LOCALITY AND THE ORDER OF ACQUISITION STEPS

JACQUELINE VAN KAMPEN

UiL OTS
Utrecht University
Jacqueline.vanKampen@let.uu.nl

Abstract
Preferably, the properties of grammar can be derived from the following factors:
(i) The primary linguistic data as they are offered to the child.
(ii) A language acquisition procedure.
Hopefully, the language acquisition procedure will be compatible with plausible assumptions about the neural abilities of human beings, but that is of no immediate concern. The interaction of the primary data and the acquisition procedure can be studied by a closer look at the order of the child's acquisition steps. What does she acquire first and why? What does she acquire later and why? My main point will be that this is empirically a promising and by no means trivial approach. At the same time, I will argue against an assumption that is quite common in computational studies and also in mere grammatical studies of child language. People from Gold (1967) to Yang (2002) assume that the acquisition procedure has simultaneous access to all data at once. My point will rather be that the acquisition procedure implies a natural selection of data. The data selection procedure must predict the actual order of the acquisition steps in the various languages.

1. Input reduction

The procedure for first language acquisition is not confronted with all grammatical option and problems at once. The child applies a radical reduction to the mother's input. The common sense background of that reduction can be formulated as in (1).

(1) *Reduction of input to intake*
  a. Leave out what you cannot fit in.
  b. Try minimal solutions for the combinatorial restrictions in the residue.

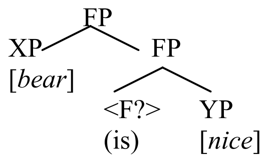
Suppose the child has reached a point at which she is able to recognize a set of separate words with denotational content {*gone, up, car, daddy, eat*} or words with an immediate pragmatic meaning {*that, wanna, no*}. The reduction procedure in (1) will then throw out all grammatical markings: articles, copulas, auxiliaries, verbal inflections, connectives. Hence, sentences like *ain't the bear nice*; *the bear is nice, isn't he?*; *I want the bear to be nice*, are all turned into [*bear nice*]. A set of binary word constructions is the result. This at least happens in the child's actual output.

Eventually, the learning procedure will identify grammatical markings between binary combinations one at a time. This stepwise learning is an important characteristic of language acquisition. So, a proposal for a language acquisition procedure should predict the reductions as they apply to an adult grammar and it should predict the linear order of acquisition steps that will follow from small, reduced utterances frames. Suppose the acquisition procedure starts with the reduction operation in (2), where <+F?> is an unidentified functional feature.

(2) *Input Reduction*
   a. *input*: substitute <+F?> for each grammatical marking still unknown.
   b. *reduction*: throw out all input sentences with more than one <+F?>.
   c. *output*: attach F_i to the selectionally dominant element to the left or right

The result satisfies the *Single Value Constraint* (Clark 1992) when it happens to be that the residue from (2)b boils down to a single set defined by F_i. The intake to the acquisition procedure is then such that one grammatical category is singled out, identified and subsequently acquired. I hope that such a property will define natural language. Let me assume it and call the convenient outcome an evidence frame, as defined in (3) and exemplified in (4).

(3) *Evidence frame*
   a. pragmatically: an intuitively understood utterance
   b. syntactically: a binary phrase structure [XP [F_i YP $]_{FP}$ $]_{FP}$
   c. semantically: fully interpretable but for a single <F?>

(4)


Functional categories are identified due to their frequency in the input and due to the fact that they cannot be and are not understood beyond grammar, i.e. beyond the grammatical relation between the phrases XP and YP.

The acquisition of the category <+F?> $\rightarrow$ F_i changes the initial state and the data reduction procedure in (2) will reapply. The next grammatical category F_{i+1} is singled out, etc.. The evidence frames do not follow from mere frequency of F_i in the raw data set. The F_i must also define a minimal frame that is fully interpretable.

The restrictive evidence frames follow from the input reduction. They offer the bootstraps for subsequent acquisition steps. It is useful, though, to realize that the evidence frames remain present and active in the adult grammar. They continue to function as the local checking domains for elementary grammatical properties. For that reason, the adult knows more or less how to schematize, c.q. creolize, his grammar. The acquisitional perspective on syntactic locality is given in (5).

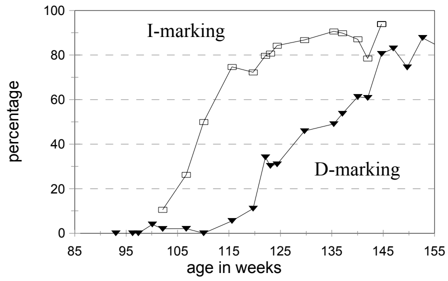
(5) a. all evidence frames in acquisition are local and binary.
   b. all grammar is acquired due to such domains.

I will present two major acquisition steps that support the reductions in (2) and (3).

2. Order of acquisition steps

Language acquisition overcomes the radical underspecifications that result from the initial data reduction. It proceeds by adding grammatical features within a local binary frame. The order of the acquisition steps can be shown by longitudinal graphs, see (6).

(6) Dutch Sarah: acquisition of I^o-marking and D^o-marking



The graphs in (6) represent the acquisition of finite verbs (I^o-marking) and determiners (D^o-marking) by the Dutch child Sarah (Van Kampen 2004). The graph for I^o-marking shows the growing percentage of grammatical predicate marking, {copula/ auxiliary/modal/finite morphology}. The graph for D^o-marking shows the growing percentage of grammatical argument marking, {article/demonstrative/ possessor/quantifier}.

Now the order of acquisition steps shows that Dutch Sarah applies systematic I^o-marking almost half a year earlier than systematic D^o-marking. The same order of appearance was found for English, French, and Rumanian. The amount of determiners outweighs the amount of finite verbs in the input data. Yet, children in various languages start to analyze predicate-argument structure by I^o-marking. The less frequent I^o-marking precedes the more frequent D^o-marking in acquisition. The order $I^o \rightarrow D^o$ must be explained. I will show how the acquisition procedure follows the *Single Value Constraint* on evidence frames as proposed in (2) when initially, sentences with both a D^o-marked noun and an I^o-marked verb are thrown out of the observation space. The feasibility of a mechanical reduction procedure was partly demonstrated by a computer simulation (Obdeijn 2004). The simulation derived child language from a child-directed input and derived an order of intake frames.

The systematic I^o-marking and D^o-marking themselves give entrance to a whole series of further acquisition steps, beginning with a grammatical decision procedure on the category membership V versus N (Van Kampen 2005). This option, chosen here for language acquisition, was implemented earlier in computational approaches to category assignment (Buszowski 1987). A general property of ‘decoding’ emerges as well. The successive evidence frames narrow down to a far more precise context and the speed of acquisition increases by an order of magnitude. The subject of the I^o-marked predicate (finite verb) initially lacks φ-features of person/number. In a subsequent step, the φ-feature content in D^o, {± person, ± number} on the subject, is figured out. However, the finite verb still doesn't show the correct agreement with the subject, see (7).

(7) de clowntjes *heb* oogjes (the clowns has eyes) Sarah. week 129

One step later, the initial I^o-marked predicate constitutes the local evidence frame for Agreement features, the copying of the φ-features on I^o. The finite verb now starts showing the correct agreement. Late acquisition of agreement has been reported for various languages.

The dense succession of the acquisition steps shows that the later steps are a matter of weeks whereas the earlier steps were a matter of months, see (8).

(8)

step I^o	step D^o	step D^o (φ)	step I^o (φ)
20 wks	25 wks	5 wks	5 wks

The more effective acquisition relates plausibly to the more precise frame that can be used to select the input. The selection of some binary combination of content signs is far more undetermined than the distributional relation between explicit grammatical markings such as φ-features and Agreement. The later set of acquisitions is supported by a lexicon with categorial marking <+I> or <+D>. I propose that after step 1 and step 2, the EPP (subject-finite verb configuration) operates as an evidence frame.

3. A discovery procedure

Generative learnability theories in the 1980^{th} were theoretical and somewhat defensive. They qualified the mathematical deduction in Gold (1967) that context-free rewriting grammars could not be identified or learned without negative data. As Wexler & Culicover (1980) argue, context free generative grammars and some transformational grammars are learnable from positive data as long as the relevant relations are sufficiently local. The main point was to argue learnability in principle for certain types of generative grammar. There was no reference to child language. The ongoing simplification of grammatical principles, pushed by Categorial Grammar, HPSG and the Minimalist Program, may re-inspire interest in their learnability. I mention four attempts into that direction. Fodor (1998), Yang (2002), Culicover & Nowak (2003) and my work with Arnold Evers (Evers & Van Kampen 2001).

Fodor (1998) and Yang (2002) assume that the child is confronted with the full variety of constructions in his language. The child meets this challenge with brilliant creativity. She comes up with all possible grammatical structures that the general theory of grammar would allow. The child's productivity in designing possible solutions is maybe comparable with his creativity in grasping visual or musical structures or maybe with the babbling phase that precedes the construction of phonological forms. Fodor as well as Yang's learner start with a variety of grammatical structures and work towards a minimal set of grammatical structures by comparing alternative solutions. Fodor's learner is sensitive to certain key-constructions (treelets) that betray the language type and Yang's learner is sensitive to rules that are too often involved in analyses that fail. The options that they compare are assumed to be a priori present from the human brain. Yang proposes an accounting system of 'penalties' for failing rules. Yang's bookkeeping of failures and Fodor's testing system could be characterized respectively as an effective *evaluation procedure* (Yang) and as an effective *decision procedure* (Fodor). Their learners start with all options offered by the theory. Subsequently, they propose computational operations that select a language-specific grammar for the input data. Both successfully simulate how the learner zeros in on the core grammar of the language.

By contrast, I propose, like Culicover & Nowak (2003), that the young learner is unaware of the grammatical alternatives that are available in the world outside. Our learning procedure could be characterized as a *discovery procedure*. My young learner must reduce its initial attention to constructions assigned to pairs of adjacent content words and so he enters a maximally reduced observation space, as formulated in (1).

A learning procedure as in (2) that adds a grammatical feature to a category moves from a less restricted superset to a more restricted subset. The learning procedure starts with underspecifications, but the associative pressure of local contexts has a healing effect. The initial underspecifications are "blocked". Blocking effects are known from the very beginning of grammatical studies (Panini, DiScullio & Williams 1987). In general, the more specified variant blocks the less specified one. Blocking in language acquisition can be traced by longitudinal graphs as we have seen. This is a contentious issue in theories of language acquisition. Some try to reconstruct child language as subset language that is extended to the correct generalizations. Others believe that child language starts with maximal generalizations and narrows down by developing subcategories (Jakobson 1942; the present perspective).

4. Perspective

The acquisition order is due to input-control, but definitely not always due to input frequency. Functional categories are acquired later than content words, yet their token frequency is 100 to 300 times higher than the token frequency of an arbitrary content word. Although highly frequent, functional categories can be learned only in constructions that contain content words. This is because a grammatical word F_i indicates a grammatical relation between two phrases $[XP\ [F\ YP]_{FP}]$ It is a word that carries no meaning beyond the syntactic relation (the word *and* does not mean 'pair', the word *but* does not mean 'objection', the word *is* does not mean 'property'). The acquisition order "content words before functional categories" is imposed by the nature of the system the child is confronted with. The same holds for I^o/predicate marking and D^o/reference marking. The I^o-markings are more diverse in form and less frequent in the raw input than the D^o-markings. Yet, I^o-marking precedes D^o-marking in acquisition.

The factual order of acquisition steps has to be established for various grammatical properties. It has to be considered whether and how that order fits the present conjecture about the hierarchy of evidence frames. For instance, the evidence frames are also effective for the subsequent learnability of scopal phenomena, like wh-marking and negation. My conjecture is that island constraints and scopal domains can be derived from the locality of the evidence frames and their dependence on the crucial terminal elements <F?> (inclusiveness). The decision proposal by Fodor and the evaluation procedure by Yang assume exclusive a priori structures, as well as procedures to compare solutions. The discovery procedure must assume that the natural input allows a reduction to local frames and a terminal string that remains informative enough in spite of the reduction. Locality and local inclusiveness of grammatical information are present to guarantee a certain type of learnability.

Bibliography

Buszkowski, W.: 1987, Discovery procedures for categorial grammars, in E. Klein & J. van Benthem (eds.), *Categories, Polymorphism and Unification*, UvA.

Clark, R.: 1992, 'The selection of syntactic knowledge' *Language Acquisition* 2(2), pp. 83-149.

Culicover, P. & Nowak, A.: 2003, *Dynamical Grammar: Minimalism, Acquisition and Change*, University Press, Oxford.

DiScullio, A.-M. & Williams, E.: 1987, *On the Definition of Word*, MIT Press, Cambridge MA.

Evers, A. & Kampen, J. van: 2001, *E-language, I-language and the Order of Parameter Setting*, Uil OTS Working Papers 00105-S-S, Utrecht.

Fodor, J.D.: 1998, Unambiguous triggers, *Linguistic Inquiry* 29: pp. 1-36.

Gold, E.M.: 1967, Language identification in the limit, *Information and Control* 10, pp. 447-474.

Jakobson, R.: 1942, *Kindersprache, Aphasie und Allgemeine Lautgesetze*. Uppsula.

Kampen, J. van: 2005, Language specific bootstraps for UG categories, *International Journal of Bilingualism* 9-2, pp. 253-277.

Neeleman, A. & Koot, H. van de: 2002, The configurational matrix, *Linguistic Inquiry* 33-4, pp. 529-574.

Obdeijn, A.: 2004, *Taalverwerving door Kinderen en Machines*, Master thesis University of Amsterdam, ILLC.

Yang, C.D.: 2002, *Knowledge and Learning in Natural Language*, U.P: Oxford.