The Role of the Input in a Connectionist Model of the Accessibility Hierarchy in Development

Hartmut Fitz and Franklin Chang
Institute for Logic, Language, and Computation, Amsterdam and
NTT Communication Science Laboratory, NTT Corp., Kyoto

1. Introduction

Language universals are important in theories of language, 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, that we examined here, is that these universals arise from the mechanisms of the language learning system. One important syntactic universal in linguistic typology is the accessibility hierarchy on 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 *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 are presented in Table 1). Keenan and Comrie (1977) sam-

Table 1. Summary of English relative clause constructions

Relativized role	Example	Label
Subject intransitive	... the boy that runs	S
Subject transitive	... 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

pled relative clause constructions from 50 languages and based on this data they formulated an implicational universal for all languages. If a language knows a 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* (AH).

Keenan and 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 adults 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 metric for the processing difficulty of 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. 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.

2. The accessibility hierarchy in development

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. For example, German relative pronouns are marked for gender, case, and number. Hence in most sentences with relative clauses, the grammatical role of the gap is resolved at the pronoun position already and the filler need not be kept in working memory. Secondly, these processing accounts
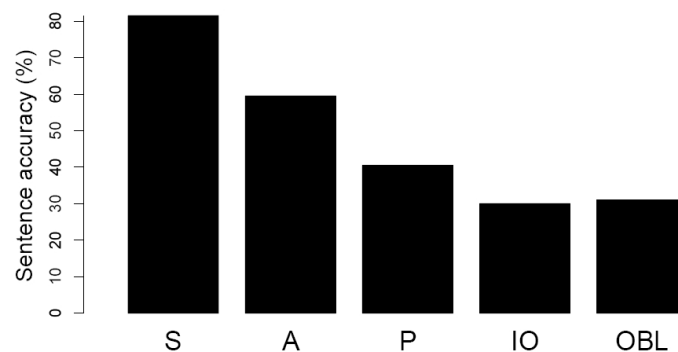

Figure 1. **Relative clause acquisition in production (Diessel and Tomasello, 2005)**

focus on comprehension, but presumably in production no filler integration is required at the gap position because the speaker's intended message is unambiguous.

Another issue that has not been examined carefully is the relationship between filler-gap accounts and acquisition. If children are not using adult-like syntactic representations, then they might not exhibit adult-like AH behavior. In a sentence

repetition study with English children [4;3-4;9], however, Diessel and Tomasello (2005) found that the order of relative clause acquisition in production matches the adult processing hierarchy (Figure 1, similar results were also found in German children). They argued that aspects of their results were not consistent with filler-gap distance processing accounts and instead 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 of the relative clause just like the sentence-initial NP in simple transitive clauses. OBL- and IO-relatives, on the other hand, are difficult because they are highly infrequent in the input.

3. The Dual-path model approach

Diessel and Tomasello's account focused on aspects of the input in explaining the AH hierarchy in development. It is difficult to experimentally link developmental behavior directly to the input, because it is difficult to manipulate a child's 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, and 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). It learned the syntax of a target language by mapping meaning representations (message) onto appropriate sentence forms. The model suggested several ways in which the input might influence AH behavior. 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 subsequences they were composed of. To examine

Figure 2.   The Dual-path model architecture

this, we manipulated the frequency of particular subsequences in the model's input to see 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
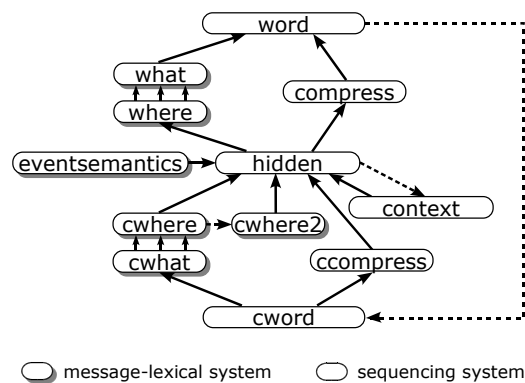
passive transitives *the dog was chased by the man*). These structures can interfere with each other and since structures in the AH differ in the number of alternations they participate in, this interference could influence the AH. By examining how frequency, interference, and meaning relate within a particular account of syntax acquisition, we hope to make more explicit how universals 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 uses three components: thematic roles (AGENT, PATIENT, RECIPIENT, etc.), concepts (lexical semantics), and event features (e.g. the number and relative prominence of participants). Before production begins, the message was encoded by binding thematic roles (WHERE) to concepts (WHAT), and the appropriate features in the EVENT-SEMANTICS were activated. We added information about the co-reference of participants in different events to the message representation of the Dual-path model. For example, the message for A-relatives (*the man that chases the dog*) contained a feature which binds 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.

4. 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). Similar to the test items in the Diessel and Tomasello study,

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

Structure	Example
Presentational	there is a boy
Transitive	the woman kick -ed the teacher
Transitive passive	the teacher was kick -ed by the woman
Intransitive	the cat was sleep -ing
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

multi-clause constructions that the model was exposed to had a relative clause attached to the predicate nominal of a presentational clause. Relative clauses were assembled from presentationals and the above structures, and all participant roles could be relativized. The head noun of dative constructions, for example, could be the agent, theme or recipient of the relative clause. The input grammar had verb tense and aspect, and these were coded by inflectional morphemes that were treated as separate words. 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 from this language, and tested periodically on 500 novel sentences after every 1,000 training items. Test sentences were randomly generated from the five sentence types which were used in the Diessel and Tomasello experiment.

5. The accessibility hierarchy in the Dual-path model

With this input language and training conditions, we replicated the relative clause hierarchy in the Dual-path model (Figure 3). Sentence accuracy was mea-
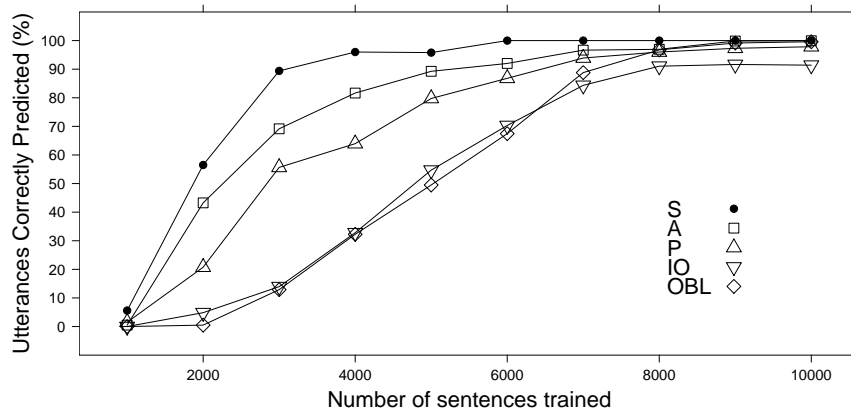
Figure 3. **The order of relative clause acquisition in the Dual-path model corresponded to the positions on the accessibility hierarchy**

sured in terms of perfect match, ignoring minor errors such as wrong determiners, verb tense and aspect. At the end of training, the model reaches an adult state where it can accurately produce all of the tested sentence structures. Thus, relative clause constructions in the model develop in the same order as in children according to the Diessel and Tomasello (2005) study.

To explore what role the input plays in creating the hierarchy, we manipulated the model's input, but used the same test set throughout. Therefore, the filler-gap distances remained the same across input manipulations. A processing account would predict that the AH should be robust over small changes in the input. 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.

5.1. 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. In the AH condition, S- and A-relatives differed on several features such as length, frequency, binding information, and participation in 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 those assumptions influenced the model's S/A difference, we equated the frequency of structures in the learning phase. Another difference between S- and A-relatives was their length. Thus, we balanced sentence length in the five test structures, e.g.

(1) *there is the man that runs in the park at night*   (S-relative)
(2) *there is a man that chases a dog down the hill*   (A-relative)

The results from both conditions are jointly shown in Figure 4. For equal fre-
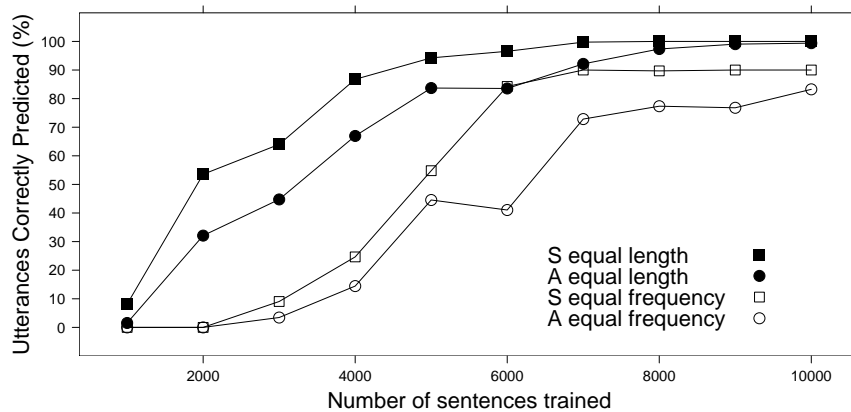
Figure 4. **The S>A difference persisted when frequency and length of all tested constructions were balanced**

quencies, 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. This suggests that the difference between S- and A-relatives is not due to overall length or frequency.

A third difference between the two structures lies in the meaning information they require. A- and P-relatives differ in terms of the position of their gap. Therefore, to be able to produce these structures correctly, there must be a feature that marks the gapped element in the message. Without this information, the model cannot decide whether to produce an A- or a P-relative. S-relatives on the other hand do not have this ambiguity. Hence, part of the S>A difference may be due to the dependence of the A-relative on meaning information. To examine how much these constructions depended on the message, we ran a condition without role and 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 finds it easier to
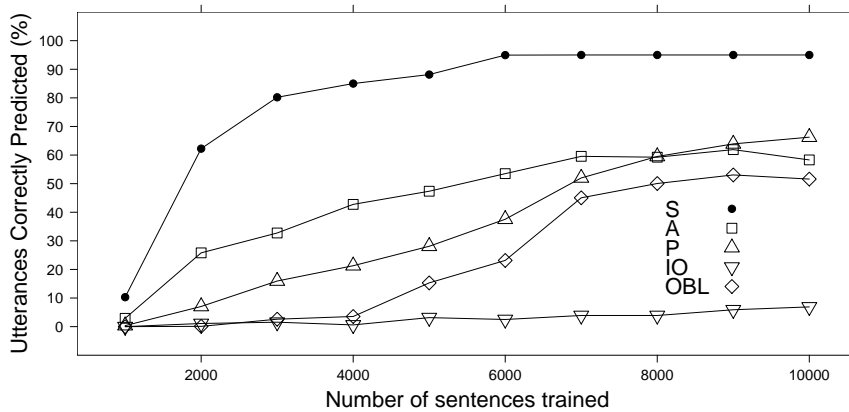
Figure 5. Removing participant roles and binding information from the message did not eliminate the S>A difference in the model

produce messages which are unambiguously associated with one structure versus those, like A-relatives, which compete with other structures in the language.

S-relative accuracy was insensitive to the message manipulation. To demonstrate that the input is critical for explaining the S>A difference, we would like to be able to remove this difference by just manipulating properties of the input. Since the S>A difference is robust over changes in the meaning 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
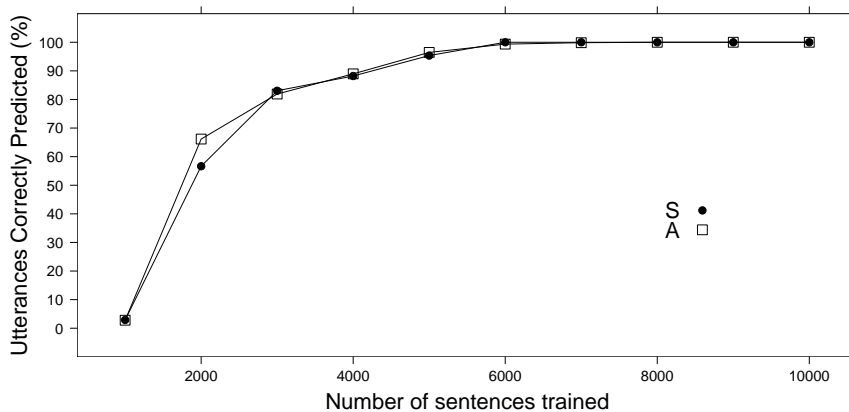
Figure 6. S-relatives equaled A-relatives when S-frequency was reduced and passive transitives were removed from the input language

A-relatives. This reflects the fact that events described by A-relatives have twice as many participants as events described by S-relatives. And we removed input

structures which make 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 has a strong bias towards S-relatives over all other structures in the hierarchy, this bias can 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, the S>A difference seems to be due to inherent factors, like the number of 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.

5.2. The A>P contrast

In the AH condition, the model performed significantly better on A-relatives than on P-relatives despite their equal frequency in the input. 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 languages. Processing accounts such as Just and Carpenter (1992) and Gibson (1998) argued that this asymmetry was due to a processing bias against object-relativized structures which require more cognitive resources.

Diessel and Tomasello (2005) suggested an alternative account of the A>P difference based on the surface sequence of syntactic categories. A-relatives contained the subsequence "THAT VERB" while P-relatives contained 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. The model should be sensitive to substructures, 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 and increased the frequency of "THAT ARTICLE NOUN" by increasing the frequency of object-relativized datives, we were able to remove the A>P difference (Figure 7). Manipulating datives allowed us to leave the transitive frequencies intact and demonstrate that it was the substructure, rather than construction, frequency that was critical for the A>P difference.

If this account is true, we can predict that "THAT VERB" substructures should 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
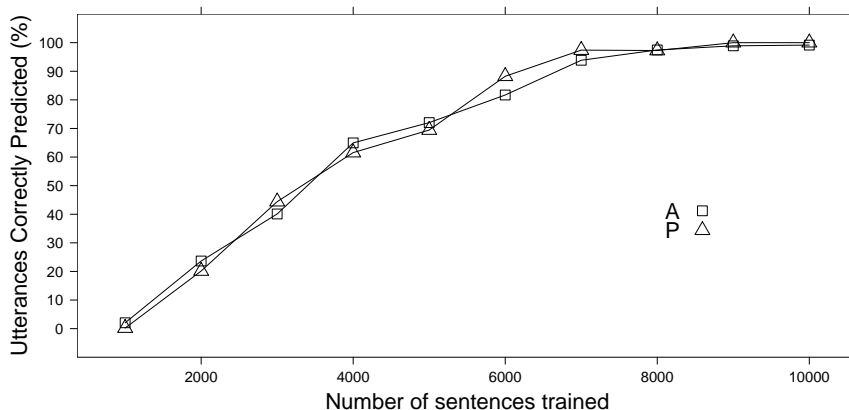
Figure 7. **A-relatives equaled P-relatives when substructure frequencies were balanced by adjusting the dative relativization ratios**

(Maslen *et al.*, 2004), we found 157 examples of "ARTICLE WORD THAT VERB" (where VERB was 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 plural agreement, "THAT VERB" is more common than "THAT ARTICLE NOUN". This provides support for the substructure account of the A>P difference and suggests that the model can be useful in determining what kinds of units to search for in a corpus analysis.

5.3. The P>IO=OBL contrasts

The performance differences for P-, IO- and OBL-relatives can be similarly reduced or even inverted by changing the input language 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 seem to have the strongest effect on each construction in the model. P-relatives were influenced by many 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 can further be amplified if we make active and passive transitives less distinct in their message representation. The result of this manipulation is shown in Figure 8 (top) after training for 5000 sentences. P-relatives go down to the level of IO- and OBL-relatives in the hierarchy condition of Figure 1. As with the P-relatives, IO-relatives were sensitive to demands of mapping similar messages onto two structures (the dative alternation). By removing the ditransitive construction (e.g.,

there is the dog that the girl gave a toy), 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. Since OBL-relatives shared semantic similarities with S-relatives, they were easily learnable in the model when frequencies of constructions were equal (Figure 8, bottom). Hence, the model's account of the low OBL-relative accuracy required that
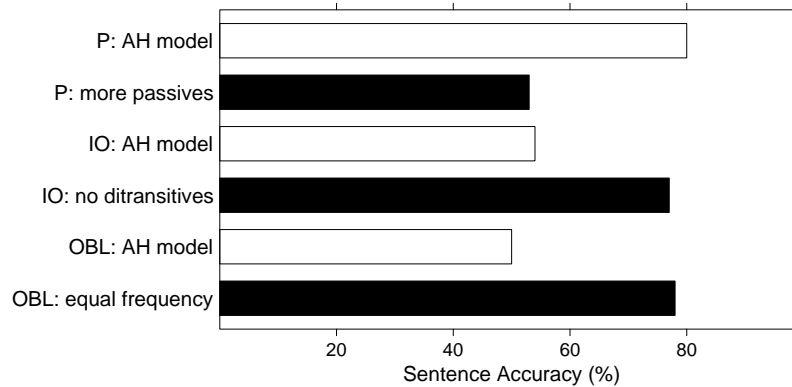

Figure 8. **Distinct factors influenced the learnability of P-, IO-, and OBL-relatives in the model after training on 5000 items**

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/A-relatives, but only 7.6% were OBL-relatives.

6. 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 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 roles. In this condition, the hierarchy disappeared (Figure 9).

This experiment shows that we controlled all the relevant factors which influence the AH over development in the model. If only the structures from the hierarchy are in the input, the same model which previously matched the order of relative clause acquisition in children now behaves entirely neutral with respect
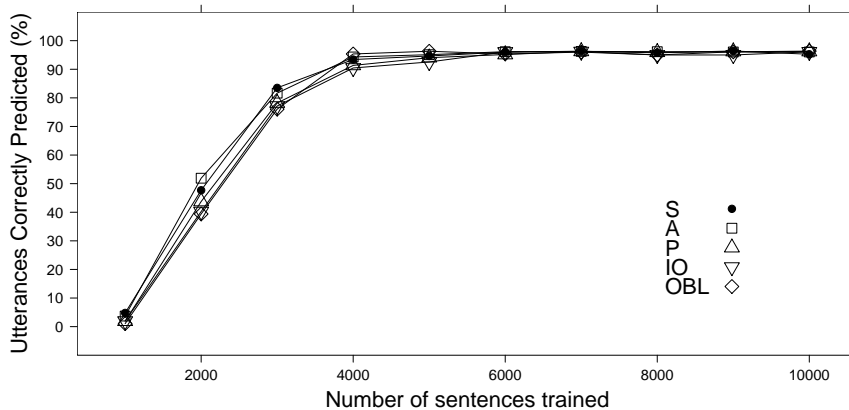
Figure 9. **When the input language did not contain alternations, and no structures with competing roles relativized, the hierarchy was erased**

to these structures. Our stepwise elimination of the hierarchy behavior suggests that patterns of interference and facilitation between the tested items and constructions in the language *outside* the test set bring about the hierarchy in development. Consequently, the processing difficulty of particular relative clause structures in acquisition can not be measured in isolation from the rest of the input language by applying some universal metric rooted in notions of syntactic complexity. Rather, it is the diversity of the total input language as filtered through the architecture of the model which makes some structures harder than others.

7. Conclusions

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 human language, the model's AH behavior was also distorted. We argued that universal 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 universal 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 and Tomasello, 2005; Brandt *et al.*, 2007) might be implemented. For example, Diessel and Tomasello 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 "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, because the model's simple recurrent network architecture attended to local statistical regularities.

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 to forms and to handle syntactic alternations, which were therefore included in our language input. But what we found was that alternations tended to complicate the generation of forms and this seemed to be important for explaining developmental patterns for different constructions. Therefore, experimental work on the AH might profit from looking at the influence of alternations.

Accounts of the universal nature 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 which learns its representations 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.

REFERENCES

Silke Brandt, Holger Diessel, and Michael Tomasello. The acquisition of German relative clauses: A case study. *Journal of Child Language*, 2007. In press.

Franklin Chang, Gary S. Dell, and Kathryn Bock. Becoming syntactic. *Psychological Review*, 113:234–272, 2006.

Franklin Chang. Symbolically speaking: A connectionist model of sentence production. *Cognitive Science*, 26:609–651, 2002.

Holger Diessel and Michael Tomasello. A new look at the acquisition of relative clauses. *Language*, 81(4):882–906, 2005.

Holger Diessel. *The Acquisition of Complex Sentences*. Cambridge Studies in Linguistics. Cambridge University Press, 2004.

Jeffrey L. Elman. Finding structure in time. *Cognitive Science*, 14(2):179–211, 1990.

Edward Gibson. Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68:1–76, 1998.

John Hale. Uncertainty about the rest of the sentence. *Cognitive Science*, 30:643–672, 2006.

John A. Hawkins. *A Performance Theory of Order and Constituency*. Cambridge University Press, Cambridge, 1994.

Marcel A. Just and Patricia A. Carpenter. A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99:122–149, 1992.

Edward L. Keenan and Bernard Comrie. Noun phrase accessibility and universal grammar. *Linguistic Inquiry*, 8:63–99, 1977.

Edward L. Keenan and Sarah Hawkins. The psychological validity of the accessibility hierarchy. In Edward L. Keenan, editor, *Universal Grammar: 15 Essays*. Croon Helm, London, 1987.

Robert J.C. Maslen, Anna L. Theakston, Elena V.M. Lieven, and Michael Tomasello. A dense corpus study of past tense and plural overregularization in English. *Journal of Speech, Language, and Hearing Research*, 47(6):1319–1333, 2004.