# Child language acquisition: Why universal grammar doesn't help

Ben Ambridge, Julian M. Pine, Elena V. M. Lieven


Language, Volume 90, Number 3, September 2014, pp. e53-e90 (Article)

Published by Linguistic Society of America
*DOI: 10.1353/lan.2014.0051*

PERSPECTIVES

# Child language acquisition: Why universal grammar doesn't help

BEN AMBRIDGE
*University of Liverpool*

JULIAN M. PINE
*University of Liverpool*

ELENA V. M. LIEVEN
*University of Manchester*

In many different domains of language acquisition, there exists an apparent learnability problem to which innate knowledge of some aspect of UNIVERSAL GRAMMAR (UG) has been proposed as a solution. The present article reviews these proposals in the core domains of (i) identifying syntactic categories, (ii) acquiring basic morphosyntax, (iii) structure dependence, (iv) subjacency, and (v) the binding principles. We conclude that, in each of these domains, the innate UG-specified knowledge posited does not, in fact, simplify the task facing the learner.

*Keywords*: binding principles, child language acquisition, frequent frames, parameter setting, prosodic bootstrapping, semantic bootstrapping, structure dependence, subjacency, syntax, morphosyntax, universal grammar

**1.** INTRODUCTION. Many leading theories of child language acquisition assume innate knowledge of universal grammar (e.g. of syntactic categories such as NOUN and VERB, constraints/principles such as structure dependence and subjacency, and parameters such as the head-direction parameter). Many authors have argued either for or against UNIVERSAL GRAMMAR (UG) on a priori grounds such as learnability (e.g. whether the child can acquire a system of infinite productive capacity from exposure to a finite set of utterances generated by that system) or evolutionary plausibility (e.g. linguistic principles are too abstract to confer a reproductive advantage).

Our goal in this article is to take a step back from such arguments, and instead to consider the question of whether the individual components of innate UG knowledge proposed in the literature (e.g. a NOUN category, the binding principles) would help the language learner. We address this question by considering the main domains for which there exists an apparent learnability problem and where innate knowledge has been proposed as a critical part of the solution: identifying syntactic categories (§2), acquiring basic morphosyntax (§3), structure dependence (§4), subjacency (§5), and binding principles (§6). We should emphasize that the goal of this article is not to contrast UG accounts with alternative constructivist or usage-based accounts of acquisition (for recent attempts to do so, see Saxton 2010, Ambridge & Lieven 2011). Rather, our reference point for each domain is the set of learning mechanisms that must be assumed by all accounts, whether generativist or constructivist. We then critically evaluate the claim that adding particular innate UG-specified constraints posited for that domain simplifies the task facing the learner.

Before we begin, it is important to clarify what we mean by 'universal grammar', since the term is often used differently by different authors. We do not use the term in its most general sense, in which it means simply 'the ability to learn language'. The claim that humans possess universal grammar in this sense is trivially true, in the same way that humans could be said to possess universal mathematics or universal baseball (i.e. the ability to learn mathematics or baseball).

Similarly, we do not use the term 'universal grammar' to mean Hauser, Chomsky, and Fitch's (2002) faculty of language in either its broad sense (general learning mechanisms; the sensorimotor and conceptual systems) or its narrow sense (including only recursion). Nor do we use the term to mean something like a set of properties or design features shared by all languages. It is almost certainly the case that there ARE properties

that are shared by all languages. For example, all languages combine meaningless phonemes into meaningful words, instead of having a separate phoneme for each meaning (Hockett 1960), though there is much debate as to whether these constraints are linguistic or arise from cognitive and communicative limitations (e.g. Evans & Levinson 2009). Finally, while we acknowledge that most—probably all—accounts of language acquisition will invoke at least SOME language-related biases (e.g. the bias to attend to speech sounds and to attempt to discern their communicative function), we do not use the term UG to refer to an initial state that includes only this very general type of knowledge.

None of these definitions seem to capture the notion of UG as it is generally understood among researchers of child language acquisition. It is in this sense that we use the term 'universal grammar': a set of categories (e.g. NOUN, VERB), constraints/principles (e.g. structure dependence, subjacency, the binding principles), and parameters (e.g. head direction, V2) that are innate (i.e. that are genetically encoded and do not have to be learned or constructed through interaction with the environment). Our aim is not to evaluate any particular individual proposal for an exhaustive account of the contents of UG. Rather, we evaluate specific proposals for particular components of innate knowledge (e.g. a VERB category, the subjacency principle) that have been proposed to solve particular learnability problems, and leave for others the question of whether or how each could fit into an overarching theory of universal grammar. Many generativist-nativist theories assume that, given the underconstraining nature of the input, this type of innate knowledge is necessary for language learning to be possible. In this article, we evaluate the weaker claim that such innate knowledge is helpful for language learning. We conclude that, while the in-principle arguments for innate knowledge may seem compelling at first glance, careful consideration of the actual components of innate knowledge often attributed to children reveals that none simplify the task facing the learner.

Specifically, we identify three distinct problems faced by proposals that include a role for innate knowledge—LINKING, inadequate DATA COVERAGE, and REDUNDANCY—and argue that each component of innate knowledge that has been proposed suffers from at least one. Some components of innate knowledge (e.g. the major lexical syntactic categories and word-order parameters) would appear to be useful in principle. In practice, however, there is no successful proposal for how the learner can link this innate knowledge to the input language (the linking problem; e.g. Tomasello 2005). Other components of innate knowledge (e.g. most lexical syntactic categories, and rules linking the syntactic roles of SUBJECT and OBJECT to the semantic categories of Agent and Patient) yield inadequate data coverage: the knowledge proposed would lead to incorrect conclusions for certain languages and/or certain utterance types within a particular language. A third type of innate knowledge (e.g. subjacency, structure dependence, the binding principles) would mostly lead the learner to correct conclusions, but suffers from the problem of redundancy: learning procedures that must be assumed by all accounts—often to explain counterexamples or apparently unrelated phenomena—can explain learning, with no need for the innate principle or constraint. We argue that, given the problems of linking, data coverage, and redundancy, there exists no current proposal for a component of innate knowledge that would be useful to language learners.

Before we begin, it is important to ask whether are setting up a straw man. Certainly, our own—of course, subjective—impression of the state of the field is that UG-based accounts (as defined above) do not enjoy broad consensus or even necessarily represent the dominant position. Nevertheless, it is undeniably the case that many mainstream

child language acquisition researchers are currently publishing papers that argue explicitly for innate knowledge of one or more of the specific components of UG listed above. For example, in a review article on *Syntax acquisition* for a prestigious interdisciplinary cognitive science journal, Crain and Thornton (2012) argue for innate knowledge of structure dependence and the binding principles. Valian, Solt, and Stewart (2009) recently published a study designed to provide evidence for innate syntactic categories (see also Yang 2009). Lidz and colleagues (Lidz & Musolino 2002, Lidz, Gleitman, & Gleitman 2003, Lidz, Waxman, & Freedman 2003, Lidz & Gleitman 2004, Viau & Lidz 2011) have published several articles—all in mainstream interdisciplinary cognitive science journals—arguing for UG-knowledge of syntax. Virginia Valian, Thomas Roeper, Kenneth Wexler, and William Snyder have all given plenary addresses emphasizing the importance of UG at recent meetings of the leading annual conference in the field (the Boston University Conference on Language Development); indeed, there are entire conferences devoted to UG approaches to language acquisition (e.g. GALANA). The UG hypothesis is defended in both recent child language textbooks (Guasti 2004, Lust 2006) and books for the general reader (e.g. Yang 2006, Roeper 2007). This is to say nothing of the many studies that incorporate certain elements of UG (e.g. abstract syntactic categories, an abstract TENSE category) as background assumptions (e.g. Rispoli et al. 2009), rather than as components of a hypothesis to be tested as part of the study. Many further UG-based proposals are introduced throughout the present article. In short, while controversial, UG—in the sense that we use the term here—is a current, live hypothesis.

**2.** IDENTIFYING SYNTACTIC CATEGORIES. One of the most basic tasks facing the learner is that of grouping the words that are encountered into syntactic categories (by which we mean lexical categories such as NOUN, VERB, and ADJECTIVE; syntactic roles such as subject and object will be discussed in the section on acquiring basic word order). This is a very difficult problem because the definitions of these categories are circular. That is, the categories are defined in terms of the system in which they participate. For example, arguably the only diagnostic test for whether a particular word (e.g. *situation*, *happiness*, *party*) is a NOUN is whether it occurs in a similar set of syntactic contexts to other NOUNS such as *book* (e.g. after a determiner and before a main or auxiliary verb, as in *the __ is*). Given this circularity, it is unclear how the process of category formation can get off the ground.

The traditional solution has been to posit that these syntactic categories are not formed on the basis of the input, but are present as part of UG (e.g. Chomsky 1965, Pinker 1984, Valian 1986). The advantage of this proposal is that it avoids the problem of circularity, by providing a potential way to break into the system. If children know in advance that there will be a class of (for example) NOUNS and are somehow able to assign just a few words to this category, they can then add new words to the category on the basis of semantic and/or distributional similarity to existing members. The question is how children break into these syntactic categories to begin with. This section considers three approaches: DISTRIBUTIONAL ANALYSIS, PROSODIC BOOTSTRAPPING, and SEMANTIC BOOTSTRAPPING.

**2.1.** DISTRIBUTIONAL ANALYSIS. In the adult grammar, syntactic categories are defined distributionally. Thus it is almost inevitable that accounts of syntactic category acquisition—even those that assume innate categories—must include at least some role for distributional analysis (the prosodic bootstrapping account, discussed below, is a possible exception). For example, as Yang (2008:206) notes, '[Chomsky's] LSLT [*Log-*

*ical structure of linguistic theory*] program explicitly advocates a probabilistic approach to words and categories "through the analysis of clustering … the distribution of a word as the set of contexts of the corpus in which it occurs, and the distributional distance between two words" (LSLT: section 34.5)'. Pinker (1984:59) argues that 'there is good reason to believe that children from 1½ to 6 years can use the syntactic distribution of a newly heard word to induce its linguistic properties' (although famously arguing against deterministic distributional analysis elsewhere, e.g. Pinker 1979:240). Similarly, Mintz (2003:112), while assuming a 'pre-given set of syntactic category labels', advocates, and provides evidence for, one particular form of distributional analysis (frequent frames). Finally, arguing for an account under which 'the child begins with an abstract specification of syntactic categories', Valian, Solt, and Stewart (2009:744) suggest that 'the child uses a type of pattern learning based on distributional regularities … in the speech she hears'.

Thus the claim that learners use distributional learning to form clusters that correspond roughly to syntactic categories (and/or subcategories thereof ) is relatively uncontroversial (for computational implementations, see e.g. Cartwright & Brent 1997, Redington et al. 1998, Clark 2000, Mintz 2003, Freudenthal et al. 2005, Parisien et al. 2008; see Christodoulopoulos et al. 2010 for a review). The question is whether, having formed these distributional clusters, learners would be helped by the provision of innate prespecified categories to which they could be linked (e.g. Mintz 2003). We argue that this is not the case, and that a better strategy for learners is simply to use the distributionally defined clusters directly (e.g. Freudenthal et al. 2005).

Although, as we have seen above, many accounts that assume innate syntactic categories also assume a role for distributional learning, few include any mechanism for linking the two. Indeed, we are aware of only two such proposals. Mintz (2003) suggests that children could assign the label NOUN to the category that contains words for concrete objects, using an innate linking rule. The label VERB would then be assigned either to the next largest category or, if this does not turn out to be crosslinguistically viable, to the category that takes NOUNs as arguments (for which a rudimentary, underspecified outline of the sentence's argument structure would be sufficient). Similarly, Pinker's (1984) semantic bootstrapping account (subsequently discussed more fully in relation to children's acquisition of syntactic roles such as subject and object) assumes innate rules linking 'name of person or thing' to NOUN, 'action or change of state' to VERB, and 'attribute' to ADJECTIVE (p. 41). Once the child has used these linking rules to break into the system, distributional analysis largely takes over. This allows children to assimilate nonactional verbs and nouns that do not denote the name of a person/thing (as in Pinker's example, *The situation justified the measures*) into the VERB and NOUN categories on the basis of their distributional overlap with more prototypical members.

A problem facing both Mintz's (2003) and Pinker's (1984) proposals is that they include no mechanisms for linking distributionally defined clusters to the other innate categories that are generally assumed as a necessary part of UG, such as DETERMINER, WH-WORD, AUXILIARY, and PRONOUN. Pinker (1984:100), in effect, argues that these categories will be formed using distributional analysis, but offers no proposal for how they are linked up to their innate labels. Thus it is only for the categories of NOUN, VERB, and (for Pinker) ADJECTIVE that these proposals offer any account of linking at all. This is not meant as a criticism of these accounts, which do not claim to be exhaustive and—indeed—are to be commended as the only concrete proposals that attempt to link distributional and syntactic categories at all. The problem is that, despite the fact that virtually all UG accounts assume innate knowledge of a wide range of categories,

there exist no proposals at all for how instances of these categories can be recognized in the input—an example of the linking problem.

In fact, this is not surprising, given the widespread agreement among typologists that—other than a NOUN category containing at least names and concrete objects—there are no viable candidates for crosslinguistic syntactic categories (e.g. Nida 1949, Lazard 1992, Dryer 1997, Croft 2001, 2003, Haspelmath 2007, Evans & Levinson 2009). For example, Mandarin Chinese has property words that are similar to adjectives in some respects, and verbs in others (e.g. McCawley 1992, Dixon 2004). Similarly, Haspelmath (2007) characterizes Japanese as having two distinct adjective-like parts of speech, one a little more noun-like, the other a little more verb-like. Indeed, even the NOUN/VERB distinction has been disputed for languages such as Salish (Kinkade 1983, Jelinek & Demers 1994), Samoan (Rijkhoff 2003), and Makah (Jacobsen 1979, Croft 2001), in which (English) verbs, nouns, adjectives, and adverbs may all be inflected for person/aspect/mood (usually taken as a diagnostic for verbs in Indo-European languages). Such considerations led Maratsos (1990:1351) to conclude that the only candidate for a universal lexical category distinction 'is between "noun and Other" ', reflecting a distinction between things/concepts and properties/actions predicated of them.

Pinker (1984:43) recognizes the problem of the nonuniversality of syntactic categories, but argues that it is not fatal for his theory, provided that different crosslinguistic instances of the same category share at least a 'family resemblance structure'. Certainly an innate rule linking 'name of person or thing' to NOUN (Pinker 1984:41) would probably run into little difficulty crosslinguistically. It is less clear whether the same can be said for the rules linking 'action or change of state' to VERB and 'attribute' to ADJECTIVE. But even if these three linking rules were to operate perfectly for all languages, crosslinguistic variation means that it is almost certainly impossible in principle to build in innate rules for identifying other commonly assumed UG categories, whether these rules make use of semantics, distribution, or some combination of the two (the problem of data coverage).

In summary, Pinker's (1984) and Mintz's (2003) proposals are useful in that they capture the insight that, in order to form syntactic categories, learners will have to make use of both semantic and distributional information. Where they falter is in their assumption that these distributional clusters must be linked to innate syntactic categories. The reason for the failure of UG accounts to propose mechanisms by which distributional clusters can be linked to innate universal syntactic categories other than NOUN is that (with the possible exception of VERB/ADJECTIVE) there ARE no good candidates for innate universal syntactic categories other than NOUN. Given that syntactic categories are language-specific, there is no alternative but for children to acquire them on the basis of semantic and distributional regularities. Indeed, even categories as (relatively) uncontroversial as English NOUN and VERB are made up of semantically and distributionally coherent subcategories such as PROPER VS. COUNT VS. MASS and INTRANSITIVE VS. MONOTRANSITIVE VS. DITRANSITIVE. Thus, even if a learner could instantaneously assign every NOUN or VERB that is heard into the relevant category, this would not obviate the need for a considerable degree of clustering based on semantic and distributional similarity. Given that such clustering yields useful syntactic categories, innate categories are redundant.

We end this section by addressing two possible objections to the claim that distributional analysis can obviate the need for innate syntactic categories. The first is that the notion of 'distributional analysis' as discussed here is ill-defined. For example, it is sometimes asked how the child knows in advance that distributional analysis must take

place at the level of the word, as opposed to the phone, phoneme, syllable, *n*-syllable sequence, and so on. The answer is that the child does not know. In fact, she will have to conduct distributional analysis at many of these levels simultaneously to solve other problems such as speech segmentation, constructing an inventory of phonemes, and learning the phonotactic constraints and stress patterns of her language. As a result of this many-layered distributional analysis, it will be noted that units of a certain size—'words'—occur more often than would be expected if speakers produced random sequences of phones (and, crucially, cooccur with concrete or functional referents in the world (e.g. 'cat', 'past-ness'). It will be further noted that these units share certain distributional regularities with respect to one another, the type of distributional analysis required for syntactic-class formation. There is no need to build in innate constraints to rule out every theoretically possible distributional-learning strategy: let the child try to perform distributional analysis based on, for example, three-syllable strings. The child will learn after a handful of exposures that these units are neither distributionally nor semantically/functionally coherent. Of course, it might turn out to be necessary to assume general constraints such as 'pay particular attention to sounds made by humans' or 'note correlations between speakers' sounds and their probable intentions', but these are not the types of constraints posited by typical UG accounts, almost all of which assume innate syntactic categories.

Note that even if one rejects these arguments entirely, the question of how the child knows to perform distributional analysis at the word level, as opposed to some other level, is equally problematic for accounts that do and do not posit innate syntactic categories, given that accounts of the former type still require word-level distributional analysis in order to assign words to the prespecified categories. This point relates to the second possible objection: that none of the distributional-analysis algorithms outlined above are unequivocally successful in grouping words into categories. While this is true, it is no argument for innate syntactic categories, as—again—accounts that posit such categories still require distributional analysis working at the single-word level (as explicitly advocated by Chomsky; see Yang 2008) in order to identify instances of these categories. Finally, note that tacit in the argument that 'distributional categories don't work' is the assumption that the categories commonly assumed by UG theories do work, an assumption that—with the possible exception of NOUN—enjoys little support crosslinguistically.

**2.2.** PROSODIC BOOTSTRAPPING. The prosodic bootstrapping hypothesis (e.g. Christophe et al. 2008) differs from the proposals above in that it does not assume that learners initially use either semantics or distributional clustering to break into the syntactic category system. Rather, children use prosodic information to split clauses into syntactic phrases (e.g. [*The boy*] [*is running*]).[1] For example, the end of a phrase is often signaled by final syllable lengthening, a falling pitch contour, and/or a short pause. Having split the clause into syntactic phrases, the child then uses 'flags' to label each phrase, and hence to assign the items to the relevant categories. For example, in this case, the child uses determiner *the* and auxiliary *is* to label the phrases as NOUN PHRASE and VERB PHRASE respectively, and hence to assign *boy* to the NOUN class and *running* to the VERB class. The advantage of the prosodic bootstrapping account is that, by using nondistri-

[1] Although these authors do not use the term 'universal grammar', some innate basis is clearly assumed. For example, Christophe, Mehler, and Sebastián-Gallés (2001:385–86) argue that the speech stream is 'SPONTANEOUSLY PERCEIVED as a string of prosodic units, roughly corresponding to phonological phrases', boundaries of which 'often coincide with boundaries of syntactic constituents' (emphasis added).

butional (i.e. prosodic) information to break into the distributionally defined system, it avoids both circularity and the problem of linking distributional clusters to UG-specified categories. Furthermore, there is evidence to suggest that even six-month-old infants are sensitive to the relevant prosodic properties. Using a conditioned-head-turn paradigm, Soderstrom and colleagues (2003) showed that infants could discriminate between two strings that were identical in terms of their phonemes, but only one of which contained an NP/VP boundary, marked by final-syllable lengthening and pitch drop.

(1) a. NO PHRASE BOUNDARY: At the discount store **new watches for men** are simple …
b. NP/VP BOUNDARY: In the field, the old frightened **gnu//watches for men** and women …

One problem facing this account is that, even looking only at the case of the NP/VP boundary in a single language (i.e. English), such a strategy would probably lead to incorrect segmentation in the majority of cases. For sentences with unstressed pronoun subjects (e.g. *He kissed the dog*) as opposed to full NPs (e.g. *The boy kissed the dog*), prosodic cues place the NP/VP boundary in the wrong place (e.g. *[$_{NP}$ *He kissed*] [$_{VP}$ *the dog*]; Gerken et al. 1994, Nespor & Vogel 1986). In an analysis of spontaneous speech to a child aged 1;0, Fisher and Tokura (1996) found that 84% of sentences were of this type. Of course, we have no idea how reliable a cue must be for it to be useful (almost certainly less than 100%). Nevertheless, it would seem difficult to argue that a cue that is not simply uninformative but actively leads to incorrect segmentation in the vast majority of cases is anything other than harmful.

The problem of the nonexistence of universal syntactic categories also clearly constitutes a problem for the Christophe et al. 2008 approach. But even if it were somehow possible to come up with a list of universal categories (as well as reliable prosodic cues to phrase boundaries), the proposal would still fail unless it were possible to identify a 'flag' for every category in every language. The outlook does not look promising, given that the possible flags proposed by Christophe and colleagues (2008) for the English NOUN and VERB categories—DETERMINER and AUXILIARY—are by no means universal. Yet even with a universal list of syntactic categories and flags to each one, children would still need an additional mechanism for recognizing concrete instances of these flags (e.g. children hear *the* and *is*, not DETERMINER and AUXILIARY). Given that there exists no proposal for a universal set of flags, the Christophe et al. 2008 account suffers from the linking problem. It also suffers from an additional problem that is common to many UG approaches. While the proposal, at its core, proposes one or two critical elements of innate knowledge (here, knowledge of prosodic cues to phrase boundaries), it requires a cascade of further assumptions that are rarely made explicit (here, observable flags for every category for every language) before it can be said to provide a potentially workable solution (e.g. Tomasello 2003, 2005).

**2.3.** INTERIM CONCLUSION. In conclusion, our goal is not to argue for an alternative account of syntactic category acquisition. Indeed, the proposals outlined here seem to us to be largely along the right lines. Learners will acquire whatever syntactic categories are present in the particular language they are learning, making use of both distributional (e.g. Mintz 2003) and semantic similarities[2] (e.g. Pinker 1984) between category mem-

[2] It would seem likely that learners make use not only of semantic but also of functional similarity between items (e.g. Tomasello 2003). For example, although most abstract nouns (e.g. *situation*) share no semantic similarity with concrete nouns (e.g. *man*), they share a degree of functional similarity in that actions/events

bers. Indeed, although there is only weak evidence for prosodic/phonological cues to category membership in English, there would seem to be no reason to doubt that, if particular languages turn out to contain such cues, then learners will use them. Where these theories falter is in their attempt to squeeze fine-grained language-specific categories, defined by distribution and semantics (and possibly also function and prosody), into a rigid framework of putative innate universal categories, derived primarily from the study of Indo-European languages. Even if these crosslinguistic categories were useful, there are essentially no proposals for how children could identify instances of them, other than by using distributional and semantics-based learning, a procedure that yields the target categories in any case. Consequently, nativist proposals for syntactic category acquisition suffer from problems of data coverage, linking, and redundancy.

**3.** ACQUIRING BASIC MORPHOSYNTAX.[3] Another task facing children is to learn how their language marks 'who did what to whom' in basic declarative sentences. For syntactic word-order languages such as English, this involves learning the correct ordering of subject, verb, and object. For other languages, this involves learning how these categories (or the equivalent) are indicated by means of morphological noun and/or verb marking. The problem is a difficult one because the notions of subject, verb, and object are highly abstract. For example, while learners of English could parse simple sentences such as *The dog bit the cat* using a basic semantic Agent-Action-Patient schema, this will not work for nonactional sentences such as *The situation justified the measures*, or sentences where the subject is more patient-like than agentive (e.g. *He received a slap* [ *from Sue*]; examples from Pinker 1984). Note also that in these nonagentive examples the subject still receives subject as opposed to object case marking (i.e. nominative *he*, not accusative *him*). This means that, just like syntactic categories such as NOUN and VERB, syntactic roles such as subject and object cannot be defined in terms of semantics, and are defined instead in terms of their place within the grammatical system of which they form a part. The only way to determine whether a particular NP is a subject is to determine whether it displays the constellation of properties displayed by other subjects (e.g. bearing nominative case, appearing first in canonical declaratives, etc.). Consequently, it has often been argued that syntactic roles (like lexical categories) are too abstract to be learned and must therefore be innately specified as part of UG. This assumption is shared by the semantic bootstrapping account and parameter-setting approaches, the latter of which additionally assume that the different word-order possibilities are, in effect, also known in advance as part of UG.

---

and properties can be predicated of both. We do not see this as a free-standing alternative account, but simply another property over which similarity-based clustering can operate. Another is the phonological properties of the word. For example, English bisyllabic nouns tend to have trochaic stress (e.g. *monkey*, *tractor*), and verbs, iambic stress (*undo*, *repeat*) (Cassidy & Kelly 2001, Christiansen & Monaghan 2006).

[3] A referee pointed out that this section addresses two distinct, though overlapping, questions in the domain of basic morphosyntax. The first (§3.1) is the question of how children learn the way in which the target language marks syntactic roles such as subject and object, whether via morphology, syntax (i.e. word order), or some combination of the two. The second (§3.2) is the question of how children acquire the order of (i) specifier and head and (ii) complement and head. In some cases, these questions overlap. For example, in word-order languages such as English, both relate to the ordering of the subject, verb, and object. In other cases, these questions are entirely distinct. For example, the ordering of specifier, head, and complementizer is both (i) irrelevant to syntactic-role marking for languages where this is accomplished entirely morphologically and (ii) relevant to phenomena other than syntactic-role marking (e.g. the ordering of the NOUN and DETERMINER within a DP/NP). Nevertheless, because both questions relate to basic morphosyntax and, in particular, because these parameters have been discussed most extensively with regard to syntactic word order (e.g. SVO vs. SOV), we feel justified in including these two separate subsections within the same overarching section.

**3.1.** SEMANTIC BOOTSTRAPPING. Pinker's (1984) semantic bootstrapping account assumes that UG contains not only syntactic roles (e.g. subject, verb, and object), but also innate rules that link each to a particular semantic role (e.g. Agent → subject, verb → Action, Patient → object).[4] Assume, for example, that the child hears an utterance such as *The dog bit the cat* and is able to infer (for example, by observing an ongoing scene) that *the dog* is the Agent (the biter), *bit* the Action, and *the cat* the Patient (the one bitten). By observing in this way that English uses Agent-Action-Patient order, and using the innate rules linking these semantic categories to syntactic roles, the child will discover (in principle from a single exposure) that English uses subject-verb-object word order. As noted in the previous section, innate rules also link 'names for people or objects' (here *dog* and *cat*) to an innate NOUN category.

An important, but often overlooked, aspect of Pinker's (1984) proposal is that once basic word order has been acquired in this way, the linking rules are abandoned in favor of (i) the recently acquired word-order rules and (ii) distributional analysis. Thus the child will be able to parse a subsequent sentence that does not conform to these linking rules, for example, *The situation justified the measures*, by using (i) the subject-verb-object rules inferred on the basis of *The cat bit the dog* and (ii) distributional similarity (e.g. if ***the*** *cat* is an NP and *cat* a NOUN, then ***the*** *situation* must also be an NP and *situation* a NOUN).

The advantage of Pinker's (1984) account is that it avoids the problems inherent in the circularity of syntactic roles by using nonsyntactic (i.e. semantic) information to break into the system. Since this semantic information is used only as a bootstrap and then discarded, sentences that do not conform to the necessary pattern (e.g. *He received a slap from Sue*; *The situation justified the measures*) do not present a problem. Although questions, passives, and other non-Agent-Action-Patient sentences would yield incorrect word-order rules (e.g. Pinker (1984:61) discusses the example of *You will get a spanking off me* yielding OVS), the suggestion is that learning is probabilistic and hence that occasional sentences of this type do not disrupt learning of the canonical pattern (Pinker 1987).

One basic problem facing Pinker's proposal is that it is unclear how the child can identify which elements of the utterance ARE the semantic arguments of the verb (Agent and Patient), and hence are available for linking to subject[5] and object, given the way that the particular target language carves up the perceptual world (Bowerman 1990). Consider, for example, the English sentence *John hit the table with a stick*. The Agent (*John*) links to subject and the Patient (*the table*) to object. As an Instrument, *the stick* links to OBLIQUE OBJECT. For English, noncanonical variations of such sentences (e.g. *John hit the stick against the table*) are presumably sufficiently rare to be disregarded. For some languages, however, the equivalent is the canonical form. Thus learners of, for example, Chechen-Ingush could perform the correct linking only if they parsed the same scene such that *the stick* (as opposed to *the table*) is the Patient, and hence links to object (the table links to oblique object).

[4] Pinker actually posits a hierarchy of linking rules (e.g. Pinker 1989:74), but since the first pass involves linking Agent and Patient to subject and object, the facts as they relate to the discussion here are unchanged.

[5] We note in passing that, exactly as for lexical categories such as NOUN and VERB, the existence of a universal crosslinguistic SUBJECT category is disputed by many typologists (e.g. Schachter 1976, Dryer 1997, Van Valin & LaPolla 1997, Croft 2001, 2003, Haspelmath 2007, but see Keenan 1976).

(2) a. English, *Chechen-Ingush: John = SUBJ, hit = VERB, the table = OBJ, stick = OBL
b. *English, Chechun-Ingush: John = SUBJ, hit = VERB, the table = OBL, stick = OBJ

It is important to emphasize that this problem is more fundamental than the problem that some languages do not map Agent and Patient onto subject and object in the same way as English (see below). The problem raised by Bowerman (1990) is that some languages do not map what English conceptualizes as Patients onto EITHER subject or object position (but rather to oblique object): a version of the linking problem.

It has been argued (e.g. by Pye 1990) that the existence of morphologically ergative(-absolutive) languages (e.g. Dyirbal) constitutes a problem for Pinker's (1984) proposal, as such languages do not map semantic roles onto syntactic roles in the same way as (nominative-)accusative languages such as English (and the majority of Indo-European languages). Languages differ in the way that they map the following semantic roles onto the morphological case-marking system.

(3) A = the Agent of a transitive verb (***The man*** *kissed the woman.*)
P[6] = the Patient of a transitive verb (*The woman kissed* ***the man***.)
S = the Single argument of an intransitive verb (***The man*** *danced.*)

Accusative languages (e.g. English) use one type of case marking (NOMinative) for A and S, and a different type of case marking (ACCusative) for P. This can be seen in English, which marks case on pronouns only, by substituting pronouns for *the man* in the sentences above: A = ***He.NOM*** *kissed the woman*, S = ***He.NOM*** *danced*, but P = *The woman kissed* ***him.ACC***. Ergative languages (remember that, for the moment, this discussion is restricted to MORPHOLOGICAL ergativity) use one type of case marking (ergative) for A and another (absolutive) for P and S.

Van Valin (1992), Siegel (2000), and Tomasello (2005) argue that particularly problematic for semantic bootstrapping are SPLIT-ERGATIVE languages, which use the nominative-accusative system in some contexts and the ergative-absolutive system in others. Languages may split according to tense (e.g. Jakaltek; Craig 1977), aspect (e.g. Hindi; Bhat 1991), an animacy hierarchy (e.g. Dyirbal; Dixon 1972), whether the morphological marking is realized on the noun or verb (e.g. Enga, Kaluli, Warlpiri, Georgian, Mparntwe Arrernte; Van Valin & LaPolla 1997), or even the particular lexical item being inflected (e.g. Tsova-Tush; Holisky 1987). Consequently, split-ergative languages have no mapping between semantic and syntactic categories that is consistent across the entire grammar.

So far we have discussed only morphological ergativity. Also argued to be problematic for semantic bootstrapping (e.g. Van Valin 1992) are languages that exhibit true SYNTACTIC ergativity (e.g. Dixon 1972, Woodbury 1977, Pye 1990). In such languages, the P role is the syntactic subject,[7] passing many traditional tests for subjecthood such as appearing in an oblique phrase in antipassives (in Dyirbal and K'iche') and being controlled by an NP in a matrix clause (in Dyirbal and Yup'ik Eskimo). The advantage of syntactic ergativity is that it allows morphologically ergative languages to maintain a

[6] Many authors use O (for Object) rather than P (for Patient). However, since the very phenomenon under discussion is that not all languages map the semantic Patient role onto the syntactic object role, this seems unnecessarily confusing.

[7] Marantz (1984) additionally proposed that the A role is the syntactic object, although such an analysis is not widely accepted.

consistent mapping between case marking and syntactic roles (similarly to nominative-accusative languages; see Pinker 1989:253). The disadvantage is that any innate rule linking Patient to object (as for English) would have to be overridden in a great many cases. One cannot solve this problem by, for example, having the learner set a parameter such that the transitive Patient links to subject rather than object: all syntactically ergative languages are split-ergative (Dixon 1994, Van Valin & LaPolla 1997:282–85), meaning that they employ nominative-accusative syntax in some parts of the system. Thus, as discussed above with regard to morphological split ergativity, linking rules must be learned on a construction-by-construction basis.

Nevertheless, the solution proposed by Pinker (1984) for noncanonical English sentences (e.g. *He received a slap off Sue*) can, in principle, be extended to deal with all types of ergativity. The solution (developed most fully in Pinker 1987) is to relegate innate linking rules to a probabilistic cue to syntactic roles that can be overruled by other competing factors, including—explicitly—distributional learning (e.g. Pinker 1987: 430, 1989:253).

While this solution potentially achieves better data coverage, it does so at the expense of redundancy, by effectively obviating the need for any innate learning mechanism (Braine 1992). This is perhaps best illustrated by split ergativity (the same problem holds for both the morphological and syntactic versions of this phenomenon). Since the mapping between semantic roles and morphological/syntactic marking changes depending on animacy, tense, aspect, and so on, there is no alternative but for children to LEARN the particular mapping that applies in each part of the system, using whatever probabilistic semantic or distributional regularities hold in that domain (e.g. animate agents are marked by a particular morpheme/word-order position, inanimate agents by another). The links between semantics and morphology/syntax that must be learned are not only complex and fine-grained but also context-dependent, varying from verb to verb, tense to tense, or human to animal. Thus, any particular set of innate linking rules would not only lead to the wrong solution in many cases, but would also be largely arbitrary (which links should we build in—those that hold for present- or past-tense marking, for humans or for animals?).

Let us conclude this section by examining which parts of Pinker's (1984) account succeed and which fail. Its first key strength is the assumption that children exploit probabilistic, though imperfect, correlations between semantic roles (e.g. Agent) and morphosyntactic marking, whether realized by word order (e.g. **[subject]** [verb] [object]) or morphology (e.g. nominative or ergative case marking). Its second key strength (as noted by Braine 1992) is the principle that 'old rules analyze new material', which allows the initial semantically based categories (e.g. Agent) to expand into syntactic categories via distributional analysis. For example, the distributional similarity between the first NPs in ***The cat*** *bit the dog* and ***The situation*** *justified the measures* allows *the situation* to be assimilated into the category containing *the cat*, even though the former is not an Agent. Although the situation is more complex for morphologically ergative languages, the 'old rules analyze new material' principle still applies, just with slightly more restrictive rules (i.e. different rules for clauses with perfective and imperfective aspect). Both of these learning procedures are extremely useful and presumably will have to be assumed, in some form or other, by any theory of acquisition. The problem for Pinker's proposal is that these learning procedures are so powerful that they obviate the need for innate linking rules (as indeed they must, given that there can be no set of rules that is viable crosslinguistically).

**3.2.** PARAMETER SETTING. An alternative UG-based approach to the acquisition of basic word order is parameter setting (Chomsky 1981b). Parameter-setting accounts assume that learners acquire the word order of their language by setting parameters on the basis of input utterances. Although perhaps as many as forty binary parameters are required to capture all crosslinguistic variation assumed within UG (Clark 1992, Baker 2001), three are particularly relevant for determining basic word order. The SPECIFIER-HEAD parameter determines, among other things, whether a language uses SV (e.g. English) or VS (e.g. Hawaiian) order. The COMPLEMENT-HEAD parameter—sometimes known simply as the HEAD-DIRECTION parameter—determines, among other things, whether a language uses VO (e.g. English) or OV (e.g. Turkish) order. The V2 parameter determines whether a language additionally stipulates that a tensed verb must always be the second constituent of all declarative main clauses, even if this means overriding the word order specified by the other parameters. Languages for which this is the case, such as German and Swedish, are said to have a +V2 setting, as opposed to the −V2 exhibited by languages such as English.

A potential problem facing parameter-setting approaches is parametric ambiguity: certain parameters cannot be set unless the child has previously set another parameter, and knows this setting to be correct (Clark 1989, 1992, Gibson & Wexler 1994). For example, suppose that a German child hears *Gestern kaufte* (V) *Hans* (S) *das Buch* (O) 'Yesterday bought (V) Hans (S) the book (O)'. Should this be taken as evidence that German has the VS and SO setting of the relevant parameters, or that the correct settings are in fact SV and VO and that the VSO word order is simply a consequence of the V2 rule? In fact, the second possibility is the correct one, but children cannot know this unless they have already correctly and definitively set the V2 parameter to +V2. In a formal mathematical analysis, Gibson and Wexler (1994) demonstrated that, in the face of ambiguous sentences of this type, there are many situations in which the learner can never arrive at the correct settings for all three parameters. This is due to the existence of LOCAL MAXIMA: states from which the learner could never reach the target grammar given the learning process assumed (or even 'archipelagos' of nontarget grammars, between which learners can move, but never escape; Frank & Kapur 1996).

Although this problem is shared by many older ERROR-DRIVEN-LEARNING approaches (e.g. Wexler & Culicover 1980, Berwick 1985, Hyams 1986), it has largely been solved by more recent work. The first solution is to propose that each parameter has a default initial state (Clark 1989, Gibson & Wexler 1994, Bertolo 1995)[8] and/or to relax the restrictions that (i) only changes that allow for a parse of the current sentence are retained (GREEDINESS) and (ii) only one parameter may be changed at a time (the SINGLE-VALUE constraint) (e.g. Berwick & Niyogi 1996, Frank & Kapur 1996). While these solutions work well for Gibson and Wexler's three-parameter space, they do not scale up to spaces with twelve or thirteen parameters (Bertolo et al. 1997, Kohl 1999, Fodor & Sakas 2004)—the approximate number generally held to be necessary for simple sentences—and/or require a prohibitively large number of utterances (Fodor & Sakas 2004). A much more successful strategy (e.g. see Sakas & Fodor 2012) is to have the parser detect ambiguous sentences. For example, Fodor's (1998a,b) STRUCTURAL TRIGGERS LEARNER attempts to parse input sentences with multiple grammars simultaneously, and discards (for the purposes of parameter setting) strings that can be successfully parsed by more than one.

[8] An alternative possibility is that UG specifies the order in which (some) parameters may be set (e.g. Baker 2001), although such proposals have been fully worked out only for phonological parameters (Dresher & Kaye 1990, Dresher 1999).

The third possible solution rejects triggering (or TRANSFORMATIONAL LEARNING) altogether in favor of VARIATIONAL LEARNING (Yang 2002:17, Pearl 2007). At any one point in development, instead of a single grammar (= array of parameter settings) that changes as each parameter is set, the learner has a population of competing grammars. When presented with an input sentence, the learner selects a grammar with probability *p* and attempts to analyze the sentence using this grammar, increasing *p* (i.e. the probability of future selection) if successful and decreasing *p* if not. Although it requires a relatively large number of utterances to succeed (Sakas & Nishimoto 2002), the variational learning model enjoys the advantages of being robust to noise (i.e. noncanonical or ungrammatical utterances) and avoiding having children lurch between various incorrect grammars as they flip parameter settings (as opposed to gradually increasing/decreasing their strength).

In short, there can be no doubt that modern parameter-setting approaches provide well-specified, computationally tractable accounts of word-order acquisition that converge quickly on the target grammar when implemented as computational models. The problem is that their success depends crucially on the assumption that the learner is able to parse input sentences as sequences of syntactic roles (e.g. subject-verb-object). Indeed, since these sequences constitute the input to computational implementations of parameter-setting models, this point is unequivocal:

> In effect, then, the [simulated] learner knows all word categories and grammatical roles in advance. In real life, such knowledge would be attained with some effort, perhaps through semantic bootstrapping and/or distributional learning (Pinker, 1984). On the other hand [real learners receive] helpful cues to syntactic phrase boundaries such as might result from prosodic bootstrapping. (Fodor & Sakas 2004:12)

The problem is that there are no successful accounts of how this knowledge could be obtained. As we argued above, semantic bootstrapping (Pinker 1984), distributional learning linked to innate syntactic categories (Mintz 2003), and prosodic bootstrapping (Christophe et al. 2008) do not work. In a variant of the prosodic bootstrapping approach, Mazuka (1996) proposed that children could set the head-direction (VO/OV) parameter on the basis of a crosslinguistic correlation between head direction and branching direction. VO languages (e.g. English) tend to be right-branching, meaning that each successive clause is added to the right of the sentence, while OV languages (e.g. Japanese) tend to be left-branching, with each successive clause added to the left. Of course, children who have yet to set the word-order parameters of their language cannot determine branching direction by parsing complex sentences syntactically. Mazuka's (1996) claim is that children can determine branching direction on the basis of purely phonological factors. For example, pitch changes are greater for subordinate → main clause boundaries than main → subordinate clause boundaries, and this could form part of children's innate knowledge. Similarly, Christophe and colleagues (2003) propose that children set the head-direction parameter using a correlation with phonological prominence. VO languages (e.g. English) tend to emphasize the rightmost constituent of a phrase (e.g. [*The* ***man***] [*kicked* ***the ball***]), and OV languages (e.g. Turkish) the leftmost.

However, it is far from clear that either correlation is universal (raising the problem of poor coverage). For example, Mazuka (1996) concedes that at least some sentence types in German and Chinese do not exhibit the phonological properties necessary for her proposed learning procedure to succeed. With regard to the proposal of Christophe et al. 2003, Pierrehumbert (2003) notes that maintaining this correlation would require somehow assigning different phonological analyses to English (SVO) and Japanese (SOV) sentences that have almost identical contours when measured objectively. Nor is

there any evidence that children are aware of such correlations where they exist. Indeed, Christophe and colleagues (2003) found that even ADULT native French speakers were able to select sentences with right- as opposed to left-hand prominence as sounding more French-like on only 65% of trials, despite an intensive training session with feedback. Note too that both proposals relate only to the setting of the VO/OV parameter and are silent on the setting of the SV/VS parameter. With regard to the third major word-order parameter, V2, prosodic bootstrapping (or, indeed, semantic bootstrapping) can offer no clue as to whether ambiguous SVO sentences (e.g. *John bought the book*) reflect the +V2 or −V2 setting (Fodor 1998b:342). Finally, Gervain and colleagues (2008) provided some preliminary evidence for the prosodic bootstrapping approach by demonstrating, using a novel grammar learning task, that Italian and Japanese eight-month-olds prefer prosodic phrases with frequent items phrase-initially and phrase-finally respectively. Given that function words are more frequent than content words, the claim is that Italian and Japanese infants have learned that their language prefers to place function words at the left vs. the right edge of the phrase respectively, and can make use of a crosslinguistic correlation between this property and various word-order phenomena (e.g. VO vs. OV respectively) to set the relevant parameters. As discussed with regard to syntactic category acquisition (§2.2), however, there is an important difference between demonstrating that infants exhibit a PREFERENCE for a particular type of stimulus (e.g. a phrase with more frequent words at the beginning) and demonstrating (i) that there exists a sufficiently robust crosslinguistic correlation between the presence of this cue and the setting of a particular parameter (e.g. VO) and (ii) that children are aware of this correlation. To our knowledge, no study has provided evidence for either of these claims.

**3.3.** INTERIM CONCLUSION. Given the problems with prosodic bootstrapping, parameter-setting accounts have never adequately addressed the linking problem. This leaves only Pinker's (1984) semantic bootstrapping account. As we argued above, however, this account also suffers from the linking problem, unless one largely abandons the role of innate semantics-syntax linking rules in favor of some form of a probabilistic input-based learning mechanism. For example, children could (i) group together items that share certain semantic regularities (e.g. acting as agents) and certain distributional regularities, and (ii) observe the ordinal positions in which these categories appear, and how this varies depending on factors such as tense, aspect, and animacy. But, as has previously been noted (e.g. Mazuka 1996, Tomasello 2003, 2005), once this has been done, children have effectively learned the word order of their language, and parameters become redundant.

As in §2 (syntactic categories), we end this section by considering the objection that, by invoking semantic and distributional analysis, we are bringing in innate knowledge by the back door. Might it be necessary, for example, to build in an innate bias to be more sensitive to certain semantic properties (e.g. Agent/Patienthood) than others (e.g. color), or to pay particular attention to the relative ordering of words, as opposed to, say, being the *n*th word? Perhaps. Certainly, it is not self-evident that this is the case. It is possible that children track all kinds of semantic and distributional properties that are rapidly discovered to be irrelevant (i.e. not to correlate with any communicative function). Indeed, given the wide range of semantic distinctions that may be encoded syntactically (e.g. humanness, animacy, evidentiality), it may be necessary for children's initial expectations to be relatively unconstrained. But even if it does turn out to be necessary to build in a bias for children to care especially about, for example, causation, this is a very different type of innate knowledge from that assumed under UG theories, in particular, innate semantics-syntax linking rules and word-order parameters.

**4.** STRUCTURE DEPENDENCE. Structure dependence has been called the 'parade case' (Crain 1991:602) of an innate principle, an 'innate schematism applied by the mind to the data of experience' (Chomsky 1971:28; see also Crain & Nakayama 1987, Boeckx 2010). Indeed, illustrations of the principle of structure dependence are often taken as the single best argument in favor of innate knowledge (e.g. Yang 2002:2). Although structure dependence applies across the entire grammar, we focus here on one domain that constitutes a particularly well-studied example of Chomsky's argument from the poverty of the stimulus.[9]

Chomsky (1980) argued that it is impossible for children to acquire the structure of complex yes-no questions from the input since they are virtually absent. Complex questions are those that contain both a **main clause** and a relative clause (e.g. ***Is the boy*** *who is smoking* ***crazy?***). Chomsky's argument runs as follows. Suppose that a child hears simple declarative/question pairs such as 4.

(4) The boy is crazy. → Is the boy crazy?

In principle, the child could formulate a rule such as 'to form a question from a declarative, move the first auxiliary to the front of the sentence'. However, this rule would generate incorrect questions from declaratives with more than one auxiliary, as in 5.

(5) The boy who is smoking is crazy. → *Is the boy who smoking is crazy?

The adult rule is 'move the auxiliary IN THE MAIN CLAUSE to the front of the sentence (or, strictly speaking, to the functional head C)'. The correct rule is STRUCTURE-DEPENDENT because it is formulated in terms of syntactic structure ('the auxiliary in the main clause') as opposed to linear order ('the first auxiliary'). Chomsky (1980:114–15) claims that children cannot learn that the structure-dependent rule, as opposed to the linear-order rule, is the correct one since 'a person might go through much or all of his life without ever having been exposed to relevant evidence [presumably complex questions, or even question/declarative pairs]'. Although this is probably an exaggeration (Pullum and Scholz (2002) find some complex yes-no questions in corpora of child-directed speech), we do not dispute the claim that they are too rare to constitute sufficient direct evidence of the correct structure (e.g. Legate & Yang 2002). Despite this paucity of evidence, even young children are able to produce correctly formed questions and avoid errors (e.g. Crain & Nakayama 1987). Chomsky (1980) therefore argues that children's knowledge of UG contains the principle of STRUCTURE DEPENDENCE (i.e. knowledge that rules must make reference to syntactic structure, not linear order).

**4.1.** COMPLEX YES-NO QUESTIONS. There are two questions at issue here. The first is how children avoid structure-dependence errors and acquire the correct generalization in the particular case of complex yes-no questions in English. The second is how children know that ALL linguistic generalizations are structure-dependent.[10]

Considering first the particular case of complex yes-no questions, there are three potential solutions that do not assume an innate principle. The first is to posit that ques-

[9] Although this argument has many different forms (Pullum and Scholz (2002) list thirteen different ways in which the child's input has been argued to be impoverished), perhaps the clearest presentation is that of Lightfoot (1989:322): 'It is too poor in three distinct ways: (a) The child's experience is finite but the capacity eventually attained ranges over an infinite domain … (b) the experience consists of partly degenerate input … (c) it fails to provide the data needed to induce many principles and generalizations which hold true of the mature category'. Lightfoot notes that '(c) is by far the most significant factor', and it this sense that we have in mind here.

[10] Or, at least, all syntactic generalizations ('grammatical transformations are invariably structure-dependent'; Chomsky 1968:61–62). There is clearly a role for linear order in, for example, phonology (e.g. the choice between *a* and *an* in English; Clark & Lappin 2011:37) and discourse structure (e.g. topic and focus; Pinker & Jackendoff 2005:220).

tions are not formed by movement rules at all, which renders moot the question of whether children might move the wrong auxiliary. Movement rules are eschewed not only by construction-based approaches (for question formation see Rowland & Pine 2000, Dąbrowska & Lieven 2005, Ambridge et al. 2006) but also by many more traditional grammars (see Clark & Lappin 2011:36 for a list). The second solution assumes that learners are sensitive to the pragmatic principle that one cannot extract elements of an utterance that are not asserted, but constitute background information (e.g. Van Valin & LaPolla 1997, a proposal that Crain and Nakayama (1987:526) also discuss, attributing it to Steven Pinker). This pragmatic principle is discussed in more detail in the following section on subjacency. For now, it suffices to note that a main clause, but not a subordinate clause, contains an assertion (which a second speaker may straightforwardly deny), as in 6, and hence, that only elements of a main clause may be extracted or questioned, as in 7.

(6) a. Speaker 1: The boy [who is smoking] is crazy.
b. Speaker 2: No, sane. *No, drinking beer.

(7) Is the boy [who is smoking] __ crazy? *vs.* *Is the boy [who __ smoking] is crazy?

While this solution in terms of a pragmatic principle is successful for complex questions (and perhaps other relative clause constructions; see §5), it has little to say about how children come to behave in accordance with the principle of structure dependence more generally.

The third potential solution for this particular case of complex English yes-no questions is that children make use of bi/trigram statistics in their input. Reali and Christiansen (2005) demonstrate that, where the correct and erroneous question forms deviate, the former contains a high-probability bigram (*Is the boy* ***who is***) while the latter contains a very low-probability bigram (**Is the boy* ***who smoking***). Consequently, a computer simulation sensitive to *n*-gram statistics predicts the correct form with higher probability than the error (Ambridge et al. 2008 also showed that this account could predict the question types for which children do, occasionally, produce such errors). Kam and colleagues (2008) and Berwick and colleagues (2011), however, showed that the model's success was due almost entirely to the fortuitous frequent occurrence of the relevant bigrams (*who* + *is*, *that* + *is*) in unrelated contexts (e.g. *Who's that? That's a rose*). That is, the bigram model succeeds only because English happens to use some homophonous forms for complementizers and WH-words/deictic pronouns. Since this is by no means a crosslinguistic requirement, Reali and Christiansen's (2005) solution is specific not only to complex yes-no questions, but also to English.

**4.2.** STRUCTURE DEPENDENCE IN GENERAL. This brings us to the more important question of how children know that syntactic rules are structure-dependent IN GENERAL. We argue that there is abundant evidence for the general principle of structure dependence not only in the language that children hear, but also in the conceptual world. With regard to the former, suppose that a child hears the following conversational fragments.

(8) a. John is smiling. Yes, he is happy.
b. The (/that/this/a etc.) boy is smiling. Yes, he is happy.
c. The tall boy is smiling. Yes, he is happy.
d. The boy who is tall is smiling. Yes, he is happy.

Such extremely simple exchanges, which occur whenever a pronoun refers back to an NP (presumably thousands of times a day), constitute evidence that strings of arbitrary

length that share distributional similarities can be substituted for one another (i.e. evidence for the structure-dependent nature of syntax). Computer models that use distribution in this way can simulate many structure-dependent phenomena, including the specific example of complex yes-no questions in English (Elman 1993, 2003, Lewis & Elman 2001, Clark & Eyraud 2007, Clark & Lappin 2011), at least to some extent. This qualification reflects the fact that a model that blindly substitutes distributionally similar strings for one another will inevitably produce a good deal of 'word salad' and uninterpretable sentences (Berwick et al. 2011).

But children—and the speakers who provide their input—are not blindly substituting phrases for one another on the basis of distributional similarity. The reason that *John*, *the boy*, *the tall boy*, and *the boy who is tall* can be substituted for one another is that all refer to entities in the world upon which the same kinds of semantic operations (e.g. predicating an action or property; being denoted as the causer of an event/state of affairs) can be performed (Tomasello 2005). The fact that, in cases such as those above, these strings may refer to the same entity presumably aids learners, but it is not crucial. The reason that languages group together concrete objects (*John*, *the boy*) with more abstract entities (e.g. *war*, *happiness*, *fighting each other*) is that all are subject to the same kinds of functional operations (e.g. predication of a property). Thus, to acquire a structure-dependent grammar, all a learner has to do is to recognize that strings such as *the boy*, *the tall boy*, *war*, and *happiness* share both certain functional and—as a consequence—distributional similarities. Whatever else one does or does not build into a theory of language acquisition, some kind of prelinguistic conceptual structure that groups together functionally similar concepts is presumably inevitable. This conceptual structure, when mapped onto language, yields a structure-dependent grammar.

This idea is not new. Returning to complex yes-no questions, Crain and Nakayama (1987, experiment 3) conducted an elicited-production study designed to test a version of this proposal formulated by Stemmer (1981). They found that children (aged 2;9–4;8) showed identical performance for questions with contentful lexical subjects (e.g. *Is* ***rain*** *falling in this picture?*) and semantically empty expletive subjects (e.g. *Is* ***it*** *raining in this picture?*), which they took as evidence against Stemmer's (1981) account. However, this finding constitutes evidence against the claim that we have outlined here only if one assumes that it is not possible that three-year-old children have done any of the following.[11]

- learned the formulaic questions *Is it raining*?, *Is there a(n) [THING]?*, and *Is it easy to [ACTION]?* (the only three items in this part of Crain and Nakayama's study)
- learned that *Is it* and *Is there* are common ways to start a question. We counted thirty-eight questions beginning *Is it* and thirty beginning *Is there* (excluding a similar number where these strings constituted the entire question) in the maternal section of the Thomas corpus (Dąbrowska & Lieven 2005, available on CHILDES). The issue is not whether this constitutes a high proportion of questions (or of all utterances), but simply whether the absolute number of these ques-

[11] Crain and Nakayama's findings arguably count against the particular stage-based account proposed by Stemmer (1981), under which children FIRST formulate a movement rule based on people, then gradually extend this to animals, objects, abstract concepts, and so forth (though, in partial support of Stemmer, Crain and Nakayama observed the worst performance for questions with abstract/actional subjects—e.g. *Is* ***running*** *fun? Is* ***love*** *good or bad?*). However, neither the movement rule nor the discontinuous stages proposed by Stemmer (1981) are a necessary part of an account based on conceptual structure.

tions (which can be estimated at around 300 and 380, respectively, under realistic sampling assumptions) is sufficient for children to learn these forms.
- generalized between dummy and lexical subjects on the basis of distributional and functional overlap (e.g. *he/it is cold*)

It is pertinent here to respond to a referee who asked where phrasal categories (e.g. NP, N′, V′, VP, CP, etc.) come from if not from UG. Although we do not wish to advocate any particular non-UG account of acquisition, if nothing else, our own informal use of such terms demands an explanation. It should be clear from the above that we use syntactic category labels (e.g. NOUN, VERB) as nothing more than a convenient shorthand for items sharing a certain degree of (sometimes) semantic, distributional, and—perhaps most importantly—functional similarity. The same is true for intermediate-level categories. For example, 'N-bar' structures like *yellow bottle* or *student of psychology* (e.g. Pearl & Lidz 2009) share a particular level of distributional similarity (e.g. *the __ is*) and functional similarity (e.g. ability to have a property predicated of them), in exactly the same way as for the simple and complex 'NPs' discussed above. We make analogous assumptions for other single and double-bar categories (e.g. 'V-bar' structures such as *chases the cat* and *causes cancer* share functional similarity in that both can be predicated of 'NOUNS'[12]). As should become clear in §§5 and 6, we view 'CP' (or clause) as reflecting an informational unit such as an assertion (main clause) or background/presupposed information (subordinate clause): hierarchical syntactic structure is a reflection of hierarchical conceptual structure.

These assumptions are less controversial than they might at first appear. Regardless of the particular theoretical background assumed, is hard to imagine any account of how children learn that (for example) *John*, *he*, and *the boy* may refer to the same entity that includes no role for semantic, distributional, or functional similarity. Indeed, many of the generativist accounts discussed in §§2 and 3 make such assumptions. Given that this type of learning yields structure-dependent generalizations, it does not seem to be such a huge step to dispense with structure dependence as an innate syntactic principle. In response to the charge that, by dispensing with innate categories (e.g. VERB) and their projections (e.g. VP, V′), we are replacing a perfectly good system with something that does not work, we would suggest that it is traditional categories (and therefore their projections) that do not work crosslinguistically (see §2), and these types of language-specific generalizations the only candidates to replace them.

Finally, we again end this section by considering the suggestion that these assumptions constitute bringing in innate knowledge by the back door. Children must learn that strings of arbitrary length upon which similar kinds of semantic/functional operations can be performed (e.g. predicating an action or property) can be substituted for one an-

[12] The referee who raised this point asked why—if syntactic structure reflects conceptual/perceptual structure—in active transitive sentences (e.g. *The boy kicked the ball*), the agent 'boy' seems to be a critical and inherent part of the conceptual/perceptual structure of the event, yet is absent from the VP. As we argue here, a VP (or V′) (e.g. *kicked the ball*) is a conceptually coherent unit in that—like *chases the cat/causes cancer*—it can be predicated of a NOUN. But, of course, this is not to say that the agent can be entirely absent. Where an action requires an agent, the VP (or V′) must indeed be combined with an obligatory NP (e.g. *the boy*) (unless it is an argument that is present in the conceptual/perceptual structure but, as an 'understood' argument, can be omitted from the syntactic structure). This NP/VP division in syntax reflects the default TOPIC/COMMENT or PREDICATE/FOCUS division in information structure. Thus, another way to think about the syntactic phrase VP (or V′) as arising from the conceptual/perceptual structure of the event is as a grammaticalization of the focus domain (a concept discussed more fully in the following section). Indeed, it has been argued that languages that do not grammaticalize the focus domain (e.g. Malayalam, Lakhota) do not make use of VPs as a unit of clause structure (Mohanan 1982:524–34, Van Valin 1987, Van Valin & LaPolla 1997:217–18).

other in many contexts. Does this require innate knowledge? Again, we would suggest that while it may or may not be necessary to assume certain very general biases (e.g. a propensity to conceptualize objects' actions and their properties as somehow similar, or to attempt to associate word strings with concepts in the world), this type of innate knowledge is qualitatively different from an innate principle of structure dependence, or an innate CP.

**5.** SUBJACENCY. Both Newmeyer (1991) and Pinker and Bloom (1990) cite subjacency (Chomsky 1973), another constraint on syntactic movement, as a prime example of an arbitrary linguistic constraint that is part of children's knowledge of UG. The standard UG assumption is that WH-questions are formed from an underlying declarative (or similar) by movement of the auxiliary (as discussed in the previous section) and, more relevant for subjacency, the WH-word (see Fig. 1 below). The phenomenon to be explained here is as follows. WH-words can be extracted from both simple main clauses and object complements.

(9) a. Bill bought a book. → What did Bill buy $t_i$?
b. Bill said [that Sue bought a book]. → What did Bill say [that Sue bought $t_i$]?

However, many other syntactic phrases are 'islands' in that WH-words (and other constituents) cannot be extracted from them (the metaphor is that the WH-word is stranded on the <island>). These include those in 10–13.

(10) (Definite) complex NPs
a. NP complements: *What$_i$ did Bill hear the rumor <that Sue stole $t_i$>?
(cf. Bill heard the rumor <that Sue stole the files>.)
b. Relative clauses: *What$_i$ did Bill interview the witness <who saw $t_i$>?
(cf. Bill interviewed the witness <who saw the files>.)

(11) Adjuncts: *What$_i$ did Bill walk home <after Sue took $t_i$>?
(cf. Bill walked home <after Sue took his car keys>.)

(12) Subjects: *What$_i$ did <Bill's stealing $t_i$> shock Sue?
(cf. <Bill's stealing the painting> shocked Sue.)

(13) Sentential subjects: *What$_i$ did <that Bill stole $t_i$> shock Sue?
(cf. <That Bill stole the painting> shocked Sue.)

Since Chomsky 1973 (though see Ross 1967 for an earlier formulation), the standard account has been the subjacency constraint, which specifies that movement may not cross more than one 'bounding node'. For English, bounding nodes are NP and S (or DP and IP), though this may vary between languages (e.g. NP/DP and S2/CP for Italian). An example of a subjacency violation is shown in Figure 1. Although this proposal has undergone some modifications (e.g. Chomsky 1986 reconceptualizes bounding nodes as BARRIERS and offers an explanation of why only certain nodes are barriers), the claim remains that some form of an innate UG island constraint aids learners by allowing them to avoid the production of ungrammatical sentences (or, in comprehension, interpretations that the speaker cannot have intended).

Our goal is not to dispute the facts regarding island constraints, which are generally well supported empirically. Nor do we argue that island constraints can be reduced to processing phenomena[13] (see the debate between Sag et al. 2007, Hofmeister & Sag 2010, Hofmeister et al. 2012a,b and Sprouse et al. 2012a,b, Yoshida et al. 2014). While

[13] These processing factors include the distance between the moved constituent and the gap (Kluender 1992, 1998, Kluender & Kutas 1993, Postal 1998), the semantic complexity of the intervening material (Warren & Gibson 2002, 2005), item and collocational frequency (Jurafsky 2003, Sag et al. 2007), finiteness (Ross 1967, Kluender 1992), informativeness (Hofmeister 2007), and ease of contextualization (Kroch 1998 [1989]).

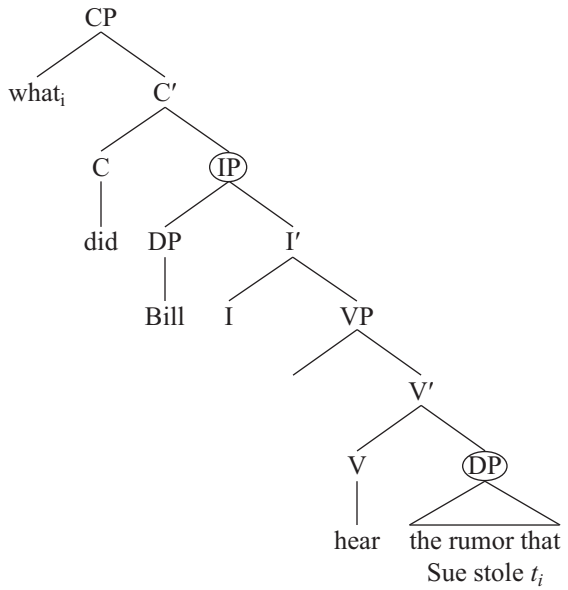


FIGURE 1. Subjacency. Extraction from a definite complex (NP/)DP (e.g. **What$_i$ did Bill hear <the rumor that Sue stole $t_i$>?*) is ruled out by the subjacency constraint, because the WH-word *what* crosses two bounding nodes (circled): (NP/)DP and IP. Note that, for clarity, movement of the subject and auxiliary is not shown.

all sides in this debate acknowledge that processing factors 'modulate the acceptability of island-violating sentences' (e.g. Sprouse et al. 2012b:404–5), processing-based accounts cannot explain equivalent constraints in WH-in-situ languages (e.g. Mandarin Chinese and Lakhota; Huang 1982, Van Valin & LaPolla 1997), where questions have the same surface structure as declaratives. (The absence of apparent movement is not a problem for grammatical accounts, however, on the assumption that movement—and hence subjacency—applies at the covert level of logical form, as opposed to the surface level of syntax; see Huang 1982). We argue, however, that an innate subjacency constraint is redundant: island constraints can be explained by discourse-pragmatic principles that apply to all sentence types, and hence that will have to be learned anyway.

The claim (see Erteschik-Shir 1979, 1998, Erteschik-Shir & Lappin 1979, Cattell 1984, Takami 1989, Deane 1991, Kluender 1992, 1998, Kluender & Kutas 1993, Kuno & Takami 1993, Van Valin 1995, 1998, 2005, Van Valin & LaPolla 1997, Goldberg 2006) is that the constituents above are islands because they lie outside the POTENTIAL FOCUS DOMAIN of the sentence. To understand this claim, a brief introduction to the notion of information structure is required (Mathesius 1928, Halliday 1967, Jackendoff 1972, Gundel et al. 1993, Lambrecht 1994, 2000). Most utterances have a TOPIC (or theme) about which some new information (the ***focus***, comment, or rheme) is asserted. In a basic declarative sentence, the *topic* is usually the subject.

(14) *Bill* bought a book.

The ***potential focus domain*** is the predicate phrase, and, under the default interpretation, is the ***actual focus*** as well (Bill ***bought a book***, rather than, say, ***ran a marathon***). However, provided that a cue such as vocal stress is used to overrule this default interpretation, the ***actual focus*** can be anywhere within the ***potential focus domain***.

(15) a. *Bill* ***bought a book***. (He didn't steal or borrow one.)
b. *Bill* ***bought a book***. (He didn't buy the particular book we had in mind, or two books.)
c. *Bill* ***bought a book***. (He didn't buy a newspaper.)

This much is uncontroversial. Also uncontroversial is the claim that children will have to learn about information structure in order to formulate even the most basic utterances. For example, most utterances require a noun phrase of some kind, and, for each, speakers must decide whether to use an indefinite NP, a definite NP, a proper name, a pronoun, or zero marking (Givon 1983, Ariel 1990, Gundel et al. 1993).

(16) [a man/the man/Bill/he/Ø] bought [a book/the book/*War and Peace*/it]

This requires an understanding of information structure. An established topic will usually be expressed by zero marking or a pronoun, and new, focal information with an indefinite NP. Violations of these information-structure principles yield infelicitous or even uninterpretable utterances.

(17) a. Speaker 1: So what did Bill do last night?
b. Speaker 2: Ate a cake./*Bill ate it.

Although young children are often assumed to have poor discourse-pragmatic skills, it has been demonstrated experimentally that even three-year-olds overwhelmingly use pronouns rather than lexical NPs to refer to a discourse topic established by an interlocutor (Matthews et al. 2006).

Returning to questions, it is clear that the questioned element is the focus of both a question and the equivalent declarative (we continue to use *italics* for the topic, ***bold italics*** for the potential focus domain, and additional ***underlining*** for the actual focus).

(18) a. *Bill* ***bought a book***.
b. ***What*** did *Bill* ***buy*** $t_i$?

The functional account of island constraints, then, is as follows: since the WH-word is the focus, it cannot replace constituents that are not in the potential focus domain. What all island constructions have in common is that the <islands> contain information that is old, incidental, presupposed, or otherwise backgrounded in some way.[14] As Van Valin (1998:232) argues:

> Questions are requests for information, and the focus of the question signals the information desired by the speaker. It makes no sense, then, for the speaker to place the focus of the question in a part of the sentence which is presupposed, i.e. which contains information which the speaker knows and assumes the hearer knows or can deduce easily.

Perhaps the clearest examples are complex NPs. Both 19a and 19b presuppose the existence of a rumor/witness, with the relative clause providing background information thereon (note that one CAN ask *What did Bill hear?* or *Who did Bill interview?* because the *rumor*/*the witness* is in the potential focus domain—indeed, the default focus).

[14] Backgroundedness is a graded notion; hence, different languages are free to 'choose' the extent to which a constituent may be backgrounded and still permit extraction. For example, Russian permits extraction from main clauses only (Freidin & Quicoli 1989), while Swedish has been described as showing no island constraints (Allwood 1976, Andersson 1982, Engdahl 1982). Hofmeister and Sag (2010:373) list Danish, Icelandic, Norwegian, Italian, French, Akan, Palauan, Malagasy, Chamorro, Bulgarian, Greek, and Yucatec Mayan as languages that exhibit 'counterexamples' to island constraints, though it may be possible to account for at least some of these cases within a subjacency framework by positing language-specific bounding nodes (as discussed in the main text with reference to Italian).

(19) a. *Bill* ***heard the rumor*** <that Sue stole the files>.
b. *Bill* ***interviewed the witness*** <who saw the files>.

Similarly, the constructions exemplified by 20a,b have the very function of emphasizing the presupposition that Bill did indeed steal the painting (more so than more usual formulations such as *Sue was shocked that Bill stole the painting*).

(20) a. <*Bill's stealing the painting*> ***shocked Sue***.
b. <*That Bill stole the painting*> ***shocked Sue***.

Adjuncts, by definition, provide background, nonfocal information (which may also be presupposed to some degree).[15]

(21) *Bill* ***walked home*** <after Sue took his car keys>.

There is a simple independent test for whether a particular constituent falls within the ***potential focus domain***: whether it can be denied (without recasting the entire phrase). The logic of the test is that it is only possible to deny assertions (not background information, presuppositions, etc.) and that assertions, by definition, constitute the potential focus domain. This test correctly predicts that 22 will not be an island in question form and that 23 will be.[16]

(22) *Bill* ***bought a book***. → No, he didn't.
(23) a. *Bill* ***heard the rumor*** <that Sue stole the files>. → No, he/*she didn't.
b. *Bill* ***interviewed the witness*** <who saw the files>. → No, he/*she didn't.
c. *Bill* ***walked home*** <after Sue took his car keys>. → No, he/*she didn't.
d. <*Bill's stealing the painting*> ***shocked Sue***. → No, it/*he didn't.
e. <*That Bill stole the painting*> ***shocked Sue***. → No it/*he didn't.

At first glance, this test—and hence the backgrounding account—appears to fail for questions with sentential complements such as **What$_i$ did Bill say (that) Sue bought?*. Since one can deny the fact but not the content of reported speech (*Bill said (that) Sue bought a book → No, he/*she didn't*), the negation test predicts, apparently incorrectly, that such questions will be blocked. In fact, not only does the negation test correctly predict the data here, but it also does so in a way that syntactic subjacency accounts cannot. The key is that both negatability/backgrounding and island status are matters of degree. Ambridge and Goldberg (2008) asked participants to rate, for particular verbs, (i) the extent to which negating the sentence entails negation of the reported speech (a measure of backgrounding) and (ii) the grammaticality of the extraction question. On these measures, *say* was rated as only moderately backgrounding the reported speech, and the extraction question only moderately unacceptable. Verbs that are informationally richer than *say* (e.g. *whisper*, *mumble*) would be expected to be rated as (i) foregrounding the speech act, hence backgrounding its content, and thus (ii) less acceptable in extraction questions. Exactly this pattern was found. Given that no subjacency viola-

[15] WH-islands are a borderline case in the subjacency literature. Huang (1982), Chomsky (1986), and Lasnik and Saito (1992) argue that weak islands (of which WH-islands are a subset; see Szabolcsi & den Dikken 2002 for a review) block adjuncts (e.g. **How did Bill wonder <whether to buy the book>?*) to a greater degree than arguments (e.g. *What did Bill wonder <whether to buy>?*). This pattern can be explained by the functional account on the assumption that the information expressed by an adjunct (e.g. *using his credit card*) is more backgrounded than that expressed by an argument (e.g. *the book*).

[16] We should acknowledge that this account (and hence this test) does not make the correct predictions for coordinate structures such as **What$_i$ did Bill eat <fish and $t_i$>?* (cf. *Bill ate <fish and chips>*) or left-branch structures such as **Which$_i$ did Bill eat <$t_i$ cake>?* (cf. *Bill ate <this cake>*). Such cases (particularly the second) seem to constitute violations of a different principle altogether: that informational units (e.g. this cake; which cake?) cannot be broken up (cf. *Which cake did Bill eat $t_i$?*).

tion occurs in any of these cases (and that such violations are binary, not a matter of degree), syntactic subjacency accounts cannot explain this graded pattern, or even why any of the sentences should be rated as less than fully acceptable. Nor can such accounts explain graded definiteness effects (*Who$_i$ did Bill read* [*a > the > ?the new >??the fantastic new history book*] *about?*). The functional account explains this pattern naturally: the more that is already known about the book (i.e. the more it constitutes background knowledge), the less acceptable the extraction question.

Do such cases mean that an innate subjacency principle could be actively harmful? After all, if learners were using only this principle to determine the grammaticality of such instances, they would incorrectly arrive at the conclusion that all were equally and fully acceptable. It seems that the only way to prevent an innate subjacency principle from being harmful to learners would be to allow the discourse-pragmatic principles discussed here to override it, rendering subjacency redundant.

This is not to deny that subjacency generally provides excellent coverage of the data. However, we suggest that the proposal is so successful because its primitives correspond to the primitives of discourse structure. For example, the principle that one can question an element of a main clause but not a relative clause or an adjunct is a restatement of the principle that one can question an assertion but not presupposed or incidental information. The very reason that languages have relative clauses and adjuncts is that speakers find it useful to have syntactic devices that distinguish background information from the central assertion of the utterance. To sum up: in order to be effective communicators, children will have to acquire principles of discourse pragmatics and focus structure. These principles account not only for island constraints, but also for some phenomena not covered by a formal subjacency account.

**6.** BINDING PRINCIPLES. Languages exhibit certain constraints on coreference; that is, they appear to block certain pronouns from referring to particular noun phrases. For example, in 24, the pronoun *she* cannot refer to Sarah, but must refer to some other (female) person who has been previously mentioned, or is otherwise available for reference (e.g. by being present in the room).

(24) She$_i$ listens to music when Sarah$_{*i}$ reads poetry.

The standard assumption of UG-based approaches is that such principles are unlearnable (e.g. Guasti & Chierchia 1999/2000:140) and must instead be specified by innate BINDING PRINCIPLES that are part of UG. The formal definition of 'binding' (e.g. Chomsky 1981a, Reinhart 1983) is that X binds Y if (i) X c-commands Y and (i) X and Y are coindexed (i.e. refer to the same entity). The notion of c-command as it relates to the three binding principles—principles A, B, and C—is explained in Figure 2.

**6.1.** PRINCIPLE C. Principle C, which rules out example 24 above, states that a R(eferring)-expression (e.g. an NP such as *Sarah* that takes its meaning directly from the world, NOT from another word in the sentence) must be free everywhere (i.e. NOT bound anywhere; Chomsky 1981a). Thus 24 constitutes a principle C violation because the R-expression *Sarah* is bound by the pronoun *She* (*She* c-commands *Sarah*, and they corefer). More informally, we can understand principle C (at least for multiple-clause sentences) by saying that a pronoun may precede a full lexical NP to which it corefers only if the pronoun is in a subordinate clause. Thus forward anaphora, where a lexical NP sends its interpretation 'forward' (i.e. left-to-right), is allowed whether the pronoun is in the main or subordinate clause.

(25) a. [$_{CP}$ [$_{CP}$ When Sarah$_i$ reads poetry] she$_i$ listens to music].
b. [$_{CP}$ Sarah$_i$ listens to music [$_{CP}$ when she$_i$ reads poetry]].

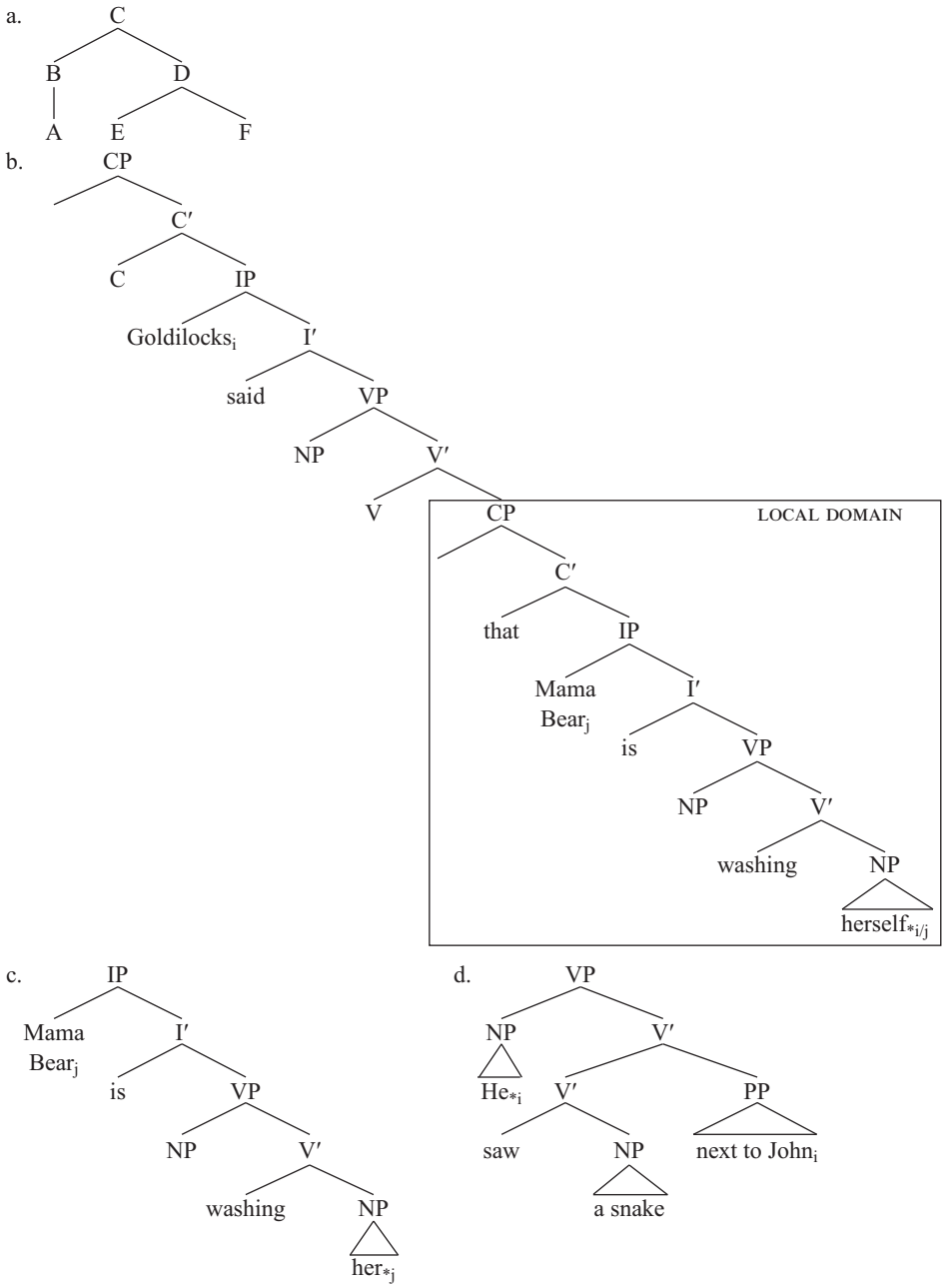


FIGURE 2. C-command and binding. Although there exist a number of different formulations of c-command (e.g. Langacker 1969, Lasnik 1976, Chomsky 1981a, Reinhart 1983), for our purposes a simple definition will suffice: 'a constituent X c-commands its sister constituent Y and any constituent Z which is contained within Y' (Radford 2004:75). A simpler way to think about c-command is to use the analogy of a train network: X c-commands any node that one can reach by 'taking a northbound train [from X], getting off at the first station, changing trains there and then travelling one or more stops south *on a different line*' (Radford 2004:75). For example, in Fig. 2a, B c-commands D, E, and F; D c-commands A and B; E and F c-command one another; A does not c-command any node. To consider some examples relevant to the binding principles, in Fig. 2b, both *Goldilocks* and *Mama Bear* c-command *herself*. Principle A stipulates that *herself* must refer to *Mama Bear*, as it is the only NP in the local domain. In Fig. 2c, *Mama Bear* c-commands *her*, meaning that, by principle B, the two cannot corefer. In Fig. 2d, *He* c-commands *John*, meaning that coreference is blocked by principle C.

Backward anaphora, where a lexical NP sends its interpretation 'backward' (i.e. right-to-left), is allowed only when the pronoun is in the subordinate clause (all examples from Lust 2006:214).

(26) a. $[_{CP}$ $[_{CP}$ When she$_i$ reads poetry] Sarah$_i$ listens to music].
b. *$[_{CP}$ She$_i$ listens to music $[_{CP}$ when Sarah$_i$ reads poetry]].

As for subjacency, we argue that the proposed UG principle—here principle C—is successful only to the extent that it correlates with principles of discourse and information structure. The functional explanation (e.g. Bickerton 1975, Bolinger 1979, Kuno 1987, Levinson 1987, van Hoek 1995, Van Valin & LaPolla 1997, Harris & Bates 2002) is as follows. As we saw in the previous section, the topic/theme is the NP that the sentence is 'about', and about which some assertion is made (the comment/focus/rheme). This assertion is made in the predicate of the main clause (e.g. *Sarah listens to music*), with subordinate clauses providing some background information. As we also saw earlier, when a particular referent is already topical (e.g. we already know we are talking about Sarah), it is most natural to use a pronoun (or null reference) as topic (*She listens to music*). Thus, when speakers use a lexical NP as topic, they do so to establish this referent as the new topic (or, at least, to reestablish a previously discussed referent as the topic of a new assertion). Once they have decided to use a lexical NP to establish a new topic, it is entirely natural for speakers to use a pronoun in the part of the sentence that provides some background information on this topic.[17]

(27) a. $[_{CP}$ Sarah$_i$ listens to music $[_{CP}$ when she$_i$ reads poetry]].
b. $[_{CP}$ $[_{CP}$ When she$_i$ reads poetry] Sarah$_i$ listens to music].

Indeed, the use of a full NP (e.g. *Sarah listens to music when Sarah reads poetry*) is so unnatural that there is a strong sense that some special meaning is intended (e.g. that Sarah is particularly obstinate in her insistence that poetry reading and music listening must always go together). Now consider cases of ungrammatical coreference.

(28) *$[_{CP}$ She$_i$ listens to music $[_{CP}$ when Sarah$_i$ reads poetry]].

In these cases, the speaker has decided to use a pronoun as the topic, indicating that the referent is highly accessible. This being the case, it is pragmatically anomalous to use a full lexical NP in a part of the sentence that exists only to provide background information. If I (as speaker) am sufficiently confident that you (as listener) know who am I talking about to use a pronoun as the topic of my main assertion (*She listens to music*), I should be just as happy (if anything, more so) to use pronouns in the part of the sentence that constitutes only background information (*when she reads poetry*). The only plausible reason for my use of a full lexical NP in this part of the sentence would be to

[17] For single-clause sentences, the discourse-functional explanation is even simpler (though, of course, there is no backgrounded clause). If a pronoun is used as the topic, this indicates that the referent is highly accessible, rendering anomalous the use of a full NP ANYWHERE within the same clause (examples from Lakoff 1968, Kuno 1987).

(i) a. *He$_i$ found a snake near John$_i$. (cf. John$_i$ found a snake near him$_i$.)
b. *Near John$_i$ he$_i$ found a snake. (cf. Near him$_i$ John$_i$ found a snake.)
c. *He$_i$ found a snake behind the girl John$_i$ was talking with. (cf. John$_i$ found a snake behind the girl he$_i$ was talking with.)
d. *He$_i$ loves John's$_i$ mother. (cf. John $_i$ loves his $_i$ mother.)
e. *John's$_i$ mother, he$_i$ adores dearly. (cf. His$_i$ mother, John$_i$ adores dearly.)

This also applies to quantified NPs (e.g. *every pirate*), as in the following examples from Guasti and Chierchia (1999/2000:131).

(ii) a. *He$_i$ put a gun in every pirate$_i$'s barrel. (cf. Every pirate$_i$ put a gun in his$_i$ barrel.)
b. *In every pirate$_i$'s barrel, he$_i$ put a gun. (cf. In his$_i$ barrel, every pirate$_i$ put a gun.)

identify a new referent. The situation is similar for so-called strong crossover questions (Chomsky 1981a).

(29) *Who$_i$ did he$_i$ say Ted criticized?

The coreferential reading (which can be paraphrased as 'Who said Ted criticized him?') is impossible for exactly the same reason that such a reading is impossible for the equivalent declarative.

(30) *He$_i$ said Ted criticized Bill$_i$.

The speaker has used a pronoun as the topic of the main assertion of the sentence (*He said X*) and so cannot use a lexical NP in a clause that provides background information (what was said) to refer to that same entity (cf. *Bill$_i$ said Ted criticized him$_i$*). (See the previous section for evidence that speakers consider the content of reported speech to be backgrounded to at least some extent.) Exactly the same situation holds for sentences with quantificational expressions (Chomsky 1981a) such as **He$_i$ said Ted criticized everyone$_i$* and *Everyone$_i$ said Ted criticized him$_i$* (which are the same sentences as the previous two examples with *everyone* substituted for *Bill*).

In general, it makes pragmatic sense to use a lexical NP (including quantified NPs like *everyone*) as the topic about which some assertion is made, and a pronoun in a part of the sentence containing information that is secondary to that assertion, but not vice versa.[18] With one exception (which we consider shortly), this generalization explains all of the cases normally attributed to principle C. Furthermore, the findings of an adult judgment study not only provide direct evidence for this backgrounding account, but also suggest that it predicts the pattern of coreference possibilities better than a syntactic account. Harris and Bates (2002) demonstrated that if a principle-C-violating sentence is manipulated such that the subordinate clause contains new information and the main clause background information (e.g. *He was threatening to leave when Billy noticed that the computer had died*), participants accepted a coreferential reading on a substantial majority of trials (75%).

An exception to this backgrounding account occurs in cases of forward anaphora from a subordinate into a main clause (e.g. *When Sarah$_i$ reads poetry, she$_i$ listens to music*). However, such examples are easily covered by the discourse-pragmatic account in general: once a speaker has already referred to an individual with a full NP, it is quite natural to use a pronoun in a subsequent clause, and indeed, unnatural not to (e.g. *When Sarah reads poetry, Sarah listens to music*). Although one might object to this order-of-mention principle as an 'add-on' to the functional account, it is equally indispensable to formal accounts, as it is necessary to account for pronominalization between sentences or conjoined clauses, to which no binding principle can apply (van Hoek 1995).

(31) a. Sarah$_i$ reads poetry. She$_i$ also listens to music.
b. *She$_i$ reads poetry. Sarah$_i$ also listens to music.
c. Sarah$_i$ reads poetry, and she$_i$ also listens to music.
d. *She$_i$ reads poetry, and Sarah$_i$ also listens to music.

Note further that this add-on to the principle C account makes reference to the same notion of information structure on which the functional account is based.[19] In order to pro-

[18] In the previous section, we discussed evidence that even three-year-olds understand the discourse-functional constraints that govern the use of pronouns vs. full NPs (Matthews et al. 2006). Thus studies that demonstrate apparent adherence to principle C at this age (e.g. Somashekar 1995) do not constitute evidence that children must necessarily be using this formal syntactic principle as opposed to discourse function.

[19] An alternative UG-based solution to the problem of intersentential pronominalization is to assume an underlying string that is present in the underlying representation, but not pronounced (Chomsky 1968, Morgan

duce even simple single-clause sentences, children need to know (and, indeed, by age three do know; Matthews et al. 2006) certain discourse-functional principles (here, when to use a lexical NP vs. a pronoun). These pragmatic principles, which must be added on to any formal account to deal with otherwise-problematic cases, in fact explain the entire pattern of the data, leaving an innate syntactic principle redundant. Again, the proposed syntactic principle offers good data coverage only to the extent that it restates these pragmatic principles. For example, the syntactic principle that one cannot pronominalize backward into a main clause (**She$_i$ listens to music* [*when Sarah$_i$ reads poetry*]) restates the pragmatic principle that one cannot pronominalize from the part of the sentence that contains the main assertion into a part of the sentence that contains only background information. Thus, in most cases, the two accounts make the same predictions. But the syntactic account is only a rough paraphrase of the functional account. When this paraphrase diverges too far from the functional account—as in Harris and Bates's (2002) sentences where the usual functions of the main and subordinate clauses are flipped—it mispredicts the data.

**6.2.** PRINCIPLES A AND B. Principles A and B (Chomsky 1981a, Reinhart 1983) govern the use of reflexive (e.g. *herself*) vs. nonreflexive (e.g. *her*) pronouns. Principle A states that a reflexive pronoun (e.g. *herself*) must be bound in its local domain. For all the cases we discuss, the local domain is the clause. Essentially, then, principle A specifies that for sentences such as *Goldilocks$_i$ said that Mama Bear$_j$ is washing herself$_{*i/j}$*, the reflexive pronoun *herself* can refer only to the NP that c-commands it IN THE LOCAL DOMAIN (i.e. *Mama Bear*). It cannot refer to an NP that (i) c-commands it but is not in the local domain (e.g. *Goldilocks*, which is in a different clause) or (ii) does not c-command it at all (e.g. another character previously mentioned in the story).

Principle B states that a nonreflexive pronoun must be free (i.e. NOT bound) in its local domain. Effectively, it is the converse of principle A: in a context where a reflexive pronoun (e.g. *herself*) must be used, one cannot substitute it with a nonreflexive pronoun (e.g. *her*) without changing the meaning. For example, for the sentence *Goldilocks$_i$ said that Mama Bear$_j$ is washing her$_{i/*j}$*, the pronoun (*her*) cannot take its meaning from *Mama Bear*.[20] If it did, this would constitute a principle B violation, since the nonreflexive pronoun (*her*) would be c-commanded in its local domain by *Mama Bear*. Note that principle B stipulates only what the nonreflexive pronoun CANNOT refer to. The pronoun may take its meaning either from the NP *Goldilocks* or from an entity in the world (e.g. *Cinderella was covered in mud. While Goldilocks read the book, Mamma Bear washed her* [Cinderella]).

---

1973, 1989, Hankamer 1979, Merchant 2005, Crain & Thornton 2012), as in the following example from Conroy & Thornton 2005.

(i) Q: Where did he$_i$ send the letter?
A: ~~He sent the letter~~ To Chuckie$_{*i}$'s house

However, this solution works by assuming that the speaker is, in effect, producing a sentence containing a pronoun topic and a coreferential NP elsewhere in the same clause. Such sentences are ruled out by the discourse-pragmatic principle outlined here (see previous footnote).

[20] It is perhaps also worth noting that the distinction between reflexive and nonreflexive pronouns emerged only relatively recently, at least in English. In Old English (i.e. before around 1000 AD), the equivalent of *Mama Bear washed her* did indeed mean 'Mama Bear washed herself'. For example, Deutscher (2005:296) cites an example from Beowulf, where the hero dresses himself for battle, but the pronoun used is *hine* 'him'. Thus if an innate principle B was selected for during evolution, it is unlikely to have been because it conferred a communicative advantage; it marks a distinction that languages seem perfectly able to do without.

Informally, principles A and B together reduce to a simple axiom: if a reflexive pronoun (e.g. *herself*) would give the intended meaning, a nonreflexive pronoun (e.g. *her*) cannot be used instead. Indeed, this is incorporated into UG accounts of binding (Grodzinsky & Reinhart 1993:79).

(32) Rule 1: NP A [e.g. *her*] cannot corefer with NP B [e.g. *Mama Bear*] if replacing A with C [e.g. *herself*], C a variable A-bound by B, yields an indistinguishable interpretation. [Chien and Wexler (1990) refer to this constraint as principle P.]

Consequently, the facts attributed to the binding principles reduce to a very simple functional explanation (Kuno 1987:67): 'Reflexive pronouns are used in English if and only if they are direct recipients or targets of the actions represented by the sentences'.

(33) a. John [killed/fell in love] with himself/*him. (target)
b. John addressed the letter to himself/*him. (recipient)
c. John [heard strange noises/left his family] behind *himself/him. (location)
d. John has passion in *himself/him. (location; cf. John sees himself as having no passion.)

A very similar formulation is that reflexive pronouns denote a referent as seen from his or her own point of view, nonreflexive pronouns from a more objective viewpoint (Cantrall 1974).

(34) I can understand a father wanting his daughter to be like himself but I can't understand that ugly brute wanting his daughter to be like him.

Since even UG-based accounts of principles A and B (e.g. Chien & Wexler 1990, Grodzinsky & Reinhart 1993) make something very similar to this assumption, additional innate principles are redundant. Furthermore, there are again cases where only discourse-functional principles offer satisfactory data coverage.

(35) Q: Who did Sue say is the cleverest girl in the room?
A: Herself (*Her)

(36) Q: Who do you think is the cleverest girl in the room?
A: Her (*Herself)

The impossible readings are not ruled out by principles A and B, which by definition cannot apply across sentence boundaries, but by the functional considerations outlined above. Principles A and B make the correct predictions only when they align with these considerations.

(37) a. $\text{Goldilocks}_i$ said that $\text{Mama Bear}_j$ is washing $\text{herself}_{*i/j}$. (= Mama Bear is the target of the washing.)
b. $\text{Goldilocks}_i$ said that $\text{Mama Bear}_j$ is washing $\text{her}_{i/*j}$. (= Mama Bear is not the target of the washing.)

There is another sentence type for which principles A and B make the wrong predictions, and this is conceded even by UG-based accounts (e.g. Chien & Wexler 1990, Grodzinsky & Reinhart 1993). These are so-called 'Evans-style' contexts (after Evans 1980).

(38) That must be John. At least, he looks like him.

While most speakers regard this sentence as acceptable, it constitutes a principle B violation since the nonreflexive pronoun *him* is c-commanded in its local domain by *he*, and both refer to the same entity. The only way to rescue principle B is to appeal to the functional explanation outlined above. The nonreflexive pronoun *him* is used because

the intended meaning is that he (the person who may be John) looks like ***him*** (John), not that he (the person who may be John) looks like ***himself*** (i.e. is the target of the resembling 'action'). Indeed, UG-based accounts propose essentially this very solution. For example, Thornton and Wexler's (1999) guise-creation hypothesis argues that listeners create two separate guises for the referents (e.g. a person who MAY BE John, and a person who IS John).

Thus we are left with exactly the same situation as for principle C: discourse-functional principles that must be included in formal accounts to explain particular counterexamples can, in fact, explain the entire pattern of data. The proposed syntactic principle is successful only to the extent that it is a restatement of the discourse-based account, and fails when it does not (e.g. for both intersentential and Evans-style contexts).

**6.3.** INTERIM CONCLUSION. For all three binding principles, there exist phenomena that—under any account, UG-based or otherwise—can be explained only by recourse to discourse-functional principles. Since these principles can explain all of the relevant phenomena, innately specified binding principles are redundant.

**7.** CONCLUSION. Many theories assume that the process of language acquisition in the face of impoverished, underconstraining input is too complex to succeed without the aid of innate knowledge of categories, constraints, principles, and parameters, provided in the form of UG. The present article has argued that, even if no restrictions are placed on the type of innate knowledge that may be posited, there are no proposals for components of innate knowledge that would simplify the learning process for the domains considered.

This is not to say that accounts in the UG tradition offer nothing by means of explanation with regard to these domains. Many of the proposals discussed are ingenious and have the advantage that they both capture aspects of the acquisition problem that might otherwise have been overlooked and identify cues and mechanisms that are likely to form part of the solution. The problem is that, without exception, each component of innate knowledge proposed suffers from at least one of the problems of linking, data coverage, and redundancy—in some cases all three. The most widespread of these problems is redundancy. For each domain, the cues and mechanisms that actually solve the learning problem are ones that are not related to UG, and that must be assumed by all accounts, whether or not they additionally assume innate knowledge. These types of learning procedures (e.g. clustering of semantically and/or distributionally similar items) and discourse-pragmatic principles (e.g. when to use a full NP vs. a pronoun; how to foreground/background particular informational units) do not constitute rival explanations to those offered by UG accounts. On the contrary, they are factors that are incorporated into UG accounts, precisely because they would seem to be indispensable to any comprehensive account of the relevant phenomenon (since, if nothing else, they are needed to account for particular counterexamples). The problem is that it is these factors that lend UG-based accounts their explanatory power. The innate categories/principles proposed are redescriptions of the outcomes of these factors. In general, they are faithful redescriptions, and hence merely redundant; occasionally they diverge and risk hindering the learning process.

Proponents of UG-based accounts may point to the fact that we have proposed no alternative to such accounts and argue that, until a compelling alternative is offered, it is logical to stick to UG-based accounts. This argument would be persuasive if there existed UG-based accounts that explain how a particular learning problem is solved with the aid of innate constraints. If there were a working UG-based explanation of, for ex-

ample, how children acquire the syntactic categories and word-order rules of their language, it would, of course, make no sense to abandon this account in the absence of a viable alternative. But as we have aimed to show in this review, there is no working UG-based account of any of the major phenomena in language acquisition; current accounts of this type explain the data only to the extent that they incorporate mechanisms that make no use of innate grammatical knowledge. Of course, we claim only to have shown that none of the categories, learning procedures, principles, and parameters proposed under CURRENT UG-based theories aid learning; we have not shown that such innate knowledge could not be useful IN PRINCIPLE. It remains entirely possible that there are components of innate linguistic knowledge—yet to be proposed—that would demonstrably aid learning. Our claim is simply that nothing is gained by positing components of innate knowledge that do not simplify the problem faced by language learners, and that this is the case for all extant UG-based proposals.

Thus, our challenge to advocates of UG is this: rather than presenting abstract learnability arguments of the form 'X is not learnable given the input that a child receives', explain precisely how a particular type of innate knowledge would help children to acquire X. In short, 'You can't learn X without innate knowledge' is no argument for innate knowledge, unless it is followed by ' … but you can learn X WITH innate knowledge, and here's one way that a child could do so'.

## REFERENCES


ALLWOOD, JENS S. 1976. *The complex NP constraint in Swedish*. (University of Massachusetts occasional reports 2.) Amherst: University of Massachusetts.

AMBRIDGE, BEN, and ADELE E. GOLDBERG. 2008. The island status of clausal complements: Evidence in favor of an information structure explanation. *Cognitive Linguistics* 19.3.49–381.

AMBRIDGE, BEN, and ELENA V. M. LIEVEN. 2011. *Child language acquisition: Contrasting theoretical approaches*. Cambridge: Cambridge University Press.

AMBRIDGE, BEN; CAROLINE F. ROWLAND; and JULIAN M. PINE. 2008. Is structure dependence an innate constraint? New experimental evidence from children's complex-question production. *Cognitive Science* 32.1.222–55.

AMBRIDGE, BEN; CAROLINE F. ROWLAND; ANNA L. THEAKSTON; and MICHAEL TOMASELLO. 2006. Comparing different accounts of inversion errors in children's non-subject wh-questions: 'What experimental data can tell us?'. *Journal of Child Language* 33.3.519–57.

ANDERSSON, LARS-GUNNAR. 1982. What is Swedish an exception to*?* Extractions and island constraints. *Readings on unbounded dependencies in Scandinavian languages*, ed. by Elisabet Engdahl and Eva Ejerhed, 33–46. Stockholm: Almqvist & Wiksell.

ARIEL, MIRA. 1990. *Accessing noun-phrase antecedents*. London: Routledge.

BAKER, MARK C. 2001. *The atoms of language: The mind's hidden rules of grammar*. New York: Basic Books.

BERTOLO, STEFANO. 1995. Maturation and learnability in parametric systems. *Language Acquisition* 4.4.277–318.

BERTOLO, STEFANO; KEVIN BROIHIER; EDWARD GIBSON; and KENNETH WEXLER. 1997. Cue-based learners in parametric language systems: Application of general results to a recently proposed learning algorithm based on unambiguous 'superparsing'. *Proceedings of the 19th annual conference of the Cognitive Science Society*, 49–54.

BERWICK, ROBERT C. 1985. *The acquisition of syntactic knowledge*. Cambridge, MA: MIT Press.

BERWICK, ROBERT C., and PARTHA NIYOGI. 1996. Learning from triggers. *Linguistic Inquiry* 27.4.605–22.

BERWICK, ROBERT C.; PAUL M. PIETROSKI; BERACAH YANKAMA; and NOAM CHOMSKY. 2011. Poverty of the stimulus revisited. *Cognitive Science* 35.7.1207–42.

BHAT, D. N. SHANKARA. 1991. *Grammatical relations: The evidence against their necessity and universality*. London: Routledge.

Bickerton, Derek. 1975. Some assertions about presuppositions and pronominalizations. *Chicago Linguistic Society (Parasession on functionalism)* 11.2.580–609.
Boeckx, Cedric. 2010. *Language in cognition: Uncovering mental structures and the rules behind them*. Oxford: Wiley-Blackwell.
Bolinger, Dwight. 1979. Pronouns in discourse. *Syntax and semantics, vol. 12: Discourse and syntax*, ed. by Talmy Givón, 289–309. New York: Academic Press.
Bowerman, Melissa. 1990. Mapping thematic roles onto syntactic functions: Are children helped by innate linking rules? *Linguistics* 28.6.1251–89.
Braine, Martin D. S. 1992. What sort of innate structure is needed to 'bootstrap' into syntax? *Cognition* 45.1.77–100.
Cantrall, William R. 1974. *Viewpoint, reflexives and the nature of noun phrases*. The Hague: Mouton.
Cartwright, Timothy A., and Michael R. Brent. 1997. Syntactic categorization in early language acquisition: Formalizing the role of distributional analysis. *Cognition* 63.2.121–70.
Cassidy, Kimberly Wright, and Michael H. Kelly. 2001. Children's use of phonology to infer grammatical class in vocabulary learning. *Psychonomic Bulletin & Review* 8.3.519–23.
Cattell, Ray. 1984. *Syntax and semantics, vol. 17: Composite predicates in English*. Orlando: Academic Press.
Chien, Yu-Chin, and Kenneth Wexler. 1990. Children's knowledge of locality conditions in binding as evidence for the modularity of syntax and pragmatics. *Language Acquisition* 1.225–95.
Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
Chomsky, Noam. 1968. *Language and mind*. New York: Harcourt Brace Jovanovich.
Chomsky, Noam. 1971. *Problems of knowledge and freedom*. London: Fontana.
Chomsky, Noam. 1973. Conditions on transformations. *A festschrift for Morris Halle*, ed. by Stephen R. Anderson and Paul Kiparsky, 232–86. New York: Holt, Reinhart & Winston.
Chomsky, Noam. 1980. *Language and learning: The debate between Jean Piaget and Noam Chomsky*, ed. by Massimo Piatelli-Palmarini. Cambridge, MA: Harvard University Press.
Chomsky, Noam. 1981a. *Lectures on government and binding*. Dordrecht: Foris.
Chomsky, Noam. 1981b. Principles and parameters in syntactic theory. *Explanation in linguistics: The logical problem of language acquisition*, ed. by Norbert Hornstein and David Lightfoot, 32–75. London: Longman.
Chomsky, Noam. 1986. *Barriers*. Cambridge, MA: MIT Press.
Christiansen, Morten H., and Padraic Monaghan. 2006. Discovering verbs through multiple-cue integration. *Action meets word: How children learn verbs*, ed. by Kathy Hirsh-Pasek and Roberta Michnick Golinkoff, 544–64. Oxford: Oxford University Press.
Christodoulopoulos, Christos; Sharon Goldwater; and Mark Steedman. 2010. Two decades of unsupervised POS induction: How far have we come? *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, 575–84. Online: http://dl.acm.org/citation.cfm?id=1870714.
Christophe, Anne; Maria T. Guasti; Marina Nespor; Emmanuel Dupoux; and Brit van Ooyen. 2003. Prosodic structure and syntactic acquisition: The case of the head-direction parameter. *Developmental Science* 6.2.211–20.
Christophe, Anne; Jacques Mehler; and Núria Sebastian-Galles. 2001. Perception of prosodic boundary correlates by newborn infants. *Infancy* 2.385–94.
Christophe, Anne; Séverine Millotte; Savita Bernal; and Jeffrey Lidz. 2008. Bootstrapping lexical and syntactic acquisition. *Language and Speech* 51.61–75.
Clark, Alex. 2000. Inducing syntactic categories by context distribution clustering. *Proceedings of the 4th Conference on Computational Natural Language Learning and of the 2nd Learning Language in Logic Workshop*, 91–94.
Clark, Alex, and Remi Eyraud. 2007. Polynomial time identification in the limit of substitutable context-free languages. *Journal of Machine Learning Research* 8.1725–45.
Clark, Alex, and Shalom Lappin. 2011. *Linguistic nativism and the poverty of the stimulus*. Oxford: Wiley-Blackwell.

Clark, Robin. 1989. On the relationship between the input data and parameter setting. *North East Linguistic Society (NELS)* 19.48–62.
Clark, Robin. 1992. The selection of syntactic knowledge. *Language Acquisition* 2.83–149.
Conroy, Anastasia, and Rosalind Thornton. 2005. Children's knowledge of principle C in discourse. *Proceedings of the 6th Tokyo Conference on Psycholinguistics*, 69–94.
Craig, Colette G. 1977. *The structure of Jacaltec*. Austin: University of Texas Press.
Crain, Stephen. 1991. Language acquisition in the absence of experience. *Behavioral and Brain Sciences* 14.4.597–650.
Crain, Stephen, and Mineharu Nakayama. 1987. Structure dependence in grammar formation. *Language* 63.3.522–43.
Crain, Stephen, and Rosalind Thornton. 2012. Syntax acquisition. *WIREs Cognitive Science* 3.2.185–203.
Croft, William. 2001. *Radical construction grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.
Croft, William. 2003. *Typology and universals*. 2nd edn. Cambridge: Cambridge University Press.
Dąbrowska, Ewa, and Elena V. M. Lieven. 2005. Towards a lexically specific grammar of children's question constructions. *Cognitive Linguistics* 16.3.437–74.
Deane, Paul. 1991. Limits to attention: A cognitive theory of island phenomena. *Cognitive Linguistics* 2.1.1–63.
Deutscher, Guy. 2005. *The unfolding of language: An evolutionary tour of mankind's greatest invention*. New York: Metropolitan Books.
Dixon, Robert M. W. 1972. *The Dyirbal language of North Queensland*. Cambridge: Cambridge University Press.
Dixon, Robert M. W. 1994. *Ergativity*. Cambridge: Cambridge University Press.
Dixon, Robert M. W. 2004. Adjective classes in typological perspective. *Adjective classes: A cross-linguistic typology*, ed. by Robert M. W. Dixon and Alexandra Y. Aikhenvald, 1–49. Oxford: Oxford University Press.
Dresher, B. Elan. 1999. Child phonology, learnability, and phonological theory. *Handbook of child language acquisition*, ed. by William C. Ritchie and Tej K. Bhatia, 299–346. San Diego: Academic Press.
Dresher, B. Elan, and Jonathan D. Kaye. 1990. A computational learning model for metrical phonology. *Cognition* 34.2.137–95.
Dryer, Matthew S. 1997. Are grammatical relations universal? *Essays on language function and language type*, ed. by Joan Bybee, John Haiman, and Sandra A. Thompson, 115–43. Amsterdam: John Benjamins.
Elman, Jeffrey L. 1993. Learning and development in neural networks: The importance of starting small. *Cognition* 48.1.71–99.
Elman, Jeffrey L. 2003. Generalization from sparse input. *Chicago Linguistic Society* 38.175–200.
Engdahl, Elisabet. 1982. Restrictions on unbounded dependencies in Swedish. *Readings on unbounded dependencies in Scandinavian languages*, ed. by Elisabet Engdahl and Eva Ejerhed, 151–74. Stockholm: Almqvist & Wiksell.
Erteschik-Shir, Nomi. 1979. Discourse constraints on dative movement. *Syntax and semantics, vol. 12: Discourse and syntax*, ed. by Talmy Givón, 441–67. New York: Academic Press.
Erteschik-Shir, Nomi. 1998. *The dynamics of focus structure*. Cambridge: Cambridge University Press.
Erteschik-Shir, Nomi, and Shalom Lappin. 1979. Dominance and the functional explanation of island phenomena. *Theoretical Linguistics* 6.41–85.
Evans, Gareth. 1980. Pronouns. *Linguistic Inquiry* 11.2.337–62.
Evans, Nicholas, and Stephen C. Levinson. 2009. With diversity in mind: Freeing the language sciences from universal grammar. *Behavioral and Brain Sciences* 32.5.472–92.
Fisher, Cynthia, and Hisayo Tokura. 1996. Acoustic cues to grammatical structure in infant-directed speech: Cross-linguistic evidence. *Child Development* 67.6.3192–218.
Fodor, Janet Dean. 1998a. Unambiguous triggers. *Linguistic Inquiry* 29.1.1–36.

FODOR, JANET DEAN. 1998b. Parsing to learn. *Journal of Psycholinguistic Research* 27.3. 339–74.
FODOR, JANET DEAN, and WILLIAM G. SAKAS. 2004. Evaluating models of parameter setting. *Proceedings of the Boston University Conference on Language Development (BUCLD)* 28.1–27.
FRANK, ROBERT, and SHYAM KAPUR. 1996. On the use of triggers in parameter setting. *Linguistic Inquiry* 27.4.623–60.
FREIDIN, ROBERT, and A. CARLOS QUICOLI. 1989. Zero-stimulation for parameter setting. *Behavioral and Brain Sciences* 12.2.338–39.
FREUDENTHAL, DANIEL; JULIAN M. PINE; and FERNAND GOBET. 2005. On the resolution of ambiguities in the extraction of syntactic categories through chunking. *Cognitive Systems Research* 6.1.17–25.
GERKEN, LOUANN; PETER W. JUSCZYK; and DENISE R. MANDEL. 1994. When prosody fails to cue syntactic structure: 9-month-olds' sensitivity to phonological versus syntactic phrases. *Cognition* 51.3.237–65.
GERVAIN, JUDIT; MARINA NESPOR; REIKO MAZUKA; RYOTA HORIE; and JACQUES MEHLER. 2008. Bootstrapping word order in prelexical infants: A Japanese-Italian cross-linguistic study. *Cognitive Psychology* 57.56–74.
GIBSON, EDWARD, and KENNETH WEXLER. 1994. Triggers. *Linguistic Inquiry* 25.3.407–54.
GIVÓN, TALMY. 1983. Topic continuity in discourse: An introduction. *Topic continuity in discourse: A quantitative cross-language study*, ed. by Talmy Givón, 1–42. Amsterdam: John Benjamins.
GOLDBERG, ADELE E. 2006. *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.
GRODZINSKY, YOSEF, and TANYA REINHART. 1993. The innateness of binding and coreference. *Linguistic Inquiry* 24.1.69–101.
GUASTI, MARIA T. 2004. *Language acquisition: The growth of grammar*. Cambridge, MA: MIT Press.
GUASTI, MARIA T., and GENNARO CHIERCHIA. 1999/2000. Reconstruction in child grammar. *Language Acquisition* 8.129–70.
GUNDEL, JEANETTE; NANCY HEDBERG; and RON ZACHARSKI. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69.2.274–307.
HALLIDAY, MICHAEL A. K. 1967. Notes on transitivity and theme in English. *Journal of Linguistics* 3.199–244.
HANKAMER, JORGE. 1979. *Deletion in coordinate structures*. New York: Garland.
HARRIS, CATHERINE L., and ELIZABETH A. BATES. 2002. Clausal backgrounding and pronominal reference: A functionalist approach to c-command. *Language and Cognitive Processes* 17.3.237–69.
HASPELMATH, MARTIN. 2007. Pre-established categories don't exist: Consequences for language description and typology. *Linguistic Typology* 11.1.119–32.
HAUSER, MARK; NOAM CHOMSKY; and W. TECUMSEH FITCH. 2002. The faculty of language: What is it, who has it, and how did it evolve? *Science* 298.1569–79.
HOCKETT, CHARLES F. 1960. The origin of speech. *Scientific American* 203.88–111.
HOFMEISTER, PHILIP. 2007. Memory retrieval effects on filler-gap processing. *Proceedings of the 29th annual conference of the Cognitive Science Society*, 1091–96.
HOFMEISTER, PHILIP, and IVAN A. SAG. 2010. Cognitive constraints and island effects. *Language* 86.2.366–415.
HOFMEISTER, PHILIP; LAURA STAUM CASASANTO; and IVAN A. SAG. 2012a. How do individual cognitive differences relate to acceptability judgments? A reply to Sprouse, Wagers, and Phillips. *Language* 88.2.390–400.
HOFMEISTER, PHILIP; LAURA STAUM CASASANTO; and IVAN A. SAG. 2012b. Misapplying working-memory tests: A reductio ad absurdum. *Language* 88.2.408–9.
HOLISKY, DEE A. 1987. The case of the intransitive subject in Tsova-Tush (Batsbi). *Lingua* 71.103–32.
HUANG, C. T. JAMES. 1982. Move *wh* in a language without *wh*-movement. *The Linguistic Review* 1.369–416.
HYAMS, NINA. 1986. *Language acquisition and the theory of parameters*. Dordrecht: Reidel.

JACKENDOFF, RAY. 1972. *Semantic interpretation in generative grammar*. Cambridge, MA: MIT Press.
JACOBSEN, WILLIAM H. 1979. Noun and verb in Nootkan. *The Victoria Conference on Northwestern Languages* (Heritage record 4), ed. by Barbara Erfat, 83–155. Victoria: British Columbia Provincial Museum.
JELINEK, ELOISE, and RICHARD DEMERS. 1994. Predicates and pronominal arguments in Straits Salish. *Language* 70.4.697–736.
JURAFSKY, DANIEL. 2003. Probabilistic modeling in psycholinguistics: Linguistic comprehension and production. *Probabilistic linguistics*, ed. by Rens Bod, Jennifer Hay, and Stefanie Jannedy, 39–95. Cambridge, MA: MIT Press.
KAM, XUÂN-NGA CAO; IGLIKA STOYNESHKA; LIDIYA TORNYOVA; JANET DEAN FODOR; and WILLIAM G. SAKAS. 2008. Bigrams and the richness of the stimulus. *Cognitive Science* 32.4.771–87.
KEENAN, EDWARD L. 1976. Towards a universal definition of 'subject'. *Subject and topic*, ed. by Charles N. Li, 303–33. New York: Academic Press.
KINKADE, M. DALE. 1983. Salish evidence against the universality of 'noun' and 'verb'. *Lingua* 60.25–40.
KLUENDER, ROBERT. 1992. Deriving islands constraints from principles of predication. *Island constraints: Theory, acquisition and processing*, ed. by Helen Goodluck and Michael S. Rochemont, 223–58. Dordrecht: Kluwer.
KLUENDER, ROBERT. 1998. On the distinction between strong and weak islands: A processing perspective. *Syntax and semantics, vol. 29: The limits of syntax,* ed. by Peter Culicover and Louise McNally, 241–79. San Diego: Academic Press.
KLUENDER, ROBERT, and MARTA KUTAS. 1993. Subjacency as a processing phenomenon. *Language and Cognitive Processes* 8.4.573–633.
KOHL, KAREN T. 1999. An analysis of finite parameter learning in linguistic spaces. Cambridge, MA: MIT master's thesis.
KROCH, ANTHONY. 1998 [1989]. Amount quantification, referentiality, and long WH-movement. *University of Pennsylvania Working Papers in Linguistics* 5.2.21–36. Online: http://repository.upenn.edu/pwpl/vol5/iss2/3/.
KUNO, SUSUMU. 1987. *Functional syntax: Anaphora, discourse and empathy*. Chicago: Chicago University Press.
KUNO, SUSUMU, and KEN-ICHI TAKAMI. 1993. *Grammar and discourse principles: Functional syntax and GB theory*. Chicago: University of Chicago Press.
LAKOFF, GEORGE. 1968. *Pronouns and reference*. Bloomington: Indiana University Linguistics Club.
LAMBRECHT, KNUD. 1994. *Information structure and sentence form: Topic, focus, and the mental representations of discourse referents*. Cambridge: Cambridge University Press.
LAMBRECHT, KNUD. 2000. When subjects behave like objects: An analysis of the merging of S and O in sentence-focus constructions across languages. *Studies in Language* 24.3. 611–82.
LANGACKER, RONALD. 1969. On pronominalization and the chain of command. *Modern studies in English*, ed. by David A. Reibel and Sanford A. Schane, 160–86. Englewood Cliffs, NJ: Prentice Hall.
LASNIK, HOWARD. 1976. Remarks on coreference. *Linguistic Analysis* 2.1–22.
LASNIK, HOWARD, and MAMORU SAITO. 1992. *Move alpha: Conditions on its application and output*. Cambridge, MA: MIT Press.
LAZARD, GILBERT. 1992. Y a-t-il des catégories interlangagières? *Texte, Sätze, Wörter und Moneme: Festschrift für Klaus Heger zum 65. Geburtstag*, ed. by Susanne Anschütz, 427–34. Heidelberg: Heidelberger Orientverlag. [Reprinted in *Études de linguistique générale*, ed. by Gilbert Lazard, 57–64. Leuven: Peeters, 2001.]
LEGATE, JULIE A., and CHARLES D. YANG. 2002. Empirical re-assessment of stimulus poverty arguments. *The Linguistic Review* 19.151–62.
LEVINSON, STEPHEN C. 1987. Pragmatics and the grammar of anaphora. *Journal of Linguistics* 23.379–434.
LEWIS, JOHN D., and JEFFREY L. ELMAN. 2001. Learnability and the statistical structure of language: Poverty of stimulus arguments revisited. *Proceedings of the Boston University Conference on Language Development (BUCLD)* 26.359–70.
LIDZ, JEFFREY; HENRY GLEITMAN; and LILA GLEITMAN. 2003. Understanding how input matters: Verb learning and the footprint of universal grammar. *Cognition* 87.3.151–78.

LIDZ, JEFFREY, and LILA R. GLEITMAN. 2004. Argument structure and the child's contribution to language learning. *Trends in Cognitive Sciences* 8.4.157–61.
LIDZ, JEFFREY, and JULIEN MUSOLINO. 2002. Children's command of quantification. *Cognition* 84.2.113–54.
LIDZ, JEFFREY; SANDRA WAXMAN; and JENNIFER FREEDMAN. 2003. What infants know about syntax but couldn't have learned: Experimental evidence for syntactic structure at 18 months. *Cognition* 89.B65–B73.
LIGHTFOOT, DAVID. 1989. The child's trigger experience: Degree-0 learnability. *Behavioral and Brain Sciences* 12.2.321–34.
LUST, BARBARA. 2006. *Child language: Acquisition and growth*. New York: Cambridge University Press.
MARANTZ, ALEC P. 1984. *On the nature of grammatical relations*. Cambridge, MA: MIT Press.
MARATSOS, MICHAEL. 1990. Are actions to verbs as objects are to nouns? On the differential semantic bases of form, class, category. *Linguistics* 28.3.1351–79.
MATHESIUS, VILÉM. 1928. On linguistic characterology with illustrations from Modern English. *Actes du Premier Congrès International de Linguistes à La Haye, du 10–15 Avril*, 56–63. Leiden: A. W. Sijthoff. [Reprinted in *A Prague School reader in linguistics*, ed. by Josef Vachek, 59–67. Bloomington: Indiana University Press, 1964.]
MATTHEWS, DANIELLE; ELENA V. M. LIEVEN; ANNA L. THEAKSTON; and MICHAEL TOMASELLO. 2006. The effect of perceptual availability and prior discourse on young children's use of referring expressions. *Applied Psycholinguistics* 27.3.403–22.
MAZUKA, REIKO. 1996. Can a grammatical parameter be set before the first word? Prosodic contributions to early setting of a grammatical parameter. *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*, ed. by James Morgan and Katherine Demuth, 313–30. Mahwah, NJ: Lawrence Erlbaum.
MCCAWLEY, JAMES D. 1992. Justifying part-of-speech distinctions in Mandarin Chinese. *Journal of Chinese Linguistics* 20.211–46.
MERCHANT, JASON. 2005. Fragments and ellipsis. *Linguistics and Philosophy* 27.661–738.
MINTZ, TOBEN H. 2003. Frequent frames as a cue for grammatical categories in child directed speech. *Cognition* 90.1.91–117.
MOHANAN, K. P. 1982. Grammatical relations and clause structure in Malayalam. *The mental representation of grammatical relations*, ed. by Joan Bresnan, 504–89. Cambridge, MA: MIT Press.
MORGAN, JERRY. 1973. Sentence fragments and the notion 'sentence'. *Issues in linguistics: Papers in honor of Henry and Renée Kahane*, ed. by Braj Kachru, 719–52. Champaign: University of Illinois Press.
MORGAN, JERRY. 1989. Sentence fragments revisited. *Chicago Linguistic Society (Parasession on language in context)* 25.2.228–41.
NESPOR, MARINA, and IRENE VOGEL. 1986. *Prosodic phonology*. Dordrecht: Foris.
NEWMEYER, FREDERICK J. 1991. Functional explanation in linguistics and the origins of language. *Language and Cognitive Processes* 11.1.3–28.
NIDA, EUGENE A. 1949. *Morphology*. Ann Arbor: University of Michigan Press.
PARISIEN, CHRIS; AFSANEH FAZLY; and SUZANNE STEVENSON. 2008. An incremental Bayesian model for learning syntactic categories. *Proceedings of the 12th Conference on Computational Natural Language Learning*, 89–96.
PEARL, LISA. 2007. *Necessary bias in natural language learning*. College Park: University of Maryland dissertation.
PEARL, LISA, and JEFFREY LIDZ. 2009. When domain-general learning fails and when it succeeds: Identifying the contribution of domain specificity. *Language Learning and Development* 5.4.235–65.
PIERREHUMBERT, JANET B. 2003. Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech* 46.115–54.
PINKER, STEVEN. 1979. Formal models of language learning. *Cognition* 7.3.217–83.
PINKER, STEVEN. 1984. *Language learnability and language development*. Cambridge, MA: Harvard University Press.
PINKER, STEVEN. 1987. The bootstrapping problem in language acquisition. *Mechanisms of language acquisition*, ed. by Brian MacWhinney, 339–441. Hillsdale, NJ: Lawrence Erlbaum.

PINKER, STEVEN. 1989. *Learnability and cognition: The acquisition of argument structure*. Cambridge, MA: MIT Press.
PINKER, STEVEN, and PAUL BLOOM. 1990. Natural language and natural selection. *Behavioral and Brain Sciences* 13.4.707–84.
PINKER, STEVEN, and RAY JACKENDOFF. 2005. The faculty of language: What's special about it? *Cognition* 95.2.201–36.
POSTAL, PAUL M. 1998. *Three investigations of extraction*. Cambridge, MA: MIT Press.
PULLUM, GEOFFREY K., and BARBARA C. SCHOLZ. 2002. Empirical assessment of stimulus poverty arguments. *The Linguistic Review* 19.9–50.
PYE, CLIFTON. 1990. The acquisition of ergative languages. *Linguistics* 28.6.1291–330.
RADFORD, ANDREW. 2004. *Minimalist syntax: Exploring the structure of English*. Cambridge: Cambridge University Press.
REALI, FLORENCIA, and MORTEN H. CHRISTIANSEN. 2005. Uncovering the richness of the stimulus: Structure dependence and indirect statistical evidence. *Cognitive Science* 29.6.1007–28.
REDINGTON, MARTIN; NICK CHATER; and STEVEN FINCH. 1998. Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science* 22.4.425–69.
REINHART, TANYA. 1983. *Anaphora and semantic interpretation*. London: Croom Helm.
RIJKHOFF, JAN. 2003. When can a language have nouns and verbs? *Acta Linguistica Hafniensa* 35.7–38.
RISPOLI, MATTHEW; PAMELA A. HADLEY; and JANET K. HOLT. 2009. The growth of tense productivity. *Journal of Speech, Language and Hearing Research* 52.4.930–44.
ROEPER, THOMAS. 2007. *The prism of grammar: How child language illuminates humanism*. Cambridge, MA: MIT Press.
ROSS, JOHN R. 1967. *Constraints on variables in syntax*. Cambridge, MA: MIT dissertation. [Published as *Infinite syntax!*, Norwood, NJ: Ablex, 1986.]
ROWLAND, CAROLINE, and JULIAN M. PINE. 2000. Subject-auxiliary inversion errors and wh-question acquisition: 'What children do know?'. *Journal of Child Language* 27.1.157–81.
SAG, IVAN A.; PHILIP HOFMEISTER; and NEAL SNIDER. 2007. Processing complexity in subjacency violations: The complex noun phrase constraint. *Chicago Linguistic Society* 43.215–29.
SAKAS, WILLIAM G., and JANET DEAN FODOR. 2012. Disambiguating syntactic triggers. *Language Acquisition* 19.83–143.
SAKAS, WILLIAM G., and EIJI NISHIMOTO. 2002. Search, structure or statistics? A comparative study of memoryless heuristics for syntax acquisition. *Proceedings of the 24th annual conference of the Cognitive Science Society*, 786–91.
SAXTON, MATTHEW. 2010. *Child language: Acquisition and development*. London: Sage.
SCHACHTER, PAUL. 1976. The subject in Philippine languages: Topic, actor, actor-topic or none of the above? *Subject and topic*, ed. by Charles N. Li, 491–518. New York: Academic Press.
SIEGEL, LAURA. 2000. Semantic bootstrapping and ergativity. Paper presented at the annual meeting of the Linguistic Society of America, Chicago, January 8, 2000.
SODERSTROM, MELANIE; AMANDA SEIDL; DEBORAH G. KEMLER-NELSON; and PETER W. JUSCZYK. 2003. The prosodic bootstrapping of phrases: Evidence from prelinguistic infants. *Journal of Memory and Language* 49.2.249–67.
SOMASHEKAR, SHAMITHA. 1995. Indian children's acquisition of pronominals in Hindi 'jab' clauses: Experimental study of comprehension. Ithaca, NY: Cornell University master's thesis.
SPROUSE, JON; MATT WAGERS; and COLIN PHILLIPS. 2012a. A test of the relation between working-memory capacity and syntactic island effects. *Language* 88.1.82–123.
SPROUSE, JON; MATT WAGERS; and COLIN PHILLIPS. 2012b. Working-memory capacity and island effects: A reminder of the issues and the facts. *Language* 88.2.401–7.
STEMMER, NATHAN. 1981. A note on empiricism and structure-dependence. *Journal of Child Language* 8.3.649–63.
SZABOLCSI, ANNA, and MARCEL DEN DIKKEN. 2002. Islands. *The second* GLOT International *state-of-the-article book*, ed. by Lisa Cheng and Rint P. E. Sybesma, 213–40. Berlin: Mouton de Gruyter.

Takami, Ken-Ichi. 1989. Preposition stranding: Arguments against syntactic analyses and an alternative functional explanation. *Lingua* 76.299–335.
Thornton, Rosalind, and Kenneth Wexler. 1999. *Principle B, VP ellipsis and interpretation in child grammar*. Cambridge, MA: MIT Press.
Tomasello, Michael. 2003. *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
Tomasello, Michael. 2005. Beyond formalities: The case of language acquisition. *The Linguistic Review* 22.183–97.
Valian, Virginia. 1986. Syntactic categories in the speech of young children. *Developmental Psychology* 22.562–79.
Valian, Virginia; Stephanie Solt; and John Stewart. 2009. Abstract categories or limited-scope formulae? The case of children's determiners. *Journal of Child Language* 36.4.743–78.
van Hoek, Karen. 1995. *Anaphora and conceptual structure*. Chicago: University of Chicago Press.
Van Valin, Robert D., Jr. 1987. The role of government in the grammar of head-marking languages. *International Journal of American Linguistics* 53.371–97.
Van Valin, Robert D., Jr. 1992. An overview of ergative phenomena and their implications for language acquisition. *The crosslinguistic study of language acquisition*, vol. 3, ed. by Dan I. Slobin, 15–37. Hillsdale, NJ: Lawrence Erlbaum.
Van Valin, Robert D., Jr. 1995. Toward a functionalist account of so-called 'extraction constraints'. *Complex structures: A functionalist perspective*, ed. by Betty Devriendt, Louis Goossens, and Johan van der Auwera, 29–60. Berlin: Mouton de Gruyter.
Van Valin, Robert D., Jr. 1998. The acquisition of wh-questions and the mechanisms of language acquisition. *The new psychology of language: Cognitive and functional approaches to language structure*, ed. by Michael Tomasello, 221–49. Hillsdale, NJ: Lawrence Erlbaum.
Van Valin, Robert D., Jr. 2005. *Exploring the syntax-semantics interface*. Cambridge: Cambridge University Press.
Van Valin, Robert D., Jr., and Randy J. LaPolla. 1997. *Syntax: Structure, meaning and function*. Cambridge: Cambridge University Press.
Viau, Joshua, and Jeffrey Lidz. 2011. Selective learning in the acquisition of Kannada ditransitives. *Language* 87.4.679–714.
Warren, Tessa, and Edward Gibson. 2002. The influence of referential processing on sentence complexity. *Cognition* 85.1.79–112.
Warren, Tessa, and Edward Gibson. 2005. Effects of NP type in reading cleft sentences in English. *Language and Cognitive Processes* 20.6.751–67.
Wexler, Kenneth, and Peter Culicover. 1980. *Formal principles of language acquisition*. Cambridge, MA: MIT Press.
Woodbury, Anthony. 1977. Greenlandic Eskimo, ergativity, and relational grammar. *Syntax and semantics, vol. 8: Grammatical relations*, ed. by Peter Cole and Jerrold M. Sadock, 307–36. New York: Academic Press.
Yang, Charles. 2002. *Knowledge and learning in natural language*. Oxford: Oxford University Press.
Yang, Charles. 2006. *The infinite gift*. New York: Scribner's.
Yang, Charles. 2008. The great number crunch. *Journal of Linguistics* 44.205–28.
Yang, Charles. 2009. Who's afraid of George Kingsley Zipf ? Philadelphia: University of Pennsylvania, ms.
Yoshida, Masaya; Nina Kazanina; Leticia Pablos; and Patrick Sturt. 2014. On the origin of islands. *Language, Cognition and Neuroscience* 29.7.761–70.

Ambridge and Pine
University of Liverpool
Institute of Psychology, Health and Society
Bedford Street South
Liverpool, L69 7ZA, United Kingdom
[Ben.Ambridge@liverpool.ac.uk]
[Julian.Pine@liverpool.ac.uk]

[Received 6 March 2013;
accepted 29 August 2013]

Lieven
University of Manchester
School of Psychological Sciences
Coupland 1 Building
Coupland Street, Oxford Road
Manchester, M13 9PL, United Kingdom
[elena.lieven@manchester.ac.uk]