Minimalism and Language Acquisition

Charles Yang & Tom Roeper*

University of Pennsylvania & UMass Amherst

Perhaps more clearly than any other field, the study of child language acquisition highlights the continuity from the Principles & Parameters framework (Chomsky 1981) to the Minimalist Program (Chomsky 1995). As is the case for all meaningful theoretical developments, under Minimalism new challenges emerge, puzzles are cast under different lights, while important insights from previous work can still be retained; this chapter provides an overview of these issues. The first part builds on the continuity from P&P to Minimalism, with focus on the role of parameters in the theory of language acquisition and the mechanisms of learning. The second part turns to the Minimalist innovations, specifically how the new formulations of the syntactic system bring new tools to the explanation of child language.

Part I. Formal Issues in Minimalism and Language Acquisition

The P&P framework, for the first time, gives a plausible solution to the logical problem of language acquisition: How the child acquires a language so rapidly and accurately

* Thanks to all of our students and colleagues. The authors' names are ordered as to reflect the organization of the article.

under limited linguistic experience? The principles, which are considered universal, are not learned, and can be expected to be operative in (early) child language; this opens up a wealth of topics for empirical research which continues in the Minimalist era. The parameter values, which vary crosslinguistically, must be learned on the basis of specific linguistic evidence. Thus, the commonalities and differences in children's acquisition of specific languages receive a unified treatment. Moreover, if the number of parameters is finite, then there is only a finite – albeit large, so it appears – number of grammars that forms the child's learning space, which at least formally sidesteps the well-known problem of inductive indeterminacy in an infinite hypothesis space associated with phrase structure rules (Gold, 1967; Chomsky, 1981).

Has Minimalism altered the fundamental problem of language acquisition? We feel that the answer is both No and Yes. No, because that Minimalism has not supplemented the basic architecture of P&P for the task for language acquisition (§1), and yes, in the sense that the Minimalist approach to the language faculty in a broad context of cognition and evolution has led to new conceptions of learning, which may provide a more complete explanation of child language acquisition (§2). These issues are closely related; for instance, the empirical evidence for or against parameters cannot be separated from the mechanisms by which parameters values are determined. For expository purposes, however, we shall discuss them in turn.

1 Parameters & Child Language

There are two senses in which the term "parameter" can be understood, and it might be useful to draw their distinctions more clearly. In the conceptual sense, parameters simply

denote the finite range of biologically possible linguistic forms, a claim about natural language upheld by most theories of grammar even though the term "parameter" is typically associated with the GB/Minimalism framework. Parameters, then, can be viewed as a type of anchor points for dividing up the linguistic space: the complex interactions among them would provide coverage for a vast array of linguistic data—more "facts" captured than the number of parameters, so to speak—such that the determination of the parameter values would amount to a simplification of the learning task. This conceptual notion of parameters goes well with the perspective of machine learning and statistical inference, where plausible learnability can only be achieved by constraining the hypothesis space within some finite dimensions (Valiant 1984, Vapnik 1995, see Nowak et al. 2002 for review). These mathematical results hold under the usual assumptions of language acquisition (e.g., the learner only receives positive data though in some cases even negative data does not make learning more tractable) but are not dependent on the nature of the specific learning algorithm or other cognitive capacities that avail to the learner. So formally, an approach to language variation and acquisition by the use of parameters remains the best, and only, game in town.

Once instantiated as specific theories about human language, parameters can be understood in the sense of empirical statements, which of course can be verified, confirmed, or rejected. The failure of certain proposals of parameters does not mean that the whole theory of parameters ought to be rejected out of hand: we might not have figured out the correct ways of dividing up the linguistic space. In fact, many specific

formulations of parameters in the theoretical literature have received support from language acquisition research, to which we turn presently.

The evidence for parameters comes in two lines. The first has been running throughout the history of P&P framework. Ever since Hyams's pioneering work (1986), parameters have been used as a tool to explain non-target grammatical patterns (see Crain & Pietroski 2002, Rizzi 2004, Roeper 2000 for recent efforts). Take, for instance, the well known case of null subjects in child English. For the first three years of life, English learning children do not use subjects consistently (Valian 1991), and objects are occasionally omitted as well. Earlier research (e. g, Hyams 1986, 1991, Hyams & Wexler 1993) has attributed these omitted arguments to an Italian type *pro-drop* grammar (Rizzi 1986) or a Chinese type *topic-drop* parameter (Huang 1986), yet the usage frequencies of subjects and objects from the studies of (both child and adult) Italian and Chinese are significantly different from those in child English (Valian 1991, Wang et al. 1992). More recently, the null subject phenomenon has been interpreted as the presence of the *topic-drop* option gradually being eliminated (Yang 2002). One of the key observations here is the striking similarity between child English and adult Chinese. For instance, the availability of subject drop in Chinese is subject to an asymmetry in topicalization.

(1) a. *Mingtian*, [___ guji [*t* hui xiayu]]. (___=John)
Tomorrow, [___believe [t will rain]]

"It is tomorrow that John believe will rain."

b. *Bill, [___renwei [*t* shi jiandie]]

(___=John) Bill, [___believe [*t* is spy]]

"It is Bill that John believes is a spy".

The main observation is that null subject, which is identified by linking to the discourse topic, is not possible when the new topic is an argument (1b) but possible when it is an adjunct (1a); see Friedmann, Belletti, and Rizzi (2009) for related theoretical considerations. Such distributional patterns are virtually perfectly replicated in child English. For instance, during Adam's null subject stage (Brown 1973), 95% (114/120) of Wh-questions with missing subjects are adjunct questions ("*Where*$_t$ ___ *going t*?"), while very few (2.8%=6/215) of object/argument questions drop subjects ("**Who*$_t$ ___ *hit t*?"). Moreover, if the Chinese type *topic-drop* option is available, and probabilistically accessed (see §2 below for the use of probabilistic learning), then a certain level of null object, which is grammatical in Chinese when the topic happens to be the object, can be expected in child English as well. And we can make the stronger prediction that the relative ratio of null objects and null subjects to be identical across English children and Chinese adults, which is confirmed with data from Wang et al. (1992). The ratio of null objects over null subjects is 0.29 (11.6%/40.6%) for Chinese adults, and 0.32 (8.3%/25.9%) for English children during the subject drop stage.

The second strand of evidence comes from the statistical correlates of parameters in child language acquisition. It builds on the observation that parameters are correctly set at different points of language development. Since parameter setting

requires language specific information, one can estimate the *amount* of necessary data for a parameter value in child directed input and relate it to the time course of parameter setting. Several parameters and their development are summarized below (see Yang 2002, 2009, Legate & Yang 2007 for additional discussion).

Parameter	Target	Requisite evidence	Input	Time of acquisition
wh fronting	English	wh questions	25%	very early[a]
topic drop	Chinese	null objects[b]	12%	very early[c]
pro drop	Italian	null subjects in Wh questions[b]	10%	very early[d]
verb raising	French	verb adverb/*pas*	7%	very early (1;8)[e]
obligatory subject	English	expletive subjects[b,f]	1.2%	3;0[c,d]
Object verb second	German/Dutch	OVS sentences[b,g]	1.2%	3;0-3;2[b,h]
scope marking	English	long-distance wh questions	0.2%	>4;0[i]

Table 1. Statistical correlates of parameters in the input and output of language acquisition. Very early acquisition refers to cases where children rarely, if ever, deviate

from target form, which can typically be observed as soon as they enter into multiple word stage of production (e.g., finite verb raising in French; Pierce 1992). Later acquisition is manifested through children's systematic use of non-target but parametrically possible options. References cited: a. Brown (1973). b. Yang (2002). c. Wang et al. (1992). d. Valian (1991). e. Pierce (1992). f. Hyams (1986). g. Lightfoot (1999), Yang (2002). h. Clahsen (1986). i. Thornton & Crain (1994).

Parameters for which the target value is expressed more frequently are learned faster by children that those which are expressed less frequently. [1] These findings provide support for the reality of parameters, adding to the traditional motivation from crosslinguistic generalizations. For an illustration that unites comparative syntax and language acquisition in a single stroke, see Snyder (2001).

[1] By implication, these findings suggest that the claims about children's adult-like linguistic competence (Pinker 1984 and Crain and Thornton (1999), and others) and very early parameter setting (Wexler 1998) must be refined; see Yang (2002, 2009) for details with specific reference to null subjects and verb second (cf. Poeppel & Wexler 1993). In any case, it is important to note that these claims are generally made without an adequate theory of *how* the learner manages to arrive at the target grammar, which does vary from language to language; see section 2.

Despite its considerable success, the parameter seems to have fallen out of favor in current Minimalist theorizing and other theoretical frameworks (Newmeyer 2004, Culicover & Jackendoff 2005). It is certainly logically possible to recast the fact of language variation without appealing to syntactic parameters; we can point to variation in the lexicon, variation in the functional projections, features, feature strengths, feature bundles, etc. to "externalize" the parametric system to interface conditions, presumably out of the syntactic system proper. For instance, consider the appeal to parameters in the acquisition of subjects above. The *topic-drop* option of the Chinese grammar (and early child English) may be construed as a discourse principle, the property of *pro-drop* can be attributed to the morphological system which may be partially connected to discrete categorization and other cognitive abilities, and the English type obligatory use of subject is a reflex at some generalized EPP feature that is realized at the PF interface. But it is also important to realize that such a move does not fundamentally change the nature of acquisition problem: the learner still has to locate her target grammar in the space of finite choices with a reasonable amount of data within a reasonable amount of time. And to the extent that syntactic acquisition can be viewed as a search among a constrained set of grammatical possibilities, the Minimalist—and indeed, non-Minimalist—alternatives to parameters ought to provide similar empirical coverage. In §2 we provide some learning theoretic considerations for a plausible theory of parameters, with the possible implication that the mechanisms of acquisition may shift some explanatory burden out of the innate UG device. Nevertheless, the empirical evidence for parameters remains, and conceptual arguments for the elimination of parameters run the risk of losing important insights and discoveries through decades of fruitful research.

2 Minimalism and Learning

One of the most revolutionary aspects of Minimalism is the consideration of the language faculty in a broad cognitive and perceptual system, which marks a significant shift from the earlier inclination to attribute the totality of linguistic properties to Universal Grammar. Viewing a theory of language as a statement of human biology, one needs to be mindful of the limited structural modification that would have been plausible under the extremely brief history of *Homo sapiens* evolution. The Minimalist program of language evolution (Hauser, Chomsky & Fitch 2002) seeks to isolate aspects of the linguistic system and identify their homologies in other domains and species. Likewise, one can raise the question how much of the Language Acquisition Device is specific to language—or acquisition.

Our review here focuses on the algorithmic mechanisms of language acquisition. First, consider the problem of parameter setting (or whatever formulation that parameters receive in the Minimalist setting). Much of the discussion in generative literature has centered around domain specific learning algorithms, the most prominent of which is the triggering model of Gibson & Wexler (1994) which is schematically illustrated below.

(2) At any time, the learner is identified with a grammar *G*, i.e. a set of parameter values

a. Upon receiving an input sentence *s*, analyze (e.g., parse) *s* with *G*

b. If success then do nothing; return to a.

c. If failure then

I. Randomly select a parameter and flips its value, obtaining a new grammar *G'*

II. Analyze *s* with *G'*

III. If success, then keep *G'*; return to a.

IV. If failure, revert back to *G*; return to a.

A model like triggering is designed to make full use of the structural properties of the linguistic systems. The so-called Single Value Constraint in (2cI), for instance, reflects the view of parameters as an intrinsically interactive system such as the modification of the grammatical hypothesis ought to be minimal, as the learner changes the value of only one parameter. It is probably fair to say that domain specific learning is still the dominant approach in the generative study of language acquisition; to wit, virtually all learning models in the Optimality Theory use some version of Constraint Demotion which takes advantage of the structures of ranked constraints—and is indeed considered as a virtue of both the theory and the learning model (Tesar & Smolensky 1998).

But the domain specific razor cuts both ways. A learning model that goes hand in hand with a linguistic theory must be modified, or completely abandoned, if the linguistic theorizing takes a different direction. Furthermore, if the defects of the learning model are revealed (see Berwick & Niyogi 1996 on the triggering model), the grammatical theory may be impeached as well (see Tauberer 2008 on OT learning). For quite independent reasons, general learning mechanisms have been applied to language acquisition in recent years (Labov 1994, Yang 2002). A prominent feature of this line of

work is the introduction of probabilistic distributions over grammatical hypotheses, which may be, and are indeed believed to be, domain specific. Consider the variational *learning* model (Yang 2002), which is borrowed from one of the earliest mathematical models of learning (Bush & Mosteller 1951)—from the behaviorist tradition, indeed—that has been observed across domains and species (Herrnstein & Loveland 1975):

(3) At any time, the learner is identified with a population of grammars with associated probabilities

a. Upon receiving an input sentence *s*, select grammar *Gi* with probability *Pi*

b. If success then increase *Pi*; return to a.

c. If failure then decrease *Pi*; return to a.

In contrast to the triggering model, the variational learner doesn't actively participate in learning: the hypotheses themselves do not change, and the only thing that changes is their distribution. The conception of learning as a gradual and probabilistic process opens up the possibility of explaining child language as quantity sensitive growth in response to the volume of necessary linguistic evidence in the environment; the developmental correlates of parameters in Table 1 are uncovered under such considerations.

The reader is directed to Yang (2002) for a more formal treatment of variational learning applied to a parametric space, and the most general convergence result can be found in Straus (2008). In some cases, the variational model is provably superior to alternatives such as triggering, with additional benefits of bringing the formal grammar model closer to the facts of child language. But there is no general result that probabilistic learning is inherently superior (see Yang 2008 for discussion). In fact, the

plausibility of a learning model depends more on the structure of the grammar space and less on the algorithmic aspects of learning. One can easily imagine the worst case scenario where parameters interact in arbitrarily complex ways such as all learning models become intractable. Here we consider the issue of plausible learning in several directions, some of them traditional while other stemming from Minimalist considerations.

First, as Chomsky noted long ago (1965, p61), an explanatorily adequate theory of grammar is one in which the hypotheses can be ``scattered", i.e., distinguished by a reasonable amount of linguistic data in a computational tractable way. (Modern theory of statistical inference speaks of a similar notion called ``shattering", which refers to the requisite amount of data capable of locating the target hypothesis in statistical classification, see Vapnik (1995)). In the generative literature, there have been several convergent lines of research that point to the advantage of a structured parametric space that favors the learner. One of the earliest efforts is the cue-based learning model of Dresher & Kaye (1990); see Dresher (1997) and Lightfoot (1999) for applications to syntax. Dresher & Kaye observed that ambiguity, which refers to the fact that an input token may be compatible with multiple grammars, could easily mislead the learner, who is presumed to make learning decisions locally. Their solution lies in a set of parameters whose values can be determined unambiguously but only following a predefined and thus presumably innate sequence. The work of Fodor (1998) and Sakas & Fodor (2001) is another response to the ambiguity problem. Here the learner hedges its bets more intelligently than the randomly guessing triggering learner. It avoids learning from input that is compatible with multiple hypotheses and only modifies the grammar on

unambiguous data. The detection of data-grammar ambiguity is achieved by trying out multiple grammars with each input token. Finally, the idea of parameter hierarchy (Baker 2002), largely motivated from a comparative/typological point of view, would set the learner on course of a sequence of decision that starts from major divisions of languages--e.g., whether a language is ergative or not—to minor ones such as the placement of adjuncts on the left periphery (Cinque 1999). The hierarchy, like cues, is conjectured to be innate and thus solves the ambiguity problem from within. The natural question, of course, is to what extent the parameters required to describe the word's languages follow the ideal expressed in these works. And to some extent, the probabilistic learning model of Yang (2002) can replicate the effect of parameter sequences and cues without assuming innate specification, but it does not remove the necessity of structured parameter space to achieve plausible ordering, especially if the parameter space is structured to simplifying learning. [2]

Second, it would be a mistake to suggest that child language can be entirely explained in terms of searching for a solution in a constrained parameter-like space. The

[2] Note that the size of the search space may not matter as much as it appears. A system with 1000 parameters seems harder to learn than one with only 50 parameters. But if the former consists only of parameters whose values can be determined independently (e.g. in the sense of Yang 2002), and the latter has massive ambiguity problem resulting from parameter interactions, then the larger space can be more plausibly learned than the latter. Recent work of Sakas and Fodor (2009) finds via computer simulation that in a linguistically realistic and complex domain of parameters, the majority of them may indeed be independent.

most obvious case can be found in morpho-phonology: innate principles of UG notwithstanding, even the most enthusiastic nativist would hesitate to suggest that the English specific rule for past tense formation ("*-d"*) is one of the options, along with, say, the *"-é"* suffix as in the case of French, waiting to be selected by the child learner. In the domain of syntax, we also find patterns of variation that may be governed by universal constraints but are realized in particular languages in highly specific and widely ranging ways such that the learner cannot but make use of (constrained) inductive learning mechanisms. An example of this type can be seen in the distribution of dative constructions such as double object and prepositional dative across languages. There is broad agreement that certain universal syntactic and semantic properties are the necessary conditions for a verb to participate in dative constructions in the first place (Pesetsky 1995, Hale & Keyser 2002, Rappaport Hovav & Levin 2008). Nevertheless, these constructions are productive in English but are limited to a lexically closed class of verbs in languages such as Korean (Jung & Miyagawa 2004) and Yaqui (Jelinek & Carnie 2003) that must be learned individually. Tellingly, English learning children show an initial stage of conservatism, that they do not generalize these constructions to novel items (and thus do not make ungrammatical errors) (Gropen et al. 1989, Snyder 2006). The detection of productivity apparently takes place at around 3;0 (Conwell & Demuth 2007; cf. Snyder & Stromswold 1997). The course of acquisition strongly resembles the phenomenon of over-regularization in morphological acquisition, suggesting that a learning mechanism capable of detecting the productivity of linguistic productivity is at play. The questions here are again traditional; the acquisition of the universal and language particular and construction specific aspects of language was once at the

forefront of language acquisition research (Fodor & Crain 1987, Pinker 1989), with special focus on the core vs. periphery distinction drawn at the outset of the P&P framework (Chomsky 1981). A recent approach (Yang 2005, 2009a) draws inspiration from the principle of efficient computation under Minimalism and develops a decision procedure by which the processing of productive and exceptional items are jointly optimized. While still preliminary, this work gives an example of how optimal design principles of language (Chomsky 2005) may be applied to language acquisition.

Finally, it is equally important to recognize the limits of general computational mechanisms in language acquisition (Gomez & Gerken 2000), which appear to have been gained popularity ever since the demonstration of statistical learning of artificial languages by infants (Saffran et al. 1996; cf. Chomsky 1955/1975). Much of this work, however, remains confined to a laboratory setting at the moment. Still less effort has been made to test whether these mechanisms scale up in a realistic learning environment, and there have been negative results (Yang 2004).

Looking more broadly, a current theme in cognitive science has advocated a data-intensive and memory-centric approach to language learning (Tomasello 2000, Bybee 2006), which leads to claims about child language as "item-based" and limited in syntactic creativity. Even though these positions have not been embraced by the Minimalist community, one does find similar stances toward the division of labor between the grammar and the lexicon. Following Borer (1984), it is assumed that language variation, and thus acquisition, can be attributed to the properties of lexical items. (Surely words have to be individually learned.) Without taking ourselves too far

afield, it is useful to discuss these issues briefly as Minimalism has forced us to reconsider the relation between the language faculty and general cognitive systems.

At the very beginning stage of acquisition, the child's grammar is necessarily restricted to specific lexical items; after all, hearing "Hi baby" once is not going to give the child the complete grammar of English. The leading questions, then, can be phrased as follows:

(4) a. From a learning perspective, how does the child go from specific instances of the data to general grammatical properties?

b. From a developmental perspective, are any major aspects of grammar (e.g., verbal syntax, noun phrase structures, as held in the item-based learning approach) actually item-based even for the youngest children that can be assessed? (The child's grammar could be off target but productively so, as in the case of Null Subjects in English acquisition.)

c. From an empirical perspective, could the item-based approach in principle offer an adequate solution for the problem of language acquisition?

No complete answers will be given here for they have not been fully explored at the present time. The generalization problem (4a) is a traditional one, as the discussion of productivity learning above indicates. To address (4b) and (4c), useful insights can be garnered from the statistical properties of the linguistic data from both adults' and children's linguistic production, which constitute the input to and the output of language acquisition. A strong and consistent pattern, one that is familiar in the field of corpus and computational linguistics, is the so-called Zipf's law (1949): most linguistic items, be

they morphemes, words, morphological rules or phrases, are used very rarely even when the amount of linguistic data is very large, and that the items that the learner receives reliable evidence in the input are relatively few, and these patterns hold for both child and adult languages (Chan 2008). Thus, regarding (4b), one cannot simply take the relatively low degree of usage diversity in child language (Tomasello 2000) to be an indication of the child's grammar as organized around specific lexical items and constructions (Yang 2009b). As of (4c), the study of the linguistic data reveals a major challenge, the so-called *sparse data problem* (Jelinek 1998), that all acquisition models must face. As the linguistic model gets more complex, the amount of data required for the instantiation of the model, i.e., acquisition, increases rapidly such that even very large samples will not be sufficient. And there is suggestive evidence from computational linguistics that piecemeal learning using lexicalized grammar models pays little dividend compared to more general and overarching rules (Bikel 2005). So we are back to the heart of generative grammar: How should a theory of grammar simplify the learner's task in order to achieve successful acquisition with a relatively small quantity of data? We hope that child language acquisition can provide stringent but revealing conditions on further developments of Minimalism, much the same way it has carried out its duty for the Principles and Parameters framework.

Part 2: Empirical Evidence of Minimalist Principles

3 How Acquisition Evidence Preceded or Illuminates Linguistic Theory

We can now ask a sharper question: does Minimalism itself offer some methods to capture the Primary Linguistic Data? Can minimalist principles of sentence construction

explain micro-steps on the acquisition path, especially those not seen in the target adult grammar? Note that one can ask a prior question: Should acquisition data be expected to map naturally onto any linguistic theory? In the 1960's, it was doubted that acquisition research was possible because performance factors so profoundly clouded what happened in a child's mind that no relevant evidence could be expected. It remains easy to attribute deviations from adult grammar to external performance factors—it can be found in every acquisition journal and is often treated as a naturally superior explanation. Nonetheless, in almost every instance the seemingly deviant acquisition data has instead proved to be a reflection of grammatical constructions found in other languages.[3]

3.1 Small Clauses

An instructive example was *small clauses*: it was often informally suggested in the early 70's that children said "it big" because they lacked the memory space to say "is" with case-assignment as a default accusative in English (as in "him big"). If factors like sentence-length (Bloom 1990) were paramount and forced deletions, then the deletions

[3] Our discussion focuses on those data where we see principles most directly, primarily English, and where the arguments can be presented efficiently. Similar claims can be made about other realms and with other languages which we have not chosen to focus upon, including root infinitives, VP-ellipsis, passive, relative clauses, articles, and quantification among others. See the articles in deVilliers and Roeper (to appear) Handbook of Generative Approaches to Acquisition.

might happen virtually randomly. If analyzed in grammatical terms--- if grammatically describable at all---they would not tell us about the acquisition path.

Radford (1990) and Lebeaux (2000) argued more interestingly that expressions like "it big" revealed that small clauses had an important status in grammar. Had the acquisition data been taken seriously right away, it might have led at a much earlier point to the recognition that small clauses are a significant form of complementation (see Moro 2000). Other examples follow.

3.2 Scope and Partial Movement

At first sentences like "Only I like milk" (meaning: "I like only milk") were taken to be a performance failure to represent scope at first. However with the advent of Logical Form, they could be analyzed as an early reflection of movement operations to a pre-sentential Logical Form position for quantificational elements (Lebeaux 2000). Similarly the unusual behavior of quantifiers in acquisition fit event-based semantic theory descriptions (Philip 1995), and has engendered a large literature of alternative syntactic and semantic analyses,[4] although at first (Crain and Thornton 1998) performance explanations were given as a way to deny the relevance of the phenomena.

Partial movement structures in acquisition ("what did she say what she wanted") (deVilliers *et al*. 1990, 2007 for comprehension and Thornton 1990, Oiry 2008, Strik 2009 for production) were analyzed at first as non-core phenomena that occurred only in dialects (McDaniel 1989) but have progressively been re-analyzed, such that they now

[4] See Philip (2004) and references in (Drozd 2001).

form support for the Strong Minimalist Thesis (as we discuss below). Had the "performance" or "marginal" explanation not seemed to have priority, spontaneous acquisition evidence (not consistent with target grammars) can be, and should have been seen, as providing UG hypotheses directly (see Boeckx 2008a,b for a compatible argument).

3.3 Principle B

Another example is the well-known delay of Principle B effect (Chien and Wexler 1990) ("John washes him" where "him" is interpreted as "John"), which has been analyzed in pragmatic terms and Optimality Theory or cognitive egocentricity perspectives (Hendricks and Spenader 2006). Recently, Elbourne (2005), Verbuk and Roeper (to appear) and Hamann (to appear) have shown that Frisian/Old English continue to allow pronouns in a single clause, suggesting that there must be a parametric account. Hamaan (to appear) and Verbuk and Roeper (to appear) have argued that the presence of sentences like "John took a wallet with him" in English, in contrast to their absence in German, indicates that even subtler parameters are at work. In German "with him" is disallowed and "mit sich" (with self) occurs, so in German Principle B applies to both arguments and adjuncts. This leads to the correct prediction that children abandon Principle B errors earlier in German than in English where a domain narrower than the clause is critical. The fact that Principle B is realized late in English is then, once again, not a reflection of a grammar-extraneous factor, but the paucity and subtlety of data needed to set a refined parameter in English as we have discussed above in Part I. This supports Kayne's (2005) claim that there are many subtle parameters (see also Lightfoot 1999, Westergaard 2009, and Bentzen et al 2009), which means, in effect, we need to have a more

microscopic view of the acquisition process (see Roeper 2009 for an overview) along many dimensions.[5]

3.4 C-Command

Acquisition evidence can play a role in overcoming the obscuring impact of social and historical factors. Boeckx (2008b) argues that it is the child data, where real support for parametric analysis belongs, not adult data which is too confounded by social and historical factors. An effort to state c-command perfectly for adult grammars (with alternatives like m-command) were proposed, but none could overcome the diversity of relevant data, so that it remains unresolved. Nevertheless, evidence for knowledge of c-command was shown by Lawrence Solan (1983), who showed that for sentences like:

5) a. the horse hit him$_i$ after the sheep$_i$ ran around.

b. *the horse$_i$ told him that the sheep$_i$ would run around.

[5] Conroy et al (2009) continue to offer performance based accounts of the evidence, but their discussion fails to include PP-pronoun violations ("he has it with him") discussed above, which changes the nature of the problem. Nevertheless they observe that the performance based accounts still do not explain even the facts they consider, which suggests that the performance variables should be seen as allowing a grammatical option to surface.

children were twice as likely to allow backwards coreference when there was no c-commanding pronoun ("him") (5a).[6] Therefore, in the spirit of Boeckx's argument, the fact that the basic notion of c-command appears clearly in the work of Solan and subsequent work shows that the principle is essentially correct, despite the fact it has been obscure in the adult grammar (with variants like m-command proposed).

4. Acquisition and Specific Minimalist Principles

The success of acquisition research in the terms of linguistic history leads, nonetheless, to a greater challenge: does it provide specific support for principles utilized in minimalism? What follows shows that Minimalism provides direct evidence for several abstract principles: Asymmetric Merge, Feature-checking, the Labelling Algorithm, the Strong Minimalist Thesis (Phase-based interpretation), recursion, and the role of Interfaces.

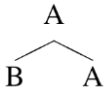
At first it might seem very unlikely that minimalism simplifies rather than complicates the acquisition problem. If a child seeks to analyze input substantively in terms of noun or verb then noun phrase or verb phrase, it would seem to be a step ahead of a child who begins only with the notion of Merge, which might seem to fit anything. A closer look reveals otherwise. In fact, minimalism allows grammatical principles, like asymmetric merge, to participate more directly in the analysis of Primary Linguistic Data than, for instance, phrase-structure rules.

---

[6] See as well Goodluck (1978) and more recent work by Crain and Thornton (1998) and references therein; see Lidz and Musolino (2002) for recent evidence of c-command effects with quantification.

4.1 Asymmetric Merge

Merge deviates in an important way from what might be called a general cognitive capacity for the act of combination or concatenation (Hornstein 2009)) which applies to almost anything in life experience. Merge requires asymmetry: a Label is chosen, usually seen as a projection of one lexical item, which allows one part of a binary Merge to dominate the other, following a Labelling Algorithm (Chomsky 2006).

6) 

In concrete terms, though hard to establish, it predicts that a child will perceive the difference between "ocean blue" which is adjectival, and "blue ocean" which is nominal, (see Roeper to appear)?

Children assemble a number of single words which involve nouns, verbs, prepositions, pronouns (although their category label may be obscure). When a child says "up" is it a preposition or really a verb? While, as with one-word utterances, many theories could explain two-and three-word expressions, the fact that we get unique combinations in early language is captured directly by the notion of Asymmetric Merge and Label[7], here with no AGR node (following Roeper and Rohrbacher (2000)):

(7) Adam 03 no play toy Adam 01 no write on there/no kick box

[7] see Roeper (1996) and Roeper and Rohrbacher (2000) for extensive further evidence.

no do suitcase/no hurt head/no have one

sarah 26: no it upsidedown

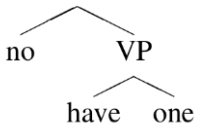
German: nein auto kaput (no car broken) nein dick baby (no fat baby)

This first representation seems to be a negative feature that subcategorizes the lexical categories of N or V or VP: Neg [V or N]. Such examples were initially—quite counter-intuitively-- analyzed as reduced forms of Negative+Sentence, where huge amounts were deleted because of performance demands or the absence of lexical items. While Asymmetric Merge might allow any combination, the absence in English and German of not (nicht) (see Deprez and Pierce (1993)) *Not run suggests that other linguistic features drawn from UG limit the range of possible two-word utterances prior to the full expression of Functional Categories. Nonetheless under minimalism it could be viewed as a grammatical expression---while earlier theories, which have been pursued over a much longer course of research, demanded either deletion or non-grammatical representations. Therefore, as a possible grammar within a theory of Multiple Grammars and competition, outlined above, it may not immediately disappear when more elaborated aspects of grammar emerge. The major question, of course, is to determine which way of interpreting children's language is better supported by empirical evidence.

An expressive pragmatic feature seems to be present as well and therefore no can be seen as an Expressive word (Potts and Roeper (2005)) that can be paraphrased as: No way yoghurt (Drozd (2001)). If it is an expressive, it may have no lexical label at first (like wow,or gee) and still be subject to asymmetric Merge. It will grow (presumably by adding features) into a NegP that is embedded within a VP.

What then are the set of possible Labels that asymmetric Merge can generate? It might, for instance, generate a Neg-feature, but possibly without all of the Functional information that languages allow, and therefore it is an incomplete Functional Category:

8) NEG [+negative, +expressive, ?+imperative]



We can guess that it may contain an Expressive feature, as we suggested above, linked to forms like wow, gee, well (Potts and Roeper (2005)) and possibly a Force feature like imperative. The question is interestingly abstract, since something like "no Yoghurt" is often associated with an imperative impulse. However wordlessly pointing a finger in a deli at a sandwich also conveys imperative import, but perhaps via a form of communication that does not invoke grammar. Therefore we do not know from this example whether we should add a Force Feature to the node (or add a higher node) or leave the imperative property initially to inferential pragmatics of the sort that interpret gestures.

The important point here is that Asymmetric Merge allows an immediate representation of a child's first utterances and, more importantly, an abstract analytic instrument that enables a child to attack in a simple way what is a very complex set of inputs, before projecting the full array of functional categories (which is not to say that

the capacity is absent). The significance of this point should be underlined: a virtue of the abstraction of minimalism is that it reduces the Primary Linguistic Data problem by giving the child representational tools.[8] that allow first stage efforts to represent linguistic forms whose full feature system has not yet been identified. In that sense, Minimalism predicts that Stages can exist.[9][10]

[8] Other abstract operations presumably delivered by UG are present in earlier theories but remain significant sources of spontaneous overgeneralization, for intance, Operators. Operators capture a wide variety of discontinuous connections in grammar. Are they among the primitives a child uses in early comprehension and production? Does a child who recognizes variation seek co-variation at a distance? Spontaneous acquisition suggests that Operators are among children's first analytic devices.

Two-year-olds say "this is to eat" which can be analyzed as [This1 is [OP1 t1 to eat t1]] with an Operator linking the *trace* after eat to the lower CP to the upper subject.

The presence of Concord like "I don't want none" without an adult model suggests that children seek and project Operator-variable relations. It is arguably present in many forms of "overgeneralization" in children's grammar, from Tense to plural to quantification: "feetses". "had came" and "both rabbits are on both sides of the fence") (Partee pc)).

[9] Lebeaux (2000) had already argued that acquisition data supported structure-building notions like Adjoin-alpha, which were more abstract than either Phrase-structure-rules or subcategorization would allow.

4.2 Pair-Merge and Set-Merge: adjunction in child grammar

Merge further divides into two kinds: Pair-Merge (adjunction) and Set-Merge (argument subcategorization). One can ask whether both kinds of Merge appear in acquisition. Lebeaux (2000) and Roeper (2003) argue that at first all attachment may be Pair-Merge or adjunction. Only upon a recursive 2nd Merge must the child decide exactly the higher label. Both Set-Merge and Pair-Merge are visibly present as soon as 3 word utterances arise (Brennan (1991) where, interestingly, prepositions are absent only with adjoined elements:

Lebeaux ((2009) and references therein)) argues that the early stages where no evidence of Functional Categories are present support a notion of subgrammars within a sequence of modules involving case-assignment and movement. He points to evidence from L2 acquisition (Vaininikka and Young-scholten (1994), (in prep) and Tag-grammars (Frank (2002)) which also are compatible with these early stages of acquisition. The acquisition data provides critical support for this nested conception of grammar (which, in turn, underlies his analysis of very complex binding structures in the adult grammar).

[10] In principle, merge might allow any combination, but in fact we neither find any combination like *"want to based on the frequent experience of ellipsis ["do you want to?"] or *"yes want to", or "said ate" = "I said I ate". Therefore more must be said about the features inside these first Merges.

8) “I cried stairs” (=on) Shirley get meat dinner (=for)

I cut it a knife (=with) Richard bring snack Shirley (=for)

feed baby fork (=with) Shirley cut fork (=with)

I sleep big bed (=in) Save some later (=for)

9) I cried [set merge => verb requires subject

cry stairs [Pair merge: stairs is adjoined to cry]

Arguments in contrast have prepositions:

9b) I played with John/ Jim was at Cooperstown/ putting Daddy in wagon

Brennan (1991) reports that there were 46 prepositions for arguments and only 3 for adjuncts, although adults and children both have more adjunct than argument PP’s. “For 3 of the 4 children studied, it was true that adjuncts never surfaced with PP’s, while the distribution of PP’s in argument position was haphazard”. The child is apparently able to adjoin stairs without any subcategorized feature linked to cry. It deviates from the adult grammar which requires a PP projection to introduce this information in order to assign case to the adjunct “stairs”. If the case-module, however, is not yet fully defined for prepositions (where transitives, intransitives, and particles must be differentiated), causing overuse of accusative Default case, then the theory predicts that such examples

should occur. This in turn reflects modularity: the emergence of language-particular features within the case module may have its own acquisition path.[11]

Assymmetric Merge also allows a direction of complementationand therefore predicts word-order invariance. Early suggestions by Bowerman (1973) that children's grammar was semantic and lacking syntactic ordering (because examples like "Adam watch" and "watch Adam" with the same meaning exist) was disproven by Bloom (1990) who showed that children at the 2 word stage never made order errors with pronouns and predication. That is they said "that big" or "it big" but never *"big that" or *"big it".

4.3 Pied-piping and Economy of Representation

Do we have evidence for Feature-checking as a motive for syntactic movement ?

Feature-checking and economy of representation receive another kind of specific support from spontaneous aspects of acquisition. If Feature-checking motivates movement, e.g., if a wh-word carries a Feature which matches a CP feature and moves to check it off, then it is only the critical feature that needs to move, not everything moved under Pied-Piping. Everything extra is a "free-rider" under a minimalist form of Feature-checking. The sharpest spontaneous evidence comes from Guasavera and Thornton (2001) who

[11] See Lebeaux (2009) for this argument. It may be a notion of Abstract Agree, see Roeper et al (2003) for related evidence from language disorders.

provide an extensive experimental evidence from at least 10 children that they will break-up “whose” and move only “who” in production:

(14) Q: John saw someone’s book. Ask him which book?

A: Who did you see t ‘s book

This is precisely what ought to occur, but children never hear direct evidence for it in English since the choice of lexical items from the Numeration offers only the contracted “whose” which drags the object along: “whose book did you see?”

Do-insertion, originally claimed to be a Last Resort phenomenon, fulfills economy of representation under Feature-checking in early acquisition in precisely the same way. Hollebrandse and Roeper (1997) find that children prefer to insert “do” rather than pied-pipe a V+Tense (as in “painted”) from the lower V-node to a higher Tense node. This occurs for brief periods in various children who spontaneously produce non-target grammar do-insertion:

15) “do it be colored”

“I did paint this and I did paint this”

Do-insertion achieves immediate Feature satisfaction without requiring percolation of the lexical feature to a higher node. (See Fitzpatrick (2005), Heck (2009), Roeper (2003), Cable (2007) for discussion of economy and Pied-piping.) From this perspective, the child resorts to do-insertion as a First Resort, preferred over pied-piping a verb and it converts do-insertion into an operation that preserves economy of representation for

Feature-checking rather than being a response to an imperfection in grammar. This can be seen as evidence that Merge is more economical than Move (Internal Merge).

4.4 Barrier Theory and the Strong Minimalist Thesis

Barrier theory formed the crux of linguistic work for a quarter century and its central tenet was clearly supported as a universal constraint in the evidence that extraction from NP's and strong islands were prohibited (Otsu (1981), deVilliers and Roeper 1990, Baauw (2000), Oiry and Demirdache 2006, Friedman et al 2009 among others; see also deVilliers and Roeper (in preparation, see for sources).[12]

Children were given a choice of "with" as part of the NP or VP in the following context (Otsu 1981):

12) "The boy fixed the dog with a broken leg with a bandage.

"What did the boy fix the dog with?"

They choose VP-with "a bandage" 90% of the time, the form consistent with an NP-barrier, where [NP_1 a broken leg [PP with [NP_2 a bandage]] prohibits extraction of one NP from inside another.

[12] See also Baauw (2000) for cases where Strong Islands seem to be violated, pointing toward pragmatic factors that cause phenomena like subjacency to be called Weak constraints. The fact that children can realize violable constraints increases the necessity for a strong innate, input oriented bias because they must in effect overlook exceptions. Work in phonology on Optimality Theory is relevant to this line of reasoning.

They likewise do not allow long-distance adjuncts in cases like (deVilliers *et al.* 1990, 2007):

(13) The boy said in the morning he balanced the ball on his nose at midnight.

When did the boy say how he balanced the ball?

Answer: In the morning. (*midnight)

Surprisingly, over 30% of children answer *how* ("on his nose"), treating *why* as a scope-marker, inviting a Partial Movement analysis. What feature-mechanism exactly blocks movement? Possibly an outgrowth of Relativized Minimality is the right path—see Grillo (2008), Friedmann et al (2009) and Schulz (2005) who posits an intermediate Focus element that is distinct from a question element to satisfy and delete a wh-feature. The strong fact remains: children easily take a medial wh- to be either a barrier or a copy of a scope-marker (Partial movement). Such findings have been reported in several other languages (Oiry (2008) and Strik (2009)). It remains to be seen how these locality effects in syntax are captured under Minimalism.

Another angle on this problem emerges below when we consider the Strong Minimalist Thesis. Chomsky (2005) has proposed that Phase theory should include interpretation. i.e. that syntax, semantics, and phonology may all be bounded within the Phase:

Strong Minimalist Thesis (SMT):

"Transfer…..hands Syntactic Object to the semantic component, which maps it to the Conceptual-Intentional interface. Call these SOs *phases*. Thus the Strong Minimalist

Thesis entails that computation of expressions must be restricted to a single cyclic/compositional process with phases." Chomsky (2005).

This principle offers a fresh perspective on Partial Movement in production and comprehension by children. We find productions such as

(16) "What did she say what she wanted?"

"What do you think which animal says "woof woof"?

"What do you think which Smurf really has roller skates?" (Thornton 1990: 246)

Here we find that the child applies the SMT to phonology and pronounces the intermediate trace. The spontaneous reflection of the SMT is underscored by the fact that Asian and Spanish second-language learners do the same (Schulz 2005, Guttierrez 2005),

In addition, interpretation of the lower clause is called for when the IP Phase Edge is met. We find the SMT interpretation for sentences like (deVilliers et al 1990, (to appear):

17) When did the clown say how he caught the ball.

where children answer the "how" question (e.g., "on his nose"), providing a plausible within-first-Phase answer. Adults delay an interpretation until the next Phase where the CP properties inherited from the higher verb make the lower clause into an Indirect Question that is not answered.

Children in fact take an extra step, interpreting not only overt wh- words, but *traces* as well:

18) [she bought a cake but said she bought paper towels]

"what did she say t she bought t?"

as if they are answering the lower clause without the impact of the Indirect Question feature on the CP inherited from say (see deVilliers et al (to appear)) while adults correctly say "paper towels". These spontaneous deviations from adult grammar are precisely what the SMT promotes---immediate interpretation within each phase--and therefore they demonstrate the core role of locality in human grammar.

How does the child eliminate an overt medial wh-word or the phasal interpretation of a trace? The answer is not yet clear but it should follow from the formal representation of opacity at LF. Roeper (2009) argues that the child must learn to alter a trace—modify the unmarked interpretation required by the SMT-- to prevent it from being interpreted in its original Phase. It changes from a Full trace to a Converted trace when evidence arises that full reconstruction delivers the wrong interpretation.[13] This is forced when the child, in the example above, recognizes that what the mother bought and what she said she bought are in conflict. Once again, it is precisely where complex constructions elicit subtle spontaneous deviations that reflect basic principles that the stunning contribution of acquisition is evident.

[13] This approach converges, from an acquisition perspective, with work on trace-conversion (Sauerland (2003)) and multi-dominance (Johnson (2009)) which is at the forefront of current syntactic research.

5 Recursion

Chomsky, Hauser, and Fitch (2002) have argued that a core feature of Minimalist representations is recursion. The operation of Merge creates recursive hierarchies in every language. Some categorical recursion is virtually invisible. Thus the fact that one article occurs inside of a structure containing another may not be detectable by the human computational system, for instance, the articles in: the man in the house.

However there are language specific forms of recursion which children do not acquire instantly and whose complexity is intuitively evident. Possessives, adjectives, and clauses require recursive generation and are systematically delayed in the grammars of children, and not uniformly present in the languages of the world. For instance children and L2 speakers find it very difficult to handle forms like:

19) Cookie Monster's sister's picture

3yr old children regularly prefer a conjoined reading Cookie Monster and sister's picture when faced with alternatives (see Roeper 2007).

Here is an illustrative dialogue, among many, where the parent does not perceive the difficulty (Childes, Brown Corpus, Sarah 039):

MOTHER: What's Daddy's Daddy's name?

SARAH: uh.

MOTHER: What's Daddy's Daddy's name?

SARAH: uh.

MOTHER: What is it? What'd I tell you? Arthur!

SARAH:Arthur! Dat my cousin.

MOTHER:Oh no, not your cousin Arthur. Grampy's name is Arthur.

Daddy's Daddy's name is Arthur.

SARAH:(very deliberately) No, dat my cousin.

MOTHER:oh. What's your cousin's Mumma's name?

What's Arthur's Mumma's name?

And what's Bam+Bam's daddy's name?

SARAH: Uh, Bam+Bam!

MOTHER: No, what's Bam+Bam's daddy's name?

SARAH: Fred!

MOTHER: No, Barney.

Where exactly then, does the difficulty lie? The answer connects to the SMT. The child must not simply grasp the fact that a category is embedded inside an identical category, but also generate an interpretation at each Phase Edge. Thus the child interprets a possessive as possessive and the next point of interpretation calls for embedding that possessive meaning inside another. The alternative is the non-embedded conjoined reading noted above (choosing Cookie Monster's and sister's picture for (19)). This shift

from conjoined to embedded occurs for every form of recursion where the Phase-Edge must be interpreted: PP, AP, CP, and DP (see Roeper 2007, 2009).[14]

Ultimately this means that just as syntax must be connected to Interface Theory, it is precisely at the point of the interface between recursive syntax and the interpretive connection to the SMT that the child experiences a challenge. These results point at the idea that interfaces must be articulated to understand UG, yet it leaves a promising challenge for future research: how exactly do recursive structures engage computational complexity?

6 Interface Connections

From an architectural perspective, there is evidence that the notion of an interface, still theoretically hazy, unlike the original theory of autonomous syntax, provides a scaffolding on which to place a crucial role for semantics and pragmatics in the emergence of grammar.

First some clarification of the concept. We take interfaces to be mechanical and biologically articulated, like the connection between the heart and the lungs. In that respect, we need to sharply differentiate an innately specified interface from a system-wide interaction in energy use that connects every part of an organism.

Acquisition theory has always tacitly assumed a rich interface whose mechanics were subtle and mysterious. There is no doubt that inferences about context must feed into

[14] See Boeckx (2008) for discussion of an Alternating Phase constraint which entails the interpretive impact of the SMT. See Hollebrandse and Roeper (to appear), Roeper (2009) for discussion and extension to acquisition.

both lexical and syntactic growth, but how? This is an old and common intuition: the challenge is to build it into a mechanism. [15] It is likely that the acquisition mechanism begins with an over-reliance on context---an interface choice--that is ultimately altered when grammatical comprehension becomes autonomous. For instance, children will misunderstand a sentence like (20a) (Bever 1972, Roeper 1981):

20) a. the mouse was eaten by the cheese

b. the cheese ate the mouse

guessing that the mouse ate the cheese for (20a) since world knowledge makes it plausible. At the same time part of their grammar is autonomous: they already know that an active sentence (20b) is nonsense. Why? If the active is in the grammar, then context does not guide interpretation, but if passive is not controlled, then the child will make a pragmatic choice—a rather common sense idea. When the child has a passive transformation as an hypothesis, then context can serve as confirmation of syntactic and semantic hypotheses:

---

[15] See Roeper (1981) and Chomsky (1980)) observes that acquisition follows "triggering experience". An example of this interface is the view by Wexler and Culicover (1980) that deep structure is recognized from context. Lebeaux (2000) argued that children's capacity to map a deep structure meaning onto surface structure serves as a means to confirm the transformational mechanism that the acquisition mechanism proposed.

21) Scene: baby drinks milk

Sentence: “the milk was drunk by the baby”

can justify the conclusion that a passive sentence is involved if one assumes these dimensions coincide:

Pragmatics: baby drinks milk

Semantics: Agent verb Theme

Syntax: object => undergoes preposing to subject

where a real event is the backdrop to comprehension. The child may need to hear quite a number of sentences before he has sufficient evidence that these interfaces all match, supporting the view that frequency of exposure will correlate with point of acquisition. When passive is entered in the grammar, then it can be immune to context, allowing the child to understand nonsense, jokes, etc. Thus context can play a role in acquisition, where it is minimized in the final grammar. In sum the pragmatics of the context is connected by innate stipulations to syntax and its semantics in a narrow range of options that the child searches through.

Outside of a role in acquisition, there are limited open parameters where the role of context participates in parameterization. Huang (1982) showed that there are language-particular choices: “hot” and “cool” languages allowing different amounts of contextual deletion. The acquisition path may allow the child to begin with a “hot” language—where there is an over-reliance on context-- and shift to a cool language just in case she is speaking an English-like language. Thus important and precise openings to “context “ are themselves part of grammar and may be linked to subtle variation in, for instance,

where object-drop is possible. We are just at the outset of discovering the acquisition path for context and pragmatic implicatures which may play a role in the recognition of syntax. These connections call for an enrichment, not a minimization, of the innate component and they feed other forms of efficient computation.

7 Some conclusions

The first half of this chapter has laid bare the mechanisms whereby, assuming UG, the child is able to analyze and compare the input data. The fact that we are able to build a fairly intimate model of how UG extends to models that accept the raw primary linguistic data is a support both for the abstractions of minimalism and the data comparison systems that utilize them.

The second half explored the promise of minimalism in the mircroscopic terrain of spontaneous acquisition. We have provided an overview of where minimalist principles are at work: Merge and Label, Merge over Move, the Strong Minimalist Thesis, and its impact upon Recursion. Constraints, like barriers, are always obeyed, but we found children not only follow the barrier-like constraints of the Strong Minimalist Thesis, but as well, show spontaneous evidence of Phase-based effects. We introduced questions about interfaces and argued that they are central to the acquisition process, allowing confirmation of syntactic analysis. It constitutes an important alternative to viewing children's language as the result of interfering performance factors. Drawn together, this evidence validates the core prediction of Minimalism: if the theory is correct, then its mechanisms should be transparent in the acquisition process.

References

Baker, M. (2002). *The Atoms of Language*. (New York: Basic Books)

Baauw, S. (2000). Grammatical Features and the Acquisition of Reference. A Comparative Study of Dutch and Spanish. Doctoral dissertation, Utrecht University.

Bentzen, K., Garbacz, P., Heycock, C., Hrafnbjargarson, G.H.. 2009. On variation in Faroese verb placement. Tromsø Working Papers in Linguistics, Nordlyd 36(2).

Berwick, R., & Niyogi, P. (1996). Learning from triggers. *Linguistic Inquiry*, 27:605-622.

Bever, T.G. (1972). The integrated study of language behaviour. In Morton, J. (Ed.), Language: *Biological and social factors* (159-206). Logos Press.

Bikel, D. (2004). Intricacies of Collins' parsing model. *Computational Linguistics*, 30, 479– 511.

Boeckx (2008a) Approaching parameters from below. In *Biolinguistic explorations*, ed. A.-M. di Sciullo and C. Boeckx. Oxford: Oxford University Press.

Boeckx, C. (2008b) On the locus of asymmetry in UG. Ms. Harvard University

Bloom, P. (1990) Syntactic distinctions in child language *Journal of Child Language*, 17, 343-355.

Borer, H. (1984). *Parametric syntax.* Foris.

Brennan, V. (1991) Formal Semantics of Telegraphic Speech. In B. Plunkett (Ed.) *Issues in psycholinguistics.* UMass GLSA

Bush, R., & Mosteller, F. (1951). A Mathematical Model for Simple Learning. *Psychological Review*, 68:313-323.

Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language*, 82, 711-733.

Cable, S. (2007). The Grammar of Q: Q-Particles and the Nature of Wh-Fronting. Ph.D. MIT.

Chan, E. (2008). Distributional and Structural Basis on Morphological Learning. Ph.D. Dissertation. Department of Computer Science. University of Pennsylvania. Philadelphia, PA.

Chien, Y.-C. and K. Wexler (1990). Children's Knowledge of Locality Conditions in Binding as Evidence for the Modularity of Syntax and Pragmatics". *Language Acquisition* 1, 225-295.

Chomsky, N. (1955/1975). *The Logical Structure of Linguistic Theory*. Manuscript, Harvard University and Massachusetts Institute of Technology. Published in 1975 by New York, NY: Plenum.

Chomsky, N. (1965). *Aspects of the Theory Of Syntax*. Cambridge, MA: MIT Press.

Chomsky, N. (1980). *Reflections on Language*. New York: Pantheon.

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

Chomsky, N. (1995). *The Minimalist Program.* Cambridge, MA: MIT Press.

Chomsky, N. (2004): Beyond Explanatory Adequacy. In: A. Belletti, *ed*., Structures and Beyond. The Cartography of Syntactic Structures (volume 3). Oxford, Oxford University Press.

Chomsky, N. (2005a): On Phases. Ms., MIT.

Chomsky, N. (2005b): Three Factors in Language Design. *Linguistic Inquiry* 36: 1-22.

Chomsky, N. (2006): Approaching UG from below. In H.-M. Gartner & U. Sauerland, eds. *Interfaces + Recursion = Language?* Gruyter, Berlin, pp. 1-30.

.

Conroy, A., Lids, J., Takahashi, E. & Phillips, C. (2009). Equal Treatment for All Antecedents: How Children Succeed with Principle B. *Linguistic Inquiry*. 40, 3.

Conwell, E. & Demuth, K. (2007). Early syntactic productivity: Evidence from dative shift. *Cognition*, 103, 2, 163-179.

Crain, S. & Pietroski, P. (2002). Why language acquisition is a snap. *Linguistic Review*, 19, 163-183.

Crain, S. and Thornton, R. (1998). Investigations in Universal Grammar. A guide to experiments on the acquisition of syntax and semantics. MIT Press.

Culicover, P. & Jackendoff, R. (2005). *Simpler Syntax*. Oxford: Oxford University Press.

de Villiers, J., Roeper, T. and Vainikka. 1990. The acquisition of long-distances rules. In *Language Processing and Language Acquisition*, L. Frazier and J. De Villiers (eds), 257-297. Kluwer Academic Publishers.

deVilliers, J. and P. deVilliers and T. Roeper (2007) What's in a Phase. Special issue on SLI. *Lingua.*

deVilliers, J. and T. Roeper (to appear) *Handbook of Generative Approaches to Acquisition.* Kluwer.

Dresher, E. (1999). Charting the Learning Path: Cues to Parameter Setting. *Linguistic Inquiry*, 30:27-67.

Dresher, E., & Kaye, J. (1990). A Computational Learning Model for Metrical Phonology. *Cognition*, 34:137-195.

Drozd, K.F. (2001). Children's weak interpretation of universally quantified sentences". In Bowerman, M. and S. Levinson (eds.) *Language Acquisition and Conceptual Development*, p. 340-376. Cambridge: Cambridge University Press.

Elbourne, P. (2005). On the acquisition of Principle B. *Linguistic Inquiry* 36: 333-365.

Fitzpatrick, J. (2005). On Minimalist Approaches to the Locality of Movement. *Linguistic Inquiry*, 33, 443-463.

Fodor, J. D. (1998). Unambiguous Triggers. *Linguistic Inquiry*, 29:1-36.

Fodor, J.D. & Crain, S. (1987). Simplicity and generality of rules in language acquisition. In MacWhinney, B. (Ed.) *Mechanisms of Language Acquisition.* Hillsdale, NJ: Erlbaum. 35-63.

Fox, D. and D. Pesetsky (2004) Cyclic Linearization of Syntactic Structure. *Theoretical Linguistics* 31, 235-262.

Frank, R. (1998) Structural Complexity and the Time Course of Grammatical Development. *Cognition*.

Friedman , N. Adriana Belletti , L. Rizzi (2009). Relativized relatives: Types of intervention in the acquisition of A-bar dependencies. *Lingua.*

Gibson, E., & Wexler, K. (1994). Triggers. *Linguistic Inquiry*, 25:355-407.

Gomez, R. & Gerken, L. A. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences.* 4, 178-186.

Goodluck, H. (1978) Linguistic Principles in Children's Grammar of Complement Subject Interpretation. Doctoral dissertation. UMass Amherst.

Grillo, A. (2008). Generalized minimality. Syntactic underspecification in Broca's aphasia. Doctoral dissertation. U of Utrecht and University of Siena

Gropen, J., Pinker, S., Hollander, M., Goldberg, R. & Wilson, R. (1989). The Learnability and Acquisition of the Dative Alternation in English. *Language*, 65 (2), 203-257.

Guasavera and Thornton (2001) Possessor extraction in child English: A Minimalist account. Paper presented at the Penn Linguistics Colloquium.

Guttierrez, M. J. (2005) The acquisition of English LD wh-questions by Basque/Spanish bilingual subjects in a school context. PhD dissertation. University of the Basque Country.

Hale, K. & Keyser, S. J. (2002). *Prolegomenon to a Theory of Argument Structure*. Cambridge, MA: MIT Press.

Hamann, C. (to appear) Binding and Coreference: Views from Child Language. In deVilliers, J. and T. Roeper (Eds.) *Handbook of Generative Approaches to Language Acquisition.* Kluwer.

Hauser, M., Chomsky, N. & Fitch, T. (2002). The Faculty of Language: What Is It, Who Has It, and How Did It Evolve? *Science*, 298 (5598), 1569-1579.

Heck, F. (2004). A Theory of Pied Piping. Ph.D. dissertation. Universität Tubingen.

Hendriks, P. and J. Spenader (2006). When production precedes comprehension: an optimization approach to the acquisition of pronouns. *Language Acquisition*, p. 319-348.

Herrnstein, R. Loveland, D. (1975). Maximizing and Matching on Concurrent Ratio Schedules. *Journal of the Experimental Analysis of Behavior*, 24, 107-116.

Hollebrandse, B. and T. Roeper (1997) The concept of DO-insertion and the Theory of INFL in Acquisition. Amsterdam Child Language Series, eds. C. Koster and F. Wynen

Hollebrandse and Roeper (to appear) Recursion and Acquisition. Proceedings of Recursion Conference UMass. Amherst, Mass.

Hollebrandse, B. and T. Roeper 2007. Recursion and Propositional Exclusivity. Ms. Submitted to *The Linguistic Review*.

Hornstein, N. 2009. A Theory of Syntax: Minimal Operations and Universal Grammar. Cambridge University Press

Huang, C.T.J. 1982. Logical Relations in Chinese and the Theory of Grammar. Doctoral dissertation. MIT.

Huang, C.-H. J. (1984). On the Distribution and Reference of Empty Pronouns. *Linguistic Inquiry*, 15:531-574.

Hyams, N. (1986). Language Acquisition and the Theory of Parameters. Dordrecht: Reidel.

Hyams, N. (1991). A Reanalysis of Null Subjects in Child Language. In J. Weissenborn, H. Goodluck, & T. Roeper (eds.) *Theoretical Issues in Language Acquisition : Continuity and Change in Development.* Hillsdale, NJ: L. Erlbaum Associates, 249-267.

Hyams, N., & Wexler, K. (1993). On the Grammatical Basis of Null Subjects in Child Language. *Linguistic Inquiry*, 24:421-459.

Hyams,N. (to appear) Missing Subjects in Early Child Language. In deVilliers, J. and T. Roeper. *Handbook of Generative Approaches to Language Acquisition*. Kluwer.

Jelinek, F. (1998). *Statistical models of speech recognition.* Cambridge, MA: MIT Press.

Jelinek, E. & Carnie, A. (2003). Argument Hierarchies and the Mapping Principle. In Carnie, A., Harley, H. & Willie, M. (Eds.) *For mal Approaches to Function*. Philadelphia: John Benjamins. 265-296.

Jung, Y.-J. & Miyagawa, S. (2004). Decomposing Ditransitive Verbs. In *Proceedings of SICGG*. 101-120.

Kayne, R. (2005). On Parameters and on Principles of Pronunciation. Ms., New York University.

Labov, W. (1994). Principles of language change : Internal factors. Oxford : Blackwell.

Lebeaux, D. (2000). *Language acquisition and the form of grammar*. Amsterdam: John Benjamins.

Lebeaux, D. (2009) *Where does Binding Theory Apply*. Cambridge, MA: MIT Press.

Legate, J. A. & Yang, C. (2007). Morphosyntactic learning and the development of tense. *Language Acquisition,* 14, 315-344.

Lidz, J., and J. Musolino (2002). Children's command of quantification. *Cognition* 84: 113-154

Lightfoot, D. (1999). *The Development of Language*. Basil: Blackwell.

McDaniel, D. (1989). Partial and Multiple Wh-Movement. *Natural Language and Linguistic Theory,* 7, 565-604

Moro, A. (2000) *Dynamic Asymmetry. Cambridge*, MA: MIT Press.

Oiry, M. (2008) L'acqujisition des Questions a Longue-distance en Francais. Doctoral dissertation. Univeriste de Nantes

Oiry, M. and H. Demirdache (2006) Evidence from L1 acquisition for the Syntax of wh-scope marking in French The Romance Turn

Otsu, (1981): reference missing.

Newmeyer, F. (2004). Against a parameter-setting approach to language variation. *Language Variation Yearbook,* 4, 181-234.

Nowak, M., Komarova, N., & Niyogi, P. (2002). Computational and evolutionary aspects of language. *Nature,* 417, 611-617.

Perez-Leroux, A. and X. Li (2001) Selectivity in the Acquisition of Complex

NP Islands in E. Klein and G. Martohardjono The Development of Second Language Grammars Benjamins

Pesetsky, D. (1995). *Zero Syntax: Experiencer and Cascade*. Cambridge, MA: MIT Press.

Philip, W. (1995). Event quantification in the acquisition of

universal quantification. Doctoral dissertation. University of Massachusetts, Amherst

Philip, W. (2004). Two theories of exhaustive pairing. Ms. Utrecht Institute of Linguistics, OTS. Utrecht University, The Netherlands.

Pierce, A. (1992). Language Acquisition and Syntactic Theor y. Boston, MA: Kluwer.

Pinker, S. (1984). *Language Learnability and Language Development*. Cambridge, MA: Harvard University Press.

Pinker, S. (1989). *Language Learnability and Cognition*. Cambridge, MA: MIT Press.

Poeppel, D., & Wexler, K. (1993). The Full Competence Hypothesis. *Language*, 69:1-33.

Potts, C. and T. Roeper (2006) "The Narrowing Acquisition Path: From

Expressive small clauses to Declaratives in eds. Ljiljana Progovac, Kate Paesani, Eugenia Casielles and Ellen Barton The Syntax of Nonsententials Benjamins

Rappaport Hovav, M. & Levin, B. (2008). The English dative alternation: The case for verb sensitivity. *Journal of Linguistics*, 44, 129-167.

Rizzi, L. (1986). Null Object in Italian and the Theory of Pro. *Linguistic Inquiry*, 17:501-557.

Rizzi, L. (2004). Grammatically based target inconsistencies in child language. Manuscript, University of Siena.

Roeper, T. (1999) "Universal Bilingualism" Bilingualism: Language and Cognition 2 169-186.

Roeper, T. (2006) “Watching NP’s Grow” in V. van Geenhoven

Semantics Meets Acquisition Kluwer (2006)

Roeper, T. 2007. *The Prism of Grammar*. Cambridge, MA: MIT Press.

Roeper, T. (2009a) “Microscopic Minimalism” BU Plenary

lecture Proceedings BU 33, Cascadilla Press

Roeper (2009b) “Vacate Phase” UMass ms in draft

Roeper, T. (forthcoming) Strict Interfaces and Three Kinds of Multiple Grammar

Roeper, T. E. Ramos, H. Seymour, L. Abdul-karim (2003). Language Disorders as a Window on Universal Grammar: An Abstract theory of Agreement for IP,DP and V-PP

Saffran, J.R., Aslin, R.N., & Newport, E.L. (1996). Statistical learning by 8-month old infants. *Science*, 274, 1926-1928.

Sakas, W. & Fodor, J. D. (2001). The structural triggers learner. In Bertolo, S. (Ed.) *Language Acquisition and Learnability*. Cambridge: Cambridge University Press. 172-233.

Sakas, W. & Fodor, J. D. (2009). Disambiguating syntactic triggers. Paper presented at Workshop on input and syntactic acquisition. University of California, Irvine.

Schulz, B. (2005) “*Wh*-scope marking in English interlanguage

grammars: Transfer and processing effects on the second language acquisition of complex *wh*-questions. Unpublished doctoral Dissertation, University of Hawai‘i.

Snyder, W. (2001). On the nature of syntactic variation: Evidence from complex predicates and complex word-formation. *Language*, 77, 324-342.

Snyder, W. & Stromswold, K. (1997). The Structure and Acquisition of English Dative Constructions. *Linguistic Inquiry*, 28 (2), 281-317.

Solan, L. (1983). Pronominal Reference: Child Language and the Theory of Reference. Dordrecht: Kluwer

Straus, K. (2008). Validations of a probabilistic model of language learning. Ph.D. dissertation, Department of Mathematics, Northeastern University.

Strik, N. (2009) Derivational Contrasts in French and Dutch Evidence from the acquisition of wh-movement. Linguistics in the Netherlands (2009)

Sugisaki, K. and Y. Otsu (to appear) Universal Grammar and the Acquisition of Japanese Syntax. In deVilliers, J. and T. Roeper (Eds.) *Handbook of Generative Approaches to Language Acquisition.*

Tauberer, J.. 2008. Learning in the face of infidelity: Evaluating the robust interpretive parsing/constraint demotion model of optimality theory language acquisition. Master's thesis, University of Pennsylvania.

Tesar, B. & Smolensky, P. (1998). Learnability in optimality theory. *Linguistic Inquiry*, 29, 229-268.

Thornton, R. 1990. Adventures in Long-Distance Moving: the acquisition of complex wh- questions. PhD dissertation. University of Connecticut.

Tomasello, M. (2000). The item-based nature of children's early syntactic development. *Trends in Cognitive Sciences*, 4, 156-163.

Vainnikka, A. and M. Young-Scholten (in preparation), The universal bare VP and its theoretical ramifications.

Verbuk, A. and T. Roeper (to appear). "How pragmatics and syntax make Principle B acquirable" Language Acquisition

Vaininikka, A. and Young-scholten (1996) Gradual Development of L2 Phrase Structure Second Language Research 12.

Valian, V. (1991). Syntactic Subjects in the Early Speech of American and Italian Children. *Cognition*, 40:21-82.

Valiant, L. G. (1984). A theory of the learnable. *Communications of the ACM.* 27, 1134–1142.

Vapnik, V. (1995). *The nature of statistical learning theory*. Berlin : Springer-Verlag.

Wang, Q., Lillo-Martin, D., Best, C., & Levitt, A. (1992). Null Subject vs. Null Object: Some Evidence from the Acquisition of Chinese and English. *Language Acquisition*, 2:221-54.

Weissenborn, J. and T. Roeper and J. deVilliers (1991) The Acquisition of WH-Movement in German and French in T. Maxfield and B. Plunkett Umass Occasional Papers in Linguistics

Westergaard, M. (2009). Word Order in Old and Middle English: The role of information structure and first language acquisition. Benjamins

Wexler,K. & Culicover, P. (1980) *Formal Principles of Language Acquisition*. Cambridge, MA: MIT Press

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*. New York: Oxford University Press.

Yang, C. (2004). Universal Grammar, Statistics, or Both. *Trends in Cognitive Sciences*, 8 (10), 451-456.

Yang, C. (2005). On productivity. *Language Variation Yearbook*, 5, 333-370.

Yang, C. (2008). The great number crunch. *Journal of Linguistics*, 44, 205-228.

Yang, C. (2009a). Three factors in language variation. *Lingua.*

Yang, C. (2009b). Who's afraid of George Kingsley Zipf? Ms. University of Pennsylvania.

Yusa (2001) Multiple Specifiers and Wh-Island Effects in L2 Acqusiition in E. Klein and G. Martohardjono The Development of Second Language Grammars Benjamins

Zipf, G. (1949). Human behavior and the principle of least effort. New York: Addison-Wiley.