Parameter Setting and Input Reduction[1]

Arnold Evers and Jacqueline van Kampen

Abstract

The language acquisition procedure identifies certain properties of the target grammar before others. The evidence from the input is processed in a stepwise order. Section 1 equates that order and its typical effects with an order of parameter setting. The question is how the acquisition procedure derives the order from input evidence. Section 2 proposes a systematic input reduction for functional categories as the key; the reduction residue contains no more than a single non-acquired functional category <F?> that is first seen as an optional element only. If that functional category has turned into the most preferred option, the input reduction shifts its acquisition focus to the next functional category. Section 3 and 4 demonstrate how quantitative proportions within the child's input reduction determine the underlying order as SOV in Dutch before the V-second shift for root sentences is derived. The child's input reductions are claimed to follow from ignorance rather than from any *a priori* information. It is argued that parameters are formal properties of the grammatical system that originate as cultural discoveries made by a reflexive mind rather than being task-specific neural *a prioris*. Section 5 suggests that this view can be extended to syntactic islands.

Keywords: cue, learnability of underling structure, acquisition of Verb-Second (V2), lexical categories, island constraints

1   Parameter Setting

The twin notions 'parameter' and 'parameter setting' (Chomsky (1981) are aimed at explaining first that grammatical properties do not vary independently of each other, and second that this fact is significant for the striking success of first language acquisition. The classic example is the property of 'subject *pro*-drop' acquired by setting a feature value in some functional category I<+agr>. Chomsky (1981: 240) analyzes *pro*-drop as possibly implying free inversion of the subject phrase, violation of the *that*-trace filter, and long *wh*-movement of a subject. An engaging perspective opened up. Suppose that all parameters are binary valued like <± subject *pro*-drop> and suppose, more daringly, that the set of possible parameters is limited and predetermined. Is it then possible that setting a limited amount of predetermined parameters might deliver all possible types of core grammars and thus reveal something about the success of spontaneous language acquisition?

Two problems with the notion of parameter setting were diagnosed in Dresher (1999): 'the epistemological problem' and 'the credit problem'. The very explanatory and abstract nature of the parameter implies that its presence cannot be noticed as a simple property in the primary data. This is the epistemological problem. The credit problem relates to the interaction of parameters. The necessary and intended interaction of parameters prevents their setting by simple observation. When combining a tentative setting of parameters yields results not confirmed by the primary data, it is unclear which parameter setting caused the uninvited

---

superset effect. Both problems for parameter setting (the epistemological problem and the credit problem) follow from the nature of the parameter itself and both seem to disqualify parameter settings as innate bootstraps for first language acquisition.

Yet, when we look at first language acquisition itself, there are reasons not to give up on the explanatory potential of the notion 'parameter setting' too easily. There is in language acquisition a stepwise appearance of grammatical properties and it would be strange if that progress were not caused by an independent identification of grammatical properties. Child language picks up the properties of a grammar in a non-arbitrary order. Parameter theory implies a hierarchy of grammatical properties as expected by Jakobson (1942) and developed by him for phonology. Parameter theory allows us to see language acquisition as a series of ever more specified grammars G_i that approach the target grammar G_n (Chomsky 1975: 119f).

$$(1) \qquad G_o \text{----} G_i \;\Rightarrow\; G_{i+1} \text{----} G_n$$

Properties acquired earlier, the left-side steps in (1), are likely to have a lasting effect upon the steps that follow. They are likely to have typological significance and may develop into what Baker dubs 'macro-parameters' (Baker 1996, this volume). The right-side steps in (1) are likely to have fewer such consequences. They will introduce variation that is more language-specific and possibly more flexible across history and dialects (cf. Panagiotidis, this volume). It is for that reason that the reconstruction of the order of acquisition steps is of the utmost importance for understanding how input data may determine the setting of a parameter.

The separate and successive identification of grammatical properties in first language acquisition reappears in computational models of the acquisition procedure (Berwick 1985, Berwick and Weinberg 1984, Clark 1992, Gibson and Wexler 1994) as the *Single Value Constraint*. The Single Value Constraint implies a parametric view of acquisition and holds that the acquisition steps that set the parameters are taken in a linear order that develops as a causal chain. Each new step $G_i \Rightarrow G_{i+1}$ shifts the focus to a new parameter. Plausibly, this happens because an orientation by G_i opens the mind to perceptions that trigger the step to G_{i+1}. In terms of data selection, G_i makes selections from the raw input and arrives at a data set D_{i+1} that will redirect the acquisition procedure towards G_{i+1}. This implies that the acquisition procedure somehow gets information about the questions in (2).

(2) a. How is data set D_{i+1} selected by G_i?
　　 b. When does a new parameter in G_{i+1} become relevant?

A mechanism to assemble that information is still needed and the present paper represents an attempt to formulate one.

1.1 Parameter setting models

Models of language acquisition by parameter setting have to deal with the problems diagnosed in Dresher (1999). They tend to do this by adding more *a priori* (innate) information about possible grammars (cf. Hale and Reiss 2003). *A priori* grammatical factors enable the acquisition procedure to construct a provisional grammar. The learner changes the provisional grammar only when it is inconsistent with the current input sentences (error-driven learning).

For example, the TLA (Triggering Learning Algorithm) model by Gibson and Wexler (1994) considers the consequences of the following speculation. Let the primary data be provisionally represented by a string of category labels as known from the adult grammars and let all parameters be given *a priori*. The learner applies a provisional parameter setting to the primary data of his language. Predictably, this will in general not work. Now the language acquisition procedure is allowed to switch one parameter value. When this move hits the jackpot, and arrives at a satisfactory analysis of (a subset of) primary data, the switched parameter is accepted as the correct value of the parameter for that grammar. This is all preliminary assumption by Gibson and Wexler in order to arrive at their genuine research question. Will this method of blind gambling in parameter values ever reach the target grammar? Gibson and Wexler try it out for an imaginary case of 5 given categories {subject, object, adverb, verb, auxiliary} and three parameters {subject-head order, complement-head order, presence or absence of a V-second rule}. The primary data are given as strings of the functional categories and the acquisition procedure must find out how to discriminate between the $2^3 = 8$ possible types of grammar allowed by the model of three binary parameters. It turns out that the method cannot guarantee an arrival at the correct target grammar. A legitimate but with hindsight premature setting of the V-second parameter will cause havoc. It may deliver distributions that require two simultaneous re-settings (instead of one). More than one parameter change for each acquisition step weakens the model as it greatly expands the search space for each step. Gibson and Wexler call such failures a *local maximum*, where the target grammar represents the absolute maximum. A subsequent study by Kohl (1999) calculated that the chance of running into local maxima rises considerably when the number of parameters rises.

However, an unambiguously faultless result is guaranteed when the acquisition procedure is informed about a linear order in which the parameters must get fixed. The demonstration case in Gibson and Wexler (1994) runs fine if an *ad hoc* decision is added to delay a decision about the V-second parameter. This is interesting since actual language acquisition does indeed show an order of acquisition steps (Brown 1973).[2] The new problem is now how the acquisition procedure can be informed about the sequencing restrictions in parameter setting. Wexler (1999) speculates quite generally that the troublemakers among the parameters will not be considered before the brain has matured sufficiently. Fodor (1998, 2001) proposes that a troublemaking parameter will not be reset from an initial default status until the learner has spotted a configuration that is an unambiguous context cue for the resetting. The successful cue has to be crafted carefully. It is supposed to be part of the innate equipment and it should protect its innocent users against all the world's grammatical distributions. Dresher (1999) even advocates a maximum use of context cues (see also Dresher and Kaye 1990, Lightfoot 1991). He argues that a sufficiently developed system of (innate) cues could do without the whole method of error-driven parameter setting.

Other attempts maintain error-driven parameter setting without cues. They evade the local maximum by adding more computational power to the acquisition procedure. For example, they

---

[2] By a surprisingly simple approach, Brown (1973) arrived at a picture of the stepwise hierarchy in morpho-syntactic progress. In attempting to explain the hierarchy, he introduced a complexity measure employed at the time, namely *Derivational Complexity*, i.e. the number of transformational steps needed for deriving the structure. This complexity measure was at the time assumed to predict measurable performance delay. When that speculation failed, however, the field of psycholinguists started without much reason to mistrust the very notion of a complexity measure and acquisitionists even gave up trying to understand the problem of order in acquisition steps. Consequently, Brown (1973) was reduced to a conglomeration of remarkable observations. With hindsight, this may have been a strategic research failure.

allow several parameter values to be switched at the same time or they allow parameters to be switched on and off for some time in order to escape the effects of a premature setting. Such measures increase the search space for possible solutions and thereby weaken the model. The most interesting variant of these is probably Yang's (2002) stochastic Variational Model (cf. also Clark 1992). Yang lets his acquisition procedure try out all kinds of parameter settings in a random way. The moment of irreversibly setting any parameter is delayed. Most such arbitrary settings will of course be widely off the mark and not succeed in deriving much of the primary data. Yet, the chance of a parameter being tried out in a plus or minus setting is enhanced by "bonus points" when that setting happened to be involved in a few successful derivations of primary data. By adding a clever arrangement for assigning bonus points, Yang's computer simulations of the acquisition procedure sift out the correct parameter settings after an acceptable amount of time (program re-runs). There is a remarkable secondary effect. Yang's system of *a priori* parameters does not include stipulations about an order of parameter setting. Yet, when applied to the real primary data, the correct parameters come up in a specific order. Some parameters assemble a decisive amount of bonus points before others.

Let us remind ourselves of the original problem. There was, as Dresher (1999) diagnosed, an inevitable and principled gap between abstract parameters and primary data. The attempts to bridge the gap between abstract parameters and plausible acquisition procedures consist so far in stipulating more innate structure, such as an (*a priori*) order of parameter setting, or (*a priori*) cues added to the parameters, with the effect that the difficulty of local maxima is nullified. Sometimes, as in the case of Yang (2002), there appears to be an innate talent for probability computations that identify abstract pattern factors. One may wonder why the human animal had to come up with languages that require such elaborate innate and task-specific provisions. There are other human activities that display kinds of rule-governed creativity. The art of playing chess in sequences of moves, the art of improvising music in sequences of striking intervals, or the art of crafting sequences of interacting tool sets are all exclusive to human beings. Do these and all other cultural activities require innate task-specific talents as well? And if not, why is language the exception?

1.2 Innate parameters?

Chomsky sometimes seems to support such skepticism about task-specific talents. He repeatedly mentions the case of arithmetic (Chomsky 1987, 2005). It has been found that counting five or six bananas is well beyond the grasp of otherwise intelligent chimpanzees. The difficulty seems to be this: it is not an inherent property of a banana to come in as first or second banana. A banana is second banana only because just before some other banana has been indicated as first banana for otherwise pretty arbitrary reasons. In order to count bananas, you must not only be attentive to bananas but also, and simultaneously, be attentive to the sequence of your own naming activities. This is more than any chimp with a healthy interest in bananas can work himself up to. The human being, by contrast, has sufficient attention and memory to spare. It keeps track of its own naming activities, sees the parallel between the naming series and the banana series, and gets the counting trick. Each banana is associated with a provisional number name, say 'six/sixth'. The name is determined by the preceding number name, say 'five/fifth'. Each number name must have a successor number name if the counting game is to go on. Inevitably, the amount of bananas will at a certain moment exceed the amount of numbers provided by some five or ten finger nursery rhyme. Other tricks must follow to provide new

number names to measure up to the amount of naming activities. Recursively raising ten spread fingers may do, as one still sees at some markets. Then and there, at the market, the world of mathematical structures opens up as a matter of achievements in cultural history. The parallel between the bananas and the mental acts is neither fruit-stuff nor (subjective) mental stuff. It is mathematical stuff. Eventually, mathematical structures can become ever more complex and surprising. It would be curious to propose that a new invention for number structures comes out of an innate neural surprise box. Yet, the inventions are predetermined. For example, the decimal notation for fractions was invented by the Dutch engineer Simon Stevin in his arithmetic book *De Thiende* (1596, 'The Tenth One'), but it had been invented earlier by Islamic astronomers (Al-Kashi *Sallam At-sama* 'The Stairway to Heaven' 1407). It answered a system-internal problem with dividing numbers. Nobody would relate that invention to until then inoperative neural constraints in the human brain. When something like the accumulative invention of system-internal solutions holds for mathematical structures, why should it not hold for grammatical structures? It seems reasonable to be skeptical of all assumptions about parameters as innate task-specific *a priori*s.

This seems close to abandoning the notion parameter as such. Yet, that is only apparently so. The parallel with arithmetic may be helpful. It is very well possible that an abstract system is learned and applied due to a set of local options that invite some trivial rules of thumb. The true effectiveness of the overall system is only understood afterwards or even not at all. The parameter is effective in the system rather than being an abstract decision by the user.

2   Input Reduction

Accumulative learning means that one begins with simple problems before one gets a clear picture of the more complex ones. We propose that the *language acquisition procedure* begins with a radical reduction of the input. The reduction yields a set of utterances that consist of two directly identifiable elements. The reduction follows from a simple common sense principle in (3). We will give it a subtler form in (6).

(3) Reduction of input to intake: Leave out what you cannot fit in.

The reduction procedure in (3) initially ignores all grammatical markings (functional categories): articles, copulas, auxiliaries, verbal inflections, connectives. This is because a functional category $<F_i>$ indicates a grammatical relation between two phrases $[XP [F YP]_{FP}]$. It is a word that carries little 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'. Although each of the functional categories in child-directed speech is 100 to 300 times more frequent than an arbitrary denotational word, functional categories cannot be identified and learned before regular combinations of denotational words are perceived and sensibly combined. Hence, sentences in the child-directed speech like *ain't the bear nice?*; *the bear is nice, isn't he?*; *I want the bear to be nice* are turned into [*bear nice*]. A set of binary word constructions is the result. They relate pragmatically as topic-comment or as operator-comment. [3] The operator-like words have an immediate situational deixis. This at least happens in the child's actual output. A longitudinal

---

[3] The example set of typical binary constructions in (4) is not exhaustive. We just give types that are relevant to the present exposition.

analysis for one language and one child is necessary to illustrate how input reduction leads to an order of acquisition steps. The present article considers the acquisition steps in Dutch taken by the child Sarah as a case study. The files are available in CHILDES (Van Kampen corpus, MacWhinney 2006). Typical two-word sentences in Dutch child language are given in (4).

(4) a. modal operator + Y
         kwi   beer   (wanna bear)
         ga   slapen   (go sleep)
         moet   op   (must on/up)
     b. deictic operator +Y
         is   drinken   (is drink)
         dis   pop   (this dolly)
         dats   beer   (that's bear)
         dit   op   (this gone)
     c   X (topic) + Y (comment)
         beer   lief   (bear nice)
         papa   slapen   (daddy sleep)
         schoen aan   (shoe on)

Most examples of pragmatic illocution operators in Dutch derive from modal verbs in the input: *kwi*(l) (wish 'I want'), *kga* (intention 'I intend/I am going to'), *moet* (requirement 'it should be/it goes'). The (later) subjects are not yet present as phrase or word. They rather are 'mode-implied', since these early operators do not vary in 1st, 2nd, or 3rd person (Van Kampen 2006b). The intention is always *I want, it must (be that)*. We feel that it would be adult over-interpretation to analyze these early pragmatic illocution operators as finite verbs. Finite verbs presuppose far more system (a verbal paradigm <±finite> and a subject-predicate construction) than is available. There is also a statement and name-giving operator *dis/dats*. It seems to fuse the demonstrative and the copula. It might once again be adult over-interpretation to analyze these early forms as phonetic fusions. They fit the binary frame more naturally as a single illocution operator unit.[4]

---

[4] In a typological study of intransitive predicates in 410 languages, Stassen (1997: 99) observes that in many languages the morphological form of the copula recalls the morphological form of demonstratives. This seems less strange if both elements are picked up in child language as an illocution operator 'statement' introducing the utterance with colons. Such an analysis is possible for (at least) Dutch (i), English (ii), and French (iii) child language.

(i)   Dutch (Sarah: week 120, Van Kampen corpus)
       a. is: gieter/bordje/water/plakkertje   (is: can/plate/water/sticker)
       b. is: heet/lekker/vies   (is: hot/nice/dirty)
       c. is: mij/jou niet   (is: mi(ne)/not you(rs))
       d. is: op   (is: on)
       e. is: koud buiten   (is: cold outside)
       f. is: deze koud buiten   (is this one cold outside)
(ii)   English (Nina: week 112; Suppes corpus)
       a. is: Mommy living room
       b. is: Mommy's living room
       c. is: a girl
(iii)   French (Grégoire 1;10.20 week 98; Champaud corpus)
       a. est: crocodi(le)   (is crocodile)

The binary word utterances have minimal grammatical structure. They relate either two denotational words to each other or they relate a denotational with a fixed operator. Relating parts of an utterance in a grammatical frame offers the same problem as banana numbering. The succession of elements reflects a structure in the situation. Again, perceiving the parallels between situational structures and grammatical markings is beyond the grasp of the nonhuman primates.

The input reduction (3) will initially set aside all functional categories as unknown, not directly interpretable <F?>. The learner need not be aware of a grammatical difference between denotational words with a direct meaning and functional elements that mark a more abstract level of meaning and categories. The distinction between denotational words and more abstract elements is simply imposed upon the learner as elements provisionally comprehensible {*beer, slapen, lief*}, {*dis, kwil*} versus high frequency elements not yet comprehensible {tense *-te*, agreement *-t*, articles *de, een*}. The words with immediate meaning can be acquired separately as names or characterizations within an immediate situation in front of you. Functional categories are left out not because they have no (semantic) meaning, but because their meaning can only be ascertained within the syntactic and morphological oppositions that are still to be acquired. It is in any case a fact that child language begins with binary utterances that lack functional categories, although these categories (articles, auxiliaries, inflectional endings) are highly frequent in the input language. This period without functional categories is the period that Lyons (1979) may have had in mind when he suggested that early child language might have proto-predication as a forerunner of predication and also that child language might have proto-reference as a forerunner of reference. We propose an interpretation of Lyons (1979) in (5).

(5)   a. proto-reference   ~   naming function
       b. proto-predication ~   comment/characterizing function
       c. proto-illocution   ~   (pragmatic) operator function

The proto-distinctions are made within a situation-bound system. Each element is supported by the situation it is used in. The relation between the elements is pragmatically guessed. There is a cognitive abstract frame that schematizes that situation-bound understanding. The elements that signal an abstract frame are present in the input, but they have been set aside as <F?> by the reduction to intake. When the relation between two denotational elements X and Y is nevertheless repeatedly and correctly guessed, the learner may observe that an abstract frame for the relation was given in the neglected elements <F?>.[5] The utterance *bear nice* allows the insertion of a copula when it is recognized as a kind of property attribution. We assume that

---

b. est: casquette d' Adrien   (is cap of Adrien)
c. est: chaussette Victor   (is sock Victor)

Stassen's parallel between the demonstrative and the copula may find its origin in proto-grammar where the distinction D^o/I^o is not yet made.

[5] It has been observed by Locke (1997) that early child language – the form without grammatical markings – is processed in the (then dominant) right brain hemisphere. The left hemisphere is more hesitant and applies no reduction of functional markings. It preserves them as phonological peculiarities. When the pragmatic understanding by the right hemisphere becomes more difficult due to the rapid expansion of the lexicon, the left hemisphere contains hints about the relations due to the functional markings it still attends to. The systematic reliance on these hints, the functional markings, represents the grammaticalization of understanding and a shift from the right to the left hemisphere. Locke points out how the shift in grammar is strikingly supported by brain scans of small children (Mills et al. 2004).

reference, predication, and illocution exist only as factors within a grammatical system. Before the grammatical structure is in place, and the functional categories are acquired, all words are used in a strictly situation-bound manner and there is no grammatically encoded reference, predication, or illocution.
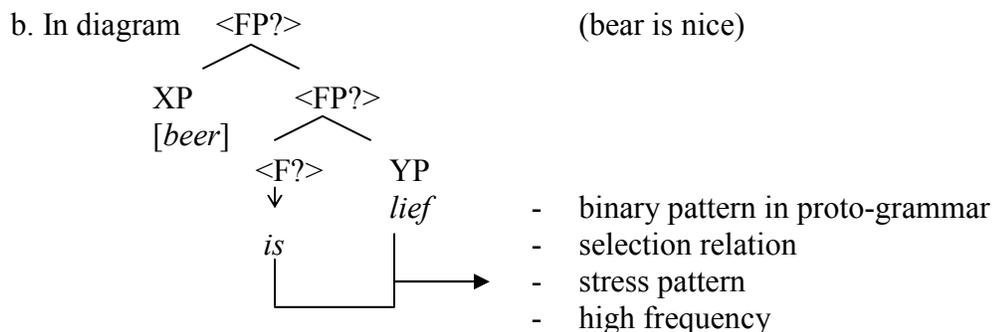
Since the adult input is dense with functional categories, all with language-specific restrictions, the acquisition procedure must apply a fixed strategy. We propose, as a specification of (3), the reduction principle in (6).

(6)   *Reduction of input to intake*
  a. first reduction: replace each functional category still unknown with <F?>.
  b. second reduction: set aside all input sentences with more than one <F?>.
  c. residue: an intake selection for the <F?>/<F_i> construction

The residue (6c) is a set of input sentences that yield an acquisitional focus for the category/parameter value <F_i>. The residue represents the minimal pattern from which the functional category <F_i> can be acquired within a certain language. It is our contention that the language acquisition procedure scans the input for such patterns. We will call these patterns *evidence frames*. The input reduction is such that the acquisition procedure concentrates on a single grammatical value <F_i> before the attention switches to the next acquisition step. Part of the acquisition procedure indeed confirms such a Single Value Constraint. Language is structured in such a way that the evidence frames must turn up in a linear order signalled by the order of acquisition steps. Note that the order of acquisition steps is not assumed to be universal. The order of acquisition steps is derived from evidence frames and the evidence frames are derived from a natural reduction of the input language. The strategy in (6) is not language-specific for getting the evidence frame, but its outcome, the evidence frame, is just that: a language specific frame.

The reduction of the input to an intake selected for a single grammatical point <F_i> confronts the acquisition procedure with a repetitive set of short sentences all containing the same functional category in the same position. In a sense, the evidence frame is a language-specific drill for the category <F_i>. An example could be the copula construction in (7).

(7) a. *Learning step*: identify the pragmatically understood <F?> as <F_i>

   b. In diagram <FP?>   (bear is nice)

      XP <FP?>
      [*beer*]
        <F?> YP
           lief - binary pattern in proto-grammar
        is   - selection relation
               - stress pattern
               - high frequency

The meaning of the binary frame in (7), 'property attribution', is pragmatically guessed and demonstrates the contribution by <F?>. We assume that all functional categories fit such an evidence frame. When <F?> is identified as the grammatical function <F_i>, a full utterance of

the input has been identified. When this acquisition step has been made, the filter in (6) may reapply and select a new intake from the same type of input ('Single Value Constraint', Berwick 1985: 108, Clark 1992: 90). The next functional category then comes into acquisitional focus. Its context, the context of $<F?/F_{i+1}>$ allows the presence of $<F_i>$. $<F_i>$ passes the filter in (6), since it has been identified and acquired. This suggests an ordering of acquisition steps. The later steps select more language-specific and more structured contexts. Turning this around, we may see the empirical presence of ordered acquisition steps as an indication that a data reduction filter (6) has been at work. If (6) selects the relevant data-set from the input, the category $<F?>$ becomes the acquisitional focus. That focus turns the diffuse stimulus of the input into the effective stimulus of a temporary intake.

The data selection function of the earlier acquisition steps has been noted in Berwick and Weinberg (1984: 284, note 2). It is proposed here that data selection D_{i+1} by the provisional grammar G_i is crucial. The unreduced input represents a diffuse stimulus, but the selectively reduced input by the procedure in (6) may prove to be a sufficiently focused stimulus. It may eliminate Chomsky's argument from the poverty of the stimulus. As a first result, the reduction of input to intake provides the learner with elementary Greenberg-type patterns that facilitate the acquisition of the major parameters of the system (Greenberg 1963). Studying the order of the child's acquisition steps provides the key towards the built-in learnability of grammar. The initial acquisition of the copula as an illocution operator for root statements (*is lief* 'is nice') subsequently allows the evidence frame [name $<F?>$ comment] to be solved (*beer is lief* 'bear is nice'). The copula now marks the comment within a topic-comment structure.

The evidence frame has been represented as the change from a binary construct 'XP (topic) YP (comment)' to a triple one 'XP [$<F?>$ → $<F_i>$copula] YP'. This is not close enough to the actual procedure. In the first place, the copula is but one of the elements that appear in the in-between position. This leads to a further, more important consideration. The various operator elements $<F_i>$ are not freely combinable with any 'comment' element. Rather, the operator elements, - the later modals, copulas, aspectuals and auxiliaries – associate with different subclasses in the complement. These stereotype selections can be modeled by a convention. The denotational element in the comment is countermarked in the lexicon as $<-F_i>$ when it combines with and is selectable by the operator $<+F_i>$. The associative counter-marking opens the general possibility that denotational elements split up into categories according to their combinatory restrictions with given functional categories.[6] The selection restrictions between operator elements and the comment do not change when the binary stage changes into more complex triple combinations, as in (7).

When the binary system contains the type of combinations in (8), the acquisition procedure is near to being able to recognize and produce the structures in (9).[7]

---

[6] The evidence frames seem close to the position of Tomasello (2003) that language acquisition is construction-based. There is an important difference, though. The evidence frames are supposed to add syntactic features $<+F_i>/<-F_i>$ to lexical items. Once they have done that, evidence frames disappear from grammar.

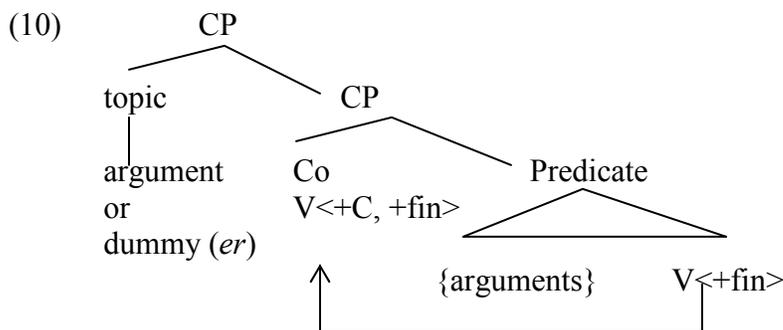[7] The term 'recognize' may be somewhat unfortunate. An experimental tradition of the last twenty years has shown that children recognize far more structure than they are in fact able to produce. This is not unlike perceiving a fugue by Bach or a perspectual effect by Esher. Perceiving something does not yet enable one to reproduce anything like it. For some fortunate reason, we can perceive and recognize more than we can produce.

(8)   a. ⎰ is lief   (is sweet)   ⎰ moet slapen (must sleep)
        b. ⎱ beer lief (bear sweet)   ⎱ beer slapen (bear sleep)

(9)   a. ⎧ is   [beer lief]   ((now) is (the) bear sweet)
            ⎩ moet [beer slapen]   ((then) must (the) bear sleep)
        b. ⎧ beer [is lief]   ((the) bear is sweet)
            ⎩ beer [moet slapen]   ((the) bear must sleep)

What happens in the three-word stage (9) is a nesting of two types of binary constructions into each other. The selection restrictions are preserved. The inner units [topic comment] in (9a) and [operator comment] in (9b) had already been constructed earlier. The binary system is preserved with either the operator in the peripheral position as in (9a) or the topic as in (9b). The patterns in (9) are reinforced by the maternal input, as indicated by the glosses, e.g. *(nu) is beer lief* '(now) is bear sweet'. These maternal structures have a subject topic (9b) or a dummy topic followed by subject-inversion (9a), but that much structure cannot yet be perceived by the child. Distinctions like (i) external argument/subject (ii) finite verb and (ii) subject-inversion belong to a more advanced system that is not yet active. It is important to notice this because an analytical perception of (9)a as free inversion of finite verb and subject is the cue in Fodor (1998 'treelet') and Lightfoot (1999) for setting a V-second parameter. Even if discussions about the A-/A-bar status of the initial phrase are relevant, they cannot be relevant at this stage of acquisition (see Van Kampen 1997: chapter 3). The acquisition procedure must first single out a class of words in predicate final position, make a distinction between the <±finite> elements in that class, introduce the category V, construe an initial head position of finite verbs and use that position to identify subjects and topics in its specifier.

3   Evidence frames for underlying directionality

The acquisition analysis below is based on the assumption that the target, adult Dutch, is an SOV language (subject precedes the predicate and object precedes the verb), with the complication added that the finite verb in root clauses is not in the SOV predicate-final position, but appears in an initial position right after the first major phrase (the V-second position). The marked and prominent position of the finite verb probably serves as a warning that the sentence has an illocutionary value (Van Kampen 1997, cf. Truckenbrodt 2006 for recent discussion). Subordinate clauses lack the illocutionary function and do not apply the V-second rule. The finite verb of the subordinate is again in predicate-final position.

(10)   CP

        topic   CP

        argument   Co   Predicate
        or   V<+C, +fin>
        dummy (*er*)
                          {arguments}   V<+fin>

The <+fin>-marked verb is associated with two positions, a predicate final position that serves to subcategorize the denotational verb for its complements, and the C^o-position that serves to mark the illocutionary function of the root clause. As in Weerman (1989) for reasons of method, and in Van Kampen (1997, 2006b) for learnability reasons, no separate I^o/T^o-position is assumed here (see also Watanabe 1994). The Germanic V-second parameter can now be presented as setting a <+C> value for <+fin>. A finite verb marked V<+fin, +C> will move into the C^o-position, not unlike the D^o-marked <+wh> will move into the SpecCP. The <+C>-value is de-activated in subordinates by the complementizer constant <+C>. It is not really the point of the present section to argue for the minimal structural analysis in (10). The basic point is that any structural analysis of the V-second clauses requires a learnability account of the kind given below. An alternative analysis, for example Zwart (1993), that Dutch is after all SVO and/or requires a host of additional verbal positions, meets with somewhat different, but no less serious, learnability problems. In all variants, a plain surface distribution must offer the learner access to underlying regularities, preferably with a minimum of *a priori* stipulations.

Gibson and Wexler (1994) confront the problem of setting the V-second parameter. They assume the Single Value Constraint, which holds that parameter setting is a stepwise procedure where each step sets a single parameter value. They show that the acquisition procedure for the Dutch type of grammar (SOV/V-second) should somehow succeed in setting the underlying Greenberg directionality parameters (subject before predicate, object before verb) before setting the V-second parameter. Yet, there is an input problem here. The V-second surface effect is present in nearly all input sentences (93% in our counts), and by consequence the underlying structure SOV is not unambiguously present. The predicate-final position of denotational verbs comes out in the subgroup of constructions where the finite verb is an auxiliary/modal and the denotational verb is transitive, see (11).

(11) a. subject Aux<+fin> [object V<+inf>]
          papa   gaat   [een boek lezen]
          daddy   goes   a book   read
          (daddy will read a book)
      b. adjunct Aux<+fin> [subject   object   V<+inf>] (narrative inversion)
          nu   gaat   papa   een boek lezen]
          now   goes   daddy   a book   read
          (now daddy will read a book)

The construction types in (11) represent only a small subset of all input sentences (15%) since explicitly transitive verbs are not that frequently present. It is a legitimate question asking why and how the acquisition procedure should select this subset to construct a grammar.

In order to make this problem more clear, Evers and van Kampen (2001, section 3.1) assembled the input given by the mother in the months just before her daughter Sarah acquired the V-second rule (Van Kampen corpus, CHILDES). The presentation below (in sections 3.1-3.2) is slightly schematized, but basically the same as earlier. There were 1017 utterances with a verb, 95% of them contained a finite verb, 93% of root sentences with the finite verb in V-second position and 2% subordinates with the finite verb in final position. Two subsets were constructed from the set defined by the presence of a finite verb, one subset containing the constructions with a finite verb and a subject, the other subset containing the construction with a finite verb and an object/complement. The question is whether the subsets could serve to identify

fixed positions for subject and complement. The actual input for subject or object identification is complicated by the fact that the denotational theta-assigning verb may move into initial position and precede rather than follow the complement and often, in case of subject-inversion, precede the subject as well, see (12).

(12) a. papa   leest   een boek t_{Vfin} (object follows the denotational verb (VO))
          daddy   reads   a book
      b. nu   leest   papa een boek t_{Vfin} (subject follows the verbal predicate (VS))
          now   reads   daddy a book

These distributional complications, consequences of the V-second rule, offers a hindrance for the early acquisition procedure. As long as the underlying structure has not yet been established, the V-second effects appear as mere distributional surface noise. The sentences in (12) have first to be analyzed as revealing an underlying structure.

3.1   Paradoxical input

The adult grammar should eventually constructs a relation between the subject and the finite verb, but the subgroup defined by the finite verb and the presence of an explicit subject shows the following puzzling percentages. Regardless of whether the finite verb is a theta-assigner or not, the subject precedes the finite verb in 30% and follows it in 70% of the constructions. The percentages 30/70 are proportions between the relevant subsets.[8]

(13) a. Subject-V<+fin>   **(30%)**
          papa moet een boek lezen   (daddy reads a book)
          papa leest een boek   (daddy reads a book)
          dat boek is leuk   (that book is funny)
      b. V<+fin>-Subject   **(70%)**
          nu *gaat papa* een boek lezen   (now will daddy read a book)
          nu *leest papa* een boek   (now reads daddy a book)
          leest papa een boek?   (reads daddy a book?)
          nu *is de sap* op   (now is the juice gone)
          op *is de sap*   (gone is the juice)

It seems then that the distinction <±finite> verb does not suffice to identify a basic subject configuration.

The object-verb directionality needed for the SOV type has its problems as well. The straight Object-V<−fin> pattern reaches 45% (14a) and trails other transitive constructions (55%). This is due to transitive constructions where the denotational verb moves by the V-second rule or the object gets topicalized (14b). Again, the percentages are proportions within the subset of transitive verb constructions in the maternal input.

---

[8] This does not contradict the count in Van Kampen (1997: 59) and in Lightfoot (1999: 153) that the finite verb is preceded by a subject in 2/3 of the cases and by a non-subject topic in about 1/3 of the cases, counting simple statements in the maternal input. The actual input is further complicated by the subject inversions due to questions and narrative inversion. Within that larger set, the percentage of subject-V<+fin> configurations lowers to 30%.

(14)   a.   XP-Aux$_{fin}$-[Object-V] **(45%)**
             papa moet *een boek lezen*   (daddy must a book read)
             nu gaat papa *een boek lezen*   (now will daddy read a book)
       b.   movement of the denotational verb **(55%)**
             papa leest een boek   (daddy reads a book)
             nu leest papa een boek   (now reads daddy a book)
             leest papa een boek?   (reads daddy a book?)
             topicalization of the object
             dat boek leest papa   (that book reads daddy)
             dat boek gaat papa lezen   (that book will daddy read)

The quantitative proportions in the input for the acquisition of Dutch seem designed to prove Chomsky's point that the input offers the acquisition procedure no more than a confusing and poor stimulus, if it offers any stimulus at all. On the other hand, no matter how much *a priori* guidance one is willing to postulate, there can be no warning in the cradle of the type "be careful, within a short time you will enter the domain of an SOV language marked by a V-second root movement". In other words, the input itself has to offer access to the system. Paradoxically enough, it seems to do so. The Dutch child begins by ignoring the massive evidence for the V-second position of <+fin>. Initially, she strongly prefers the so-called 'root infinitives'. This will be shown by longitudinal graphs in the next section. By that approach, she immediately opts for the object-verb order, in spite of the fact that this order constitutes only 45% of the transitive constructions in the input. Finally, the rise of finite verbs in child Dutch correlates with a systematic appearance of the subject, although the subject shows no obvious distributional preference for a standard position in the specifier of the finite phrase (CP).

3.2   Intake by evidence frames

Let us consider now how systematic input reduction and its resulting language-specific *evidence frames* control the acquisition procedure. As observed by Lebeaux (1988:11f), the child's initial structures for two denotational words in proto-grammar represent minimal theta-assigning structures of sister government. If the acquisition procedure is attentive to sister government structures, content elements are perceived as belonging to linearly fixed frames for theta-assignment. The choice of the frame follows from the most frequent PF predicate form.

    Let the acquisition procedure focus on adjacency when setting up the theta-frames for lexical items. Distributions of transitive structures for deriving the OV order are given in the bracketed phrase in (15). Keep in mind that the child is as yet unaware of the V-second rule/finiteness marking. Hence, she may also consider the bracketed phrase in (15)b as potential evidence for a theta-frame. The remaining transitive structures (15)c are irrelevant and neglected due to the initial data reduction by the adjacency requirement. The OV/VO relations in (15)c are not string adjacent. The subject is in between.

(15)   Frames for Object-Verb/Verb-Object
       a.   *Evidence*: adjacent Object-Verb **(78%)**
             papa moet een [*boek lezen*]   (daddy must a book read)
             nu gaat papa een [*boek lezen*]   (now will daddy read a book)
             dat [*boek leest*] papa

b. *Counter-evidence*: adjacent Verb-Object   (**22%**)
   papa [*leest* (een) *boek*]   (daddy reads a book)
c. *No evidence*
   nu leest <u>papa</u> een boek   (now reads daddy a book)
   leest <u>papa</u> een boek?   (reads daddy a book?)
   dat boek moet <u>papa</u> lezen   (that book must daddy read)

A parallel evidence filter can be used to determine prospective subject directionality, where the predicate functions as the theta-assigner. Subject directionality holds for all predicates, verbal and non-verbal (copular constructions). Within the present view, early child grammar cannot distinguish these two kinds of predicates, since the verbal paradigm is not yet available. Therefore early child language automatically addresses the more abstract, less specified subject-predicate directionality of the input rather than subject-finite verb directionality.[9]

(16)   Frames for Subject-Predicate/Predicate-Subject
        a. *Evidence*: adjacent Subject-Predicate   (**68%**)
           nu gaat [papa [(een) boek lezen]]   (now will daddy book read)
           [papa leest] een boek   (daddy reads a book)
           nu (is) de [sap op]   (now is the juice gone)
           ik vind [papa lief]   (I find daddy nice)
        b. *Counter-evidence*: adjacent Predicate-Subject   (**32%**)
           nu [leest papa] een boek   (now reads daddy (a book))
           [leest papa] een boek?   (reads daddy (a book)?)
        c. *No evidence*
           papa <u>moet</u> een boek lezen   (daddy must a book read)
           dat boek <u>is</u> leuk   (that book is funny)
           op <u>is</u> de sap   (gone is the juice)

The theta evidence frames for the subject argument will place brackets as in (16)a, setting aside all the functional categories (auxiliary, modal verb and copula) that precede the subject. That reduction allows the acquisition procedure to tap the rich amount of subject inversion structures. These support the acquisition of theta frames for denotational verbs. Some counter-evidence is brought in by subject-inversion with a denotational verb, (16)b. These constructions in (16)b where the denotational verb is finite, must either enter the system as fixed idioms or remain unanalyzed. The prospective subject is now determined by lexical structure. In (16)c the copula or modal verb is between the subject and predicate. For that reason, these sentences are irrelevant and not considered as long as the acquisition procedure is still controlled by Lebeaux' filters of theta-assignment under strict adjacency (Lebeaux 1988).

The bracketed evidence frames in (15) and (16) now yield (17) as input evidence for argument order.

---

[9] The subject is string adjacent to the theta assigner within the evidence frame when the early object verb frames in (16)a are read as incorporations (Van Kampen 1997). All early theta assignments are ambiguous in status between theta-frame and stereotype lexical association as Lebeaux (1988: 13) observes. They certainly fit Tomasello's construction-based grammar in early child language. Explicit argument marking by D^o appears later, after V-second, see the graphs in (18) section 4. The lexical associations in proto-grammar do exhibit the dominant SOV order and they initially obviate the V-second pattern as shown by the <+C, +fin> graph for V-second in (18).

(17)   Directionality: Input percentages on string adjacent evidence frames

| *Parameter*	Ratio in input [+ evidence]	Ratio in input [−evidence]
Subject is adjacent to and precedes theta-assigner	**68** %	32 %
Complement is adjacent to and precedes theta-assigner	**78** %	22 %

The bracketed evidence frames in (15)a and (16)a suggest that the acquisition procedure could set the argument frame as a property of the lexical item that assigns the theta roles by the following practical decision. If there is support from some 2/3 of the distributional evidence within the relevant frame, the directionality parameters are set for the underlying SOV order in Dutch. The stage for the acquisition of the V-second rule has now been set as will be seen in the next section.

4   Order of acquisition steps

When a previously unknown category <F?> has been identified as <F_i> , the acquisition procedure has two expressions at its disposal [X + Y] (*bear nice*) and [X + <F_i> + Y] (*bear is nice*). Eventually, the more specified adult variant will block the less specified child variant. Blocking in language acquisition takes some time and it can be traced by longitudinal graphs (Van Kampen 1997). The acquisition graph identifies a 'parameter setting' by its blocking effect on the less-specific predecessor. Graphs for different parameters reveal the order of acquisition steps. See the graphs in (18) for Dutch for predicate marking by V-second and argument marking by D^o (Van Kampen 2004). The graphs are constructed from the Sarah corpus available in CHILDES (MacWhinney 2006). The V-second graph establishes the marking of a predication by a <+C, +finite> illocution operator in the C^o position.[10] The C^o<+fin> factor generalizes over a variety of categories {copula, auxiliary, modal, finite morphology}. D^o<+det>-marking is the explicit marking of arguments for reference. The reference factor also generalizes over a variety of categories {article, demonstrative, possessor, quantifier}.

---

[10] We prefer C^o<+finite> rather than C^o<+tense>. The opposition <±finite> is present in the child's early grammatical system, whereas the opposition <±tense> comes in later, when the V-second rule is already acquired. The question is not whether the child is cognitively aware of time. Any conscious being is aware of time. The question is when such awareness is systematically expressed in forms of language. The oppositions <±finite> and <±tense> are more problematic in the acquisition context for English. We are happy to leave that aside here.

(18)   Dutch Sarah: the acquisition of C^o<+fin>-marking and D^o-marking
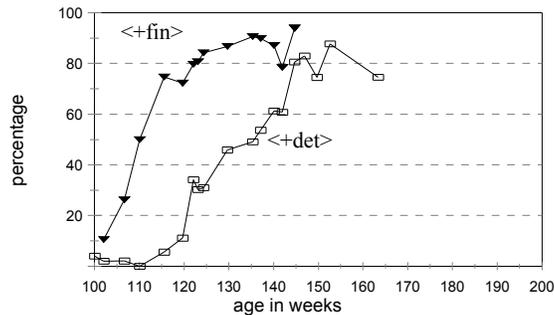
Dutch Sarah applies systematic <+C, +fin>-marking (V-second) almost half a year earlier than systematic D^o-marking (articles and their like). She acquired C^o<+fin>-marking at week 120 and D^o-marking at week 145.[11] The same order of appearance, the acquisition of finite verbs preceding the acquisition of determiners, was found for French (Van Kampen 2004), English (Brown 1973) and for Rumanian (Avram and Coene 2004, this volume).

   The C^o<+fin>-marking graph of Dutch Sarah in (18) simply lists the number of elements in the first or second position that could be seen as a finite verb in the adult system. What actually happens in the rise of the graph is more intricate. The lower part of the graph can be covered by a more simple system. There are designated elements that function only as pragmatic illocution operators in the initial utterance position. These are the later modal verbs. They do not yet require the presence of a full predication with a topic and a subject. When the graph rises above 50%, finite forms of denotational verbs come in. The second part of the <+C, +fin> graph rises due to the operators that are derived from denotational verbs. At the same time, the finite form begins to imply a full predication frame (topic, subject and arguments).

   The child's optional infinitive constructions (*papa boekje lezen* 'daddy booklet read') disappear either for an explicit modal verb (*papa moet boekje lezen* 'daddy must booklet read') or for a finite denotational verb (*papa leest boekje* 'daddy reads booklet'). The last construction (the V-second construction) is anomalous in two ways, see the diagram in (19).

(19)
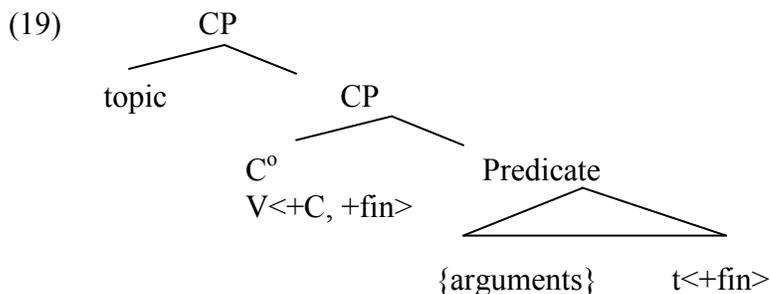
There is no designated illocution operator present in C^o and there is no theta-assigner present in the predicate final position. The theta-assigner is a morphologically marked <+fin> and it

---

[11] It is true that there are examples with a D^o-marking and without a C^o<+fin>-marking for Dutch Sarah, especially around week 120, as the graphs in (9) suggest. Note that this is not relevant to the question of why systematic C^o<+fin>-marking precedes systematic D^o-marking .

appears in the C^o position of the illocutionary operator. The set of theta-assigners that allow this switch form a category 'verb'. Some notation is needed to express that the theta/argument frame known from the <−fin> variant reappears for the <+fin> variant. The derived secondary effect of the theta structure is expressed by a movement of the <+fin> form from the predicate final position into the C^o-position. If movement rules are not an option in a grammatical model (e.g. Categorial Grammar, HPSG, LFG), a derived theta frame for <+finite/illocutionary> could be expressed by a separate frame for the <+C, +fin> form. The important point for the acquisition procedure is that the V-second rule identifies the V^o in C as a distributional variant of the V^o in predicate-final position. It is not the task of a grammar or an acquisition procedure to merely list variants. They have to capture the underlying identities.

One may see the acquisition of the V-second rule in the following way. There was originally an option in proto-grammar. The topic and/or the illocution marker could be left out. That option gets marginalized for a norm that requires both the topic and the illocution marker to be realized. Constructions where a suitable modal or copula is lacking bring in a finite form of the denotational verb as illocution marker and preserve the verb's full argument structure. Constructions where a suitable topic is lacking fit the adult pattern with a dummy topic or adverb.[12] See the examples in (20) and (21). Proto-forms are given in (20). Forms with the first grammatical devices appear in (21), where (21)a realizes all options from the proto-grammar. (22)b realizes the V-second shift of a denotational verb, (22)c the use of a dummy topic and (22)d all the previous devices together.

(20)　a.　beer slapen　　　(bear sleep)
　　　　b.　gaat slapen　　　 (goes sleep)
　　　　c.　(in) bed slapen　　((in) bed sleep)

(21)　a.　beer　　gaat　　in bed slapen
　　　　　　bear　　goes　　in bed　sleep
　　　　b.　beer　　slaapt　in bed　−
　　　　　　bear　　sleeps　in bed
　　　　c.　er/d'r　gaat　　een beer　slapen
　　　　　　there　goes　　a bear　　sleep
　　　　d.　er/d'r　slaapt　een beer　in bed　−
　　　　　　there　sleeps　a bear　　in bed

The systematic C^o<+fin>-marking and the D^o-marking of arguments lead to further acquisition steps, beginning with a grammatical decision procedure on the category membership V versus N (Van Kampen 2005).

4.1　Acquisition steps due to local evidence frame IP/DP

The systematic marking of comment parts by a variety of <F?> elements {copula, modal, auxiliary, inflection} divide the denotational elements of the comment into a subclass V and a

---

[12] V-second and dummy topics are not simultaneously acquired. The systematic introduction of dummy topics does not take place before the acquisition of <± definite> articles on arguments and the acquisition of D^o-marking follows the acquisition of V-second, see (18). This ordering "D^o-marking on arguments precedes the introduction of dummy topics" is predictable, since topic selection is sensitive to the ± definite distinction.

subclass non-V. When a denotational element has access to all the C^o<+fin>-markings in (22), it is a verb. Verbs and nouns are considered here as derived from the grammatical system. They are not considered to be innately or cognitively given.

(22) X^o = V   when it may enter the C^o contexts in a, b and c

$$
\left\{
\begin{array}{ll}
\text{a.} & [<+C, +modal>] \quad ..... \quad [\underline{\qquad\qquad}] \\
& \qquad\qquad\qquad\qquad\qquad\qquad <+ \text{infinitive}> \\
\text{b.} & [<+C, +fin>] \qquad ..... \\
\\
\text{c.} & [<+C, +aux>] \qquad .... \quad [\underline{\qquad\qquad}] \\
& \qquad\qquad\qquad\qquad\qquad <+\text{past participle}>
\end{array}
\right\}
$$

The somewhat elaborate context condition in (22) simply states that a denotational element is V when it fits the verbal paradigm.[13] The parts of the paradigm (22) refers to, are present in early Dutch child language. Other parts of the verbal paradigm, especially the <+finite> oppositions for tense and number, play no part in the acquisition of the category V in Dutch. They appear later around week 140, as we will show in section 4.2. The intention of (22) is to stress that the paradigm is learnable from the C^o evidence frame. The acquisition of the category V by means of the evidence frame in (22) may take up to half a year in Dutch child language. This amounts to some high six-digit number of short parallel sentences to get the verbal paradigm and the category V. It was Briscoe (2000) who pointed out that the speed of acquisition steps is to be measured by estimates of the number of learning experiences. Within this perspective, language acquisition is the most intensive onslaught on our brain we have ever experienced. Grammar results from an unrelenting round-the-clock course that continues for years, almost a brainwashing.

Let the rise of the category V be simultaneous with the rise of <+C> by the stipulation that <+C> automatically causes a countermarking in its denotational partner <+C\–>. The countermarking simply indicates "fits the <+C> paradigm in (22)". The <+C\–> is better known as V. This morphological partner marking eventually reinterprets all auxiliaries as verbs. The identification of the category V by means of C^o<+fin>-marking refers to various constructions. Sometimes the <+C, +fin> factor is a morphological part of the V^o-element, sometimes it is a factor outside the verbal phrase at a distance from the V^o-element.

---

[13] This view on parts of speech is not uncommon. Nouns and verbs are identified as different kinds of words by their morphological paradigms. In as far as a language does not offer such paradigms, it will offer functional words <auxs> type C^o/I^o for verbs and <articles> type D^o for nouns. Our main point has been that the grammatical markers that lead up to V and N in syntax will do so in acquisition as well. Baker (2003) argues that, as far as comparative grammar is concerned, one would be better off to see it differently. Many a language lacks paradigms as well as C^o/I^o and/or D^o projections. Baker (2003) considers verbs as inherently predicative (type <e,t>), and nouns as inherently referential (type <e>). The functional categories have, in Baker's view, the more circumstantial task of indicating types of verbal predication {± modal, ± tense, ± aspect, ± negation} and types of nominal reference {± countable, ± gender classifiable, ± definite}. More importantly, their presence is an addition that can in principle be left out, and is often left out in certain languages. We must at least weaken our statement and claim that our acquisition procedure may succeed even when explicit markings of the verbal paradigm (I^o) or the nominal paradigm (D^o) are present as options only. If so, optional particles for mood, negation and illocution should be sufficiently present to identify V^o, and occasional demonstratives, possessives and classifiers frequent enough to identify N^o.

The identification of the category N by means of D^o-marking looks, by contrast, like a quite simple affair, that involves no paradigms or phrasal networks when the grammar is case-free, as it is in Dutch.

(23) X^o = N when it may enter the D^o context $[D^o \text{ --- }]_{DP}$

The difficulty with D^o and N^o is more with the acquisition of D^o. D^o implies two relations, one within the pragmatic discourse network and one within the network of sentence grammar (Van Kampen 2004, cf. also Őztürk, this volume, Ramchand and Svenonius, this volume, and Avram and Coene, this volume). The discourse D^o varies over two values <+definite> and <-definite>, which translates into 'previously mentioned' and 'not previously mentioned'. Sensibly adding the D^o-markers implies a daunting pragmatic task: check your frame for presupposition and check your immediate discourse context. This means reference-tracking, something far more complex than name-giving. The D^o-marked expression is not a new name. It rather represents the requirement for D^o<+definite> that the temporarily built discourse network should contain a previous mention of the same argument, and for D^o<−definite> that this should not be so. The D^o-relation is non-local and determined by pragmatic felicity conditions. In addition, the D^o-marked phrase is to be placed in a local syntactic network, where it is connected with the predicate head by means of a standard theta role. As such, one can also see D^o-marking as a grammatical expression of UTAH (Baker 1988). Consequently, a D^o-marked phrase includes a pragmatic index for reference tracking and an argument index for theta-role assignment (Baker 2003).

There are arguments that the referential as well as the case-marking function belong to the non-denotational D^o, rather than to the denotational N^o, although this position is rejected in Baker (2003). In the first place, when the referential index is expressed by articles, one expects that the case endings are in principle on the article as well, as in German, and not, or only marginally, on the noun. In the second place, when there are no articles in the language, one expects case-marking on the nouns as if the denotational N's had incorporated a D^o, since in such languages, the same case-marking will appear on the non-denotational free anaphors. This suggests that the indices for reference and theta role are carried by a non-denotational factor, *in casu* D^o (Van Kampen 2004, 2006a, cf. also Őztürk, this volume). The non-denotational free anaphors carry the local theta index, as well as the pragmatic reference index. All of this underlines Postal's view (1966) that personal pronouns and articles are deeply identical and both D^o.

There is a third argument that articles and pronouns are of the same type and it derives from language acquisition. The acquisition graph for the insertion of obligatory articles and the acquisition graph for free anaphoric pronouns coincide for Dutch Sarah in (24). See Van Kampen (2004) for an elaboration.

(24) *Sarah* (Dutch): Acquisition of D°<−pro> (articles) and D°<+pro> (3rd p. pronouns)
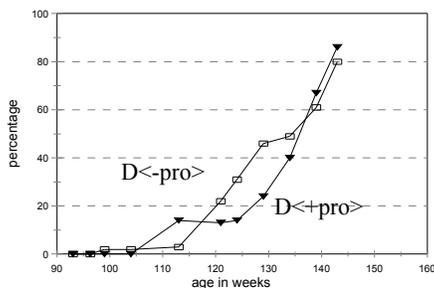
The coincidence of the two graphs is natural when both represent the acquisition of the double indexing system of the D°. Due to D°-marking, the language of the child becomes discourse-bound rather than situation-bound. As such, the rise of the two D°-graphs (<±pro>) has a certain magic. It shows how space is created for the full human flexibility with respect to situations. The argument for the D<±pro> identity in Dutch can be repeated for the acquisition of articles and free anaphoric clitics in Romance languages. In child French, the acquisition of the clitic system immediately follows the acquisition of articles, see Van Kampen (2004 and references cited there).

4.2 Local evidence frames outweigh mere input frequency

We saw in section 3 that frequency plays a role in deriving the basic word order, but that V-second is delayed in spite of its 93% presence in input utterances. Something like this also holds for subject-verb agreement.

The initial evidence frame for finite verbs leaves out the φ-features of person/number. Finiteness is first represented by the 3rd person singular only. This is also the default in comparative grammar (Benveniste 1966: 228ff, 255f). The acquisition of D°-marking leads to a subsequent step that brings in the φ-feature content of D° {± person, ± number}. The acquisition of φ-features on the finite verb <+C, +fin> follows the acquisition of φ-features on D°.

There is a measurement problem with 'correct agreement'. Irregular paradigms may take more than a year and there are sometimes difficulties with identifying the phonological forms on recordings. What has been done in (25) for φ-agreement in the speech of Dutch Sarah is a measurement of clear mistakes in plural agreement. Sarah went down from 7 cases of 'wrong' agreement (<+singular>) for 12 plural subjects in weeks 130-135, to 0 cases of 'wrong' agreement for 13 plural subjects in week 140. See Van Kampen (2006a) for an elaboration. Late acquisition of agreement has also been reported by Ferdinand (1996, for French), Avram and Coene (2003, for Rumanian).

The Agr-φ steps are a matter of weeks whereas the earlier <+C, +fin> and <+D, +det> steps were a matter of months. See (25) for Dutch Sarah.

(25)	step C°<+fin>	step D°	step D°-φ	step Agr-φ
	20 wks	25 wks	5 wks	5 wks

The more effective acquisition plausibly relates to the more effective evidence frame. The φ-feature acquisitions are supported by a lexicon with categorial marking <+fin> and <±plural>.

The EPP, the nearly obligatory presence of the subject in <+fin>-marked predicates, operates as a frequent and effective evidence frame for φ-agreement. The input of adult child-directed speech has of course not been lacking in φ-features on <+C, +fin> and D^o at all; rather the φ-features could not become part of the intake before the acquisition of <+C, +fin> (V-second) and the EPP had been established. This is a matter of clearly perceiving initial structure before it can become a quantitative factor in acquisition. The preceding EPP structure is needed to spot the relevant points. The EPP is acquired before agreement marking. From an acquisitional point of view, the later φ-agreement therefore appears as a final touch rather than as a structural underpinning.

5.   Movement structures

The preceding section introduced a movement rule. The movement rules reorder an underlying array of heads and phrases in order to arrive at the perceived surface structure. At least two learnability problems have to be dealt with.

(26) a. *The perception of underlying structure*
         How can a phrase position be perceived as an antecedent or a gap due to a movement rule?
     b. *The distance problem*
         How are syntactic islands learned as phrases that do not allow the relation of an internal gap to an outside antecedent?

Both problems look more manageable in unification-based approaches that trade in the movement rule for a lexical feature matching between two sister-constituents. Neeleman and Van de Koot (2002) derive such an approach from Minimality principles. The left-hand sister X in (27) below is grammatically marked <F$_a$, F$_b$>. These features <F$_a$, F$_b$> need a licensing context. For example, *wh*-phrases in the SpecCP position are case-marked and preposition-marked as if they occupy an argument position. The right-hand sister Y in (27) should contain a grammatically definable gap based on the same licensing features <–/gap F$_a$, F$_b$>.

(27)   <+C> movement structure

    X-phrase<F$_a$, F$_b$>   ←→   Y-phrase

                                        </gap, F$_a$, F$_b$>
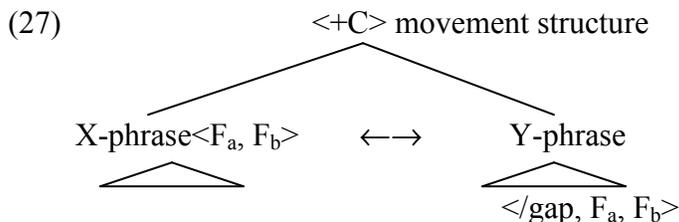
Along the same lines, see (10) section 3, the Dutch finite verb in the C^o position in (10) carries properties for licensing complements as if it were in the predicate final head position on the right.

   Obviously, the learnability of the antecedent~gap relation is on a promising track when the learner already commands a grammar that:

(28) a. spots the markings F_a, F_b orphaned in the C-projection on the left and spots their absence in the complement of C on the right.

b. projects the licensing markings <F_a, F_b,> according to existing head-complement conventions.

Fortunately, it is as required in (28)a. The grammatical licensing markings <F_a, F_b> that define the antecedent on the left in (27), have been acquired earlier on the right-hand side in non-gapped structures marked for preposition, case, and theta role. The acquisition of argument licensing in the Dutch [complement+V] structures precedes the acquisition of the V-second and the *wh*-movement distributions as we have seen in section 3 for V-second. The acquisition order is an empirical point and it fits (28)a. The procedure to project grammatical features, required in (28)b, was acquired when heads were subcategorized for the grammatical categories of their complements. This fits point (28)b. The subcategorization of heads works for complements, but not for subjects and adjuncts. The latter, subjects and adjuncts, are in attached non-subcategorizing, non-complement positions. This seems to be the heart of the matter. Subcategorization is learned by assigning to lexical heads frames that specify the order and the categorial properties of their complements. Non-subcategorizing positions are not represented in such frames for lexical heads. If licensing features are necessary, as in (27) above, it follows that gaps in non-complements will not be related to an antecedent structure. A positive definition of the antecedent-gap relation was already given in Kayne (1981) as the government projection path. Note that arrangements as in Kayne (1981), or more recently Neeleman and van de Koot (2002), evade the notion 'island'. If Kayne's projection path, or some equivalent of it, is learnable, the learnability of islands need no longer be considered.

The learnability of island constraints has been a topic of debate (Chomsky 1975, 1980: 319, Crain and Nakayama 1987, *Linguistic Review* 2002). In (20), the copula from the main clause is fronted. Copula movement out of a subject relative like (20) is not possible (subject island and 'complex NP').

(29) *Is$_i$ any ape that is brainy t$_i$ talkative?*
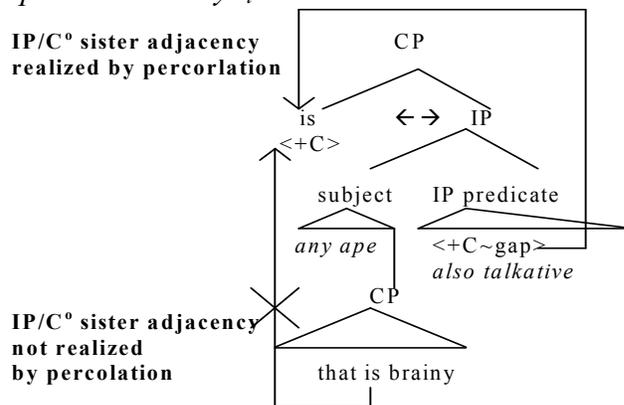
The point of discussion has been how children learn that the movement is structure dependent, where the raw input consists of strings only. Crain and Nakayama (1987), as well as Legate and Yang (*Linguistic Review* 2002), Fodor and Crowther (*Linguistic Review* 2002) relate this acquisitional fact to innate principles, whereas Pullum and Scholz (*Linguistic Review* 2002) prefer to see a hint in input percentages as sufficient. In the present view, island effects need

neither follow from innate principles, nor from input frequency. It is rather a consequence of a feature projection system that has been acquired earlier. The relative clause [*that is brainy*] does not subcategorize for the noun '*ape*'. It is not a complement and must be attached to the DP phrase [*any ape*]. By consequence, it will be an island by virtue of not being head-governed. In the same vein, the subject phrase does not subcategorize for the copula. It must be attached, in English at least, to the copula phrase. By consequence, it will not be head-governed and it will automatically be an island. A gap within a subject phrase cannot and will not be noted in the projection line that the subject phrase is attached to. If it has already been established in earlier acquisition steps that the projection-line subcategorizes for the grammatical properties of its complements and that the subject is defined by an attachment to the marked predicate, the rest follows. Suppose some overzealous and slightly unethical psycho-linguist were to prescribe a daily menu of nursery rhymes for a whole Kindergarten class, and suppose the rhymes were maliciously construed with gaps in the subject phrase. Then we may now predict in advance that these gaps would not be acquired. They can only be acquired when the pupils first unlearn the idea that subjects are in a non-subcategorizing attached position. Unlearning that property comes close to abandoning the (English) notion of subject**.**

In short, it is a misconstrual of the acquisition problem to worry about the learnability of islands or the learnability of long *wh*-movement. These phenomena are not learned at all. They rather follow from the combined effect of more elementary properties. Just like other parametric consequences, they are a surprise effect. Such effects are a characteristic of combinatory systems, where early elementary decisions have complex and unforeseen consequences.

6. Conclusion

Parameters**,** not unlike Jakobson's (1942) feature-oppositions, should clarify how the oppositions and implications in comparative grammar can be analyzed as resulting from choices in language acquisition. As pointed out by Dresher (1999) there are, however, clear difficulties faced by parametric approaches to language acquisition. UG parameters, as linguists generally conceive of them, are far removed from the primary data in language acquisition. An acquisition procedure based on parameter setting must be able to analyze the primary data in quite an abstract way (subject, object, predicate, auxiliary, complementizer, X-bar principles, islands, etc.). This seems impossible without already knowing the very grammar that is still to be acquired. That is Dresher's epistemological problem. At the same time, the acquisition procedure must be attentive to the fact that the raw data follow from an interaction of several parameters, Dresher's credit problem. There have been several attempts to bridge the gap. For the epistemological problem, Dresher (1999) proposed that each parameter was extended with a cue − a property more directly recognizable in the primary data (cf. also Lightfoot 1999). The cue would obligatorily trigger parameter setting and reduce the search space. As far as the credit problem is concerned, ordering of parameter setting is a key property. Ordered parameter setting is also postulated in Gibson and Wexler (1994) and it likewise appears spontaneously in the Variational Model of Yang (2002). What is significant about the idea of sequential ordering of parameter setting in the theoretical models of Gibson and Wexler (1994) and Yang (2002) is that language acquisition does indeed show a linear order of steps.

The present paper directs the attention to the actual order in parameter setting. It proposes that the order of acquisition steps in child language can be predicted from reductions of the adult

input language. The adult input continues to be recognized as a confused and poor stimulus, but reduction according to the acquisitional principle (6) yields a series of intakes that are focused on single values $<F_i>$. As such principle (6) is highly effective. A prime example has been the paradoxical acquisition of the underlying word order and the V-second rule in Dutch.

In general, when reduced structures can be extended to a real input structure by the addition of one functional category $<F_i>$, they start to function as an *evidence frame* for that category $<F_i>$. Evidence frames remind one of the cues in Dresher's view on parameter setting. Yet, the evidence frame is derived from the reduced input. It is not an *a priori* given. The evidence frame is close to Tomasello's (2003) 'linguistic construction', but also different from it. The evidence frame is aimed at building abstract combinatorial categories in the lexicon. The reduction method (6) is universal and not language-specific, but neither the evidence frame itself, nor the functional category it identifies are given *a priori*. The reduction procedure directs the focus of the acquisition procedure to a single functional category. It is a naturally occurring 'structure drill'. The fact that many languages make use of the same parameters need not be derived from a language-specific genetic endowment. Rather, parameters represent parallel solutions that appear in other cultural constructs as well. Elementary constructions, typological distinctions and steps in language acquisition remain closely related in the following three points.

1. The present approach to parameter setting obviates both the epistemological problem and the credit problem highlighted by Dresher (1999). The reduced structures focus on a specific acquisitional step $<F_i>$, a 'parameter setting', solving the epistemological problem. Setting the parameter is a causal effect and irreversible. The interaction of parameter settings will lead to superset effects, but not to credit problems. The unintended superset effects are rather the merit of the whole approach. The superset effects must be either correct or easily correctable by an additional micro-parameter. Otherwise the system will not survive.

2. The longitudinal analysis of child language may show how macro- and micro-parameters appear as a hierarchy of acquisition steps due to successively weaker input reductions. The Greenberg-type implications for universals follow from that hierarchy. The earlier steps (macro-parameters) set the stage for the later ones (micro-parameters).

3. There may be a type of combinatorial (generative) system that allows natural decoding and that enables a learner to acquire the system from scratch. The question is what such systems actually entail and whether the core grammars of human language are such systems. In the best of all possible worlds, minimal properties of grammar like 'locality' and 'inclusiveness' are stable, because they guarantee the survival of the system by keeping it learnable due to parameter setting.

The usual contention that parameters are somehow *a priori* options that require specific postulations about the brain has here been given up. A grammar is rather seen as a highly learnable cultural invention, not unlike number systems. This entails a change of focus. The number of (micro)-parameters is the major point of interest; their ranking in the acquisition hierarchy becomes the main issue.

References

Avram, L. and Coene, M. 2003. Why is it difficult to reach Agreement?. *Romance Linguistics. Theory and Acquisition*, A.T. Pérez-Leroux and Y. Roberge (eds), 247-262. Amsterdam: Benjamins.

Avram, L. and Coene, M. 2004. What early clitics can tell us about early subjects. In *Proceedings of GALA 2003*, Van Kampen, J. and Baauw, S. (eds), 93-102. Utrecht: LOT Series.

Avram, L. and Coene, M. This volume Can children tell us anything about parameter clustering?

Baker, M.C. 1988. *Incorporation: A Theory of Grammatical Function Changing.* Chicago: University of Chicago Press.

Baker, M.C. 1996. *The Polysynthesis Parameter.* New York: Oxford University Press.

Baker, M.C. 2003. *Lexical Categories. Verbs, Nouns and Adjectives.* Cambridge: Cambridge University Press.

Baker, M.C. This volume. The macroparameter in a microparametric world.

Benveniste, E. 1966. *Problèmes de Linguistique Générale*. Paris: Gallimard.

Berwick, R. 1985. *The Acquisition of Syntactic Knowledge*. Cambridge, Mass: MIT Press.

Berwick, R. and Weinberg, A. 1984. *The Grammatical Basis of Linguistic Performance.* Cambridge, Mass: MIT Press.

Briscoe, E.J. 2000. Grammatical acquisition: inductive bias and co-evolution of language and the Language Acquisition Device. *Language* 76(2): 245-296.

Brown, R. 1973. *A First Language: the Early Stages.* Cambridge, Mass: Harvard University Press.

Chomsky, N. 1975. *Reflections on Language.* New York: Pantheon.

Chomsky, N. 1980. *Rules and Representations.* Oxford: Blackwell.

Chomsky, N. 1981. *Lectures on Government and Binding.* Dordrecht: Foris.

Chomsky, N. 1987. *Language and Problems of Knowledge. The Managua Lectures*. Cambridge, Mass: MIT Press.

Chomsky, N. 2005. Three factors in language design. *Linguistic Inquiry* 36 (1): 1-22.

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

Crain, S. and Nakayama, M. 1987. Structure dependence in children's language. *Language* 62: 522-543.

Dresher, B. E. 1999. Charting the learning path: cues to parameter setting. *Linguistic Inquiry* 30: 27-67.

Dresher, B. E. and Kaye, J.D. 1990. A computational learning model for metrical phonology. *Cognition* 34:137–195.

Evers, A. and Kampen, J. van. 2001. E-language, I-language and the order of parameter setting. *UiL OTS Working Papers* 00105-S-S. Utrecht: UiL. **http://www-uilots.let.uu.nl/research/Publications/Publications.htm**

Ferdinand, A. 1996. The Development of Functional Categories. The Acquisition of the Subject in French. Ph.D. Dissertation, University of Leiden.

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

Fodor, J.D. 2001. Setting syntactic parameters. In *Handbook of Contemporary Syntactic Theory*, Baltin, M. and Collins,C. (eds.), 730-767. Oxford: Blackwell.

Fodor, J.D. and Crowther, C. 2002. Understanding stimulus poverty argument. *LinguisticReview* 19(1/2): 105-145.

Gibson, E. and Wexler, K. 1994. Triggers. *Linguistic Inquiry* 25(3): 407-454.

Greenberg, J.H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In *Universals of Language*, J.H. Greenberg (ed.), 73-113, Cambridge, Mass: MIT Press.

Hale, M. and Reiss, C. 2003. The subset principle in phonological acquisition: why the *tabula* can't be *rasa*. *Journal of Linguistics* 39: 219-244.

Jakobson, R. 1942. *Kindersprache, Aphasie und Allgemeine Lautgesätze*. Upsala: Uppsala Universitets Årsskrift.

Kampen, J. van. 1997. *First Steps in Wh-movement*. Delft: Eburon.

Kampen, J. van. 2004. Learnability order in the French pronominal system. In *Selected Papers from Going Romance* 2002, Bok-Bennema, R., Hollebrandse, B.,. Kampers-Manhe, B. and Sleeman, P. (eds.), 163-183. Amsterdam: Benjamins.

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

Kampen, J. van. 2006a. The acquisition of the standard EPP in Dutch and French. In *Studies on Agreement*, Costa, J. and Figueiredo Silva, M.C. (eds.), 99-121. Amsterdam: Benjamins.

Kampen, J. van . 2006b. Early operators and late topic-drop/*pro*-drop. In *The Acquisition of Syntax in Romance Languages*, Torrens, V. and Escobar, L., 203-223. Amsterdam: Benjamins.

Kayne, R. 1981. ECP extensions. *Linguistic Inquiry* 12: 93-133.

Kohl, K. 1999. An Analysis of Finite Parameter Learning in Linguistic Space. MA thesis, MIT.

Lebeaux, D. 1988. Language Acquisition and the Form of Grammar. Ph.D. Dissertation, University of Massachusetts.

Legate, J.A. and Yang, C. 2002. Emperical re-assessment of poverty arguments. *LinguisticReview* 19(1/2): 151-162.

Lightfoot, D. 1991. *How to Set Parameters*. Cambridge, Mass: MIT Press.

Lightfoot, D. 1999. *The development of language: acquisition, change and evolution*. Oxford: Blackwell.

Locke, J. L. 1997. A theory of neurolinguistic development. *Brain and Language* 58(2): 265-326.

Lyons, J. 1979. Deixis and anaphora. In Myers, T. (ed.), *The Development of Conversation and Discourse*, 88-103. Edinburgh: Edinburgh University Press.

MacWhinney, B. 2006. *The CHILDES Project: Tools for Analyzing Talk*, 3rd edition. Mahwah NJ: Lawrence Erlbaum.

Mills, D.L., Prat, C.., Zangl, R., Stager, C.L., Neville, H.J. and Werker, J.F. 2004. Language experience and the organization of brain activity to phonetically similar words: ERP evidence from 14- and 20-month-olds. *Journal of Cognitive Neuroscience* 16: 1452-1464.

Neeleman, A. and Koot, H. van de. 2002. The configurational matrix. *Linguistic Inquiry* 33(4), 529-574.

Őztürk, B. This volume. Non-configurationality: free word order and argument drop in Turkish.

Panagiotidis, P. This volume. Diachronic stability and feature interpretability.

Postal, P. 1966. On so-called pronouns in English. In *Monograph Series on Language and Linguistics* 19, Dinneen, F. (ed.), 177-206. Georgetown: Georgetown University Press.

Pullum, G. and Scholz, B. 2002. Empirical assessment of stimulus poverty arguments. *Linguistic Review* 19(1/2): 9-50.

Stassen, L. 1997. *Intransitive Predication*. New York: Oxford University Press.

Tomasello, M. 2003. *Constructing a Language. A Usage-Based Theory of Language Acquisition.*

Cambridge, Mass: Harvard University Press.

Truckenbrodt, H. 2006. On the semantic motivation of syntactic verb movement to C in German. *Theoretical Linguistics* 32(3): 257-306.

Watanabe, A, 1994. The role of triggers in the extended split Infl hypothesis: Unlearnable parameter settings. *Studia Linguistica* 48: 156-178.

Weerman, F. 1989. *The V2 Conspiracy*. Dordrecht: ICG.

Wexler, K. 1999. Maturation and growth of grammar. In *Handbook of Child Language Acquisition*, Ritchie, W.C. and Bhatia, T.K. (eds.), 55-105. San Diego: Academic Press.

Yang, C. 2002. *Knowledge and Learning in Natural Language*. Oxford: Oxford University Press.

Zwart, C. J.-W. 1993. Dutch Syntax. A Minimalist Approach. Ph.D. Dissertation, University of Groningen.