) *Head-driven phrase structure grammar*. University of Chicago Press, Chicago
- Potthast R, beim Graben P (2009) Inverse problems in neural field theory. *SIAM J Appl Dyn Syst* (in press)
- Prince A, Smolensky P (1997) Optimality: from neural networks to universal grammar. *Science* 275:1604–1610
- Siegelmann HT, Sontag ED (1995) On the computational power of neural nets. *J Comput Syst Sci* 50(1):132–150
- Smolensky P (1986) Information processing in dynamical systems: foundations of harmony theory. In: Rumelhart DE, McClelland JL, the PDP Research Group (eds) *Parallel distributed processing: explorations in the microstructure of cognition*, vol I. MIT Press, Cambridge, pp 194–281 (Chap 6)
- Smolensky P (1990) Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artif Intell* 46:159–216
- Smolensky P (2006) Harmony in linguistic cognition. *Cogn Sci* 30:779–801
- Smolensky P, Legendre G (2006a) *The harmonic mind. from neural computation to optimality-theoretic grammar*, vol 1: cognitive architecture. MIT Press, Cambridge

- Smolensky P, Legendre G (2006b) The harmonic mind. From neural computation to optimality-theoretic grammar, vol 2: linguistic and philosophic implications. MIT Press, Cambridge
- Stabler EP (1997) Derivational minimalism. In: Retoré C (ed) Logical Aspects of computational linguistics, springer lecture notes in computer science, vol 1328. Springer, New York, pp 68–95
- Stabler EP (2000) Minimalist Grammars and recognition. In: CSLI (ed) Linguistic form and its computation, Rohrer, Rossdeutscher and Kamp, pp 327–352
- Stabler EP (2004) Varieties of crossing dependencies: structure dependence and mild context sensitivity. *Cogn Sci* 28:699–720
- Stabler EP, Keenan EL (2003) Structural similarity within and among languages. *Theor Comput Sci* 293:345–363
- Staudacher P (1990) Ansätze und Probleme prinzipienorientierten Parsens. In: Felix SW, Kanngießer S, Rickheit G (eds) Sprache und Wissen. Westdeutscher Verlag, Opladen, pp 151–189
- Tabor W (2000) Fractal encoding of context-free grammars in connectionist networks. *Expert Syst Int J Knowl Eng Neural Netw* 17(1):41–56
- Tabor W (2003) Learning exponential state-growth languages by hill climbing. *IEEE Trans Neural Netw* 14(2):444–446
- Tabor W, Tanenhaus MK (1999) Dynamical models of sentence processing. *Cogn Sci* 23(4):491–515
- Tabor W, Juliano C, Tanenhaus MK (1997) Parsing in a dynamical system: An attractor-based account of the interaction of lexical and structural constraints in sentence processing. *Lang Cogn Process* 12(2/3):211–271
- Traxler M, Gernsbacher MA (eds) (2006) Handbook of psycholinguistics. Elsevier, Oxford
- Vosse T, Kempen G (2000) Syntactic structure assembly in human parsing: a computational model based on competitive inhibition and a lexicalist grammar. *Cognition* 75:105–143
- Vosse T, Kempen G (this issue) The Unification space implemented as a localist neural net: predictions and error-tolerance in a constraint-based parser. *Cogn Neurodyn*
- Weyerts H, Penke M, Muentz TF, Heinze HJ, Clahsen H (2002) Word order in sentence processing: an experimental study of verb placement in German. *J Psycholinguist Res* 31(3):211–268