

## Macquarie University ResearchOnline

---

**This is an article from the following conference:**

Thornton, Rosalind and Tesan, Graciela (2006) Parameter setting and statistical learning *Annual Meeting of the Australian Linguistic Society*, (7 - 9 July 2006 : Brisbane)

**Access to the published version:**

[http://espace.library.uq.edu.au/eserv/UQ:13242/Thornton-RJ-ALS2006\\_1\\_.pdf](http://espace.library.uq.edu.au/eserv/UQ:13242/Thornton-RJ-ALS2006_1_.pdf)

# Parameter Setting and Statistical Learning

## Authors

Rosalind Thornton  
Linguistics Department and Macquarie Centre for Cognitive Science  
Macquarie University  
North Ryde, NSW 2109

Rosalind.Thornton@ling.mq.edu.au

Graciela Tesan  
Macquarie Centre for Cognitive Science  
Macquarie University  
North Ryde, NSW 2109

gtesan@maccs.mq.edu.au

## *Abstract*

Three main models of parameter setting have been proposed: the Variational model proposed by Yang (2002; 2004), the Structured Acquisition model endorsed by Baker (2001; 2005), and the Very Early Parameter Setting (VEPS) model advanced by Wexler (1998). The VEPS model contends that parameters are set early. The Variational model supposes that children employ statistical learning mechanisms to decide among competing parameter values, so this model anticipates delays in parameter setting when critical input is sparse, and gradual setting of parameters. On the Structured Acquisition model, delays occur because parameters form a hierarchy, with higher-level parameters set before lower-level parameters. Assuming that children freely choose the initial value, children sometimes will mis-set parameters. However, when that happens, the input is expected to trigger a precipitous rise in one parameter value and a corresponding decline in the other value. We will point to the kind of child language data that is needed in order to adjudicate among these competing models.

## *Key Words*

child language development, Universal Grammar, parameter setting, statistical learning, triggering model, negation

## 1. Introduction

The last thirty years have seen remarkable advances in linguistic theory, and corresponding advances in our understanding of how children acquire language. Advances on both fronts have resulted in large part, in our view, because of a shift from the 1980s rule-based theories of grammar to the current Principles and Parameters approach (e.g., Chomsky 1981, 1995). The Principles and Parameters approach enabled researchers in language development to make many new and far-reaching predictions about the course of language acquisition. According to this framework, children were no longer expected to accrue individual rules for the local language being spoken around them, as in the earlier versions of linguistic theory. The initial state of the language faculty continued to embody universal principles that establish boundary conditions on children's linguistic hypotheses, and children were not expected to deviate from these principles in the course of language development (see, e.g., Atkinson 1992, Crain 1991, Guasti 2002).

But in addition to linguistic universals, certain aspects of language variation took on a new look in the Principles and Parameters approach. Many differences across languages were taken to be encoded in the language faculty as innately specified parameters, where the parameters establish (typically binary) choices among linguistic properties of particular natural languages. The introduction of an innately specified system of parameters in Universal Grammar was motivated by the desire to ensure that language learning was less burdensome for the learner than it would be otherwise (Chomsky 2002). The new look learner is seen as navigating through an innately specified parameter space that is made available by Universal Grammar; learning is largely replaced by (or reduced to) parameter setting (cf. Clahsen 1990). This assisted the theory of Universal Grammar in meeting its overarching goal of 'explanatory adequacy,' i.e., to explain children's rapid mastery of the grammar of any natural language (Chomsky 1965; 1986).

In the theoretical literature, parameter setting was originally conceived to be executed by a linguistic mechanism that resided in the language acquisition device. Each time the mechanism was engaged, it had immediate and far-reaching consequences throughout a learner's grammar. A metaphor for this mechanism was that of a switch – where the learner simply flicked a switch to one setting or the other in response to some 'triggering' experience that was readily observable in the primary linguistic data. The switch metaphor suggested that, at some circumscribed period during the course of development, the setting of a parameter would be decisively triggered, with one value being adopted rather than the other (Fodor 1998, Gibson and Wexler 1994, Hyams 1986, Roeper 2000). To continue with the metaphor of setting a switch, if the switch was set one way, then the child's grammar took one form, and if the switch was set the other way, the child's grammar took another form. Parameter setting was seen to set in motion radical changes in children's grammars, for example from a grammar with null subjects to one with overt subjects, or from a grammar without Wh-movement to one with Wh-movement, and so on. It was suggested, moreover, that setting a single parameter might induce the introduction of a cluster of properties into children's emerging grammars. The paradigm case exemplifying this was the null subject parameter studied by Hyams (1986; 1987; 1989) (cf. Rizzi 1992). Other work, such as Snyder's (2001) work on acquisition of complex predicates and word-formation has followed in this tradition.

Although parameters were, admittedly, fixed on the basis of input, it was generally assumed that the ambient input sufficed for ‘early parameter setting’ (see, e.g., Borer and Wexler 1987, Wexler 1998). Nothing in the theory itself prevented parameters from being set early, so if it turned out that they were not set early, then something outside the theory must be responsible for any delay in parameter setting. Therefore, it was the ‘null hypothesis’ that parameters were set early. Finally, researchers working within the parameter-setting framework assumed that children were initially free to pick one or the other setting, unless a subset problem would arise if one particular setting were adopted, rather than the other. The possibility of ‘default’ settings was available, in principle, but there was no reason to suppose *a priori* that there were default settings. Another view, advanced by Lebeaux (1988), was that children begin with both parameter values being operative, with one of them taking priority in response to input from the local language. This conception of parameter setting has been resurrected in the Variational model (cf. Yang 2002; see below).

The observation that children could set parameters to either value immediately raised the expectation that children could initially ‘misset’ parameters. That is, the learner could initially adopt a value that was inconsistent with the local language. The mismatch would presumably be easily detected, and soon set straight. Still, it could take a child some amount of time to reset a parameter, and during the period of parameter resetting, the child would be speaking a fragment of a ‘foreign’ language. Therefore, the investigation of children’s early productions promised, potentially, to offer empirical support for the parameter-setting approach. On other approaches, the learner was seen to be attempting to match the input, by accruing rules or constructions on the basis of positive examples.

The earliest empirical support for the principles and parameters approach was one such case of apparent parameter missetting. As reported in Hyams (1986), English-speaking children have been found to omit subject NPs early in the course of development. Hyams argued that such omissions were the result of children having adopted a parameter setting according to which subject NPs could be omitted, as in Italian, but not in English. Over the years, there have been a number of other reports of misset parameters, where children were found to be projecting parameter values, rather than being directly guided by the input in language development. Empirical data along this line can be found in Armon-Lotem, Crain and Varlokosta (2006), Becker (2000), Hyams (1987; 1989). Of course, children eventually converge on a grammar that is equivalent to that of adult speakers of the local language, so parameter resetting must be responsive to the input.

Although nothing in the theory of Universal Grammar specifies precisely how parameter setting might unfold in real time, the ‘null hypothesis’ was that parameter setting (and even parameter resetting) would take place early in the course of language development, triggering immediate and far-reaching changes from one kind of grammar to another. However, the empirical data have not unequivocally supported the null hypothesis. There are several ways to explain the lack of ‘fit’ between theory and data. One way for triggering models to explain the recalcitrant data is to invoke performance factors to account for children’s unexpected behavior. Another response is to invoke maturation for late-developing grammatical properties (Borer and Wexler 1987; 1992, Wexler 1994; 1998). Another kind of response to the recalcitrant data is to bring statistical learning mechanisms into play, in addition to the principles and parameters of Universal Grammar. We will scrutinize this last approach, focusing on one important model of parameter setting augmented by statistical learning, advanced first in Yang (2002).

## 2. Three models of parameter setting

Currently, there are three main parameter-setting models. One is the Very Early Parameter Setting model (Wexler 1994; 1998). The Very Early Parameter Setting model (VEPS) (Wexler 1994; 1998) postulates: (a) parameters are independent (*ordering*), (b) children initially begin with a single parameter value, but may adopt either value, unless this would lead to subset problems (*starting point, initial value*), (c) grammar formation is characterized by abrupt changes in grammars (*trajectory*), (d) differences in the primary linguistic data have little impact on the observed course of parameter setting (*requisite input*), so no special (e.g., statistical) learning mechanisms are needed to assist in parameter setting, and (e) since parameter setting is completed very early, little individual variation will be observed (*conformity*). The VEPS model has little room to maneuver in response to apparent delays in parameter setting. Maturation is one possibility. Late emergence could also be interpreted as evidence that some phenomenon does not properly count as a parameter. This is the approach taken by the VEPS model for the so-called optional infinitive (OI) stage of language development.

The second model is the Structured Acquisition model (Baker 2001; 2005), based largely on the “implicational universals” proposed in Baker (2001; 2005). On this model parameters are ordered in a hierarchy, with large-scale parameters at the top of the hierarchy, including the polysynthesis parameter and the head directionality parameter. These parameters are presumably set early and have significant impact on the overall form of the language that is acquired. Smaller-scale parameters reside lower in the hierarchy, and they are not set early because they must await the decisions about parameters that are more dominant in the hierarchy. On the Structured Acquisition model: (a) parameters are interlocked (*ordering*), (b) children initially begin with a single parameter value, though either value may be selected (*starting point, initial value*), (c) grammar formation is characterized by abrupt changes in grammars (*trajectory*), (d) differences in the primary linguistic data have little impact on the observed course of parameter setting (*requisite input*), so no special (e.g., statistical) learning mechanisms are invoked in parameter setting, and (e) setting some parameters can only occur once others have been set, and since children may adopt different starting values, different children may set the same parameters at different times (*conformity*), giving rise to individual variation. The first two models are similar in character. Both of these models assume that parameter setting is straightforward for learners, and does not require specialized statistical learning mechanisms. However, the Structured Acquisition model introduces an ingredient beyond that of the Very Early Parameter Setting model, namely parameter ordering. Parameter ordering leads to far-reaching empirical predictions that distinguish the Structured Acquisition model from of the Very Early Parameter Setting model. The third model, the Variational model, introduces statistical learning into parameter setting. The assumption that statistical mechanisms play a critical role in development has taken a strong hold in the field, so it is instructive to explore the proposal that statistical mechanisms are engaged by learners in parameter setting.

The third model is the Variational model (Yang 2002; 2004, Legate and Yang 2005). On this model: (a) (most) parameters are independent of each other (*ordering*), (b) children initially begin with competition among parameter values (*starting point*) (c) if children do not initially favor the value that is associated with the target language, then grammar formation is characterized by gradual changes in development (*trajectory*), (d) differences in the primary linguistic data determine the observed course of parameter setting (*requisite input*), because stochastic learning mechanisms determine the course of parameter setting, and (e) since input

is assumed to be uniform across children, individual differences are not anticipated for children who begin with the same level of competition among parameter values (*conformity*). In contrast to VEPS, the Variational model sees the optional infinitive stage of development as falling within its purview. In fact, optionality in children's behavior is probably the principle motivation for the assumption that parameter values initially compete against each other (*starting point*).

What course of acquisition is expected on a statistical learning model of parameter setting, such as the Variational model? This model supposes that children initially attempt to parse the linguistic input using two 'grammars', one with each value of the parameter operative in it. If one of these competing grammars parses the input successfully, that grammar is 'reinforced', increasing the probability that it will be used in the future. Assuming that the grammar with the alternative parameter value is unable to parse the same input, then that grammar is 'devalued', and its probability of being selected in the future is correspondingly reduced. Gradually, probability weights are adjusted until the grammar with the non-target parameter value is no longer a contender, and becomes obsolete. On this model, then, quantitative data from input frequencies can be used to estimate whether a parameter setting will be consolidated early or late. The precise course of development depends on the initial weights of the competing values of a given parameter.

Initially, some children will strongly favor the correct setting for the target language, and will rapidly converge on the adult grammar rapidly. Other children will initially favor the incorrect setting. These children will only gradually converge on the correct setting, based on the properties of the input. But such children should be the exception, and not the rule. Most children will, presumably, start with the competing values hovering around 50/50, though to a lesser extent, other children will adopt other 'weights' (60/40, 70/30, and so on), for the competing values. These differential probabilities for the competing values are not expected on the other models of parameter setting.

In the next section we turn to an evaluation of parameter setting models, using data from negation. From this point, we will eliminate the Very Early Parameter Setting model (VEPS) from our consideration. As Yang (2002; 2004) points out, there are child data that reveal children exercising a non-target parametric option quite late in language development. For example, some children use a *wh*-copying structure (as in questions like *What do you think what Cookie Monster eats?*) until they are three or four years old (Crain and Thornton 1998). If this is considered to be a parameter of natural language, then it is a clear counterexample to the proposal that parameters are set perceptually, before the onset of speech. From this point then, we will restrict our evaluation to Yang's Variational model and the Structured Acquisition model.

### **3. An illustration: A parameter for negation**

From this point forward, our goal is to explain certain non-adult properties of children's speech. By way of illustration we will focus on a parameter that concerns the placement of negation across languages. It is well known that cross-linguistically, the lexical item for negation can be positioned in the **head** of the Negation phrase (NEGP), or in the **specifier** position. Here, we frame this as a parametric choice between the two structural positions in which negation can reside. The choice between head or specifier interacts with the verb movement properties of the language. Roughly, negative expressions that are heads can move along *with* the verb to a higher structural position in verb raising languages. But a

verb cannot raise *over* a negative head in such languages. This contrasts with the status of negative expressions that are specifiers in a language; such expressions do not block the movement of a verb, which is free to raise or lower over expressions in specifier position. We will illustrate these parametric options with cross-linguistic examples.

Spanish is a verb raising language, and the negative item *no* is a head. Because the form of negation is a weak form, a clitic, it raises along with the verb to the INFL position, as shown in (1). In the example, the origins of the elements that have been moved are indicated with strikethrough font.

- (1) [IP Juan ~~no~~ habla [<sub>NEG</sub> ~~no~~ [<sub>VP</sub> ~~habla~~ italiano]]]  
 Juan NEG speak-3sg Italian  
 “Juan doesn't speak Italian.”

In French, as in Spanish, the weak clitic form of negation (*ne*) is a head (although this form is often omitted in colloquial language). It also raises with the verb to INFL. But there is a second form of negation in French, namely the lexical item *pas*. The expression *pas* is obligatory in negative sentences and is positioned in specifier position. The example in (2) illustrates a sample derivation in which the main verb raises to pick up the negative element *ne* in the head position, and passes over *pas* as it raises to INFL.

- (2) [IP Jean ~~ne~~-parle [<sub>NEG</sub> pas ~~ne~~ [<sub>VP</sub> ~~parle~~ grec ]]]  
 Jean NEG -speak not greek  
 “Jean doesn't speak Greek.”

Another language that positions negative expressions in the specifier position is Swedish. However, in Swedish, the verbal affix lowers over the negative item *inte* to merge with the main verb. This is most transparent in embedded clauses. The example in (3) (taken from Tesan 2005) illustrates the word order of Swedish in embedded sentences.

- (3) ...att Lena inte köpte en ny bok igår  
 ...that Lena not bought a new book yesterday  
 ...”that Lena didn't buy a new book yesterday.”  
 (adapted from Vikner 1995: 45, (28))

Turning to English, the negative expressions *not* and *n't* are usually categorized as heads (cf. Chomsky and Lasnik 1993). It is useful to compare these expressions with the negative adverb *never*, which resides in the specifier position of the negation projection and, therefore, functions much like *pas* in French and *inte* in Swedish. Since *never* is a specifier, the verbal affix can lower across the adverb *never* and attach to the verb. This is illustrated in (4) where the 3<sup>rd</sup> person agreement affix *s* lowers onto the main verb to yield ‘He never speaks French’.

- (4) [IP He ~~s~~ [<sub>NEG</sub> never [<sub>VP</sub> speak-s French ]]]

The verbal affix cannot lower across the head *not*, however. This prevents the derivation of utterances like *He not speaks French*. A rescue operation, ‘do-support’, is needed to salvage the derivation. Insertion of *do* provides a host for the stranded affix that is prevented from lowering to the verb. This yields acceptable utterances such as ‘He does not speak French’.

(5) [IP He doe -s [NEG not [VP speak French ]]]

The most common form of negation in English is *n't*. This contracted form of negation is a clitic that joins with a host auxiliary verb or modal (e.g., *doesn't*, *can't*, *haven't*, *isn't* etc.), as in (6) (cf. Zwicky 1983).

(6) [IP He doe -s -n't [NEG n't [VP speak French ]]]

With this survey of parametric options in mind, let us see how the negation parameter is set in the grammars of English-speaking children. The fact that *n't* is attached to modals and auxiliaries (*can't*, *shouldn't*, *haven't*, *isn't* etc.) informs English-speaking children that this form of negation is a head. However, the knowledge that *n't* is a head doesn't help children implement this value of the negation parameter in sentences with main verbs. Children must be exposed to the specific lexical item *doesn't* to see that *n't* must remain higher than the verb phrase, so the negation parameter may not be set by children until they discover the lexical item *doesn't*.<sup>1</sup>

Having established that *doesn't* constitutes unambiguous data for learners to set the negation parameter, we can use the frequency of occurrence of *doesn't* in the input to establish whether the parameter is expected to be acquired early or late. To obtain an estimate of the frequency of *doesn't* in the input, we conducted a search of the adult input to Adam and Eve in the CHILDES database (MacWhinney and Snow 2000). Of the 30,882 adult sentence utterances that were checked, only 296 (0.95%) contained *doesn't*. According to the statistical learning model proposed by Yang, parameters whose unambiguous input appear with a frequency of occurrence of 1.2% or less are expected to be set late in the course of acquisition. The expectation is, therefore, that the negation parameter will be set late.

To test this expectation, an elicited production experiment was conducted with four English-speaking children who attended our language acquisition laboratory starting at about age 2, and continued to visit the lab every two weeks for roughly a year, at which point the verbal morphology of all of the children was close to adult-like. The elicited production experiment encouraged children to use constructions that they would otherwise have avoided. This is particularly true of negative utterances, which are sparse in the CHILDES database. For example, Harris and Wexler (1996) searched the transcripts of 10 children who ranged in age from 1;6 to 4;1, and found 52 sentences with 3<sup>rd</sup> person subjects in structures that contained *no* or *not* and a main verb (cf. H&W, Table 5, p. 16). Our study evoked 204 comparable structures from the four two-year-old child subjects over a considerably shorter period of time (see Tesan 2005 for more details).

Procedures to evoke negative sentences included a range of games to see, for example, where various objects would fit. For example, a puppet might try to complete a puzzle, but would end up putting pieces in the wrong place (*"It goes here!"*). The child was encouraged

---

<sup>1</sup> Of course, the discovery that *n't* is a head still does not guarantee that the negative morpheme *not* is also a Head; it could be a Specifier. Therefore, children could use *doesn't* in the same way as adults do but, at the same time, they could analyze *not* as a specifier. Our empirical findings suggest that once children acquire *doesn't*, they cease to use *not*, at least for a time. For now, we will simply assume that *doesn't* is the critical data that children need, and leave the continuing status of *not* in children's grammars as an open question.



to correct the puppet (“*It not goes there!*”). Or, the child subject was assigned the task of performing ‘experiments’ with a group of various objects, to see if they float, or squeak, or would stick to a magnetic board, and so forth (“*This one squeaks. This one not squeaks*”). The inclusion of these elicitation procedures resulted in a robust set of data for each child.

Of the four children, one child, SL, initially selected the correct ‘adult’ value of the parameter, with negation residing in head position. Since SL began with the adult value, no parameter resetting was needed so this child’s pattern of behavior was essentially flat (although the lexical item *doesn’t* still had to be acquired). By contrast, the other three children began with the other value, with negation in specifier position. This means that these three children treated *not* like *never*, and produced utterances like *He not fits*, with an inflected main verb. The pattern of behavior across time was quite different for the three children who initially adopted the non-target value of parameter. These children exhibited an abrupt change in values, with different children initiating and completing their own precipitous change at different ages. There was no indication that the statistical distribution of structures or lexical items in the input was responsible for the trajectories of any of the children. As illustration, Figure 1 provides the data for CM, one of the three children who initially misset the negation parameter, and treated *not* as a specifier. CM initially used only this option in the session at 1;10;27, when she produced three instances of sentences like *He not fits*. About two weeks later, at 1;11;11, CM produced 50% sentences like *He not fits*, in which negation was treated as a Specifier, and 50% adult-like sentences, with do-support, in which *n’t* was behaving as a head. In fact, in the first half of the session at 1;11;11, CM produced six utterances like *He not fits* and, in the second half of the session, CM produced six adult-like utterances, with *doesn’t*. In other words, there was an abrupt change from one parameter value to the other -- within a single testing session in the laboratory. In short, grammatical change was initiated almost immediately by CM, and CM’s non-adult treatment of negation rapidly disappeared, and did not reappear after age 2;1.

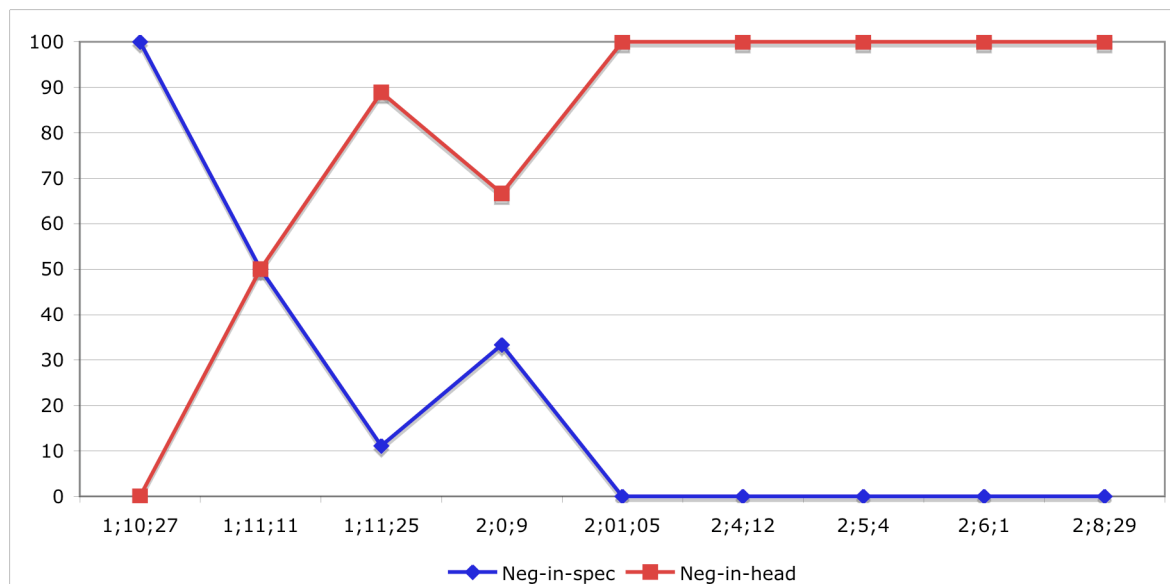


Figure 1: The trajectory for the two negation values in CM’s data

This brings us back to a consideration of the alternative models of parameter setting.

There are two main predictions of the Variational model, as we understand it. One is that the two values of a parameter should be in conflict early in the course of language development, at least for the majority of children. Second, to the extent that the incorrect value is favored over the correct value, the trajectory of development should be gradual. By contrast, the Structured Acquisition model anticipates that children will start with one value or the other, and that grammatical change, once it begins, will be rapid. The findings from the present study clearly favor the Structured Acquisition model on both counts. First, there was no evidence of children beginning with both values being weighed against each other. Second, children's grammatical change was precipitous, not gradual. Furthermore, the three children who misset the parameter did not all show similar trajectories as they moved toward the adult grammar.

#### 4. Conclusion

Beginning with Saffran, Aslin and Newport (1996), the last decade has witnessed a series of research studies showing that children are endowed with statistical learning mechanisms that are sufficiently powerful to assist them in word segmentation, and even in the detection of phrasal units (Saffran 2001; 2002). Yang (2002; 2004) has proposed that such learning mechanisms can be paired with Universal Grammar to assist language learners in setting parameters. Granting that learners employ a statistical learning mechanism for certain tasks, we investigated children's acquisition of a negation parameter, to see whether the learning path in child language development assumed the gradual curve associated with statistical learning over time or, instead, if the path of language development resembled the sharp edges associated with setting and resetting parameters. The empirical findings from a longitudinal study of four children's development of negation do not support the proposal that statistical learning is driving children's parameter setting. Our empirical findings show, instead that (a) children initiate grammatical change at some point in time, (b) change takes hold quickly, and (c) change is brought to closure quickly. We do not fully understand the mechanisms that set grammatical change in motion, but they are apparently responsive to the child's internal grammatical development, and do not always reflect children's linguistic experience.

#### References

- Armon-Lotem S, S Crain and S Varlokosta 2004 'Interface conditions in child language: Cross-linguistic studies on the nature of possession' **Language Acquisition** 12: 171-217.
- Atkinson M 1992 **Children's syntax: An introduction to principles and parameters theory**. Blackwell Cambridge, MA.
- Baker M 2001 **The atoms of language: The mind's hidden rules of grammar**. Basic Books New York.
- Baker M 2005 'Mapping the terrain of language acquisition' **Language learning and development** 1: 93-129.
- Becker M 2000 'The development of the copula in child English: The lightness of *be*' Doctoral dissertation, UCLA.
- Borer H and K Wexler 1987 'The maturation of syntax' In T Roper and E Williams (eds) **Parameter setting** Reidel Publishing Company Dordrecht.
- Borer H and K Wexler 1992 'Bi-unique relations and the maturation of grammatical principles' **Natural Language and Linguistic Theory** 10: 147-190.

- Chomsky N 1965 **Aspects of the theory of syntax** MIT Press Cambridge MA.
- Chomsky N 1981 **Lectures on government and binding** Foris Publications Dordrecht.
- Chomsky N 1986 **Knowledge of language: Its nature, origin and use** Praeger New York.
- Chomsky N 1995 **The minimalist program** MIT Press Cambridge MA.
- Chomsky N 2002 **On nature and language** Cambridge University Press Cambridge.
- Chomsky N and H Lasnik 1993 'The theory of principles and parameters' In J Jacobs, A von Stechow, W. Sternefeld and T. Vennemann (eds) **Syntax: An international handbook of contemporary research** De Gruyter Berlin.
- Clahsen H. 1990 'Constraints on parameter setting: A grammatical analysis of some acquisition stages in German child language' **Language Acquisition** 1: 361-391.
- Crain S 1991 'Language acquisition in the absence of experience' **Behavioral and Brain Sciences** 14: 597-650.
- Crain S and R Thornton 1998 **Investigations in universal grammar: A guide to experiments of the acquisition of syntax and semantics** MIT Press Cambridge MA.
- Fodor J 1998 'Unambiguous triggers' **Linguistic Inquiry** 29: 1-36.
- Gibson, E and K Wexler 1994 'Triggers' **Linguistic Inquiry** 25: 355-407.
- Guasti, MT 2002 **Language acquisition: The growth of grammar** MIT Press Cambridge MA.
- Harris T and K Wexler 1996 'The optional infinitive stage in child English: Evidence from negation' In H Clahsen (ed) **Generative perspectives on language acquisition** John Benjamins Amsterdam/Philadelphia.
- Hyams N 1986 **Language acquisition and the theory of parameters** D. Reidel Dordrecht.
- Hyams N 1987 'The theory of parameters and syntactic development' In T Roeper and E Williams (eds) **Parameter setting** D. Reidel Publishing Company Dordrecht.
- Hyams N 1989 'The null subject parameter in language acquisition' In O Jaeggli and K Safir (eds) **The null subject parameter** (pp 215-238) Kluwer Academic Publishers Dordrecht.
- Lebeaux D 1988 'Language acquisition and the form of the grammar' Doctoral dissertation, University of Massachusetts Amherst.
- Legate J and C Yang 2005 'Morphosyntactic learning and the development of tense' Ms University of Delaware and Yale University.
- MacWhinney B 2000 **The CHILDES project: Tools for analyzing talk** Third Edition Lawrence Erlbaum Associates Mahwah NJ.
- Rizzi L 1990 **Relativized minimality** MIT Press Cambridge MA.
- Roeper T 2000 'Universal bilingualism' **Bilingualism: Language and Cognition** 2: 169-186.
- Saffran J 2001 'The use of predictive dependencies in language learning' **Journal of Memory and Language** 44: 493-515.
- Saffran J 2002 'Constraints on statistical language learning' **Journal of Memory and Language** 47: 172-196.
- Saffran J R Aslin and E Newport 1996 'Statistical learning by 8-month old infants' **Science** 274: 1926-1928.
- Snyder W 2001 'On the nature of syntactic variation: Evidence from complex predicates and complex word-formation' **Language** 77: 324-342.
- Tesan, G 2005 'What do children have in their heads? Functional heads and parameter setting in child language' Doctoral dissertation, University of Maryland.
- Vikner S 1995 **Verb movement and expletive subjects in the Germanic languages** Oxford University Press New York.
- Wexler K 1994 'Optional infinitives, head movement, and the economy of derivation in child language' In D Lightfoot and N Hornstein (eds) **Verb movement** Cambridge University Press Cambridge (pp 288-315).

- Wexler K 1998 'Very early parameter setting and the unique checking constraint: A new explanation of the optional infinitive stage' **Lingua** 106: 23-79.
- Yang, C 2002 **Knowledge and learning in natural language** Oxford University Press New York.
- Yang, C 2004 'Universal grammar, statistics, or both?' **Trends in Cognitive Science** 8: 451-456.
- Zwicky A and G Pullum 1983 'Cliticization vs. inflection: English *n't*' **Language** 59: 502-513.