

**Syntactic structure is NOT simply data-driven learning:
Reply to Akhtar 1999**

One of the most basic tasks in constructing a mental grammar is to determine the syntactic encoding of the relations among people, objects, and events. There seem to be two basic devices available in natural language grammars that encode the grammatical aspect of object permanence, namely *morphological case marking* and the *linear sequencing of phrasal constituents* (i.e., *word order*; henceforth WO).

An acquisition theory within the generative linguistic tradition (Theory of U(niversal) G(rammar)) has been concerned with the acquisition of WO for more than twenty years now (Chomsky 1981, 1986). It relies on the empirical findings of Greenberg (1963), providing a strong thesis about the consistency of the relative position of a head and its complement at a phrase level. In his discussion on language universals, Chomsky (1981) suggests that the relative position of heads and complements for all phrases needs to be specified only once in a given language. Rather than a long list of individual rules specifying the position of the head in each phrase type, a single generalization (a) or b)) suffices: a) *A head comes first in a phrase*, b) *A head comes last in a phrase*.¹

From the learnability point of view, a genetically given UG (as part of a L(anguage) A(cquisition) D(evice)) provides a child with a set of structural parameters that impose strict constraints on the range of structural variation permitted in natural languages and that need to be set on the basis of the linguistic input. The linguistic data hence include triggers that help the child set the parameter. The *Head Parameter*, for example, which conditions the canonical WO in a given language, is believed to come with a binary choice (cf. above) and the child sets it according to the linguistic environment (e.g., SVO in English or French, but SOV in Turkish or Japanese).

¹ Note that besides a powerful generalization, such an approach to head directionality is also *elegant*, which is desired to be a virtue of any scientific theory. It is also simple (and thus plausible) from a learnability perspective, though, of course, simplicity in theorizing does not entail easier and faster learning. Constructionists, on the other hand, as we will see later, do not care about elegance and simplicity at all.

Data-driven theories, on the other hand, place more emphasis on the environmental input rather than biology. They claim that children basically start learning grammar from scratch (a kid as *tabula rasa*) and that learning a language is a long(er) and slow(er) process of inducing (general) grammatical patterns from specific examples (*schema/rote learning*). This is also the approach adopted by Akhtar (1999), which examines children's production of novel and non-novel WOs in child English.

The main claim that Akhtar makes in her article is that “young English-speaking children do not have a general understanding of the significance of SVO order in reversible sentences...[but that] they seem to rely on verb-specific formulas...” (p. 339). She bases her claim on the findings of a study where three groups of kids (two-year olds, three year-olds, and four-year-olds) were asked to comprehend and produce the English and two non-English word orders (SVO vs. SOV/VSO). This paper questions her study on both *conceptual* and *methodological* grounds, arguing against the data-driven approach in child language acquisition in general.

The first conceptual argument against Akhtar's data-driven approach, which claims that child language acquisition is a *slow* and *gradual learning process*, comes from the *speed of acquisition and the nature of errors and non-errors* in child language development. Virtually all studies in child language acquisition conclude that kids seem to acquire the basic word order *very early* and *extremely fast* (cf. Pinker 1984, 1994, Radford 1990, among others). In fact, children produce the right word order *right from the beginning* as soon as they start forming multi-word utterances. Radford (1990) observes that English children as young as 18 months appear to set the head parameter at its appropriate head-first setting from the earliest multiword utterances they produce. As early as at a two-word stage, children produce sentences such as *Mommy eat* or *Eat lunch* (Gleason and Ratner 1998: 366). They never produce sentences such as *Eat mommy* or *Lunch eat* for *Mommy eats/is eating lunch* or *I eat/am eating/want to eat lunch* respectively².

² Children, however, are prone to produce utterances such as *car red*, which might imply at a first glance that they have not set the head parameter correctly or have no head parameter to begin with. However, Bloom 1990 (reference?) found out that such constructions are finite predicative clauses (e.g., *The car is red*), showing different distributional properties than simple noun phrases, e.g., (a) *red car*. When kids do say 'red car', it refers to an NP in an error-free fashion, i. e. with the right word order

Can 2003 observes that Turkish children start off with the correct SOV word order right from the beginning. Akhtar might argue that this ‘big fat’ fact still doesn’t provide a conceptual counterargument to her theory since “...the verbs used in practically all existing [generative] studies...” were always *familiar* verbs which could still be learned on “a verb-by-verb basis” (p.341, 342). However, an extensive line of research using the so-called *preferential-looking technique* (Hirsh-Pasek & Golinkoff 1996, Golinkoff 2002) reports that even with novel verbs, children seem to know which argument in sentences containing transitive verbs has an agentive, patient, or recipient role³. These robust findings suggest that kids at a very young age *must* have some kind of abstract notion of theta-roles and word order.

Interestingly, Akhtar and her colleagues believe that children do *learn* a notion of ‘semantic roles’, associated only with “...*a much more specific schema* that is centered around the particular verb used.” (p. 341, emphasis mine), probably acquired via general cognitive mechanisms on the basis of induction (problem solving)⁴. However, it is *not* intuitive to me where this semantic ‘nounness’ (in the sense of O’Grady 1997) comes from⁵. In other words, *how* is it acquired and *how many trials* are needed for it to be acquired. How does a child know, for example, that the ‘*tamee*’ in *Big Bird is taming Cookie Monster* should be *nominal* at all if he has never heard the verb *to tam* and cannot possibly have a template, such as NP_{tamer} – TAM – NP_{tamee}.⁶ If such schemas or templates truly exist in a child’s grammar, they should be *extremely general* (contrary to Akhtar’s introductory theorizing) and should not be sensitive to specific verb tokens. This conceptualization of *generality*, however, leads us to the knowledge of categories,

³ However, Thornton (p.c.) reminds me that one has to be careful with their data as kids look in both directions – the ‘right’ one and the ‘wrong’ one. Golinkoff and her colleagues usually report a significant statistical difference in kids’ looking in the correct directions as opposed to looking in the wrong direction. This is taken to be evidence that kids know WO, though the data is, in fact, *musty* sometimes.

⁴ For example, a child would acquire the word order on the basis of ‘semantic templates’, e.g., NP_{pusher}-form of the verb PUSH-NP_{pushee} (p.339)

⁵ O’Grady (1997), however, suggests that his ‘general acquisition device (GAD) position’ does presuppose the existence of *categories* and *basic relations between objects, people, and places* (e.g., ‘semantic roles’); though Akhtar and her colleagues (e.g., Tomasello) posit no language device at all and argue against any kind of innate knowledge of categories or relations (Tomasello 2003). All these are learned specifically on a verb-by-verb-basis.

⁶ This is Akhtar’s example from Akhtar & Tomasello (1997) (p. 342). Akhtar’s current paper *never* actually provides the test items used in the study except for a couple of examples (out of 80 trials!). Hence, Akhtar would do us all a favor if she had attached an appendix with all the test items.

semantic roles, and syntactic structure (WO) - in other words, it takes us back to our good old theory of UG.

Another conceptual counterargument immediately comes to mind. If learning truly proceeds on a verb-by-verb basis, how does a child at around two years of age (i.e., in the period of the so-called *vocabulary spurt* when a child can learn as many as 30 words per week, mainly verbs and nouns) *memorize* a humongous number of specific verb formulas and come to the *right* conclusion that each verb token is used in the 'head-first' rather than head-last' fashion? From a psychological point of view, this task would be a very tough one as such learning would be extremely burdensome (from learnability perspective) and would require a lot of memory resources.

Furthermore, I also question the study on methodological grounds, particularly the nature and design of the study and the conceptual grounding behind the study.

First, the instructions in the comprehension task were not controlled for comprehension *at all*. They were presented in the following form: *Make Big Bird dack Cookie Monster* (p.342), expecting a child to *act out* the command with a set of toys. Two-year olds were apparently unable to use the WO prompt correctly, frequently reversing the order of arguments in their interpretations. However, Akhtar does not provide any account of control for this experiment. In other words, we do not know if children *really understood* the structure used on the test items. The structure containing *Make + NPsubj + V[-inf] + NPobj* is a pretty complex structure for two-year olds, requiring quite a lot of processing load. If the kids did not comprehend this structure, we cannot conclude that they are simply 'reversing' the head parameter – they might simply not understand the instruction sentence. Another possible explanation would be that the kids are simply holding non-parsable sentences in working memory, and hence mixing up the WOs. Hence, a failure to produce the right WO could follow naturally from the fact that the child is not generating the sentences as a product of his or her own grammar, but is simply memorizing them.

Second, Akhtar sneakily avoids using *auxiliary* verbs and *animate nouns for object NPs*. One can only imagine what kind of 'messy data' she would have gotten had she used auxiliaries. In the SOV WO, a child would need to come up with some *wild* structure such as *Cookie Monster Big Bird taming is*. Now, Akhtar would certainly not

want this as she would not get such structures from the kids. Moreover, she keeps giving ‘test’ examples from other studies (e.g., Tomasello 1992, Akhtar & Tomasello 1997), containing the schema NP_[animate] – VERB – NP_[inanimate], though she did not use such a schema in her current study *at all*. One can only imagine what kind of results she would have gotten with two animate NPs – again, it probably wouldn’t do her any good. Here’s the big crux: imagine a sentence like *Mary is taming John*. If a kid realizes that *Mary* is the ‘tamer’, how will he recognize that *Mary* in *John is taming Mary* is (suddenly) the ‘tamee’? In other words, how does a child figure out the thematic relations between the arguments, assuming that a child has *no* notion of thematic relations and the word order to begin with. Again, the acquisition of argument structure and the WO seems pretty much *impossible* if the whole burden of language development rests on the input alone.

Thirdly, her study reports some extremely fishy results (p. 346) compared to her initial predictions. Though she reports that “...the two younger groups were equally likely to match the order modeled in the non-SVO conditions as to correct to SVO order...”, she mentions only in passing that *some* 2-year olds (when analyzed separately) showed great sensitivity (and consistency) to the SVO WO. This, I think, is simply not what her theory would predict, let alone Tomasello’s ‘stricter’ *Verb Island Hypothesis* (Tomasello 1992).

The last methodological-conceptual problem has to do with the *matching* technique. Suppose kids would want to please the examiner or make fun of him by mimicking him. Suppose kids *do* want to mimic the original sentence, be it SVO, SOV, VSO or any other possible word order. Young children’s non-adult behavior hence might be attributed to *priming effects* (e.g., kids would *imitate* the examiner’s sentence either for the sake of fun or to attract the examiner’s attention).⁷ When one critically evaluates the entire design and methods used in the study, one is almost led to believe that kids are playing some *mimicking game* rather than showing their knowledge of WO since the input is ungrammatical.

⁷ Though Akhtar claims this option is excluded since children who used non-SVO orders with novel verbs did not do so when a familiar verb was presented in a non-SVO sentence (e.g., *Elmo the car pushing*) (p. 351). However, she doesn’t say whether this occurred on the first or the second task and if it was across the board. In other words, we are forced to believe that this was the kids’ behavior and that the condition of mimicking was well-controlled.

Remember that some two-year olds did show sensitivity to non-SVO WOs and that their three-year olds did a much *better* job at correcting the two non-SVO orders (Table 3, p. 350). Furthermore, there are practically *no* errors found in four-year olds (Table 1, Fig. 1, p. 346). Kids do make errors and we all know that. Akhtar’s data-driven approach now faces one big challenge in this respect, namely why do kids overgeneralize, producing non-adult structure, if these are not part of their linguistic input⁸.

Where does all this leave us? It leaves us with the conclusion that input matching simply *cannot* explain the acquisition of word order *at all* and that children *know much more* than the proponents of data-driven approaches assume⁹. The UG-driven parameterized syntax approach to acquisition gives us a powerful tool to predict both *grammaticality* and *non-adult utterances (errors and/or overgeneralizations)*. When very young children make (occasional) errors, they do so because either a certain parameter has not been fully set yet (perhaps due to *maturation*, as in Borer & Wexler 1987 or Wexler 1998) or due to some kind of processing deficit responsible for children’s non-adult utterances (as in Bloom 1990). In either case, one is forced to consider some kind of nativist theory, proposing a language faculty that facilitates language acquisition.

References:

⁸ Moreover, it faces a more general conceptual problem (in language acquisition), hinted at already above, namely *how many instances* of a particular verb does a child need to hear in order to acquire the right schema for that particular verb, or what is the role of frequency in child language acquisition in general. According to Akhtar, *frequency* (or maybe also *consistency*) would need to play an extremely important role since all the burden of acquisition is placed on the P(rietary) L(inguistic) D(ata). It is empirically impossible to arrive at some consistent and reliable measure of frequency in the PLD. Suppose a child heard a verb Vx twice and correctly used it in a construction such as NP_{agent} + Vx + NP_{patient}. Would this suggest the child has *learned* the schema for Vx or did he just ‘map’ Vx to his ‘general abstract notion’ of structure, following the Head Parameter? (by triggering as in the sense of Gibson & Wexler 1994)

⁹ In her Discussion section (p.352), however, Akhtar does mention that Tomasello’s “[p]revious conclusions...[claiming] that young children tend to be quite conservative and use verbs only (or mainly) with the argument structures...” they have heard were probably too strong and that it follows from her study that children’s knowledge of argument structure is *more general* than initially believed in the data-driven approaches. I personally find this embarrassment for the theory that she is proposing. In other words, she is saying that Tomasello’s conservationist view was *wrong* in the first place. The same conclusion is arrived at in Ninio 2003, a study that analyzed the Travis corpus for verb antecedents of 3-word utterances, concentrating on novel verbal argument constructions. It reported that 40% of the verbs in novel SVO patterns had no verb antecedents (i.e., no familiar verbs) and none of the VIO construction had such an antecedent in the corpus. Ninio concludes that this is a strong evidence against Tomasello’s hypothesis of child’s grammar being a collection of verb-islands, and evidence for a consistent “system from day one”.

- Akhtar, N. 1999. Acquiring basic word order: evidence for data-driven learning of syntactic structure. *Journal of Child Language* 26. Cambridge: CUP.
- Berko Gleason, J. & Bernstein Ratner, N. eds. 1998. *Psycholinguistics*. 2nd Ed. London & New York: Harcourt Brace.
- Bloom, P. 1990. Subjectless sentences in child language. *Linguistic Inquiry* 21: 491-504.
- Borer, H. & Wexler, K. 1987. *The maturation of syntax*. In Roeper, T. & Williams, E., eds., Dordrecht: Reidel.
- Can, C. Head Parameter Setting in the Acquisition of Turkish as a First Language [Retrieved on the internet on Oct 10 2003:
<http://www.ekmekci.com/Publicationdocs/AcqTur/Head%20Parameter.doc>]
- Chomsky, N. 1981. *Lectures on government and binding*. Dordrecht: Foris.
- Chomsky, N. 1986. *Knowledge of Language*. New York: Praeger.
- Crain, S. 2003. *Post-modern de-constructionism*. In-class presentation at UMD College Park.
- Golinkoff, R.M. 2002. A Talk given at BUCLD 27. Boston University, Boston MA.
- Greenberg, J. H. (1963). *Universals of language*. Cambridge, Mass.: MIT Press.
- Hirsh-Pasek, K. & Golinkoff, R.M. 1996. *The origins of grammar: evidence from early language comprehension*. Cambridge, MA: MIT Press.
- Lardiere, D. 2003. Class notes on *Generative Approaches to Language Acquisition*. Spring 2003, Georgetown University, Washington DC.
- Ninio, A. 2003. No verb is an island: Negative Evidence on the Verb Island hypothesis. *Psychology of Language and Communication* 7 (1).
- O'Grady, W. 1997. *Syntactic Development*. University of Chicago Press.
- Pinker, S. 1984. *Language Learnability and Language Development*. Cambridge, MA: Harvard University Press.
- Pinker, S. 1994. *The Language Instinct*. London: Penguin.
- Tomasello, M. 2003. *A Usage-Based Account of Early Syntactic Development*. Talk given at GURT 2003, Georgetown University Round Table on Languages and Linguistics - Language in Use: Cognitive and Discourse Perspectives on Language and Language Learning. Washington DC.
- Wexler, K. 1998. Very early parameter setting and the unique checking constraint: A new explanation of the optional infinitive stage. *Lingua* 106: 23-79 [also published in Sorace et al., eds., 1999. *Language Acquisition: Knowledge Representation and Processing*. Amsterdam: North Holland].