

John Benjamins Publishing Company



This is a contribution from *The Acquisition of Relative Clauses. Processing, typology and function.*

Edited by Evan Kidd.

© 2011. John Benjamins Publishing Company

This electronic file may not be altered in any way.

The author(s) of this article is/are permitted to use this PDF file to generate printed copies to be used by way of offprints, for their personal use only.

Permission is granted by the publishers to post this file on a closed server which is accessible to members (students and staff) only of the author's/s' institute, it is not permitted to post this PDF on the open internet.

For any other use of this material prior written permission should be obtained from the publishers or through the Copyright Clearance Center (for USA: www.copyright.com).

Please contact rights@benjamins.nl or consult our website: www.benjamins.com

Tables of Contents, abstracts and guidelines are available at www.benjamins.com

A connectionist account of the acquisition and processing of relative clauses

Hartmut Fitz, Franklin Chang and Morten H. Christiansen

Relative clause processing depends on the grammatical role of the head noun in the subordinate clause. This has traditionally been explained in terms of cognitive limitations. We suggest that structure-related processing differences arise from differences in experience with these structures. We present a connectionist model which learns to produce utterances with relative clauses from exposure to message-sentence pairs. The model shows how various factors such as frequent subsequences, structural variations, and meaning conspire to create differences in the processing of these structures. The predictions of this learning-based account have been confirmed in behavioral studies with adults. This work shows that structural regularities that govern relative clause processing can be explained within a usage-based approach to recursion.

1. Introduction

Relative clauses have been an important source of data for psycholinguistic theories since the inception of its modern incarnation in the 1950s. In particular, the processing problems associated with multiple center-embedded object relative clauses, such as (1), have figured prominently in accounts of adult sentence processing (e.g., Church 1982; Gibson 1998; Just & Carpenter 1992; Kimball 1973; Marcus 1980; Miller & Chomsky 1963; Reich 1969; Stabler 1994). Similarly, the problems that children experience with relative clauses during acquisition have also informed theories of language development (for a review, see O'Grady 1997). More recently, research on relative clauses has resurged within usage-based approaches to language acquisition and processing (e.g., Diessel 2009; Diessel & Tomasello 2005; Kidd *et al.* 2007; MacDonald & Christiansen 2002; Reali & Christiansen 2007). This work has suggested that the ability to process relative clauses emerges gradually, primarily as a function of experience with language, and that the effect of exposure to these constructions continues to be important for processing even in adulthood.

- (1) The cat that the dog that the mouse bit saw ran away.

Connectionist modeling provides a natural computational framework within which to explore the role of experience in the acquisition and processing of relative clauses. These models generally consist of layers of nodes connected to one another by way of weighted links (by analogy to the strength of synaptic connections between neurons in the brain). Importantly, such networks learn from repeated exposure to input-output examples and are able to generalize to novel examples not presented during learning. As such, connectionist models have been successfully applied to the modeling of many aspects of language acquisition and processing, including language impairments (for reviews, see Christiansen & Chater 2001; Gupta 2008). In this chapter, we outline a connectionist approach to the acquisition and processing of relative clauses. The main focus of the paper is a usage-based connectionist model of relative clause acquisition, building on prior work on sentence production (Chang, Dell & Bock 2006). The model demonstrates that exposure to different types of relative clause constructions is key to understanding the patterns of linguistic behavior observed in children. In the context of a connectionist usage-based model of adult recursive sentence processing (Christiansen & MacDonald 2009), we then argue that the role of experience continues to be a key factor in explaining relative clause processing in adults. Finally, we discuss the relationship between our connectionist approach and other recent usage-based computational models, and sketch future directions for our approach.

2. The relative clause accessibility hierarchy

Cross-linguistic patterns have traditionally played an important role in theories of language – often couched as “language universals” – because they seem to require some innate endowment. Theoretical accounts of language universals sometimes argue that they arise from the nature of an innately-specified language processor. Another possibility, which we pursue here, is that these cross-linguistic patterns may arise from the various mechanisms that support language acquisition and processing, but these mechanisms may not be unique to language. One important cross-linguistic pattern in linguistic typology is the accessibility hierarchy of relative clause constructions. English relative clause constructions can be distinguished based on the grammatical function of their head noun in the relative clause. For example in the sentence fragment “the boy that runs”, the constituent “boy” functions as the subject of the intransitive clause and we label this as an S-relative (other types of relative clauses and their labeling are presented in Table 1).

Table 1. Summary of English relative clause constructions

Relativized role	Example	Label
Intransitive subject	... the boy that runs	S
Transitive subject	... the boy that chased the dog	A
Direct object	... the cat that the dog chased	P
Indirect object	... the girl who the boy gave the apple to	IO
Oblique object	... the boy who the girl played with	OBL

Keenan & Comrie (1977) sampled relative clause constructions from 50 languages and based on this data formulated an implicational universal for all languages. If a language has a syntactic construction to relativize subjects (S + A) and any other grammatical role in the ordering

$$(S + A) > P > IO > OBL$$

then it can relativize any position in between using the same construction. In typology this ordering is known as the *accessibility hierarchy* (henceforth: AH).

Keenan & Hawkins (1987) speculated that this hierarchy may be rooted in processing difficulties. They conducted an experiment in which subjects had to first comprehend and then reproduce different relative clause types. They found that the order of difficulty in English-speaking adults and children qualitatively matched the AH ordering. Several processing accounts have been proposed to explain this data, based on the syntactic structure of relative clauses and/or working memory limitations. For instance, Hawkins (1994) defined a measure of processing difficulty for relative clause types in terms of phrase-structure tree complexity. According to Hale (2006), the AH in sentence processing can be explained as a function of entropy reduction in incomplete parse trees. One of the most influential proposals, the dependency-locality theory of Gibson (1998), argued that the hierarchy can be accounted for by combining two factors, the distance between filler and gap and the number of incomplete syntactic dependencies at each sentence position. Gibson's theory would predict, for instance, that S-relatives ("the man that _ runs") are easier to process than P-relatives ("the man that a dog chases _"), because the distance between the head noun of the relative clause (called 'filler', in this case "man"), and the canonical position of the head noun in the relative clause (called 'gap', indicated by the underscore in the examples) is larger in P-relatives than in S-relatives. Although these approaches differ substantially in how they account for differential relative clause processing, they all suggest that a metric can be found that assigns a numeric score to individual structures which provides a measure of processing difficulty. The hierarchy then reflects differences in these scores. In this chapter we provide an alternative account in which

differential processing is a consequence of learning. The hierarchy arises mainly from complex interactions between structures in the input and is only partially due to intrinsic properties of structures in isolation. There is some evidence that this hierarchy may not be a behavioral universal (Hsiao & Gibson 2003; Ozeki & Shirai 2007; Chen & Shirai 2010), but this fact does not distinguish between processing and learning-based accounts, since both posit that language-specific features can influence the hierarchy. Given our limited space here, our discussion of the hierarchy will be mainly focused on English data.

2.1 The accessibility hierarchy in development

As Diessel & Tomasello (2005) observed, there are several aspects of AH behavior which are not addressed by filler-gap distance processing accounts. First, these accounts may not make the right cross-linguistic predictions. German relative pronouns, for example, are marked for gender, case, and number. Hence in most sentences with relative clauses, the grammatical role of the gap is already resolved at the pronoun position and the filler need not be kept in working memory until it can be integrated. Secondly, processing accounts have focused on comprehension, but presumably in production no filler integration is required at the gap position because the speaker's intended message is unambiguous. Unlike listeners, a speaker is not in a state of uncertainty about the semantic role of the relative clause head noun in a complex utterance.

Another issue that has not been examined carefully is the relationship between filler-gap accounts and language acquisition. If children are not drawing upon adult-like syntactic representations, they might not exhibit adult-like AH behavior in development. Yet, in a sentence repetition study with English children [4;3–4;9], Diessel & Tomasello (2005) found that the order of relative clause acquisition in production matched the adult processing hierarchy (Figure 1, similar results were also found in German children).

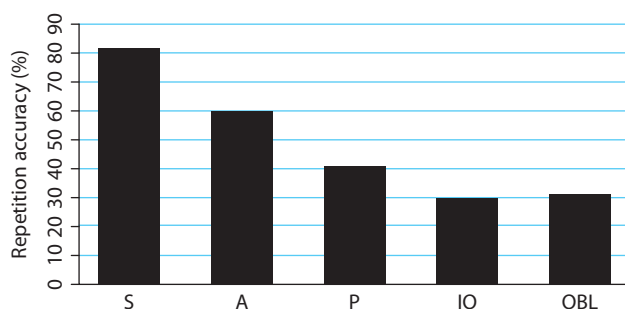


Figure 1. Relative clause acquisition in elicited production (figure based on data presented in Diessel and Tomasello 2005)

They argued that aspects of their results were not consistent with filler-gap distance processing accounts. For instance, children found S-relatives easier to process than A-relatives, despite identical filler-gap distances. Instead, they proposed an account where the frequency of structures and the similarity between structures in the input were responsible for creating the hierarchy in development. For example, subject relatives (S + A) are easier than P-relatives, they claim, because the head noun expresses the actor (or agent) of the relative clause just like the sentence-initial NP in simple transitive clauses and these structures are highly frequent in child-directed speech. OBL- and IO-relatives, on the other hand, are difficult because they are highly infrequent in the input.

3. Modeling the acquisition of relative clauses

Diessel and Tomasello's account focused on aspects of the input in explaining the hierarchy in development. It is challenging to experimentally link developmental behavior directly to the input, because it is difficult (and unethical) to manipulate a child's natural input over development. Hence, we examined how the input might influence the AH within a computational model of syntax acquisition. The model we used was the Dual-path sentence production model of Chang, Dell & Bock (2006). This connectionist model was built from a simple recurrent network (Elman 1990) augmented with a second processing pathway in which the sentence message was represented for production (Figure 2).

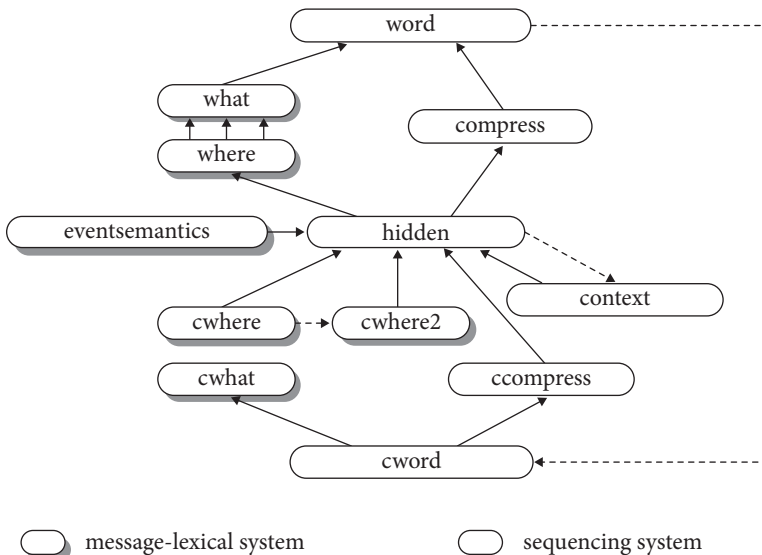


Figure 2. The Dual-path model architecture (Chang, Dell and Bock 2006)

In training, the model was exposed to sentences paired with their meanings. It learned the syntax of the input language by mapping these meaning representations (message) onto the target sentence forms. The model suggested several ways in which the input might influence AH behavior. For example, the model's simple recurrent network was sensitive to subsequences of syntactic categories (e.g., "THAT ARTICLE NOUN") and, therefore, performance differences between relative clause constructions could be due to the fact that they are composed of different subsequences occurring with different frequencies. To examine this, we selectively manipulated the frequency of particular subsequences in the model's input to determine how they related to the AH. Another feature of the model is that it was designed to learn syntactic alternations, where two surface structures are associated with a similar meaning (e.g., active transitives "the man chased the dog" and passive transitives "the dog was chased by the man"). These structures can interfere with each other because similar messages mapped to different sentence forms and the model had to encode both mappings in the same set of connection weights in order to learn the transitive alternation. Since structures in the AH differ in the number of alternations in which they participate (e.g., none for intransitives, one for transitives and two for datives), this interference could influence the model's AH behavior. By examining how frequency, interference, and meaning relate within a particular account of syntax acquisition, we hope to make more explicit how cross-linguistic regularities like the AH might be influenced by the input.

For the current task we extended the Dual-path model to accommodate multi-clause utterances. The message input to the model used three components, thematic roles (AGENT, PATIENT, RECIPIENT, etc.), concepts (lexical semantics), and event features to signal the number and relative prominence of event participants. Before production began, the message was encoded by binding thematic roles (WHERE layer) to concepts (WHAT layer), and the appropriate features in the EVENT SEMANTICS layer were activated. An important aspect of relative clauses is that participants can have different roles in the main clause and the embedded clause. For example, in the sentence "the cat that the dog chased climbed the tree", the "cat" is the agent subject of climbing and the patient object of chasing. To encode the different roles of the "cat", separate clause-specific roles (e.g., AGENT1, PATIENT2) were linked to the concept CAT. To signal that the same cat individual is involved in both events, a special set of event semantic features was used. For example, the message for A-relatives ("the man that chases the dog") contained a feature which bound the head noun "the man" in the main clause event to the transitive agent of the subordinate clause event. In a P-relative ("the man that the dog chases"), another feature bound "the man" to the patient role in the relative clause. This co-reference information uniquely encoded the meanings

of the target utterances from the AH. It helped the model to identify which event participant in the message was modified by a relative clause in the target sentence. In similar vein, it informed the model which event participant was relativized in the embedding and thus allowed it to produce word sequences corresponding to a subject or object relative clause, respectively (for details about this type of coding, see Chang 2009 and Fitz 2009).

3.1 Language and method

The language we used to train the model contained the basic structures needed to reproduce the processing hierarchy, including transitive and ditransitive alternations (Table 2). Active and passive transitives were distinguished in the message by differences in the activation values of the participant features for the agent and patient in the model's EVENT SEMANTICS. For passive sentences the patient feature was more active than the agent feature to bias the model towards early production of the patient noun phrase. The dative alternation was encoded on the theme and recipient features in a similar way. Like the test items in the Diessel and Tomasello study, the multi-clause constructions that the model was exposed to had a relative clause attached to the predicate nominal of a presentational clause (e.g., "There is a boy that runs."). Sentences with relative clauses were assembled from presentationals and the structures listed in Table 2. In relative clauses all participant roles of the underlying construction could be relativized. The head noun of dative constructions, for example, could be the agent, theme or recipient of the relative clause:

- (2) There is a girl that throws the cat the stick. (agent)
- (3) There is the stick that the girl throws the cat. (theme)
- (4) There is the cat that the girl throws the stick. (recipient)

The input grammar had verb tense and aspect, and these were coded by inflectional morphemes that were treated as separate words (e.g., give -ing, run -ed). The lexicon contained 56 words in 14 categories which allowed the creation of roughly 2.4×10^6 different sentences.

The model was trained on a set of 10,000 sentences that were randomly selected from this artificial language and tested periodically on 500 novel sentences after every 1,000 training items. Test sentences were randomly generated from the five relative clause construction types (100 each) which were used in the Diessel and Tomasello experiment (see Table 1).

Table 2. Basic construction types in the language to train the Dual-path model

Structure	Example sentence
Presentational	There is a boy
Intransitive	The cat was sleep -ing
Active transitive	The woman kick -ed the teacher
Passive transitive	The teacher was kick -ed by the woman
Prepositional dative	A girl throw -s the stick to the cat
Double object dative	A girl throw -s the cat the stick
Oblique	The nurse is play -ing with a dog
Relative clause	There is a boy that the woman chase -s

3.2 Modeling results

Using this input language and training conditions, we replicated the relative clause hierarchy in the Dual-path model (Figure 3). The x-axis here represents the number of trained sentences, the y-axis indicates the model’s accuracy in producing the five tested structures.

Sentence accuracy was measured in terms of perfect match, ignoring minor errors such as wrong determiners, or verb tense and aspect. Thus, performance was evaluated in a similar way as in Diessel & Tomasello (2005). All results reported here graphically are averaged over 10 different model subjects where each subject was exposed to a different randomly generated training set. As Figure 3 indicates, relative clause constructions in the model developed in the same order

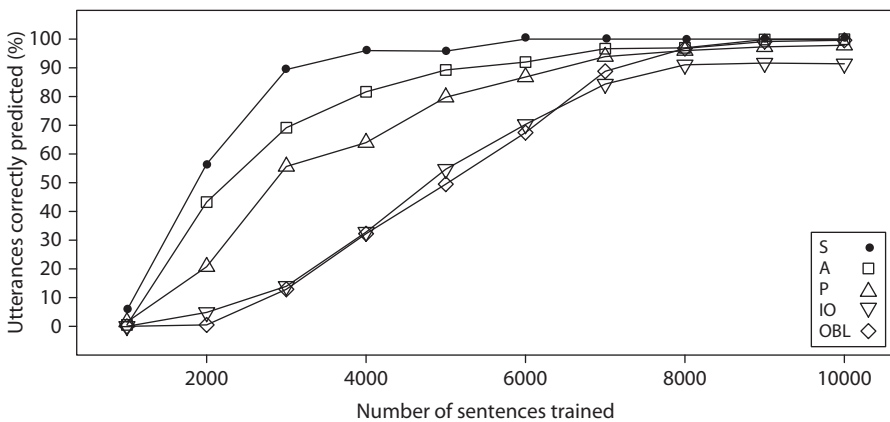


Figure 3. The order of relative clause acquisition in the Dual-path model corresponded to the positions on the accessibility hierarchy

as in children according to the Diessel and Tomasello study. At the end of training, the model reached an adult state where it could accurately produce all of the tested sentence structures (>90%).

To explore what role the input played in creating the hierarchy, we manipulated the model's learning environment, but used the same test set throughout. Therefore, the filler-gap distances in the tested items remained the same across input manipulations. A processing account would predict that the AH should be robust over small changes in the input since processing difficulty on this view depends on structural properties of the test items rather than linguistic experience. If it is possible, however, to change the AH in the model, then learning might play a larger role in the development of the AH than previously thought.

3.2.1 *The S>A contrast*

First, we focused on the contrast between S- and A-relatives in a model which was trained on the full language, i.e., containing all permissible relative clause types. In the AH condition, S- and A-relatives differed on several features such as length, frequency, binding information, and participation in syntactic alternations. If we can determine which of these features are important in the model's S>A behavior, that might indicate how the human syntax acquisition system could be influenced by these factors. Input in the hierarchy condition of Figure 3 made several assumptions about the frequency of different structures. To see how these assumptions influenced the S/A difference in the model, we equated the frequency of structures in the learning phase. Another difference between S- and A-relatives was their length; S-relatives tended to be shorter. Thus, we balanced sentence length in the five test structures, e.g.,

- (5) There is the man that runs in the park at night. (S-relative)
- (6) There is a man that chases a dog down the hill. (A-relative)

The results from training the model in both input conditions are jointly shown in Figure 4. For equal frequencies, S-relatives were still learned significantly faster than A-relatives. When sentence length was balanced, we found a similar pattern, except that the learning of both structures was delayed and the end-state accuracy decreased slightly for both structures. This suggests that the difference between S- and A-relatives observed in the condition of Figure 3 was not due to overall length or frequency.

A third difference between the two structures lies in the meaning information they require to be learned. A- and P-relatives differ in terms of the relative clause word order and the canonical position of the putatively gapped element. Therefore, in order for the model to be able to produce these structures correctly, there

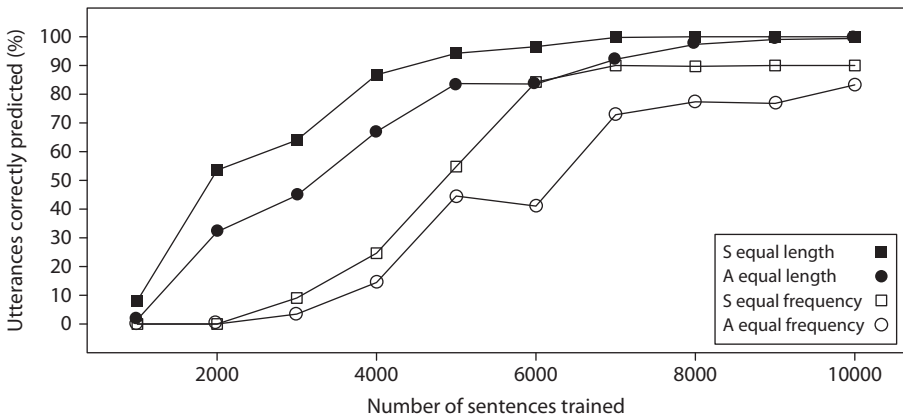


Figure 4. The S>A difference persisted when frequency and length of all tested constructions were balanced

had to be a feature that marked the gapped element in the message. Without this information, the model could not decide whether to produce an A- or a P-relative. S-relatives on the other hand do not exhibit this kind of ambiguity. Hence, part of the S>A difference may be due to the dependence of the A-relatives on additional meaning information. To examine how much these constructions depended on the message, we ran a condition without any role or co-reference information in the EVENT SEMANTICS. As shown in Figure 5, this model had trouble learning most of the constructions, except for S-relatives which were still learned to an adult degree. This suggests that the model found it easier to convey messages which were unambiguously associated with one structure versus those, like A-relatives, which were competing with other structures in the language.

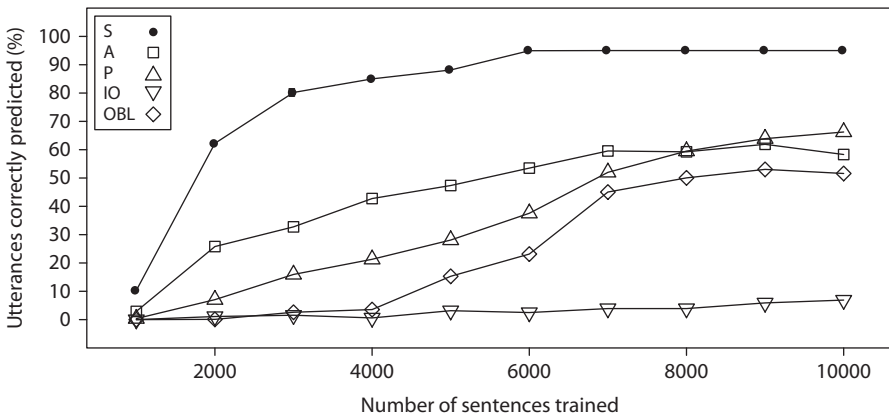


Figure 5. Removing participant roles and binding information from the message did not eliminate the S>A difference in the model

The accuracy of S-relatives was rather insensitive to this message manipulation. To show that the input is critical for explaining the S>A difference in the model, we would like to be able to remove this difference by just manipulating properties of the input. Since the S>A difference was robust over changes in the message and when length and frequency were equated, a more radical manipulation of the input was needed. First, we reduced the frequency of S-relatives to half of the frequency of A-relatives. This reflects the fact that events described by A-relatives have twice as many participants as events described by S-relatives (two versus one). Secondly, we removed input structures that made A-relatives difficult to learn, namely passive transitives. Passive transitives complicate the meaning-to-form mapping the model has to acquire in that they invert the sequence of event participants in the active sentence surface form. When both factors were combined, the model learned A-relatives as fast as S-relatives (Figure 6). Hence, even though the model had a strong bias towards S-relatives over all other structures in the hierarchy, this bias could be erased by manipulating the model's input distribution. This demonstrates that the S>A difference in development may not be maintained in a learning system if the input does not also support that difference.

To summarize, we found that the S>A difference was not due to frequency or sentence length alone. Instead, the S>A difference seemed to be due to inherent factors, like the number of participant roles, but also due to the learning problem posed by the existence of multiple ways of conveying the same meaning, as in the active/passive transitive alternation. When both these factors were controlled for in the input distribution, the S>A difference could be eliminated.

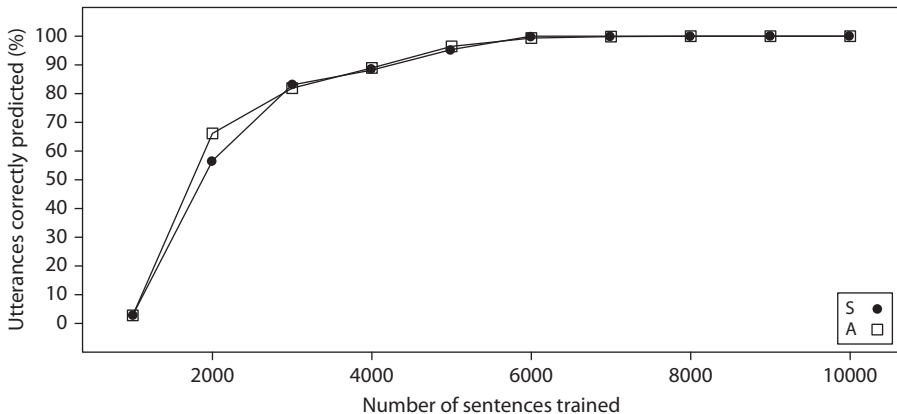


Figure 6. S-relatives equaled A-relatives when S-frequency was reduced and passive transitives were removed from the input language

3.2.2 *The A>P contrast*

In the input condition that created the AH in development (Figure 3), the model performed significantly better on A-relatives than on P-relatives despite their equal frequency of occurrence in the training set. This behavior is in line with many comprehension studies which have found that object-relativized structures are harder to process than subject-relativized structures, both for adults and children across many languages. Processing accounts such as Just & Carpenter (1992) and Gibson (1998) argued that this asymmetry was due to a processing bias against object-relativized structures which require more cognitive resources to integrate the head noun at the gap position.

Diessel & Tomasello (2005) suggested an alternative account of the A>P difference based on the surface sequence of semantic roles in these structures. In our account we focused on differences in sequences of word categories. A-relatives contain the subsequence “THAT VERB” while P-relatives contain the subsequence “THAT ARTICLE NOUN”. Since all of the relative clause structures can relativize subjects, “THAT VERB” substructures might be more common than “THAT ARTICLE NOUN” in a learner’s linguistic environment. If speakers are sensitive to the frequency of substructures, this could help explain the A>P difference. To explore how substructure frequencies relate to the A>P difference, we manipulated these frequencies in the model’s training set. The model should be sensitive to substructure frequencies, because it used a simple recurrent network architecture that learned statistical relationships between sequences of adjacent syntactic categories (Elman 1990; Chang 2002). When we reduced the frequency of “THAT VERB” by reducing the frequency of subject-relativized datives (see sentence (2) above) and increased the frequency of “THAT ARTICLE NOUN” by increasing the frequency of object-relativized datives (see sentences (3) and (4) above), we were able to remove the A>P difference (Figure 7). Manipulating dative frequencies only allowed us to leave the transitive frequencies intact and demonstrate that it was the substructure, rather than construction, frequency that was critical for creating the A>P difference.

If this model-based account is viable, we can predict that “THAT VERB” substructures should also be more frequent than “THAT ARTICLE NOUN” in the input to English speaking children. In our analysis of the mother’s speech in a dense English corpus (Lieven *et al.* 2003), we found 157 examples of “ARTICLE WORD THAT VERB” (where the category VERB contained only verbs morphologically marked by -ed or -es). But when we searched for cases like “ARTICLE WORD THAT ARTICLE”, we found only 67 instances. Therefore, even without auxiliaries and verbs marked for plural agreement, “THAT VERB” is more common than “THAT ARTICLE NOUN” in this corpus of child-directed speech. This

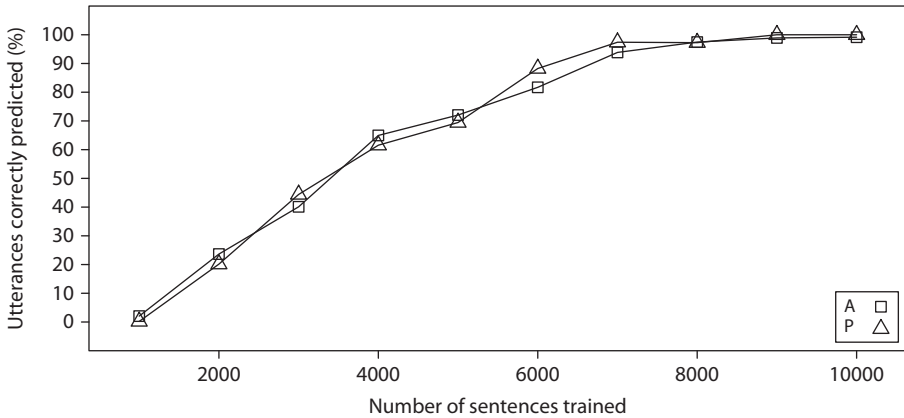


Figure 7. A-relatives equaled P-relatives when substructure frequencies were balanced by adjusting the dative relativization ratios

provides support for the substructure account of the A>P difference and suggests that the model can be useful in determining what kinds of substructure units to search for in a corpus analysis.

3.2.3 *The P > IO = OBL contrasts*

The performance differences for P-, IO- and OBL-relatives of Figure 3 could be similarly reduced or even inverted by changing the model's input distribution. Each of these constructions was influenced by several distinct factors in complex ways. Since these constructions were not significantly different from each other in the Diessel and Tomasello data, we only report the factors which seemed to have the strongest effect on each construction in the model's behavior. P-relatives were influenced by several of the factors we have mentioned in earlier sections, but in addition, they were also strongly influenced by the frequency of subject-relativized passives (e.g., "there is a man that was chased by a dog"). Although these structures are infrequent in child-directed speech, children must hear them or related structures in order to acquire an adult grammar. We found that increasing the frequency of subject-relativized passives reduced the accuracy of P-relatives. This effect could further be amplified when we made active and passive transitives less distinct in their message representation (see Section 3.1). The result of this manipulation is shown in Figure 8 (top) after training for 5000 sentences. P-relatives went down to the accuracy level of IO- and OBL-relatives in the hierarchy condition of Figure 3. As with the P-relatives, IO-relatives were sensitive to the demands of mapping similar messages onto two structures (the dative alternation). By removing the

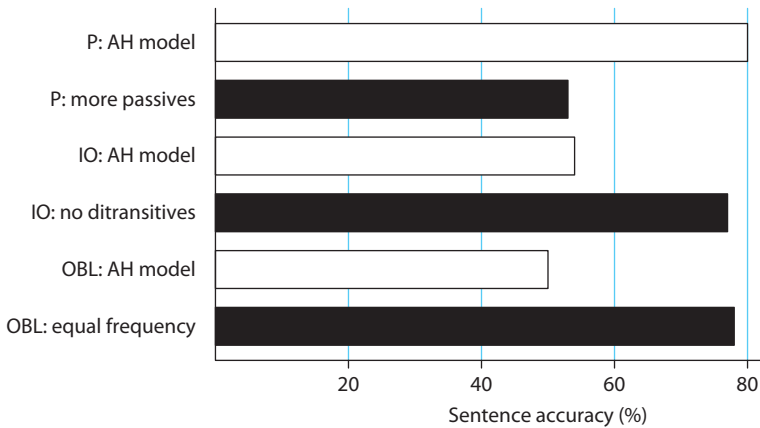


Figure 8. Distinct factors influenced the learnability of P-, IO-, and OBL-relatives in the model after training on 5000 items

ditransitive construction (e.g., “there is the dog that the girl gave a toy”) from the input, we increased the accuracy of IO-relatives to the level of P-relatives (Figure 8, middle).

The OBL-relative construction, on the other hand, was most sensitive to frequency because it is not in direct competition with other input structures; there is no syntactic alternation for obliques. Since OBL-relatives shared semantic similarities with S-relatives in the message, they were easily learnable by the model when the frequencies of these constructions were equal (Figure 8, bottom). Hence, the model’s account of the low OBL-relative accuracy required that these structures are much less frequent than S-relatives in the input. Support for this account comes from a corpus study by Diessel (2004) which found that out of all of the relative clauses in a corpus of child-directed speech, 35.6% were S- or A-relatives, whereas only 7.6% were OBL-relatives.

3.2.4 *Eliminating the relative clause hierarchy*

If filler-gap distances are not crucial for creating the hierarchy, we should be able to find an input condition in which the model learns a language that does not display the AH in development. We achieved this by creating an input environment with only single-clause utterances and sentence tokens of the five tested structures in training. This manipulation removed any effect of syntactic alternations on AH behavior and limited the relativization possibilities by removing subject-relativized obliques and subject and theme-relativized prepositional datives. To equate for the number of roles in the embedded clause, we made the frequency of each relative clause construction in the input proportional to the number of its participant roles. When trained in

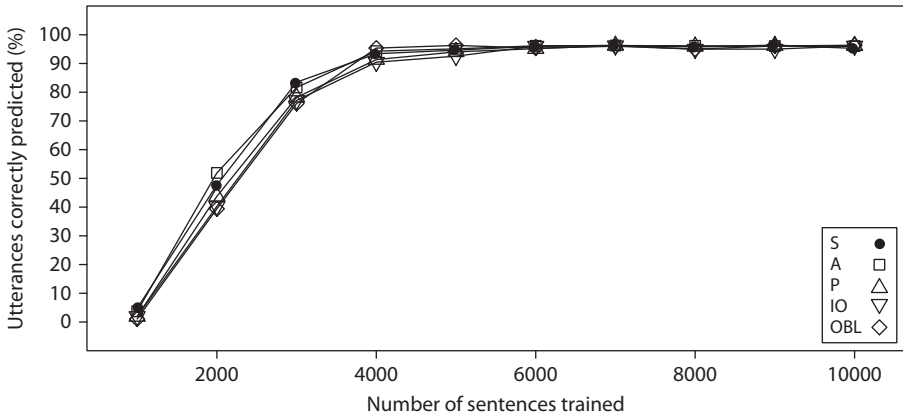


Figure 9. When the input language did not contain alternations, or structures which relativized competing roles, the hierarchy was erased

this condition the hierarchy disappeared (Figure 9), i.e., the model acquired all structures in the AH at the same rate and reached the same end-state accuracy.

This experiment shows that we controlled all the relevant factors which influence the AH over development in the model. When only the structures from the hierarchy were in the input, the same model which previously matched the order of relative clause acquisition in children (Figure 3) now displayed no bias towards any particular structure. The stepwise elimination of the hierarchy behavior indicates that patterns of interference and facilitation between the tested items from the AH and constructions in the language outside the test set brought about the hierarchy in development. Such interactions suggest that the processing difficulty of specific relative clause structures in acquisition may not be quantifiable in isolation from the rest of the input language by applying some universal metric rooted in notions of syntactic complexity. Rather, it was the diversity of the total input language as filtered through the architecture of our connectionist model which made some structures harder to learn and process than others.

4. From acquisition to adult processing

We showed that a neural network model of syntax acquisition and sentence production was able to exhibit evidence of the AH in syntactic development when given English-like input. However, when that input language was distorted, such that it no longer resembled a natural language, the model's developmental behavior was also distorted. We argued that key properties of human languages, such as

the existence of structural alternations, similarity in meaning between different constructions, and consistent frequency across different languages, may play a part in making the AH a cross-linguistic feature of human languages.

In addition to providing an account for AH behavior in development, the model suggests how the mechanisms proposed in experimental work (Diessel & Tomasello 2005; Brandt, Diessel & Tomasello 2008) might be implemented. Diessel and Tomasello, for example, explain structural errors in their data by stipulating that S/A-relatives are easier to activate than other structures. The model suggests that the frequency of the substructure “THAT VERB” over “THAT ARTICLE NOUN” across all of the constructions in the language is partially responsible for the ease of activating S/A-relatives. These types of substructure representations were learned from examples, because the model’s simple recurrent network architecture attended to local statistical regularities in word sequences.

The model not only implements mechanisms that have been proposed in the literature, but also emphasizes factors in the AH that have not been considered important. One such factor is syntactic alternations. The model was designed to map from meaning representations to sentence forms and to handle syntactic alternations, which were included in our language input. But it was found that alternations tended to complicate the generation of forms and this seemed to be important for explaining developmental patterns for different constructions. Evidence for the hypothesis that constructions can mutually support or hinder each other in acquisition has also been found by Abbot-Smith & Behrens (2006) for the German passive and in a connectionist model of the development of grammatical relations (Morris *et al.* 2000). Therefore, experimental work on the AH structures might profit from looking at the influence of syntactic alternations on relative clause development.

Accounts of the AH have focused on processing difficulty as the driving force behind the hierarchy. But work with the Dual-path model, which is a sentence processor with a limited capacity memory, indicates that the AH is not an inevitable consequence of sentence processing. No matter how complex a structure is, a model that learns its representations from examples can recode this structure in a way that requires a minimal amount of memory. This suggests that the learning mechanism may play an important role in determining the complexity of syntactic representations.

4.1 Experience and relative clauses processing

The model we have presented has stressed the role of experience and the interaction between relative clause constructions in explaining developmental patterns of processing found in the literature. But what is the implication of this view for adult processing? To answer this question we draw on a related connectionist model of

recursive sentence processing in adulthood by Christiansen & MacDonald (2009). This model shares many similarities with the previously described Dual-path model in that it relies on a simple recurrent network (Elman 1990) to predict words in an utterance (but without the key production component of the Dual-path model). Compared with the AH model, this model was trained on a relatively more complex grammar involving different kinds of recursive structure in the form of prenominal possessive genitives, relative clauses, sentential complements, prepositional modifications of NPs, and NP conjunctions. The grammar also incorporated subject noun/verb number agreement and three verb argument structures (transitive, optionally transitive, and intransitive). The model developed a usage-based notion of constituency that allowed it to generalize agreement patterns in sophisticated ways. It also acquired a usage-based notion of recursion, whereby its ability to process recursive structure was highly affected by their presence in the input and the interaction between such constructions (similarly to what was observed in the Dual-path model above).

Importantly, the model made a number of predictions regarding the processing of recursive structure – predictions which were subsequently confirmed by human experimentation. Of particular relevance to this chapter were predictions regarding the processing of sentences involving doubly center-embedded object relative clauses, such as (7) and (8):

- (7) The apartment that the maid who the service had sent over was cleaning every week was well decorated.
- (8) *The apartment that the maid who the service had sent over was well decorated.

The sentence in (8) is identical to the one in (7) but with the middle VP deleted (i.e., “was cleaning every week”). In an off-line grammaticality rating study, Gibson & Thomas (1999) found that even though (8) is ungrammatical, it was rated no worse than the grammatical version in (7). However, Christiansen & MacDonald’s (2009) model suggested that (8) should actually be rated as better than (7) because the network experienced less processing difficulty for the ungrammatical sentence compared to the original version. They then tested this prediction in an on-line grammaticality rating experiment, in which participants read sentences such as (7) and (8), one word at a time, and then rated their goodness (as sentences of English) at the end of each sentence. The results showed that subjects under such on-line processing conditions rated the ungrammatical version (8) as being significantly better than its grammatical counterpart (7). Christiansen and MacDonald further replicated the study with stimuli controlled both for length and semantic cues, as in (9) and (10), and obtained the same results, thus again confirming the predictions from the connectionist model.

- (9) The chef who the waiter who the busboy offended appreciated admired the musicians.
- (10) *The chef who the waiter who the busboy offended frequently admired the musicians.

The usage-based perspective on relative clause processing embodied in both models discussed in this chapter predicts that there should be substantial differences in relative clause processing in development as well as adulthood, and that such differences are likely to have a substantial experiential factor. Consistent with these predictions, Roth (1984) demonstrated in a training study that preschoolers' ability to comprehend relative clauses could be improved considerably by way of increased exposure to such constructions. Using a simple recurrent network model, MacDonald & Christiansen (2002) further demonstrated the importance of linguistic experience for the processing of object relative clauses in particular, because these constructions require direct exposure to be processed efficiently, whereas subject relative sentence can piggyback on the processing of the simpler, and very frequent, transitive sentences. This predicted asymmetry with regard to the role of direct experience for subject and object relative clauses was subsequently confirmed in a training study with human subjects (Wells *et al.* 2009). More generally, as a proxy of linguistic experience, educational background is strongly correlated with the ability to comprehend complex recursive constructions, including those involving relative clauses (Dabrowska 1997).

Our usage-based connectionist approach also suggests that linguistic experience may shape the processing system itself because the same weights are used both to represent the network's experience with language and to process such language. We might therefore expect that such a close link between experience and processing ability could lead to considerable differences in relative clause processing across languages with different distributional properties (Engelmann & Vasishth 2009; Hakuta 1981). As a case in point, Hoover (1992) found that comparable doubly center-embedded object relative clauses were more easily processed in Spanish than in English (even without morphological cues in Spanish). Similarly, substantial differences in perceived processing difficulty for the exact same types of relative clause structure have been demonstrated across a variety of other languages – including English, German, Japanese, and Persian (Hawkins 1994) – again highlighting the importance of linguistic experience in explaining recursive processing abilities. Thus, as originally suggested by Stolz (1967), the ability to use specific relative clause constructions recursively appears to be acquired in an item-based manner.

5. Conclusion

In this chapter, we have outlined a connectionist perspective of the acquisition and processing of relative clauses within a usage-based framework. English relative clause constructions give rise to similar orderings of differential processing in adult comprehension and language production in development. This pattern matches the typological cross-linguistic pattern called the noun phrase accessibility hierarchy (Keenan & Comrie 1977). We have proposed an input-based explanation of this data here. The Dual-path model displayed the ordering of the hierarchy in syntactic development when it was exposed to plausible input distributions. But it was possible to manipulate individual contrasts and completely remove this ordering by varying properties of the input from which the model learned. This suggests, we argued, that patterns of interference and facilitation between input structures can explain the accessibility hierarchy in processing and development when all structures are simultaneously learned and represented over a single set of connection weights in a neural network model. Many approaches to explaining relative clause acquisition and processing behavior agree that it is necessary to specify how the language system deals with the syntactic representations for the particular relative clause structure that is being processed. What is unique to our approach is the idea that one cannot understand the processing of a particular structure without considering the whole space of other structures that are similar to it. Thus in contrast to the existing tendency to compare the processing of two structures in acquisition or adults, our approach suggests that more attention needs to be paid to interactions between structurally similar word sequences and constructions.

Our approach also highlights the important link between linguistic experience and processing ability. Consequently, this perspective may allow us to both account for the cross-linguistic patterns of processing embodied in the accessibility hierarchy and the patterns of differences observed across individuals as well as across languages in processing specific relative clause constructions. How does this connectionist perspective compare with other computational approaches to usage-based modeling? Recently, rational, Bayesian, and information-theoretic approaches have become a popular means of modeling language phenomena (Hale 2006; Demberg & Keller 2008). These models are similar to the connectionist models presented here in that they learn from examples and are able to capture structural and word-based expectations. They are analytically more tractable and allow for a broader coverage of linguistic data than most connectionist models. Interesting findings with these models, however, often depend on human-labeled syntactic categories in corpora and this can be a barrier in using these models to study syntax acquisition which, at least in part, involves discovering such categories in the first place. Connectionist models like the Dual-path model and the simple recurrent network, on the other

hand, learn syntactic categories as well as high level relationships between categories and words from experience. Direct comparisons between these approaches are few, but in some cases, it seems that connectionist approaches provide a better match to the human processing data than symbolic approaches (Frank 2009). It would be desirable to combine the strengths of both connectionist and information-theoretic approaches into one modeling framework.

The connectionist approach we presented emphasizes the role of linguistic experience in the acquisition and processing of relative clauses but also the distributional make-up of that experience. The Dual-path model account of relative clause acquisition in English suggested that it was not intrinsic syntactic complexity, biological constraints on working memory, or innate language universals that explained the order of acquisition. Rather, it was brought about by distributional properties of the input. This explanation is supported by two recent studies on the processing of English object relatives (Kidd *et al.* 2007; Reali & Christiansen 2007). They found that production and comprehension were facilitated when the relative clause head noun was inanimate and the relative clause subject a pronoun, as in

(11) There is the toy that he bought at the supermarket today.

and they argued that these effects were due, at least partially, to frequency factors. In the present study, we did not manipulate animacy and pronominality. We predict, however, that our model should display similar differential behavior when inanimate heads are strongly linked to object relatives and pronominal subjects in object relatives are more frequent in the language input than full NP subjects.

Future work should seek to validate this input-driven explanation cross-linguistically for languages with different relative clause systems. German relative pronouns inflect according to gender, case and number and German relative clauses are typically verb-final (in contrast to single-clause sentences). The surface sequence of word categories in transitive subject and object relatives is identical and, in contrast to English, does not signal the grammatical role of the head noun in the relative clause. Instead, the role is marked on the pronoun and this can create ambiguities. In Japanese, relative clauses are pronominal and pronouns are not used. Transitive subject and object relatives differ only with respect to a case particle in their surface form. Across languages, differences in word order, case marking, and relativization strategies will create very different distributional information from which our connectionist models learn. Thus, it remains to be tested whether the input factors we identified for English have a similar explanatory value in the usage-based acquisition of relative clauses in typologically-distinct languages.

References

- Abbot-Smith, K. & Behrens, H. 2006. How known constructions influence the acquisition of other constructions: The German passive and future constructions. *Cognitive Science* 30: 995–1026.
- Brandt, S., Diessel, H. & Tomasello, M. 2008. The acquisition of German relative clauses: A case study. *Journal of Child Language* 35: 325–348.
- Chang, F. 2002. Symbolically speaking: A connectionist model of sentence production. *Cognitive Science* 26: 609–651.
- Chang, F. 2009. Learning to order words: A connectionist model of heavy NP shift and accessibility effects in Japanese and English. *Journal of Memory and Language* 61: 374–397.
- Chang, F., Dell, G.S. & Bock, K. 2006. Becoming syntactic. *Psychological Review* 113: 234–272.
- Chen, J. & Shirai, Y. 2010. The development of relative clauses in spontaneous child speech in Mandarin. Paper presented at the 34th Annual Boston University Conference on Language Development.
- Christiansen, M.H. & Chater, N. 2001. Connectionist psycholinguistics in perspective. In *Connectionist Psycholinguistics*, M.H. Christiansen & N. Chater (eds), 19–75. Westport CT: Ablex.
- Christiansen, M.H. & MacDonald, M.C. 2009. A usage-based approach to recursion in sentence processing. *Language Learning* 59 (Suppl.1): 126–161.
- Church, K. 1982. *On Memory Limitations in Natural Language Processing*. Bloomington IN: Indiana University Linguistics Club.
- Dabrowska, E. 1997. The LAD goes to school: A cautionary tale for nativists. *Linguistics* 35: 735–766.
- Demberg, V. & Keller, F. 2008. Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition* 109: 193–210.
- Diessel, H. 2004. *The Acquisition of Complex Sentences*. Cambridge: CUP.
- Diessel, H. 2009. On the role of frequency and similarity in the acquisition of subject and non-subject relative clauses. In *Syntactic Complexity: Diachrony, Acquisition, Neuro-cognition, Evolution* [Typological Studies in Language 85], T. Givón & M. Shibatani (eds), 251–276. Amsterdam: John Benjamins.
- Diessel, H. & Tomasello, M. 2005. A new look at the acquisition of relative clauses. *Language* 81: 882–906.
- Elman, J.L. 1990. Finding structure in time. *Cognitive Science* 14: 179–211.
- Engelmann, F. & Vasishth, S. 2009. Processing grammatical and ungrammatical center embeddings in English and German: A computational model. In *Proceedings of 9th International Conference on Cognitive Modeling*, A. Howes, D. Peebles & R. Cooper (eds.). Manchester.
- Fitz, H. 2009. Neural Syntax. PhD dissertation, University of Amsterdam.
- Frank, S.L. 2009. Surprisal-based comparison between a symbolic and a connectionist model of sentence processing. In *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, N.A. Taatgen & H. van Rijn (eds.), 1139–1144. Austin TX: Cognitive Science Society.
- Gibson, E. 1998. Linguistic complexity: Locality of syntactic dependencies. *Cognition* 68: 1–76.
- Gibson, E. & Thomas, J. 1999. Memory limitations and structural forgetting: The perception of complex ungrammatical sentences as grammatical. *Language and Cognitive Processes* 14: 225–248.
- Gupta, P. 2008. The role of computational models in investigating typical and pathological behaviors. *Seminars in Speech and Language* 29: 211–225.

- Hale, J. 2006. Uncertainty about the rest of the sentence. *Cognitive Science* 30: 643–672.
- Hakuta, K. 1981. Grammatical description versus configurational arrangement in language acquisition: The case of relative clauses in Japanese. *Cognition* 9: 197–236.
- Hawkins, J.A. 1994. *A Performance Theory of Order and Constituency*. Cambridge: CUP.
- Hoover, M.L. 1992. Sentence processing strategies in Spanish and English. *Journal of Psycholinguistic Research* 21: 275–299.
- Hsiao, F. & Gibson, E. 2003. Processing relative clauses in Chinese. *Cognition* 90: 3–27.
- Just, M.A. & Carpenter, P.A. 1992. A capacity theory of comprehension: Individual differences in working memory. *Psychological Review* 99: 122–149.
- Keenan, E.L. & Comrie, B. 1977. Noun phrase accessibility and universal grammar. *Linguistic Inquiry* 8: 63–99.
- Keenan, E.L. & Hawkins, S. 1987. The psychological validity of the accessibility hierarchy. In *Universal Grammar: 15 Essays*, E.L. Keenan (ed.), 60–85. London: Croon Helm.
- Kidd, E., Brandt, S., Lieven, E. & Tomasello, M. 2007. Object relatives made easy: A crosslinguistic comparison of the constraints influencing young children's processing of relative clauses. *Language and Cognitive Processes* 22: 860–897.
- Kimball, J. 1973. Seven principles of surface structure parsing in natural language. *Cognition* 2: 15–47.
- Lieven, E., Behrens, H., Speares, J. & Tomasello, M. 2003. Early syntactic creativity: A usage-based approach. *Journal of Child Language* 30: 333–370.
- MacDonald, M.C. & Christiansen, M.H. 2002. Reassessing working memory: A comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review* 109: 35–54.
- Marcus, M. 1980. *A Theory of Syntactic Recognition for Natural Language*. Cambridge MA: The MIT Press.
- Miller, G.A. & Chomsky, N. 1963. Finitary models of language users. In *Handbook of Mathematical Psychology Vol. 2*, R.D. Luce, R.R. Bush & E. Galanter (eds.), 419–492. New York NY: Wiley.
- Morris W.C., Cottrell, G.W. & Elman, J.L. 2000. A connectionist simulation of the empirical acquisition of grammatical relations. In *Hybrid Neural Systems*, S. Wermter & R. Sun (eds.), 175–193. Berlin: Springer.
- O'Grady, W. 1997. *Syntactic Development*. Chicago IL: The University of Chicago Press.
- Ozeki, H. & Shirai, Y. 2007. Does the noun phrase accessibility hierarchy predict the difficulty order in the acquisition of Japanese relative clauses? *Studies in Second Language Acquisition* 29: 169–196.
- Real, F. & Christiansen, M.H. 2007. Processing of relative clauses is made easier by frequency of occurrence. *Journal of Memory and Language* 57: 1–23.
- Reich, P. 1969. The finiteness of natural language. *Language* 45: 831–843.
- Roth, F.P. 1984. Accelerating language learning in young children. *Child Language* 11: 89–107.
- Stabler, E.P. 1994. The finite connectivity of linguistic structure. In *Perspectives on Sentence Processing*, C. Clifton, L. Frazier & K. Rayner (eds.), 303–336. Hillsdale NJ: Lawrence Erlbaum Associates.
- Stolz, W.S. 1967. A study of the ability to decode grammatically novel sentences. *Journal of Verbal Learning and Verbal Behavior* 6: 867–873.
- Wells, J., Christiansen, M.H., Race, D.S., Acheson, D.J. & MacDonald, M.C. 2009. Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology* 58: 250–271.