Sequential Structure Suffices to Solve Nativist Puzzles

Bod, R.

Publication date
2013

Document Version
Final published version

Published in
Cooperative Minds: Social Interaction and Group Dynamics

Sequential Structure Suffices to Solve Nativist Puzzles

Rens Bod (rens.bod@uva.nl)
ILLC, University of Amsterdam
Science Park 107, Amsterdam, Netherlands

Keywords: language modeling; computational models of language learning.

Nativism versus Empiricism

The debate between hierarchical versus sequential structure in language acquisition has recently flared up again (cf. Frank, Bod & Christiansen 2012; Pesetsky 2013). Roughly, the nativist view on language endorses that human language acquisition is guided by innate rules that operate on hierarchical structures. The empiricist view assumes that language acquisition is the product of abstractions from empirical input but leaves it as an open question whether sequential or hierarchical structure is needed. Some empirical models use sequential structure (e.g. Reali & Christiansen 2005) while other models are based on hierarchical structure (Bod 2009; Bod & Smets 2012).

Much work in empirical language acquisition has focused on a relatively small set of phenomena such as *auxiliary fronting*. For example, Reali & Christiansen (2005) argued that auxiliary fronting could be learned by linear models based on sequential structure, though Kam et al. (2008) showed that the success of these models depend on accidental English facts. Other empiricist approaches have taken the notion of structural dependency together with a combination operation as minimal requirements (e.g. Bod 2009), which overcomes the problems raised by Kam et al. (2008).

In Bod and Smets (2012) it was shown that a much larger set of phenomena can be learned by an empiricist computational model. These phenomena are well-known in the generativist literature (Ross 1967; Adger 2003) and are related to *wh-questions, relative clause formation, topicalization*, *extraposition* and *left dislocation*. It turned out that these hard cases can be learned by an unsupervised tree-substitution grammar induction algorithm that returns the sentence with the best-ranked derivation for a particular phenomenon, using only a very small fraction of the input a child receives.

However, Bod and Smets (2012) also observed that these nativist cases were learned by using relatively shallow structures with little or no hierarchy. This raised the question as to how much structure is actually needed to learn these syntactic constraints. In the current paper, we present a very simple model that reduces all syntactic structuring to concatenations of substrings without any hierarchy. We show that almost all results obtained by the hierarchical grammar in Bod & Smets (2012) can also be learned by means of a sequential grammar using substring-concatenation only.

It should be stressed that the essence of the debate between nativism and empiricism lies often in the relative contribution of prior knowledge and linguistic experience (cf. Lidz et al. 2003; Ambridge & Lieven 2011; Clark and Lappin 2011). Following the nativist view, the linguistic evidence is so hopelessly underdetermined that innate components are necessary. This Argument from the Poverty of the Stimulus can be phrased as follows (see Pullum & Scholz 2002 for a detailed discussion):

(i) Children acquire a certain linguistic phenomenon
(ii) The linguistic input does not give enough evidence for acquiring the phenomenon
(iii) There has to be an innate component for the phenomenon

In this paper we falsify step (ii) for a number of linguistic phenomena that have been considered "parade cases" of innate constraints (Crain 1991; Crain and Thornton 2006). We will show that even if a linguistic phenomenon is *not* in a child's input, it can be learned by a sequential model using only a tiny fraction of child-directed utterances, i.e. the Adam corpus in Childes (MacWhinney 2000).

Methodology

Our methodology is very simple: by means of concatenations of substrings (of parts of speech) of any length from the Adam corpus, we compute from the alternative sentences of a syntactic phenomenon (reported in the generativist literature) the sentence with the *most probable shortest concatenation*. Next, we check whether this sentence corresponds with the grammatical sentence. The shortest concatenation is defined as consisting of the minimal number of substrings (smoothed by the n-shortest concatenations, similar as in Bod and Smets 2012). In case there is more than one shortest concatenation, the most probable one is computed by multiplying the (smoothed) relative frequencies of these substrings in the corpus. For example, given a typical nativist problem like auxiliary fronting, the question is: how do we choose the correct sentence from among the alternatives (0) to (2):

(0) is the boy who is eating hungry?
(1) *is the boy who eating is hungry?
(2) *is the boy who is eating is hungry?

According to Adger (2003), Crain (1991) and others, this phenomenon is regulated by an innate principle. In our approach, instead, we produce all concatenations of substrings that generate (the pos-sequences corresponding to) those sentences. Next, the sentence generated by the most probable shortest concatenation is compared with the grammatical expression.

An Example and Overview of the Results

As an example we will look into the Left Branch Condition (Ross 1967; Adger 2003). This condition has to do with the difference in grammaticality between (3) and (4):

(3) which book did you read?
(4) *which did you read book?

When we let our model generate these two sentences by the shortest combinations of substrings from Adam, we get the respective concatenations (3') and (4'), where for reasons of readability we substituted the pos-tags with the words:

(3') [which book] + [did you read]
(4') [which] + [did you] + [read book]

In this case the shortest concatenation already breaks ties, thus we do not have to compute the most probable shortest concatenation (the latter actually being the typical case).

Table 1 gives an overview of the syntactic constraints/phenomena we have tested so far, and whether these can be successfully explained by the most probable shortest concatenation. The table shows that with only a tiny fraction of a child's input (i.e. just the sentences from the Adam corpus) the correct sentence can be predicted by our simple model for all but two of the phenomena. Our result approaches Bod and Smets (2012) which missed only one phenomenon rather than two, but which relied on a much more complex hierarchical model that induced full-fledged probabilistic tree-substitution grammars. In the future we will therefore also test with larger corpora in Childes.

Table 1: overview of phenomena tested

Phenomenon	**Succesful?**
Subject Auxiliary Fronting	yes
WH-Questions	
Unbounded Scope	yes
Complex NP Constraint	yes
Coordinate Structure Constraint	no
Left Branch Condition	yes
Subject WH-questions	yes
WH in situ	yes
Superiority	yes
Extended Superiority	yes
Embedded WH-questions	yes
WH-islands	yes
Relative Clause Formation	
Complex NP Constraint	yes
Coordinate Structure Constraint	yes
Sentential Subject Constraint	yes
Left Branch Condition	yes
Extraposition from NP	yes
Topicalization	
Complex NP Constraint	yes
Coordinate Structure Constraint	yes
Sentential Subject Constraint	yes
Left Branch Condition	yes
Left Dislocation	
Coordinate Structure Constraint	no
Restriction	yes

References

Adger, D. (2003). *Core syntax: A minimalist approach*. Oxford University Press..

Bod, R., Hay, J. & Jannedy, S. eds. (2003). *Probabilistic Linguistics*, The MIT Press.

Bod, R. (2006). Exemplar-Based Syntax: How to Get Productivity from Examples. *The Linguistic Review*, 23(3), 291-320.

Bod, R. (2009). From Exemplar to Grammar: A Probabilistic Analogy-based Model of Language Learning. *Cognitive Science*, 33(5), 752-793.

Bod, R. & Smets, M. (2012). Empiricist Solutions to Nativist Problems using Tree-Substitution Grammars, *Proceedings Cognitive Models of Language Acquisition and Loss, EACL 2012*, 10-18.

Clark, A. and Lappin, S. (2011). *Linguistic Nativism and the Poverty of the Stimulus*, Wiley-Blackwell.

Crain, S. (1991). Language acquisition in the absence of experience. *Behavorial and Brain Sciences*, 14, 597-612.

Crain, S. & Thornton, R. (2006). Acquisition of syntax and semantics. In M. Traxler and M. Gernsbacher, editors, *Handbook of Psycholinguistics*. Elsevier.

Frank, S., Bod, R. & Christiansen, M. (2012). How Hierarchical is Language Use? *Proceedings of the Royal Society B*, 297(1747), 4522-4531.

Frank, S. & Bod, R. (2011). Insensitivity of the Human Sentence-Processing System to Hierarchical Structure, *Psychological Science*, 22(6), 829-834.

Kam, X., L. Stoyneshka, L. Tornyova, J. Fodor & W. Sakas (2008). Bigrams and the Richness of the Stimulus. *Cognitive Science, 32*, 771-787.

Lidz, J., S. Waxman and J. Freedman (2003). What infants know about syntax but couldn't have learned: experimental evidence for syntactic structure at 18 months. *Cognition*, 89, B65–B73.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk*. Mawah, NJ: Erlbaum.

Pesetsky, D. (2013), *What is to be done?*, Plenary talk, Linguistic Society of America, 1/4/2013.

Pullum, G. & Scholz, B. (2002). Empirical assessment of stimulus povery arguments. *The Linguist Review, 19(2002)*, 9-50.

Reali, F. & Christiansen, M. (2005). Uncovering the richness of the stimulus: structure dependence and indirect statistical evidence. *Cognitive Science, 29*, 1007-1028.

Ross, J. (1967). *Constraints on variables in syntax*. PhD thesis, Massachusetts Institute of Technology.