Verb Inflections in German Child Language: A Connectionist Account

Gert Westermann
Department of Computer Science
Technical University of Braunschweig
38116 Braunschweig, Germany
westerma@ibr.cs.tu-bs.de

Risto Miikkulainen
Department of Computer Sciences
The University of Texas at Austin
Austin, TX 78712 USA
risto@cs.utexas.edu

Abstract

The emerging function of verb inflections in German language acquisition is modeled with a connectionist network. A network that is initially presented only with a semantic representation of sentences uses the inflectional verb ending `-t` to mark those sentences that are low in transitivity, whereas all other verb endings occur randomly. This behavior matches an early stage in German language acquisition where verb endings encode a similar semantic rather than a grammatical function. When information about the surface structure of the sentence is added to the input data, the network learns to use the correct verb inflections in a process very similar to children's learning. This second phase is facilitated by the semantic phase, suggesting that there is no shift from semantic to grammatical encoding, but rather an *extension* of the initial semantic encoding to include grammatical information. This can be seen as evidence for the strong version of the functionalist hypothesis of language acquisition.

Introduction

It is widely assumed in the study of language acquisition that children's early speech is not based on an encoding of grammatical relations (such as subject and object) but semantic properties (such as agent and patient; see e.g. Bates and MacWhinney, 1979; Slobin, 1981; MacWhinney, 1987). The children generally use the same coding elements found in adult speech (such as inflectional verb endings, case markers, and word order) for encoding semantic properties, and therefore often produce grammatically incorrect utterances that are nevertheless consistent from the semantic viewpoint. For example, Clahsen (1986) showed that German children initially use the inflectional verb ending `-t` to encode low transitivity of a sentence. Only at later stages of language acquisition they discover the notion of grammatical subject, which enables them to use the `-t` in its correct function to encode a third person singular or second person plural subject.

The goal of this study is to give a computational account for this phenomenon using a connectionist model. The model is trained to map a sentence representation to the appropriate inflectional ending of its main verb. The network is initially presented with sentence representations that do not convey grammatical information but only encode surface-semantic properties. This representation corresponds to the semantic encoding of language at the early stages of German language acquisition. At a later stage, the input data is augmented with information about the surface structure of the sentences, simulating the child's discovery of grammatical properties of language. The behavior generated by the network is found to match the behavior observed by Clahsen (1986) in German children: When only semantic information is available, the network produces the verb ending `-t` only for those sentences that are low in transitivity. When information about the surface structure is added, the network gradually learns the correct verb inflections in an order that matches Clahsen's observations. Analyzing the network weights shows that acquiring the correct verb-inflection paradigm is supported by the existing semantic representations, particularly the agency of the subject. This seems to indicate that there is no developmental shift from a semantic to a grammatical encoding, but instead learning is based on extending the semantic encoding to include surface-structure information. This result provides a computational justification for the strong version of the functionalist hypothesis of language acquisition (Bates and MacWhinney, 1979), which states that semantic and pragmatic representations are instrumental in learning grammatical concepts.

Below, the German verb inflection paradigm is first briefly reviewed, and Clahsen's observations on how children acquire it are outlined. Our experimental setup, including the data and the network architecture, are then described, followed by a detailed explanation of the results. Analysis of the implications to the study of language acquisition and the functionalist hypothesis concludes the paper.

Verb inflections in German

In the German language, verbs are inflected according to the person and number of the subject in the sentence (*subject–verb agreement*). The inflectional formatives are affixed to the root of the verb. For the present tense, there are five suffixes encoding the following functions:

`-e` :	first person sg. for full verbs.
`-st` :	second person sg. for full verbs and modal verbs.
`-t` :	third person sg. for full verbs, second person pl. for full verbs and modal verbs.
`-∅` :	(null) first and third person sg. for modal verbs.
`-en` :	first and third person pl. for full and modal verbs.

Clahsen (1982, 1986, 1988) described the process of German language acquisition based on a longitudinal study of three children. These subjects were observed for two

To appear in the *Proceedings of the 16th Annual Conference of the Cognitive Science Society*, 1994.

	Criterion	High transitivity	Low transitivity
1.	Participants	two or more (agent and object)	one participant (agent)
2.	Agency	Agent is high in potency	Agent is low in potency
3.	Kinesis	action	nonaction
4.	Volitionality	volitional	nonvolitional
5.	Animacy of object	human, animate	inanimate
6.	Affirmation	affirmative	negative

Table 1: Semantic features that determine the transitivity of a sentence (adopted from Hopper and Thompson, 1980).

Indicator for low transitivity	Stage II	Stage III
Participants: one	84.0%	71.0%
Agency: low	95.5%	69.0%
Kinesis: nonaction	82.5%	65.7%
Volitionality: nonvolitional	90.5%	74.7%
Animacy: low	79.0%	62.0%
Affirmation: negative	20.1%	29.3%

Table 2: The percentage of `-t` inflections produced by German children for sentences that match the different indicators for low transitivity.

years, starting at the age of 14 months when their speech consisted mostly of single word utterances, and ending after they had acquired the full subject–verb agreement paradigm.

Clahsen (1986) distinguished between five developmental stages extending from one-word utterances (I) to the use of embedded sentences (V). Verbs are first used at stage II. At stage III (indicated by an increase of correct verb inflections and by overgeneralization of the `-e` inflection), children start to discover grammatical aspects of language, which leads to acquisition of the full subject–verb agreement system at stage IV.

According to Clahsen (1986), children at stage II encode their speech according to the semantic property of *transitivity*. The transitivity of a given sentence can be measured according to the six criteria listed in table 1. For example, the sentence `Ich gebe Peter ein Buch` (I am giving a book to Peter) is highly transitive: There are two participants (`Ich` and `Peter`), `Ich` is a subject high in agency, the verb `geben` has a high kinesis and is a volitional act, the object `Peter` is a specified human being, and the sentence is not negated. By contrast, the sentence `Das Buch liegt nicht hier` (The book does not lie here) is very low in transitivity: There is only one participant (`das Buch`) which is low in agency, the verb `liegen` describes a state and not an action and is therefore low in kinesis, the act is not volitional, and the sentence is negated.

Clahsen showed that most of the sentences where the children used the `-t` ending matched the criteria for low transitivity (table 2). The `-t` ending was preferred with every single indicator for low transitivity (except affirmation, which was not significant because of the low number of negated sentences in the data). The preference was stronger at stage II than III, indicating that at stage III the children are beginning to turn away from pure semantic encoding.

The conclusion is that at early stages of language acquisition, the `-t` ending encodes not the person/number agreement with the subject, but the low transitivity of the whole sentence. At stage III the use of `-e` emerges and the semantic function of `-t` begins to fade out, and at stage IV all inflectional endings are used correctly. The experiments reported in this paper aimed at verifying Clahsen's theory computationally, by simulating the transition from semantic to grammatical encoding of sentences and observing its effect on the use of different verb inflections.

Experiments

There is strong evidence that inflectional affixes are attached to words as needed during speech production (Aitchison, 1987). The neural network used in our experiments is intended to model only the specific part of the human language production system which is responsible for this process. The input to the network comes from other parts of the language production system. Since syntactic information is not available to the child at early stages of language acquisition, the input during modeling stage II will be restricted to a semantic encoding of sentences. This encoding allows the network to develop representations for semantic concepts without access to the specific constituents of the sentences. During later stages children start to encode their speech according to non-semantic, or grammatical, concepts. This change will be modeled by augmenting the input information with an abstract representation of the words in the sentences. The network should be able to develop grammatical rules based on regularities in the word occurrencies. The task of our network, therefore, is not to motivate the change in the input representation from stage II to stages III and IV (this change is initiated by other parts of the language production system), but to *react* to this change by choosing different cues in order to determine the correct verb inflection.

Data

A set of sentences was generated according to templates in table 3. The vocabulary consisted of nouns in various semantic categories, transitive and intransitive verbs, modal verbs, and the negation particle `nicht`. In order to obtain a good variety of sentences, many of the nouns could be used in the subject as well as in the object position (e.g. `Mama sieht Papa` (AV_fO) – `Papa sieht Mama` (AV_fO)), and many verbs could optionally take an object and a beneficiary (such as `Wir lesen` (AV_f) – `Wir lesen Buch` (AV_fO), and `Ich sage das` (AV_fO)

Example sentence	Sem.rep.	Surface-structure representation	Inflection
`Ich gebe Papa Apfel nicht`	111100	110010010000100100000010111001000000	10000 (-e)
`Du willst Mama Bild nicht zeigen`	110100	000111100101110001011111111001100000	01000 (-st)
`Ihr baut Lego`	111101	101100100111101000000000000000000000	00100 (-t)
`Papa mag Apfel`	110001	100100000101000010000000000000000000	00010 (-∅)
`Wir lesen`	011101	010101001100000000000000000000000000	00001 (-en)

Table 4: Example input sentences and their semantic and surface-structure encodings.

Template	Example sentence
AV_f	`Ich lese.`
AV_fN	`Ich lese nicht.`
AMV_i	`Ich will lesen.`
$AMNV_i$	`Ich will nicht lesen.`
AV_fO	`Ich lese Buch.`
AV_fON	`Ich lese Buch nicht.`
$AMOV_i$	`Ich will Buch lesen.`
$AMONV_i$	`Ich will Buch nicht lesen.`
AV_fBO	`Mama gibt Papa Ball.`
AV_fBON	`Mama gibt Papa Ball nicht.`
$AMBOV_i$	`Mama will Papa Ball geben.`
$AMBONV_i$	`Mama will Papa Ball nicht geben.`

Table 3: Sentence templates. In the template names, A = agent, V_f = finite verb, V_i = infinite verb, O = object, B = beneficiary, M = modal verb, N = negation.

– `Ich sage Mama das` (AV_fBO)). Articles were omitted because children do not use them at early stages of language acquisition. In German, the article disambiguates the grammatical role of a noun (subject, object, beneficiary etc.), which allows for a relatively free variation of word order. The word order in our data was fixed, and therefore there was no need for such disambiguation.

A representative data set consisting of 200 sentences for each of the five inflectional verb endings was randomly extracted from the full set of 6610 generated sentences. The resulting set was randomly divided into 667 sentences for training and 333 for testing the model. The sentences were encoded in two complementary ways: a "*semantic*" encoding that did not convey any information about the sentence structure, and a "*surface-structure*" encoding where the words were represented as abstract tokens without any semantic information (table 4).

In the semantic encoding, each sentence was represented by a vector coding the presence (`1`) or absence (`0`) of the six semantic features (table 1): more than one participant, high agency of the subject, high kinesis of the verb, volitionality of the act, animacy of the object, and affirmation. The number of `1`s in the representation therefore directly corresponds to the degree of transitivity of the sentence.

In the surface-structure encoding, each word was represented by a random bitstring, which stands for an abstract pointer to whatever non-semantic encoding of the word first becomes available to the child. There were 57 words in the lexicon, so each word could be represented in six bits. All inflectional forms of a verb were encoded by the same bitstring. Since the maximum length of a sentence was six words, the full surface-structure representation of a sentence consisted of 36 bits. If a sentence had fewer than six words, zeroes were added to the right.

Network architecture

A feedforward backpropagation network with one hidden layer was trained to map sentence representations to the correct inflectional verb endings (figure 1). The network had 42 input units: the leftmost six received the semantic encoding of the sentence, and the remaining 36 the surface-structure encoding. There were 8 hidden units and five output units, one for each possible inflectional verb ending. The output unit with the highest activity indicated the network's choice of ending. An output would be regarded as incorrect if the activation of the strongest output unit was less than 0.5.

The simulation experiment was divided into two phases: in the first, only the semantic encoding was presented to the network while the word units were set to zero, thus simulating stage II of language acquisition. In the second phase, the word units received proper input, allowing the network to exploit both semantic and surface-structure cues in learning the mapping task. This simulated the discovery of grammatical information by the child at stages III and IV.

Modeling stage II of language acquisition

After 25,000 training epochs with the semantic input, the network was able to produce the correct verb-ending only in about 10% of the test sentences. This is because the semantic representation is highly ambiguous: the same bitstring can represent many different sentences requiring different verb inflections. For example, the semantic encoding `011101` (meaning 1 participant—high agency—action verb—volitional—no animate object—not negated) represents the sentence `Wir lesen`, which requires the inflectional ending `-en`, the sentence `Ich esse`, requiring the ending `-e`, the sentence `Mama läuft` with the inflectional ending `-t`, and many more. Although the correct verb inflection cannot be determined from the semantic representation alone, this is what children apparently attempt at early stages of language acquisition.

Most significantly, the network behavior was quite similar to that of children at stage II of German language acquisition. Out of 333 test sentences, the network produced a `-t` ending for 15 sentences, or 4.5%, which matches the 5% Clahsen (1986) observed. As in children's speech, almost all of them were produced for sentences on the lower side of the transitivity scale (figure 2). However, no correlation between the degree of sentence transitivity and any of the other inflectional

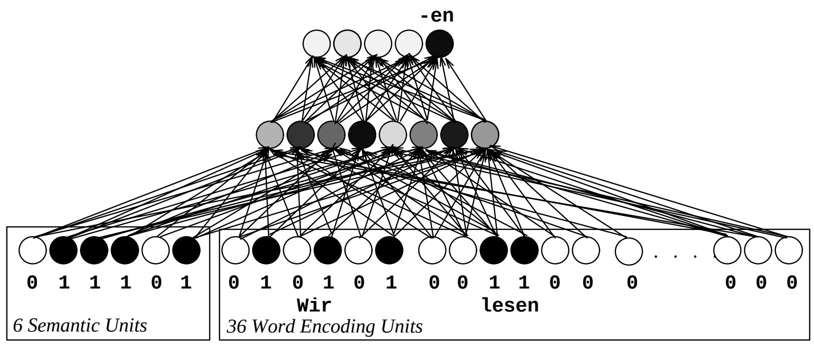

Figure 1: The network architecture. When modeling stage II, the 36 Word Encoding Units were set to zero. All of the units were used in modeling stages III and IV.

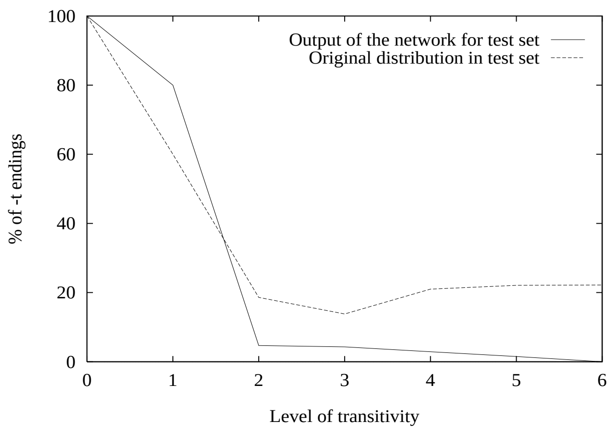

Figure 2: The percentage of `-t` endings out of all endings produced by the network for each level of transitivity. The horizontal axis indicates the number of high-transitivity cues that were on in the input representation (i.e number of ones). The dashed line indicates the (practically identical) distributions in the training and testing sets, and the solid line shows the percentage generated by the network for the testset.

endings `-e`, `-st`, `-∅`, and `-en` could be detected. In particular, no verb–ending was found that would encode a *high* degree of transitivity. This result is also in line with Clahsen's (1986) observations: "[...] *word order* is initially used as a grammatical coding device in transitive clauses and [...] there is no need for the child to further grammaticize these clauses in the early developmental phases" (p. 111, emphasis by G.W.). Since the word order was fixed in our experiments, this hypothesis could not be tested.

Why does the network favor the `-t` ending for sentences with low transitivity? In trying to find a mapping from the semantic representation to the inflectional ending, the network is confronted with an impossible task due to the ambiguity of the input. Yet the relation between the transitivity of a sentence and the inflectional ending is not entirely arbitrary. In the input data, as in real speech (Clahsen, 1986), `-t` occurs more often with sentences of low transitivity. The network learns to exploit these correlations, even amplify them. No such correlations exist for the other verb endings, or the other levels of transitivity, and the other verb endings are generated seemingly randomly. This explanation agrees with Clahsen's (1986, 1988) theory, and confirms it computationally: "I suggest that `-t` is the optimal candidate available in the German verb inflection paradigm to grammatically encode the intransitive construction. Note that the other overt formatives, `-e` and `-st`, are encodings of the 1st and 2nd person subjects in German. 'I' and 'you' are, however, prototypical agents, but the children are looking for a grammatical expression for nonagents in the intransitive construction. Thus, `-t` appears to be the best choice in regard to what the children are looking for, since `-t` often refers to inanimate subjects in the adult language, and, moreover, `-t` does not necessarily encode agents" (Clahsen, 1986, p. 113). And in (Clahsen, 1988, p. 100): "The child finds in the input data frequent confirmation for the equation ...`-t` = [+intransitive] [...]. For the case ...`-(en)` = [-intransitive] children find less evidence, because transitive verbs can occur with different grammatical subjects. It therefore may be assumed that in a transitory developmental phase there exist certain word–specific paradigms of the type [+`intransitive` = `-t`; `-intransitive` = `-en`], in which the right side of the equation [-`intransitive` = ?] remains empty" (Translation by G.W.)

Modeling stages III and IV of language acquisition

At stage III children begin to discover grammatical categories and encode their speech accordingly. In our experiments this was simulated by continuing to train the network from the first phase of the experiment, but now also including the surface-structure encoding in the input. This new information enabled the network to acquire the grammatical category of subject, and it learned the correct verb inflections very fast. After 50 epochs the network produced 48 endings correctly, 34 of which were first person singular (`-e`). On the other hand, `-e` was overgeneralized to 47 sentences in which it was not

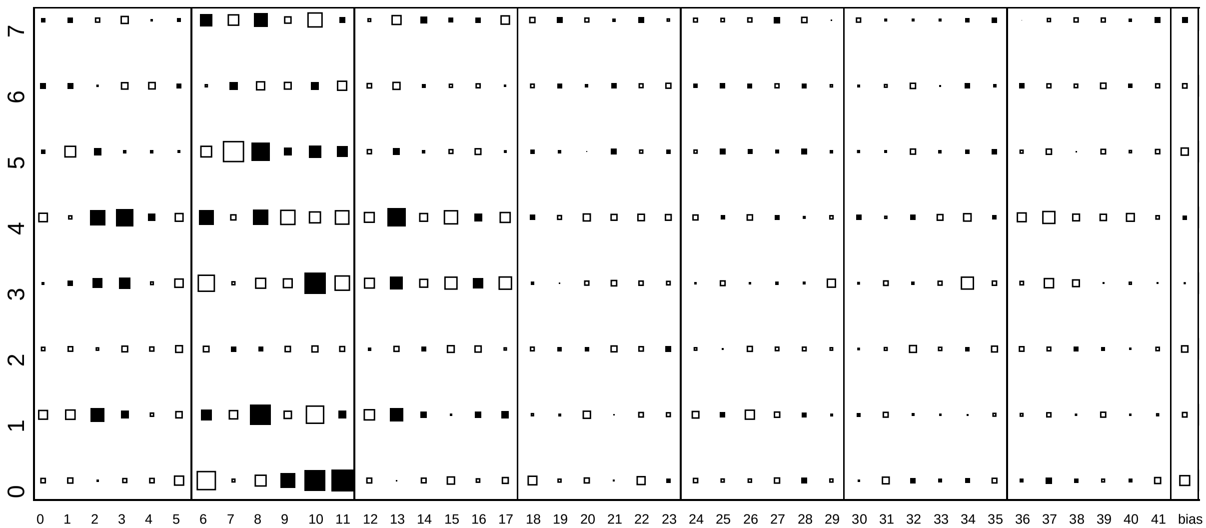

Figure 3: Weights from the input (horizontal axis) to the hidden layer (vertical axis) after training. The first six units (0-5) correspond to the semantic input, the next (6-11) to the first word etc. Dark boxes indicate positive and hollow boxes negative weights, and the box size corresponds to the absolute value of the weight. The average norms of the weight vectors were 3.182 for the semantic units (8.348 for units 2 and 3 encoding kinesis and volitionality), 11.417 for the units representing the first word, and 2.645, 1.042, 1.467, 1.552, and 0.712 for those representing the remaining words.

correct. This corresponds to stage III as described by Clahsen (1986). After 8,000 epochs, the network generated the correct output for all 333 test sentences. Corresponding to stage IV, the subject–verb agreement had been fully acquired.

Figure 3 shows the weights from the input layer to the hidden layer of the network after the training was complete. The network is using the first word (units 6 to 11) as the most important cue in determining the verb inflection, which makes sense because in our data the first word is always the subject. However, the network also utilizes semantic cues, especially the features *kinesis* (unit 2) and *volitionality* (unit 3). This is surprising because the verb endings can be completely determined based on the grammatical properties of the subject (its person and number), and its semantic features should not matter.

In order to better understand this result, two contrasting training strategies were compared: In the first, both the semantic and the word encoding were presented to the network from the start. In the second, the semantic input was set to zero through the whole run. In the second experiment, therefore, the network had to pick up the mapping solely from the surface structure of the sentences, whereas in the first it could exploit semantic cues as well.

The network that could use both encodings learned the task significantly faster (32 epochs vs. 139 epochs, averaged over 10 trials each). A weight analysis confirmed the results: the network with both encodings made good use of the semantic features kinesis and volitionality. This is particularly interesting because it seems to comply with the functionalist notion of *agency*. In functional grammar, agency is restricted to an entity controlling an action that has the features [+dynamic] and [+controlled] (Dik, 1978). In our experiments, kinesis and volitionality carry similar meaning, and therefore it seems that the network learns to use the implicit semantic property of agency to help determine the inflectional ending of the verb. Although not entirely reliable and not at all necessary, this is nevertheless a helpful strategy because high-agency sentences are often associated with `-e` and disassociated with `-t` , `-∅` and `-en` endings. After discovering agency, the network quickly learns to pay attention to the first word, which in effect corresponds to discovering the grammatical notion *subject*.

Discussion

Our results have interesting implications on the study of language acquisition. They suggest that the transition from stage II to stage IV, in which the subject-verb agreement system is acquired, does not constitute a developmental shift from a semantic to a grammatical encoding, but an *extension* of the semantic representation to include surface information. In the first phase of our experiment, corresponding to stage II of language acquisition, the network builds internal sentence representations based on semantics that only need to be modified, not replaced, when the surface-structure input is added. Useful features such as kinesis and volitionality serve as a good starting point, and the network learns to concentrate on the first word (i.e the subject) in a sentence faster.

This can be seen as evidence for the strong version of the functionalist hypothesis of language acquisition (Bates and MacWhinney, 1979). In contrast to theories postulating autonomous syntax (e.g., Chomsky, 1975), this hypothesis maintains that "grammatical forms are 'determined' and 'maintained' by [...] communicative functions and processing constraints. The strong version leads to a developmental model in which children discover the structure of grammar through their experience with competing communicative factors" (Bates and MacWhinney, 1979, p. 174). Specifically, "The strong functionalist view predicts that children will show evidence for the intention to encode agent and/or topic *before* they evidence control of the surface role of subject" (p. 189). In our model, the low transitivity of a sentence (or more specifically, nonagency) was encoded with the verb ending `-t` even before word information was available to the network. Once the surface structure became available, the verb inflections started to encode agents as well, which eventually led to forming the concept of subject. In this sense the network behavior followed the functionalist hypothesis quite closely.

Conclusion

Connectionist modeling of learning verb inflections in German confirmed Clahsen's (1986) hypothesis computationally. The initial semantic sentence representation resulted in inflections very similar to those produced by children at stage II of language acquisition, in which they have access only to semantic information and not to the grammatical concept of subject. Adding information about the surface structure of the sentences led to a gradual acquisition of the subject–verb paradigm through stages similar to those observed in children during the transition from stage II to III, and from III to IV. At these later stages the network still utilized the semantic notion of agency, which can be seen as support for the strong version of the functionalist hypothesis.

The current model acquires the grammatical concept of subject based on the semantic notion of agency as well as surface-structure cues. It would be interesting to see if the model could be extended to learn other grammatical roles based on semantic notions as well, such as object (based on the notion of patient), or verb (based on action). The success of such extensions depends largely on whether it will be possible to provide an input representation rich enough in (realistic) cues. This will be our main direction of future research.

References

Aitchison, J. (1987). *Words in the Mind: An Introduction to the Mental Lexicon.* Oxford, UK; New York: Blackwell.

Bates, E., and MacWhinney, B. (1979). A functionalist approach to the acquisition of grammar. In Ochs, E., and Schieffelin, B. B., editors, *Developmental Pragmatics*, 167–211. New York, San Francisco, London: Academic Press.

Chomsky, N. (1975). *Reflections on Language.* New York: Pantheon.

Clahsen, H. (1982). *Spracherwerb in der Kindheit.* Tübingen: Narr.

Clahsen, H. (1986). Verb inflections in german child language: Acquisition of agreement markings and the functions they encode. *Linguistics*, 24:79–121.

Clahsen, H. (1988). *Normale und gestörte Kindersprache.* Amsterdam, Philadelphia: John Benjamins Publishing Company.

Dik, S. C. (1978). *Functional Grammar.* Amsterdam; New York: North–Holland.

Hopper, P. J., and Thompson, S. A. (1980). Transitivity in grammar and discourse. *Language*, 56:251–299.

MacWhinney, B. (1987). The competition model. In MacWhinney, B., editor, *Mechanisms of Language Acquisition.* Hillsdale, NJ: Erlbaum.

Slobin, D. I. (1981). The origin of grammatical encoding of events. In Deutsch, W., editor, *Behavioural Development*, 185–199. London, New York, Toronto: Academic Press.