8. Project description

8.1 General description

Chomsky (1975:119f) has proposed to think about first language acquisition as a set of transitions between intermediate grammars.

\[ G_0, \ldots, G_i, \ldots, G_f \]

Each of these transitions would add a language-specific characteristic to a general frame (UG). \( G_0 \) would be the initial grammar, possibly generating no more than lexical projections. \( G_f \) stands for the final grammar that is eventually acquired. The status of \( G_i \) is more enigmatic. It stands for the intermediate grammars. The successive transitions should more and more approach the final grammar \( G_f \) by a selection of as yet unknown steps and twists. The present research proposal takes the abstraction in (1) as its starting point.

The transitions in (1) represent a series of successive learning steps. Each transition can in principle be seen as a computational step \( f_i \).

\[ f_i (G_{i-1}, D_i) = G_i \]

The input of learning-step \( f_i \) is the currently acquired grammar \( G_{i-1} \) and a triggering data set \( D_i \). The output is the new intermediate grammar \( G_i \). All intermediate grammars fit UG, that is the set of grammatical principles that hold for all human grammars (Chomsky 1981, 1995). The parameter setting functions \( f_i \) hang together in a language acquisition ‘device’ (LAD). This LAD may be represented as in (3).

\[ \begin{align*}
\text{a.} & \quad \text{UG (principles, parameters)} \\
\text{b.} & \quad G_{i-1} \quad \text{(the current intermediate grammar)} \\
\text{c.} & \quad f_i (G_{i-1}, D_i) = G_i \quad \text{(the current parameter setting function)}
\end{align*} \]

The content of the LAD keeps changing at each step \( f_i \) in the direction of the language-specific grammar \( G_f \). Each time a new parameter has been set, and a new intermediate grammar \( G_{i-1} \) has been reached, it is predicted that the next parameter setting function will get activated. The resulting order of parameter setting functions should reveal a learnability hierarchy among the properties of \( G_f \).

It has been shown by Gibson & Wexler (1994) (in a theoretical exercise) that random setting of parameters would imply the unlearnability of several well-known languages types. Hence, some ordering of parameter setting will be necessary. Pursuing the same line of research, Fodor (1998, in press) argues that an unfortunate course of parameter setting can be avoided if parameter setting were dependent upon quite specific structural properties. At this moment the following three points need to be investigated: (i) The order of learning steps has to be established empirically; (ii) The actual order of learning steps has to be derived from the factors in (3): UG (3a), the intermediate grammar (3b) and the intake procedure (3c). At present we have no more than some global impressions about the quantity of the input cues and their triggering effects. (iii) Only after the order of learning steps has been reconstructed from these factors, will it become possible to simulate the learning procedure. The present research proposal intends to focus upon the research points in (i) and (ii) in order to open the way for (iii) in follow-up studies.

Ad (i). The transition from \( G_{i-1} \) to \( G_i \) can be studied by the construction of longitudinal graphs. Longitudinal analyses (Van Kampen 1997, Wijnen 1997) show how a newly emerging construction may appear for some ten weeks or longer in less than 10% of the relevant environments. At a certain point this period of latency is followed by a steep and irreversible change. The proportion of the new construction rises from less than 10% to more than 90% within a certain period. The length of the period, c.q. the steepness of the graph, indicates the robustness of the input cue. The dramatic change may be abstractly represented by a three-part linear graph as in (4). See for some examples of longitudinal graphs and their interpretation Van Kampen (1997: 73-75).
Ad (ii). The second aim of this research proposal is the reconstruction of the actual order of learning steps as a set of decisions (computations \( f_i \)) that could be made on the basis of the available input evidence \( D_i \). The core of the problem, but not its solution, was put forth by Gibson & Wexler (1994). The input to the acquisition procedure is not ordered. It contains the evidence \( (D_i) \) for all learning steps \( (f_i) \) simultaneously. Nevertheless, the learning procedure is highly selective towards the available evidence. This may be illustrated by the following example. Preliminary research (by Van Kampen) has shown that 93% of the input structures demonstrate that Dutch main clauses have the finite verb sentence initially (in ‘V(erb)-second’ position). Transitive structures containing a verb and its object constitute 33% of the input. Only 15% of the input structures present the transitive structures as straight OV (object-before-verb) order. In spite of the 93% versus 15% difference, all children seem to start with the OV parameter setting, without paying attention to the overwhelming V-second evidence. The selection of the evidence \( (D_i) \) needs to be explained, partly in terms of UG (not language-specific), partly in terms of the currently available grammar \( (G_i) \).

The present proposal will establish interaction between acquisition studies and research on formal learnability. Conceptual tools for the reconstruction of successive learning steps are developed in theories of formal learnability. Concepts like text presentation (learning decisions are made from positive evidence alone), text efficiency (learning decisions are made on robust evidence), set-drivenness (learning decisions are made from an unordered data set), incremental learning (learning decisions are made memory-less on the basis of the current \( G_i \) and the current \( D_i \)), subset learning (the learning function \( f_i \) maps the input data \( D_i \) to that value of a parameter that generates the smallest among the languages compatible with the data), conservative learning (learning decisions are error-driven, i.e. only made if the current \( G_{i+1} \) is inconsistent with the current \( D_i \), contextual category identification (learning decisions for category identification are made from syntactic context information alone) play a part in formal learnability as well as in empirical (linguistic) inquiry (see Saleemi 1992 for an overview of the two fields). However, the two fields are not well informed about each other’s perspectives. The present proposal aims to establish an effective interaction between these two fields by interpreting real acquisition steps in terms of learnability theory. This approach has already been proved to be fruitful in Van Kampen (2000). The concept of contextual category identification used by Buszkowski’s (1987) algorithm for finding a categorial grammar has been applied in Van Kampen (2000) to derive a learning strategy for lexical categories in a generative grammar.

Ad (iii). When the actual order of the learning steps has been rationally reconstructed it will become possible to consider how the grammar acquisition procedure can be simulated by a computer program. This simulation is not part of the present proposal, whereas filling out the ordering of learning steps (i) as well as a computational reconstruction (ii) are. The computer simulation is mentioned nevertheless, because a group of researchers at CUNY led by Fodor, Teller and Sakas is setting up computer simulations for a 10 parameter space, including the parameter space studied by Gibson & Wexler (1994). At the moment these computer simulations can only be run.
on input samples from imaginary languages. The Fodor-project is a theoretical exercise on the learnability of language types, whereas the present proposal focuses on the learnability of Dutch and English taking into account natural language data. Both projects cover the same parameter space and will therefore benefit from each other. Cooperation between both projects is part of the research plan.

8.2 Research design

Main lines. The present project will specify the schema in (4) by constructing the graphs for some ten acquisition steps, five related to the illocution marking of the sentence (so-called C-properties) and five related to referential phenomena (so-called D-properties). Construction of these acquisition steps should result in: (a) ten cases of actual parameter setting, each specified for its robustness or speed and placed in a linear order with respects to the others; (b) principles that derive the order and speed of the acquisition steps from the currently available grammar and the available input data.

The acquisition of C-properties. In order to derive the acquisition steps related to the illocution marking of the sentence, the actual temporal relations between the graphs in (5) and their relative speed will be compared.

(5) Graphs for acquisition steps related to C-properties (sentence/illocution marking)
   (i) Directionality parameters and V-second
   (ii) V-second and Auxiliary placement
   (iii) V-second and Wh-movement/topicalization
   (iv) V-second and its absence in subordinates
   (v) V-second and residual V-second

The sentence-initial position of the finite verb in main clauses is known as V(erb)-second, and holds for all Germanic languages except English. The acquisition of V-second has been studied extensively in work on real-time acquisition (De Haan 1987, Clahsen 1991, Poeppel & Wexler 1993, a.o.). It was the central point in the study of Gibson & Wexler (1994) and in the learnability discussion that has followed since (Berwick & Niyogi 1996, Frank & Kapur 1996, Fodor 1998). The phenomenon constitutes two main problems.

(i) How is it possible to acquire the underlying directionality parameter object-before-verb and subject-before-predicate (example 6a) from a language that consistently (>90%) moves the finite verb into a derived position (V-second position, example 6b)?

   (6) a. pappa boek lezen  (daddy book read)
   b. pappa leest een boek // dat boek leest pappa // nu leest pappa een boek
      daddy reads a book // that book reads daddy // now reads daddy a book

This was the crucial, but unsolved, point in Gibson & Wexler (1994).

(ii) How is the category V(erb) acquired as a generalization over lexical and auxiliary verbs? De Haan (1987) argued that the generalization is made after the acquisition of V-second. Poeppel & Wexler (1993) argued that the category V is established instantaneously. Neither of them constructed longitudinal graphs.

Other phenomena related to the central V-second graph are the acquisition of (iii) question words (‘wh-movement’) and topicalization, (iv) subordinate order, (v) residual V-second in English and French (in questions only (English/French), auxiliaries only (English)). Preliminary research by means of longitudinal graph construction (Evers & Van Kampen 1999, 2000) has shown that the acquisition of C-related rules is radically different in English from what is found for Dutch.

The acquisition of D-properties. In order to derive the acquisition steps related to reference marking, the actual temporal relations between the graphs in (7) and their relative speed will be compared.

(7) Graphs for acquisition steps related to D-properties (reference marking)
   (i) D-insertion and Pied-piping
   (ii) D-insertion and Definiteness
   (iii) D-insertion and Object shift
   (iv) D-insertion and er-insertion
   (v) D-insertion and Number marking/full agreement

Determiners are at first absent in child language, later they are used optionally, whereas they are obligatory in the adult language. (i) One of the properties expected to be dependent upon the obligatory presence of determiners and other D-elements (demonstratives, question words) is pied-piping of the nominal complement. Dutch child
language shifts the D-element in questions and topicalizations, while stranding the nominal complement, as illustrated in (8) (Hoekstra & Jordens 1994, Van Kampen 1994).

(8)  a.  dat wil ik [∅ boek ] (that want I [∅ book ])
    b.  welk wil jij [∅ boek ]? (which want you [∅ book ]?)

Later the stranding of the nominal element is given up for the adult constructions in (9) with pied-piping.

(9)  a.  dat boek wil ik (that book want I)
    b.  welk boek wil jij? (which book want you?)

Van Kampen (1994, 1997) has proposed that obligatory pied-piping is a consequence of the obligatory presence of D, since D would represent abstract case (cf. Lebeaux 1988). It is expected then that pied-piping will follow the acquisition of ‘D-insertion’. The fact as such, the speed of the acquisition of D-insertion, the speed of the acquisition of pied-piping and the distance between the two graphs need to be settled and explained.

Other grammatical properties related to the acquisition of D-insertion are the acquisition of (ii) the definite/indefinite distinction, (iii) direct object shift (‘scrambling’), (iv) insertion of the dummy subject er (‘there’), (v) number marking and full agreement. Several of these properties have been studied experimentally and longitudinally (Eissenbeiss & Penke 1996, Schaeffer 1997, Krämer 2000), but not by means of graphs.

8.3 The longitudinal method

The above proposal contains a fact-finding part in addition to an analysis of parameter setting as a linear order of learning steps that is partly predictable from the nature of the input. The actual order of parameter setting is found out by a quantitative method of longitudinal graphs that tracks the development of single children over a period of several years. The quantitative method will be applied to fairly large single child corpora of Dutch and English adult-child conversations recorded at regular intervals for an extended period (>3 years; >10,000 utterances). The CHILDES archive (see MacWhinney 1991) offers such corpora. For English: Adam 45,185 and Sarah 36,755 (Brown), Naomi 12,000 (Sachs); for Dutch: Laura 26,220 and Sarah 19,954 (Van Kampen). These corpora offer the phenomena mentioned under 8.2 in large quantity. In addition, we will use some smaller, but very dense, corpora, for several cases mentioned in (5) and (6), e.g. the Utrecht corpus (Elbers/Wijnen): 71 hours audio recording of two Dutch children between 2:3 and 3:1; and the Groningen corpus (Bol/Krikhaar/Wijnen): 170 hours audio recording of seven Dutch children between 1:5 and 3:7.

The data will initially be modeled by truncated S-curves, with three linear parts. These will respectively represent the period before, during and after the parameter setting. The simplification by linear graphs is sufficient and effective in order to answer the primary questions about order and speed of parameter setting. Later, the same data will be brought into more advanced logarithmic models e.g. such as those that have been used by Kroch (1989) in his work on the historical and sociolinguistic change of language and by Ruhsland (1998) from a developmental psychology perspective. The main reason for this adaptation is to remain in line with the modeling of order and speed in parameter setting that will be developed by the Fodor-group at Cuny.

The present research proposal adopts a longitudinal method instead of an experimental or cross-sectional method. The aim is to get direct answers for the questions that are posed (order and speed of parameter setting). For example, if an experiment were set up to measure the average progress in V-second parameter setting for a cohort of 100 children, and this experiment were repeated at two months intervals, one would get many data and these would show for the entire cohort a progressive setting of the V-second parameter over the entire year. The more children are investigated, the smoother the continuous setting of the V-second parameter. The fact that each of the individual children sets the parameter in a dramatic switch of some 10 weeks would, however, get lost. The more data, the more likely this crucial point would be missed. If, by contrast, a parameter setting order between two parameters were confirmed by tracking the grammatical development of three children longitudinally, one might direct further research in two directions: either a reconstruction of the interaction between input data, UG principles and current grammar (cf. (3)), or an extension of the data base. If the former option is taken, one enters a new level. A rational reconstruction of the acquisition procedure may subsequently be used in a computer simulation of the actual acquisition process. This is a way to add higher quality to less data and this is the direction that is presently proposed.

8.4 Scientific significance

The central aim of the proposal is to develop an analysis of parameter setting as a linear order of learning steps. Such an analysis serves a theoretical purpose, as well as a practical one.

The theoretical purpose is pursued by separating in a more substantive way a priori elements in language acquisition (UG principles and parameters) from a posteriori control by language-specific input. The usual assumption in generative grammar has been that lexical elements might be a language-specific invention, whereas all grammatical peculiarities would be due to a selection made from a priori parameter possibilities. The
The present proposal will inquire to which extent the order of acquisition and its speed follow from input evidence. The relation between a priori frames and a posteriori input needs to be stated in terms of successive learning steps and to be related to models of formal learnability. The need for this type of empirical research has been pointed out in studies of formal learnability (see Osherson, De Jongh, Martin & Weinstein 1997: 745).

The practical purpose is served by the ordering the learning steps in the course of time. The type of research proposed here will result in a far more detailed picture of first language acquisition. A more precise picture of language acquisition in normal children is crucial to more precise diagnostics of delayed language acquisition (see Schlichting 1987). Existing methods for remedial intervention make use of analytical tests based on normal language development. At the moment there are several of those tests (e.g. TARSP, GRAMAT). These tests characterize the acquisition of a phenomenon by means of the notion ‘first appearance’. It has been shown that ‘first appearance’ is a sign of acquisition in rare cases only (Van Kampen, project Wetenschapswinkel). The specific claim of the present research proposal is that not first appearance, but the more elaborate method of growth curve construction establishes the point of acquisition. Moreover, the present proposal intends to identify the cue/trigger of the learning step. In that sense, remedial teaching is offered a hint of where to focus and how to focus. The aims of the intervention by remedial teaching can get a sharper grammatical characteristic and that will subsequently allow a better measurement of the effectiveness of the intervention. This line of research builds on the work that the proposed postdoctoral researcher has done for the ‘Wetenschapswinkel’ in 1997 and 1998.

8.5 Institutional environment

The project will be carried out at the Utrecht institute of Linguistics OTS, more specifically within the theme group of language development as well as the ‘broad strategy program’ entitled Language in Use. UiL OTS hosts an active group of researchers on first language acquisition (S. Avrutin, S. Baauw, E. Blom, P. Coopmans, A. Evers, J. Evers-Vermeul, J. de Jong, R. Kager, W. Philip, M. Verrips, F. Wijnen, W. Zonneveld), statistics (H. van den Bergh, G. de Krom) and computational linguistics (C. Costa Florenzio, M. Moortgat, W. Vermaat). It is this last group in connection with learnability theorists at Amsterdam that will provide the formal expertise required for this project. UiL OTS has recently set up a new and fully equipped laboratory for experimental research and natural language data collection in the area of language acquisition. The project will profit largely from the cooperation of researchers, both national and international, who are working on the type of questions addressed here: J. Fodor (parsing/learnability), R. van Hout (statistics), C. Jakubowitz (acquisition of syntax), T. Roepers (acquisition of syntax).

References