

**ADVENTURES WITH *CAMILLE*:
INVESTIGATING THE ARCHITECTURE OF THE LANGUAGE FACULTY
THROUGH COMPUTATIONAL SIMULATION¹**

Peter W. Culicover
The Ohio State University

Andrzej Nowak
University of Warsaw

Wojciech Borkowski
University of Warsaw

Katherine Woznicki
The Ohio State University

A foundational issue in cognitive science is the extent to which the properties of particular mental faculties are the product of general capacities that hold for cognition in general. The debate has been especially lively in the case of language, where the particular properties appear to have no counterpart in other cognitive domains, and are therefore good candidates for being specific to the language faculty. If they are specific to language, the argument goes, it is not necessary to explain them in terms of how cognition works in general; they are presumably simply the product of evolution.²

On the other hand some of these properties are so specific and apparently so unrelated to functionality to that it is reasonable to question why evolution would have given rise to them. For example, in standard varieties of English it is not possible to have a gap corresponding to a displaced constituent immediately following the complementizer *that*:

¹ This report extends the discussion of the computational simulation of language acquisition reported on in Culicover and Nowak 2003.

² For a recent exchange on this general issue, see Hauser et al. 2002 and Pinker and Jackendoff 2005.

- (1) Who_i did you expect (*that) ____i would win.
cf. I expected (that) Fred would win.

It is not clear what evolutionary advantage would follow from a constraint that rules out **that* ____i, especially given that there are languages and even varieties of English that allow it.

A somewhat less categorical view is that certain properties arise from the interaction between the structures of language and the requirement that they be computed in real time by speakers, hearers and learners. For example, the hearer is faced with the task of determining the meaning of an expression on the basis of its form, and in certain cases, the complexity of the form may pose particular challenges for the computational device that determines the meaning of an expression. This is almost certainly true for well-known cases of multiple center embedding, as in (2), but it may be the basis of an explanation in other cases as well, such as Ross' (1967) island constraints.³

- (2) The man that the criminal that the cop arrested mugged was my friend.

For the learner, the processing task is similar. On the basis of examples of form/meaning correspondences, the learner must construct general rules that say what the possible structures are, and how they are mapped into meaning. Again, it is not implausible that certain systems of realizing such mappings are more complex than others, and pose difficulty for learners or even render learning impossible.⁴

Finally, we come back to the view that the observed properties of language are not specific to language itself. Depending on the property in question, it is possible to find a range of positions under this general rubric. A representative view is that of Tomasello (2003), who claims that a substantial number of properties that theoretical linguists have posited as universals resident in the language faculty are emergent in the knowledge acquired by the learner.

Much of the debate in the literature has turned on points of logic and rhetoric. In part the likely reason for this is that it is impossible to demonstrate strictly on empirical grounds that a particular property of language is **not** specific to the language faculty. In the absence of a fully worked out alternative explanation, it is as plausible that the impossibility of a gap after the complementizer *that* in (1) is due to a specific property of the language faculty, or to the complexity of processing such structures, or to the difficulty of learning a language that treats such sentences as grammatical.

While some of the properties of language are relatively specific, and turn out not to be found in all languages, such as the one exemplified in (1), others are very general

³ For considerable discussion of this general idea and some specific proposals, see Hawkins 1994, 2004.

⁴ See Wexler and Culicover 1980.

and appear to be universal. For example, all languages appear to have nouns and verbs, all languages appear to distinguish Subject and Object, many languages can highlight constituents of a sentence by locating them in designated positions (usually clause-initial position), and all languages appear to be able to express the same range of communicative functions, such as statements, questions, requests, and so on.⁵ For some if not all of these properties it is at least plausible that they are not explicitly represented in the architecture of the language faculty. Rather, they are part of the cognitive/social environment in which humans communicate. Hence they are exemplified in the linguistic experience of the learner and emerge in the learner's grammar in the course of learning.

In order to explore these issues constructively we have been developing a computational simulation of a language learner, called *CAMILLE* (*Conservative Attentive Minimalist Language Learner*). The idea behind this simulation is to endow a learner with strictly general computational capacities for identifying patterns, expose it to data about a natural language, and see what it is able or unable to accomplish.

Assuming that the simulation itself is well-constructed, there are two types of outcomes that are useful, success and failure. If the learner is successful, we have a demonstration that a learner with a particular computational capacity is able to formulate correctly hypotheses about the grammar of the language without the benefit of specific a priori knowledge about the structures. If the learner is not successful, we have reason to believe that some a priori knowledge may be necessary in order for learning to take place.

To make the discussion more concrete, consider the rule of wh-movement, which derives wh-questions in languages like English.

- (3) Who_i did you call ____i?
 What_i are you talking about ____i?

If the simulation is able to learn such a rule on the basis of exemplars in which it has applied without there being specific knowledge built into the learner that such a rule is possible, this constitutes the basis for an argument that this knowledge does not have to be part of the language faculty. On the other hand, if the simulation is unable to learn wh-movement without knowing that in principle a language may have such a rule, then that consists the basis of an argument that such knowledge must be part of the language faculty.

Of course, in practice matters are typically not as straightforward as this, and the reasons for success or failure may not be of the sort that will allow us to draw firm conclusions about the architecture of the language faculty. Nonetheless, a computational simulation holds out the promise of allowing us to determine, for each putative component of the language faculty, whether it is necessary for the successful acquisition

⁵ Everett 2005 argues that the Amazon language Pirahã lacks many of the expressive capacities of other languages.

of knowledge of language, or whether it can be dispensed with in favor of general computational principles that are not specific to language.⁶

In this paper we describe the basic architecture and capacities of our simulation, *CaMiLLe*, and summarize what it is able to do and what it is not able to do. Because *CaMiLLe* is a simulation of a **minimalist** learner, as its name suggests, it has little prior knowledge about the structure of language. On the basis of its successes and failures, we draw some tentative conclusions about the architecture of the language faculty, arguing that it must have some specific knowledge of linguistic structure beyond what we have endowed our computational simulation with, although perhaps not as much as is often claimed in the literature.⁷ Moreover, on the basis of the apparent successes of our minimalist learner, we offer a hypothesis about the nature of early language development that is to some extent consonant with those who have argued against a highly structured language faculty.

1. Grammatical preliminaries

We adopt an overall perspective on grammar that addresses not only the very general and universal or quasi-universal phenomena that are found in natural language, but also the idiosyncratic and exceptional (see Culicover 1999; Jackendoff 2002; Culicover and Jackendoff 2005). Our view about the goals of syntactic theory is the following, from Culicover and Jackendoff 2005:

Simple(r) Syntax Hypothesis (SSH):

The most explanatory syntactic theory is one that imputes the minimum syntactic structure necessary to mediate between sound and meaning.

On this view,

- The job of grammar is to describe the sound-meaning correspondences.
- Some of these correspondences are unanalyzable; that is, they are individual words that correspond to primitive concepts.
- Some have linguistic structure but are simple or not entirely transparent on the meaning side (idioms) (i.e. there are no nice structure/meaning matchups).
- Some have structure and are transparent on the meaning side (i.e. there is a compositional semantics that interprets canonical phrase structure).
- Some are a combination of the above ('constructions'), ranging from quasi-idioms, double-objects, movement along a path expressions, 'syntactic nuts' (see

⁶ Moreover, a simulation can be very helpful in investigating the behavior of a very complex system. Admittedly, it is sometimes possible to make analytic arguments for the necessity of some mechanisms. But it is well documented that simple rules interacting with each other may result in the emergence of unexpected properties that can be investigated only through computational simulation.

⁷ Culicover and Nowak 2003 offers a detailed discussion of *CAMiLLe*'s design and some preliminary conclusions regarding the architecture of the language faculty based on its performance. Our conclusions here are based on those of Culicover and Nowak 2003 but go beyond them in a number of respects.

above), various operator-trace binding constructions, etc. Each has some degree of predictability and generality, and some idiosyncrasies.

This approach to grammar is a ‘constructionalist one’, in two senses. On the one hand, it assumes that in some cases the best account of the sound/meaning correspondence is one in which meaning is not determined compositionally by the individual words. On the other hand, it assumes that the grammatical knowledge of a language learner is to some extent constructed on the basis of evidence, and is not predetermined.⁸

The evidence that a more nuanced approach to the sound/meaning correspondence is plausible is the following.

First, many words are unanalyzable correspondences between sound and meaning. Some (e.g. Hale and Keyser 2002) have argued that words with complex meanings are syntactically complex and are the product of derivations involving movement and deletion.⁹ However, Culicover and Jackendoff 2005 show that the full range of lexical phenomena requires that the morphological and semantic idiosyncrasies of words be irreducible – they must be stated explicitly and individually in any characterization of grammatical knowledge, and cannot be derived from general principles.

Second, idioms have recognizable syntactic structure but unpredictable meaning, and there are vast numbers of non-idiomatic but nevertheless not strictly transparent expressions in natural languages whose meanings have to be at least in part explicitly associated with them. Some typical examples that suggest the range of possibilities are the following; they can be multiplied almost endlessly.

- (4) by and large
 lo and behold
 beat a dead horse
 make amends
 cast aspersions on (*at / *to)
 a flash in the pan
 put up with
 have a problem with

⁸ Here we have in mind a variant of the view expressed by Quartz and Sejnowski 1997 as a ‘Constructionist Manifesto’: “In contrast to learning as selective induction, the central component of the constructivist model is that it does not involve a search through an a priori defined hypothesis space, and so is not an instance of model-based estimation, or parametric regression. Instead, the constructivist learner builds this hypothesis as a process of activity-dependent construction of the representations that underlie mature skills.”

⁹ Typical cases are words such as the verb (*to*) *shelve*, which means ‘put on a shelf’. The issue is whether there is a syntactic representation that contains the formatives *put* and *on* that maps into this meaning, or whether the meaning is directly associated with the lexical entry of the verb.

Go Bucks!

Third, there are numerous constructional idioms that have partially transparent interpretations whose meanings are in part associated with the entire structure.

- (5) a. *Way*-construction (Jackendoff 1990, Goldberg 1995):
Elmer hobbled/laughed/joked his way to the bank.
(Lit. ‘Elmer went/made his way to the bank hobbling/laughing /joking.’)
- b. *Time-away* construction (Jackendoff 1997):
Hermione slept/drank/sewed/programmed three whole evenings away.
(Lit. ‘Hermione spent three whole evenings sleeping/drinking/sewing /programming.’)
- c. *Sound+motion* construction (Levin and Rappaport Hovav 1995):
The car whizzed/rumbled/squealed past Harry.
(Lit. ‘The car went past Harry, making whizzing/rumbling/squealing noises.’)
- d. *Resultative* construction
The chef cooked the pot black.
(Lit. ‘The chef made the pot black by cooking in/with it.’)

The constructions in (5) share the same basic syntax (not surprisingly, since they are all English VPs); what is idiosyncratic is the way in which their meanings are related to the meanings of the parts and to the structure in which they (the parts) appear.

Finally, there are the general rules of language, such as those expressed by phrase structure rules like $VP \rightarrow V NP$, where it may be presumed that there is a corresponding rule of interpretation that composes the interpretation of the head with the interpretation of the argument to form an interpretation of the phrase.

Given this range of sound/meaning correspondences that a learner must acquire, the question naturally arises, How does the learner know where on the spectrum a given correspondence falls? What is it about a particular piece of linguistic experience that tells the learner that it is an idiom, or an expression with some idiosyncratic meaning components, or a general construction, or a fully general rule? In our view, the answer is that there is no way *a priori* for the learner to know where on the spectrum a correspondence really is. The conservative strategy is to start at the word/idiom end, and then move away from the maximally specific as the weight of the evidence warrants generalization.¹⁰

Our general view can thus be summarized as the following: **Construction of language produces constructions in language.** This means that as knowledge of language is constructed dynamically by a learner, what emerges are constructions that

¹⁰ Tomasello 2003 argues that this is the way that language learners in fact proceed.

may ultimately become ‘rules’, but only if given enough evidence and a suitable generalization mechanism; otherwise, they remain constructions.

Our simulation of language acquisition thus explores the question of what specific prior knowledge of language the learner **requires** in order to be able to acquire the full range of grammatical phenomena found in a language. Note that we emphasize the word ‘requires’. We can, if we choose, build into a learner specific knowledge about some grammatical phenomenon, and tell the learner how to identify whether a given language contains this phenomenon. It does not necessarily follow that a learner will be able to correctly identify that the language in fact contains this phenomenon.¹¹ But if a learner can perform the identification, this does not mean that the specific knowledge is necessary. Since the crucial question for us is what must be part of the language faculty, the way to approach the question is to begin by assuming that the learner’s prior knowledge is not specific and see what kinds of failures, if any, this assumption produces.

2. *CAMiLLe*

Our computational simulation, *CAMiLLe*, is conservative, in the sense that it does not generalize beyond what the evidence justifies. Hence it is different from a learner that chooses a grammar from a set of predetermined alternatives on the basis of selected ‘triggering’ data, as in the Principles and Parameters idealization of language acquisition.¹² Along related lines, *CAMiLLe* is attentive, in that every piece of data is relevant to the construction of a grammatical hypothesis, and not just particular triggering data. *CAMiLLe* is minimalist, in the sense that it is endowed with the minimal knowledge about linguistic structure that will allow it to form any hypotheses at all, and no more.

CAMiLLe’s task is to acquire a set of form/meaning correspondences. The data that *CAMiLLe* is exposed to consists entirely of pairs of sentences and conceptual structure representations. We assume that the sentences are strings of words and formatives, and the meanings are expressions in a simple attribute-value language.¹³ E.g.,

(6) John touch ~s the cat = TOUCH (\$AGENT:MAN, \$THEME:ANIMAL, \$TIME:NOW)

Relations that are typically expressed by verbs are represented as constants with an associated argument structure (e.g. TOUCH). Arguments are given as thematic roles with

¹¹ See, for example, Wexler and Culicover 1980, Gibson and Wexler 1994, Berwick and Niyogi 1996.

¹² See Fodor 1998.

¹³ It is important to note that even the written input to *CAMiLLe* is a significant idealization and simplification of what is actually presented to a human learner in a real learning context. One of the most salient differences is that the written input is segmented into words, but it is not in the real input. Among other things, *CAMiLLe* does not have to deal with variations among speakers, false starts, and environmental sounds.

their values (like $\$AGENT:MAN$). We assume that the meaning that *CAMiLLe* is presented with contains only primitives that are cognitively accessible to *CAMiLLe* at a given stage of development. For example, *John touches the cat* could have the meaning shown in (6) at an early stage, or even just $ANIMAL$ at an even earlier stage. Meanings may become more sophisticated as a consequence of development of cognition and perception. E.g., later, the learner may perceive that there is John, a distinct male person, that there is a particular type of animal (a cat), that both are singular in this context, and that they participate in this relation.

- (7) $TOUCH(\$AGENT:JOHN(\$TYPE:PERSON, \$GENDER:MALE, \$NUM:SG),$
 $\$THEME:CAT(\$TYPE:ANIMAL, \$NUM:SG), \$TIME:NOW)$

CAMiLLe tries to figure out how the parts of the string of words corresponds to the parts of the meaning. *CAMiLLe* does not know whether each word in the string is independently meaningful, or whether there are parts of the string that are idiomatic, in that the words together correspond to a single unanalyzed meaning. So at the outset, *CaMiLLe* hypothesizes all possible correspondences between the string of words and the meaning.

To illustrate, suppose that we expose *CAMiLLe* to the pair in (8).

- (8) that's a bunny = $BUNNY(\$TYPE:ANIMAL)$

CAMiLLe will form all of the possible hypotheses to account for the meaning $BUNNY(\$TYPE:ANIMAL)$. In this case there are six such correspondences.

- (9) 1. that's a bunny $\Leftrightarrow BUNNY(\$TYPE:ANIMAL)$
 2. that's a $\Leftrightarrow BUNNY(\$TYPE:ANIMAL)$
 3. a bunny $\Leftrightarrow BUNNY(\$TYPE:ANIMAL)$
 4. bunny $\Leftrightarrow BUNNY(\$TYPE:ANIMAL)$
 5. that's $\Leftrightarrow BUNNY(\$TYPE:ANIMAL)$
 6. a $\Leftrightarrow BUNNY(\$TYPE:ANIMAL)$

In each case but the first, some of the string is taken to be meaningful and the rest is treated as noise.

Each of these rules has an equal weight (.166) when it is first created. However, there will be other sentences with the word *bunny* and the corresponding meaning $BUNNY(\$TYPE:ANIMAL)$ (such as, *Do you want to pet the bunny?*). The more specific the rule is, the harder it is to support it further because it is more likely to be inconsistent with future experience, unless of course it is an exactly correct hypothesis about an idiom. So rule 1 will be lost unless the learner experiences many instances of *That's a bunny*, and the weight of rule 2 relative to the total number of examples exemplifying $BUNNY(\$TYPE:ANIMAL)$ will decrease over time while the rules 3 and 4 will grow. If there are examples with *the bunny*, *that bunny* and *this bunny* in the strings that the learner is exposed to, and $BUNNY(\$TYPE:ANIMAL)$ in the meanings, then the strongest correspondence will be expressed by rule 4.

The results of an experiment in which Camille is presented with the sentences in (10) containing the word *bunny* are given in (11).

- (10) that's a bunny = BUNNY (\$TYPE:ANIMAL)
 that's a nice bunny = BUNNY (\$TYPE:ANIMAL)
 see the bunny = BUNNY (\$TYPE:ANIMAL)
 that bunny is very soft, yes = BUNNY (\$TYPE:ANIMAL)
 do you want a bunny ? = BUNNY (\$TYPE:ANIMAL)
 show me the bunny = BUNNY (\$TYPE:ANIMAL)
- (11) 1. [15] BUNNY (\$TYPE:ANIMAL) ⇔ bunny
 2. [6] BUNNY (\$TYPE:ANIMAL) ⇔ a bunny
 3. [2] BUNNY (\$TYPE:ANIMAL) ⇔ that's+1->a

Bunny appears in every sentence, the string *a bunny* appears twice, *that's a* appears twice, and *the bunny* appears twice. The first three of these are hypothesized to possibly correspond to the meaning BUNNY (\$TYPE:ANIMAL). As the input to the learner becomes more complex and more diverse, many such hypotheses are formed and entertained.

In our implementation of *CAMILLE* it is possible to limit the number of rules that are entertained at any one time. This allows us to filter out highly implausible correspondence rules when there are more plausible alternatives available. This feature of *CAMILLE* may be viewed as a variant of the idea of markedness discussed in Chomsky 1965, whereby less complex rules are favored over more complex rules, other things being equal. In our case, the measure of markedness is simply the weight of the rule that *CAMILLE* acquires through experience with positive exemplars.

If there are two words that appear in identical linguistic expressions and the meanings of these two expressions are identical except for the meanings of the two words, *CAMILLE* will form a category consisting of the two words. Given the input in (12), *CAMILLE* forms the rules in (13).

- (12) eat bunny = EAT (\$THEME:BUNNY)
 eat doggie = EAT (\$THEME:DOG)
 eat kitty = EAT (\$THEME:CAT)
 eat bunny = EAT (\$THEME:BUNNY)
 eat doggie = EAT (\$THEME:DOG)
 eat kitty = EAT (\$THEME:CAT)
- (13) 1. [62] EAT (\$THEME:[BUNNY; CAT; DOG;])
 ⇔ eat+1->[bunny; doggie; kitty;]
 2. [51] EAT ⇔ eat

These 'single difference' rules can be formed even when there are several differences in a string. For example, if we have *Kitty eat bunny* and *Doggie eat kitty*, then if there are

enough examples, *bunny* and *kitty* can be put into the same category on the basis of their co-occurrence with *eat*.¹⁴ It is not surprising to learn that such distributional evidence is neither necessary nor sufficient for accurate category formation; however, the implications of this observation are far from trivial, as we discuss in §4.3.

3. Templates

3.1. A minimum condition for finding minimal structure

As in the case of *eat bunny*, etc., when presented with more complex data, *CAMiLLe* is able to separate the constants from the variables. *CAMiLLe* was exposed to naturally occurring English spoken to children from the Childe database (MacWhinney 1995). Some examples of the correspondences that *CAMiLLe* forms are shown in (14).

- (14) a. GO(\$AGENT:WE) ⇔ are<-1->we going+1->to
 b. GO(\$AGENT:[WE; YOU;])
 ⇔ are<-1->[we; you;] are+1->going going+1->to
 c. THINK(\$EXPERIENCER:I) ⇔ 1.I 2.think 3.that
 d. BE(\$THEME:[WHAT; WHO;]) ⇔ 1.[what; who;] 2.is 3.that 4.?
 e. BE(\$THEME:[HE; HERE; IT; THERE; THIS;])
 ⇔ 1.[he; here; it; there; this;] 2.is 3.a
 f. [BABY; BALL; BED; BOOK; BOY; BUNNY; CAR; CHAIR; COOKIE;
 DUCK; HOUSE; NOSE; ONE; THAT; THIS; TRUCK;](\$REF:[\$DEF;
 \$INDEF;])
 ⇔ [a; that;] [?; baby; ball; bed; book; boy; bunny; car; chair; cookie;
 duck; house; nose; truck;]
 g. BE(\$PRED:[BOX; BUNNY; COLOR; DARK; FACE; FUNNY; GOOD;
 HOUSE; IT; LETTER; ONE; RIGHT; ROOM; TAPERECORDER; THAT;
 THERE; TOOTH; YOU;])
 ⇔ [box; bunny; face; funny; good; house; it; letters; one; right; room;
 tape; teeth; there; what;]+1->.

The rules in (14a,b) show that *CAMiLLe* has extracted the essential correspondences of *we are going to* and *are we going to*. The notation *are<-1->we* means that *are* and *we* appear adjacent to one another in both orders. (14b) shows that *we* and *you* have a similar distribution and so form a small category with respect to these expressions. These are typical examples of what we call **templates**, that is, restricted expressions with variable slots that correspond to particular meanings.

¹⁴ While it is possible to get nice results when the data is constructed by hand, as it is here, the kind of distribution illustrated in (12) does not occur in naturally occurring speech to language learners. *CAMiLLe* does not take into account similarity of meaning; hence even if *CAMiLLe* knows *eat X* is possible when *X* refers to an animal, the fact that *pig* is an animal does not allow it to hypothesize that *eat pig* is possible without actually encountering *eat pig*. This is a matter of implementation, not principle.

Example (14c) shows the emergence of another template, *I think that*. Example (14d) shows a fixed expression, *what/who is that?* and its corresponding meaning. Example (14e) illustrates the template *X is a* where *X* is a pronoun. Example (14f) is the template for [*a, that*] *X*, where *X* is a noun and the meaning is annotated for definiteness.¹⁵

This last template illustrates the fact that *CAMILLE* is capable of correlating properties of a concept (e.g. definiteness of a object) with specifiers and modifiers of the head. In order for *CAMILLE* to be able to do this it is critical that there be prior knowledge that such a relation may exist. The relation in question is one in which the specifier/modifer-head relation in the syntax corresponds to an attribute-head relation in the meaning. Our experiments in the early development of *CAMILLE* suggested that without the knowledge that these relations exist and that there are correspondences between them, *CAMILLE* cannot discover them. On this basis, we posit our first assumption about the architecture of the language faculty.

Architectural Feature 1. There are corresponding specifier/modifer-head relations in the syntax and attribute-head relations in the meaning.

Finally, in (14g) *CAMILLE* has hypothesized that a noun in sentence-final position (immediately before ‘.’) is interpreted as predicational. Hence if asked to produce an utterance with the meaning “That’s a bunny”, *CAMILLE* would simply say “Bunny.” (and if possible, point).

In general we find these results to be typical of *CAMILLE*’s behavior. In the face of the very diverse input found in naturally occurring talk to children, *CAMILLE* forms numerous correspondences of this type. If presented with constructed input that more systematically reveals grammatical relationships in a language, *CAMILLE* is capable of extracting more sophisticated templates.

We have presented *CAMILLE* with constructed input for several reasons. First, limited samples of naturally occurring speech to children may not provide sufficient examples for *CAMILLE* to be able to form a reasonable hypothesis. Second, it is technically difficult to provide satisfactory meanings for large amounts of naturally occurring speech to children. Third, some relationships require morphological analyses that are not available in transcripts of naturally occurring speech. Fourth, our implementation of *CAMILLE* does not provide it with the capacity to construct sufficiently general categories that can form the basis of general rules. Since these are all simply a matter of implementation and not of principle, construction of the input data allows us to explore *CAMILLE*’s capacities more effectively. It is critical that if

¹⁵ Of interest is the fact that *CAMILLE* includes the question mark ? in the possible strings. We attribute this to the fact that there are many questions of the form *What is that? Is that a bunny?*, etc. in the input. Owing to the way that meanings were assigned to the strings, there is a degree of error in the input that leads *CAMILLE* to formulate correct hypotheses (for *CaMILLE*) that appear to us to be errors.

CAMiLLe fails under these highly controlled circumstances, we are able to draw the firmest conclusions about what such a learner can and cannot do.

In general, we find that *CAMiLLe* is capable of learning those relations that are strictly local in the string. For example, an imperative in English typically lacks an overt subject adjacent to the verb, a question with inversion has the first auxiliary verb immediately preceding the subject instead of following it, and so on. Here we consider in somewhat greater detail what *CAMiLLe* does when confronted with examples of these constructions.

3.2. Imperatives

Superficially, an imperative sentence in English is of the form

- (15) V ...
[e.g. pet the doggie!]

Typically, the imperative lacks a subject. The form of the verb is virtually identical to the form that is used in the non-third person singular present tense, with the exception of *be*.

- (16) Be quiet!
*Am/are quiet!

It is questionable whether a learner is aware of either characteristic during the earliest stage of learning. Given the overwhelming number of imperatives in speech to children, it would not be surprising if learners hypothesized that the citation form of a verb is its form in the imperative. It is plausible that at some point a learner becomes aware that imperatives differ from their declarative counterparts in that they lack a subject NP in the position where an NP might normally appear. At some later point, the learner would become aware that the form of the verb is the ‘bare’ form, in contrast with inflected forms in the paradigm.

We simulated the effects of assuming different sequencing of the analysis of the input data to the part of the learner, corresponding to different hypotheses about what the learner is capable of understanding about the structure of the input.

We presented *CAMiLLe* with a set of positive declarative and imperative sentences, with enough information so that the program could confidently identify the meanings of the noun phrases referring to actors, the verbs, and negation.

First we gave *CAMiLLe* examples of positive imperatives, such as

- (17) be quiet! = \$IMP (BE (\$THEME:YOU, \$PRED:QUIET))

The results are the following rules.

- (18) a. \$IMP([BUY; FIX; GIVE; GO; KEEP; LISTEN; MAKE; RECYCLE;
SECURE; SELL; SING; SKI; SPY; WATCH;])

- ⇔ 1.[buy; fix; give; go; keep; listen; make; recycle; secure; sell; sing; ski; spy; watch;]
- b. [BUY; FIX; GIVE; GO; KEEP; LISTEN; MAKE; RECYCLE; SECURE; SELL; SING; SKI; SPY; WATCH;] (\$AGENT:YOU)
- ⇔ 1.[buy; fix; give; go; keep; listen; make; recycle; secure; sell; sing; ski; spy; watch;]

Representing the verbs list in square brackets as the category V, rule (18a) says that V in first position corresponds to $\$IMP(V)$, where V is the meaning associated with V. Rule (18b) says that a verb in first position corresponds to a meaning in which the agent is YOU. Both of these rules are correct empirical generalizations. Neither requires that there be an empty subject represented in the input to the learner, nor does the learner posit a virtual empty subject as it computes the correspondence. In other words, *CaMiLLe* acquires the imperative as a construction.

Next we presented *CAMiLLe* with negative imperatives of the form *Don't be*. The correspondence rules are as follows.

- (19) a. [BUY; DESTROY; EXTRACT; FEAR; FIX; GO; INVITE; KILL; MARRY; READ; SELL; SING; SPY;] (\$AGENT:YOU)
- ⇔ 2.[buy; destroy; extract; fear; fix; go; invite; kill; marry; read; sell; sing; spy;] don't don't+1->[buy; destroy; extract; fear; fix; go; invite; kill; marry; read; sell; sing; spy;]
- b. \$NEG([BUY; DESTROY; EXTRACT; FEAR; FIX; GO; INVITE; KILL; MARRY; SELL; SING; SMELL; SPY;])
- ⇔ 2.[buy; destroy; extract; fear; fix; go; invite; kill; marry; sell; sing; smell; spy;] don't don't+1->[buy; destroy; extract; fear; fix; go; invite; kill; marry; sell; sing; smell; spy;]
- c. \$NEG ⇔ 1.don't
- d. \$IMP(*NULL*:\$NEG) ⇔ 1.don't
- e. \$IMP ⇔ 1.don't

What these rules say is that the V in second position may have the interpretation \$AGENT:YOU, and *don't* immediately preceding second position V has the negative interpretation scoping over the interpretation of the V. Rules c-e express the correspondences of initial *don't* with \$NEG, \$IMP(\$NEG, and \$IMP.

If we mix the positive and negative imperatives, *CAMiLLe* constructs all of the preceding correspondences.

3.3. Inversion

Inversion occurs in English yes-no and wh-questions, and in some less frequent constructions. *CAMiLLe* was presented with sets of sentences in which the meaning of the sentence with inversion contained a representation of the fact that it is a question. To simplify *CAMiLLe*'s processing, simple subjects consisting of one or two words were used.

In the analysis of inversion in contemporary linguistic theory, the observation has been made that the auxiliary verb moves into an empty head position that is also the position occupied by the complementizer *that* in embedded sentences. The structure is that of (20), where the head of IP is Tense. In order for the auxiliary verb to move into initial position, it must therefore first move to Tense. This style of analysis is called ‘head-to-head movement’.

$$(20) \begin{array}{l} [_{CP} C [_{IP} DP \text{ Tense } [_{VP} V \dots]]] \Rightarrow \\ [_{CP} C [_{IP} DP V+\text{Tense } [_{VP} __ \dots]]] \Rightarrow \\ [_{CP} V+\text{Tense } [_{IP} DP __ [_{VP} \dots]]] \end{array}$$

Our experiments show that *CAMiLLe* deals with inversion simply by correlating the initial position of the auxiliary verb with the interrogative interpretation. In the first experiment, we did not provide *CAMiLLe* with information about the morphological structure of the verb, and for simplicity used only the verb *is*.

$$(21) \text{\$YNQ}(*\text{NULL*}:\text{BE}) \Leftrightarrow 1.\text{is}$$

When we introduce *do/does* into the data, *CAMiLLe* determines that the auxiliary in first position and the verb in third position correlate with the interrogative interpretation.

$$(22) \text{\$YNQ}(*\text{NULL*}:[\text{LIKE}; \text{PLAY};]) \Leftrightarrow 1.\text{does } 3.[\text{like}; \text{play};]$$

In the second experiment, we provided *CAMiLLe* with information about the morphology – the tense inflection is represented as a separate element in the string. In the case of inversion, the sequence *V tense* is in sentence-initial position, so *V* is in first position and tense is in second position. *CAMiLLe*’s correspondence rules reflect these generalizations.

$$(23) \begin{array}{l} \text{\$YNQ} \Leftrightarrow 2.\sim\text{tense} \\ \text{\$YNQ}(*\text{NULL*}:\text{HATE}) \Leftrightarrow 2.\sim\text{tense } 4.\text{hate} \end{array}$$

As long as *CAMiLLe* pays attention to position in the string relative to the beginning of the string, templates like those in (23) will be formed. Within linguistic theory, position in the string is not linguistically significant, unless it is first or second position.¹⁶ Certainly mention of fourth position *per se* does not appear to have linguistic relevance.¹⁷ It is also possible to represent the template in (23) equally in terms of relative position, where *~tense* is two to the left of *hate*. But in more complex data sets, generalizations in terms of cardinal position cannot be sustained because there is too

¹⁶ For some recent proposals regarding what constitutes second position in a sentence, see the papers in Halpern and Zwicky 1996.

¹⁷ Any sentence will have a first position (and implicitly, a second position). But not every sentence will have a fourth position.

much variability in the position of the elements with respect to the beginning of the string, and too much intervening material of indeterminate length.

In this case, use of string position allows *CAMILLE* to capture accurate generalizations about subsets of possible strings, but not about the language as a whole. In fact, the standard approach to teaching syntactic theory begins by demonstrating that there are no valid generalizations about language that mention absolute position (except perhaps first and second), because of the fact that there are phrases within the string whose length cannot be bounded. But, we suggest, such templates based on a restricted subset of the language may be correct characterizations of learners' early hypotheses about the language (see Tomasello 2003).

The failure of *CAMILLE* to capture generalizations about complex data in terms of cardinal position comes as no great surprise. However, we do want *CAMILLE* to be able to recognize second position, and also to recognize adjacency (one position to the left or the right). Thus we gave *CAMILLE* the capacity to count, expecting that hypotheses formulated in terms of absolute or relative position based on counting would eventually disappear, and generalizations that do not involve counting would emerge. *CAMILLE* did demonstrate that as long it does not have to deal with variable length phrases, it is capable of formulating relatively narrow but serviceable template correspondences in terms of position.

If the learner also has the capacity to generalize over variable length substrings (that is, phrases), then the templates may contain variables. Consider the string

(24) 1.does 3.[like; play;]

from (22). If position 2 can be an NP of arbitrary complexity, and position 3 is generalized to all verbs, then this template is adequate for a substantial body of the grammatical cases. The cases that are presumably excluded are the NPs that contain relative clauses, since these presuppose not just that the phrasal category NP has been identified, but that VP or S have been identified as possible constituents of larger phrases. If *CAMILLE* could process up to the NP level of structure, then we would be able to claim with some justification that *CAMILLE* is a realistic model of acquisition by an early learner, one very similar to that described by Tomasello (2003). In many respects this 'phrasal' *CAMILLE* would give the impression of knowing certain constructions of English. In actual fact *CAMILLE* would only have acquired templates with NP variables that give the illusion that it has formulated grammatical rules such as inversion. E.g., *CAMILLE* would simply have a template for yes-no questions in which the auxiliary verb is positioned in initial position, to the left of the subject NP.

It can be argued that a complete set of such templates, that is, constructions in the sense discussed earlier, if generalized sufficiently, is sufficient to give the impression that a learner has acquired the grammar of a language. If the templates are sufficiently elaborated, it may be that constructing them is in fact extensionally equivalent to having acquired the grammar, thereby opening the question of whether what is acquired is a grammar in the more traditional sense (see Culicover 1999). But a closer examination of

CAMiLLe's limitations, to which we turn next, shows that the templates that would be required to demonstrate knowledge of English require capacities that go beyond what *CAMiLLe* is presently capable of. Identification of these capacities is critical to our goal of determining what must be in the language faculty.

4. Some limitations and their significance

While there are many things that *CAMiLLe* does not do, either for principled or practical reasons, we highlight three here as central to our investigation into what a learner must know in order to acquire knowledge of a natural language:

- *CAMiLLe* does not form general phrase structure rules of the type $VP \rightarrow V NP$.
- *CAMiLLe* does not identify filler-gap relations, e.g. between a fronted wh-phrase and its corresponding gap.
- *CAMiLLe* does not form supercategories, grouping all of the nouns into one category, all the verbs into another category, and so on.

4.1. From templates to rules

In our discussion of templates above we contrasted the situation where what is learned is a fixed string of constant forms (corresponding to some meaning), and a string of constants that contains one or more variables. An example of each is given in (25).

- (25) a. gimme that
b. gimme NP

At first glance, it might appear that making the transition from a fixed template to one with variables would be straightforward for a learner like *CAMiLLe*. Suppose that instead of *gimme that*, the learner hears a large number of expressions like *gimme the book*, *gimme a kiss*, *gimme a ball*, *gimme a red hat* and so on. If the meaning representation contains simply an element that corresponds to the head noun, then the correspondence rules will identify *gimme a* as a constant string, and the noun will be the variable. On the basis of such very systematic input *CAMiLLe* will hypothesize a number of plausible rules, including the template *gimme a N*. The following shows the results of an experiment on input of the form *gimme a N*.

- (26) \$IMP (*NULL*:GIVE) ⇔ gimme
 \$IMP (*NULL*:GIVE) ⇔ gimme+1->a
 GIVE (\$AGENT:YOU) ⇔ gimme
 GIVE (\$RECIPIENT:ME) ⇔ gimme
 GIVE (\$AGENT:YOU) ⇔ gimme+1->a
 GIVE (\$RECIPIENT:ME) ⇔ gimme+1->a
 GIVE (\$THEME:[BOOK; KISS; PEAR; PENCIL;])
 ⇔ gimme+2->[ball; kiss; pear; pencil;] a+1->[ball; kiss; pear; pencil;]

CAMiLLe in fact associates the meaning $GIVE(\$AGENT:YOU, \$RECIPIENT:ME)$ with *gimme (a)*, and picks out the word to the right of *a* as corresponding to the $\$THEME$.

The next natural step would appear to be one in which sequences of the form *a book*, *a kiss*, *a pear*, and so on are recognized as units, that is, as phrases. Let us suppose for the sake of illustration that this occurs when the meaning of *a* is known. On the basis of recognizing that *a* contributes to the meaning, the phrase would be parsed into the head and the correspondence between *a* and its meaning checked off. That is, *CAMILLE* would carry out the following reduction.

(27) *gimme a ball* =
`$IMP (GIVE ($AGENT:YOU, $RECIPIENT:ME, $THEME:BALL ($REF:$INDEF)))`
 \Rightarrow
gimme ball =
`$IMP (GIVE ($AGENT:YOU, $RECIPIENT:ME, $THEME:BALL ($REF:$INDEF)))`

Then the more general correspondence would be formed of the form *gimme N*.

We say ‘*CAMILLE* would’ because in fact *CAMILLE* does not do this. The reason is instructive. On the basis of sentences of the form

(28) *gimme a N*

CAMILLE can identify the overt N as corresponding to the `$THEME` of `GIVE`, as we have seen. And *CAMILLE* is able to form a correspondence rule in which *a N* corresponds to `N ($REF:$INDEF)`. But there is nothing in the input tells *CAMILLE*, first, to treat *a N* as a unit headed by the N, and second, to take this abstract N, call it $N^@$, as corresponding to the `$THEME` of `GIVE`. While N is concretely present in the input, $N^@$ is not. It must be created by *CAMILLE*, and then *CAMILLE* must know what to do with it.

CAMILLE is not helped if we provide it with concrete information about other possible complements of *gimme*, e.g. *gimme that*, *gimme money*, etc. In the absence of the capacity to posit headed phrases, such input simply makes *CAMILLE* more confused about the combinatorial possibilities for *gimme*, since now it must deal with *gimme N* and *gimme a N*.

Similar problems arise if we ask *CAMILLE* to deal with phrases consisting of more than one specifier/modifier of a head, such as *the angry dog*. Suppose that *CAMILLE* knows that *angry dog* is an instance of an $N^@$. Given this, *CAMILLE* is then faced with the string *the dog[@]*, where the [@] here is our notation to indicate that this instance of *dog* is actually not original in the string but is arrived at by parsing *angry dog*. What *CAMILLE* needs to know now is that having parsed *angry dog*, it must now parse *the dog[@]*. Again, this is not something that is implicit in the computation of string/meaning correlations, and does not suffice even to form templates of the form C_i NP C_j for constant strings C_i and C_j . And it is not something that *CAMILLE* will figure out on its own.

Finally, suppose that we make *CAMiLLe* able to deal iteratively with the output of replacing a substring with a constant. In some cases, the result is not a well-formed sentence of the language. For example, if the input is *give a book to Chris*, and we parse *a book* into *book*, the resulting string is ungrammatical.

(29) *give book to Chris

But if *CAMiLLe* is able to treat this on a par with original input, *CAMiLLe* will acquire incorrect knowledge of language. Therefore *CAMiLLe* must be able to distinguish between original and derived strings.

On the other hand, (29) is well-formed if the direct object is a mass noun, like *money*. So while *CAMiLLe* should not use (29) as a basis for deciding whether a count noun can appear as the direct object without a specifier, it should use (29) to establish and strengthen the correspondence between *give NP* and $\text{GIVE}(\$THEME:NP)$.

In sum, in order to arrive at the appropriate generalization, *CAMiLLe* must be equipped with the following three features.

Architectural Feature 2. If there is a string MH and M corresponds to a modifier of H' , then MH can be replaced by H° in the string.

Architectural Feature 3. Process derived strings as though they are original strings.

Architectural Feature 4. Derived strings can be the basis for learning syntactic correspondences but not for learning the properties of lexical items.

Or, to put it another way, to go beyond rigid idioms and fixed templates a language learner needs to learn to parse the input. The parser manages the correspondence between sound and meaning at the point at which generalizations begin to emerge, such that some correspondences become nested within other correspondences.

There are many objections that can be raised against this observation, from different quarters. On the one hand, it might be objected that this conclusion, i.e. that *CAMiLLe* must be able to do parsing as well as pattern extraction and correlation, is a completely trivial one, since everyone knows that natural languages have this level of structure. Our response is to emphasize that by withholding this capacity from *CAMiLLe*, we are able to see what *CAMiLLe* can do without it. Without the ability to find structure, *CAMiLLe* can nevertheless acquire a set of correspondences that gives the appearance of knowing something about a language. We have suggested that this may be what very early learners are doing. Whether this means that they are not actually able to parse input at an early stage, or whether there is not enough evidence to tell them that parsing is necessary, is an open question.

Another objection is that we have not made *CAMiLLe* sophisticated enough, in comparison, for example, to machine learning approaches that have demonstrated the possibility of discovering linguistic structure through unsupervised learning. Yuret (1998) reports “I developed an unsupervised language acquisition program that learns to identify linguistic relations in a given sentence. The only linguistically represented linguistic

knowledge in the program is lexical attraction. There is no initial grammar or lexicon built in and the only input is raw text.”

Yuret’s program determines the cooccurrence properties of pairs of words in strings, and on this basis posits structure. Crucially, Yuret assumes that there exist syntactic relations in language, and that syntactic structure is a reflex of these relations. So the goal of his program is to discover the correlations that may be taken as evidence of syntactic relations: “Lexical attraction is the likelihood of a syntactic relation” (22).

His program finds likely dependencies within strings, and ranks them with respect to one another. The ranking in part is determined by the alternative structures that the links give rise to: if two links cross one another, the stronger one wins. Eliminating crossing links gives rise to a clean parse of the sentence where every word is an immediate constituent of a phrase, and the heads are linked to one another. The algorithm assumes right branching, since the language under investigation is English.

These characteristics of Yuret’s program are in fact particular realizations of the assumption that the language learner needs to be able to parse the input. Precisely how to do this depends on the properties of the input and other assumptions built into the learner, but the core assumption is that there is a structure with certain properties that needs to be discovered. And there is a general characterization of the properties of this structure that guides the construction of the parse.

In general, we suggest that it is impossible for a learner to get structure of the sort that occurs in natural language out of unstructured input unless the learner is looking for the structure and knows at least a minimum about what its properties are. The question for the theorist, it seems to us, is not whether the learner knows that there is structure, but how much the learner knows *a priori* about the properties of the structure. Our hypothesis about *CAMILLE*, which remains to be tested, is that the knowledge of structure embodied in Architectural Features 1-4 is sufficient for the idioms, constructions and phrase structure of a natural language, assuming that the morphology is properly dealt with.

4.2. Unbounded dependencies and gaps

A second area in which *CAMILLE* falls far short of the capacity of a human learner to acquire language involves sentences in which there is a long-distance dependency between two parts of the sentence. Before we discuss these in some detail, we will contrast them with local dependencies.

Not surprisingly, local dependencies are not a problem for *CAMILLE*, since in the simplest case they involve adjacency. We gave *CAMILLE* a set of sentences of the form

(30) (N-<1/2/3>-<sg/pl>) <1/2/3>-<sg/pl>-V ...

where the <1/2/3> is the person and <sg/pl> the number of the preceding N (which is the subject) that agrees with what is marked on the verb. Whether or not there is a subject,

the morphological number corresponds to a number feature in the meaning. The results are along the lines of (31).

- (31) SG \Leftrightarrow ~sg
 PL \Leftrightarrow ~pl
 [HE; I; YOU_SG;] (\$NUM:SG) \Leftrightarrow [\sim 1; \sim 2; \sim 3;]+1->~sg
 [THEY; WE; YOU_PL;] (\$NUM:PL) \Leftrightarrow [\sim 1; \sim 2; \sim 3;]+1->~pl

We would expect similar results where the adjective and determiner of NP agree with the N in feature like number, gender and case, as long as there is a correspondence with some feature of the meaning.

For unbounded dependencies we have two cases.

Case 1. There is a dependent overt element that agrees in some way with an antecedent. An example is left dislocation in English.

- (32) Sally_i, I would say that everyone thinks that she_i is a great teacher.

The key property of such a sentence is that the pronoun identifies the grammatical function and thus the semantic role played by the dislocated NP, while the NP provides the identity or ‘index’ of the individual. The NP is said to be in a non-argument or A’ position.¹⁸ The identity of the individual that has the semantic role identified by the pronoun cannot be determined unless the dislocated NP is linked to the pronoun.¹⁹ The sentence also has special discourse properties (Prince 1987, 1998). We mark the discourse function here as the feature \$_{DISC:TOPIC}, as illustrated in (33) for the sentence *Sally, she eats pizza*.

- (33) Sally-3-sg, she-3-sg eats pizza =
 EAT (\$AGENT:SALLY (\$DISC:TOPIC) , \$THEME:PIZZA)

CAMiLLe’s job in this case is to figure out on the basis of the morphological agreement that SALLY is the \$AGENT of EAT, and that moreover an NP in this topicalized position has the feature \$_{DISC:TOPIC}. Along with the simple example of (33), **CAMiLLe** is also presented with sentences in which there is no left dislocation (*Sally eats pizza*) and those in which there is non-adjacent left dislocation. In order to make the task as easy for **CAMiLLe** as possible, we used a wide variety of examples, all of which illustrate the point that the topicalized NP, *Sally* or *pizza*, has the topic discourse function.

¹⁸ Strictly speaking the NP in left dislocation has no grammatical function at all, while in a wh-question or topicalization it does have a grammatical function, although only in virtue of forming a chain with an empty position in the sentence.

¹⁹ We are adapting here the analysis developed in Culicover and Jackendoff 2005.

Even with this very redundant information, *CAMiLLe* is unable to form the generalization that the left dislocated phrase followed somewhere by a pronoun corresponds to the argument in CS indicated by the syntactic function of the pronoun. The closest *CAMiLLe* comes is the following.

- (34) a. PIZZA(\$DISC:TOPIC) ⇔ pizza -> ~3 -> ~sg ->it
 b. [EAT; WEAR;] (\$THEME:PIZZA)
 ⇔ pizza+4->[eat; wear;] ~3+3->[eat; wear;] ~sg+2->[eat; wear;] ~3+4->it ~sg+3->it [eat; wear;]+1->it

The first rule shows that there were enough examples containing the sequence *pizza-3-sg it* for *CAMiLLe* to hypothesize a correspondence between this sequence and the discourse function. The second rule shows that *CAMiLLe* was able to see that in the sequence *pizza-3-sg <wear/eat> it* the \$THEME role is assigned to PIZZA.

What *CAMiLLe* cannot see is that this possibility for interpreting *pizza* does not depend on fixed length expressions, but holds across arbitrarily long spans of a string. This observation takes us back to the discussion in §4.1. It appears that if *CAMiLLe* was able to treat arbitrarily long spans of string as though they were of fixed length, *CAMiLLe* would be able to deal with left dislocation. The way to make an arbitrarily long string be of fixed length is to iteratively reduce it to the heads of phrases by parsing out the adjuncts and arguments.

But, crucially, the pronoun that is to be linked to the left dislocated NP cannot simply be parsed like a normal argument, since it will be lost in the intermediate strings and not available at the end of the parse. The presence of the pronoun in the string has to be carried along in the parse. Let us work through a simple example. Suppose that the sentence is *Pizza, I would say that everyone likes it*. Assume for simplicity that V is the head of S. When the pronoun is parsed it is encoded as a feature on the verb, which is passed up through the parse. The sequence of reductions is shown in (35).

- (35) Pizza, I would say that everyone likes it
 Pizza, I would say that everyone likes-(it)
 Pizza, I would say that likes-(it)
 Pizza, I would say likes-(it)
 Pizza, I would say-(it)
 Pizza, I say-(it)
 Pizza, say-(it)

This technique is that of passing features through a parse tree proposed originally by Harman (1963), introduced by GPSG (Gazdar et al. 1985) and implemented quite generally in HPSG (Pollard and Sag 1992). As we can see, the pronoun will be either adjacent to the left dislocated constituent, or one element away from it, at some stage of the parse. If the parsing is done in this way by *CAMiLLe*, then a rule such as (34), suitably generalized, will suffice.

Let us suppose that this is the correct way to characterize left dislocation. The next question is, How does *CAMiLLe* acquire it? Clearly there are several characteristics of this construction that could tell a learner that there is something special going on: the fronted NP is not in a position where it gets assigned a grammatical role, and hence a semantic function, while there is a pronoun in the position that identifies the grammatical role.

However, although it is straightforward for us to characterize what is going on, *CAMiLLe* cannot figure out the correspondence without being afforded specific knowledge about how to deal with this type of construction. The example provides evidence that the learner must have (i) the capacity to recognize that an expression lacks a grammatical function and a corresponding thematic interpretation, (ii) the capacity to recognize that this expression must have a grammatical function and a corresponding interpretation, (iii) the capacity to recognize that a proform agrees with such an expression, and (iv) the ability to link the unincorporated expression with the proform. Points (i) and (ii) are related to what has been called the θ -Criterion in GB Theory, which is, informally, that every phrase in a sentence must have a grammatical function and be interpreted; points (iii) and (iv) constitute binding, in an informal sense.

Architectural Feature 5. The θ -Criterion must hold for all expressions of a language.

Architectural Feature 6. Binding may be used to satisfy the θ -Criterion.

As in the case of the features that we have already discussed, these do not come for free but must be built into the learner.

Case 2.

Although left dislocation is a problematic case of unbounded dependency for *CAMiLLe*, it is by no means the most problematic such case. The most familiar phenomenon of ‘unbounded movement’, in which the fronted constituent is bound to a gap, is far beyond *CAMiLLe*’s capacities.

The typical case is that of a simple wh-question, such as (36).

(36) What are you looking at ___ ?

The argument of *looking at* is not in its canonical position, it is in an A', sentence-initial position. The same considerations that led us to posit that the pronoun in left dislocation is carried through the parse leads to the conclusion that there must be a similar feature that identifies that there is a gap in the parsing of a sentence such as (36). This feature must be attached to the head of which it is an argument, and it must be carried through the parse, so that it can be bound by the fronted wh-phrase. Again, this is the approach to movement originally proposed by Harman (1963).

The problem for *CAMiLLe* is to identify the gap and determine that it has to be bound by the moved constituent. As before, some of the cases involve NPs that cannot be assigned a grammatical function; these must be bound to a missing NP position. One way of characterizing this relation is to say that there is a chain consisting of the moved

constituent and an invisible placeholder, a trace in GB Theory. Linking of the two elements in principle accounts for the fact that the moved constituent is interpreted as though it is in the position occupied by the trace.

Experiments with the current implementation of *CAMiLLe* show, not surprisingly, that *CAMiLLe* cannot do this. *CAMiLLe* has no idea that there can be gaps in a string that have some syntactic reality. There are a variety of technical means for representing the trace of movement,²⁰ but they are equivalent in the sense that none of them is something that *CAMiLLe* is able to invent simply on the basis of sound/meaning pairs. The notion that there can be a missing argument, and that a dislocated NP can supply this missing argument, is something that has to be built into *CAMiLLe*.

As is well known, the problem is actually somewhat more complex than the way that we have just characterized it, because of the fact that constituents other than argument NPs can be moved. The following examples illustrate the fact that prepositional phrases, adverbs, and adjectives can be moved.

- (37) a. *Under which table* did you find the money __?
 b. *How quickly* do you think they will let you know the results __?
 c. *How tall* is your child __?

The point for each of these is the same as it is for NPs; these phrases are in a position where their grammatical function cannot be determined. If they can be linked to a suitable empty position in the string, their grammatical function can be determined. So, using the notation of *t* for trace, (37b) must have an analysis equivalent to the following –

- (38) *How quickly*_{*i*} do you think [they will let you know the results *t*]

– where the trace is in the position occupied by an adverb modifying the verb phrase of the embedded clause. Finding the gap in such cases is non-trivial, because of the fact that the adverb is not selected by the verb and thus the gap cannot be projected locally – it must be projected as a function of the unincorporated fronted adverb.

None of this is news, and it is standard in the analysis of wh-questions in which movement is involved. What is crucial to the present discussion is that the capacity to recognize the fact that the fronted constituent is not incorporated, the capacity to posit the corresponding trace (or equivalent feature) in the relevant position, and the capacity to link the two into a chain are all things that do not appear to emerge naturally from simply finding correlations between patterns of strings and patterns of meaning. It appears that

²⁰ The standard approaches are movement (which leaves a trace as a copy), binding a empty NP, passing a feature corresponding to a trace up through a phrase marker, and passing a feature corresponding to the selectional requirements of a head up through a phrase marker.

knowledge that movement constructions may occur is something that has to be built into *CAMiLLe* and something that *CAMiLLe* has to be seeking, in order to be able to find it.²¹

We stress that this is not a matter of a particular implementation of the movement relation, or a particular way of representing the trace. It is more fundamental than that. The following must be assumed to be a part of *CAMiLLe*'s architecture.

Architectural Feature 7. Constituents may be in an A' position and form a chain with an agreeing gap.

4.3. Categories

The last limitation of *CAMiLLe* that we discuss here is how it forms lexical categories. The questions that we are faced with are whether *CAMiLLe* can determine what the lexical categories are, and what the membership of each category is, simply on the basis of the properties of the sound/meaning correspondences that are exemplified for it in the input.

It is generally assumed that languages have categories, such as Noun and Verb. These categories transcend semantic categories in that it is impossible to give a semantic criterion that is sufficient to identify a word as a Noun, or a Verb, etc.²² To take Noun, as an example, some nouns refer to things that can be individuated (*book, unicorn*), some refer to substances (*water, sincerity*), some refer to places or times (*New York, tomorrow*), some refer to properties or dimensions (*sincerity, height*), and so on.

The distributional criteria that *CAMiLLe* has available to it are (i) context in the string, (ii) morphological form.²³ Staying with the example of Noun, in a language like English a string context might be that of following a determiner or adjective, as in

- (39) the dog
silly kitty
two chipmunks
every student
her sincerity
his persistence

Another would be that the word serves as the subject of a sentence, e.g. –

²¹ J. Feldman (p.c.) points out that it is not necessary that this particular knowledge be built in explicitly. It is conceivable that some other knowledge can be built in that will permit *CAMiLLe* to discover the existence of chains.

²² This is a long-standing problem in the field. For a recent review of efforts to define syntactic categories in semantic terms, see Baker 2003. See Culicover 1999 and Croft 2001 for arguments that categories cannot be general but must be defined in terms specific to individual constructions.

²³ In practice we reduce (ii) to (i) because it simplifies the implementation.

- (40) John shaves.
Mary is singing.
Dogs bark.

or as the object of a sentence –

- (41) I admire sincerity.
 We like dogs.

Morphological cues to the category Noun, for English at least, would be the plural marking *-s*, which works for most count nouns but not for mass nouns or proper nouns.

While distributional and morphological criteria such as these can be used to define categories, they do not define a single category Noun. In fact, the situation is very much the same as what we find when we considered semantic criteria, only worse. For a given language, like English, the only morphological criterion is singular/plural, and it only serves to distinguish the count nouns. For a language like Russian, there are several classes of nouns that are morphologically distinct; hence the endings that appear on members of one class can be entirely different from the endings that appear on members of another class. For a language like Chinese, on the other hand, there are no morphological criteria that will distinguish nouns from verbs.

These considerations suggest that morphological criteria cannot in general be used to define syntactic categories. This leaves distributional criteria. Because of the count/mass distinction, only count nouns can appear in NPs with count quantifiers, like *every* and *two*, while only mass nouns and plural count nouns can appear with mass quantifiers, like *a lot of*. The indefinite determiner appears only with count nouns. The definite determiner *the* and possessives appear with count nouns and mass nouns, but they do not typically appear with proper nouns.

- (42) the dog
 her sincerity
 *the Mary
 *my Fred

It does appear that all nouns may be preceded by an adjective, in English, even proper nouns.

- (43) furry dogs
 characteristic sincerity
 long tall Dexter

However, this raises the question of the distributional criteria that can be used to define Adjective. The property of preceding a noun is obviously not workable, for reasons of circularity. Moreover, non-adjectives can appear in the same position, e.g., in compound noun constructions such as

- (44) military training
table manners
computer software

The property of being the complement of *be* is not sufficient, because noun and verbs can also appear in this position.

- (45) We are <happy/athletes/sleeping>.

And not all adjectives can appear in this position.

- (46) former mayor ~ *be former
perfect idiot ~ *be perfect [in the intended sense]

Summarizing to this point, it appears that at best the phrase-internal distributional criteria can be used to define subcategories, but are not sufficient to define supercategories such as Noun and Verb. The other distributional criteria that we might appeal to are connected to grammatical function. A noun phrase typically can function as the Subject of a sentence or as the complement of a verb or preposition. Assuming that we cannot use distributional criteria to determine that something is a verb or a preposition, we can appeal only to the knowledge that a given phrase has a particular grammatical function, or that it has a particular semantic interpretation.

Distributional and morphological criteria are further suspect because by definition they cannot be cross-linguistic. The intuition that all languages have the category Noun, for example, cannot be sustained by distributional tests because there can be no tests that are valid in more than one language. As much as these tests fail even for a single language, they cannot even be envisioned as applying across languages.

Having ruled out distributional, morphological, and semantic criteria, only grammatical function is left. If, for example, N is the head of a phrase that can be Subject (or Object, etc.), and only N can be the head of such a phrase, then we might have a basis for distinguishing N from other categories. This generalization is not true, however, because phrases with heads other than N can be Subjects.²⁴

- (47) [For you to do that] would bother me.
[That Sandy is rich] is obvious.
[Visiting relatives] turns out to be unpleasant.

But then the question arises, how can grammatical function be determined independent of the grammar having been acquired in the first place? For example, in

²⁴ It is possible to stipulate that these phrases have empty N heads. It would be impossible for *CAMILLE* to find this out through inspection of the evidence, but *CAMILLE* could make the inference that they do if confronted with overwhelming independent evidence that only NPs can be Subjects. Or *CAMILLE* could use the evidence to amend the generalization to allow for sentential and VP subjects.

order for *CAMILLE* to know that a word is a noun or that a phrase is an NP in virtue of its grammatical function, *CAMILLE* must be presented with the information that it is the Subject of the sentence. While it can be argued that such grammatical information must play a role in the mapping between sound and meaning (see Culicover and Jackendoff 2005), it has never been demonstrated, or even argued, that learners are presented with this type of information as part of their primary linguistic experience.

The question is, then, Can a learner such as *CAMILLE* determine, on the basis of semantic and distributional information, that a phrase has a particular grammatical function in a sentence? At this point we must leave this question open, since we do not see how to define the basic grammatical functions in semantic terms. Possibly Subject and Object are bootstrapped from core (or default) cases on the basis of meaning, and subsequently become syntactically autonomous. If this was possible, then on the basis of grammatical function *CAMILLE* could ultimately posit broader syntactic categories. And in principle it might be possible to account for the apparent appearance of the same categories across languages. But then we require an additional architectural feature.

Architectural Feature 7.

- a. There is an a priori set of grammatical functions.
- b. There is a default linking between semantic roles and grammatical functions.

If knowledge of grammatical functions, cannot be extracted from the primary linguistic experience, and we reject the assumption that they are given a priori, the conclusion would be that *CAMILLE* is in principle incapable of constructing broader syntactic categories.²⁵ This is arguably an empirically correct conclusion, since with suitably well-defined subcategories, the learner's lack of broader syntactic categories will not be seen in its linguistic behavior. It will correctly produce and understand sentences, and it will function as predicted in experiments that ask it to generalize (e.g. "if this is a 'wug', what are two of them called?"). It will only be possible for us to see that the learner deviates from contemporary linguistic theorizing or to normal intuition by examining its internal representations. Whether or not this is an acceptable outcome, it is worth pointing out *CAMILLE*'s rather striking limitation of not being able to form broad syntactic categories.

5. Conclusion

²⁵ Crucially, we are not suggesting that *CAMILLE* is incapable of forming categories, only categories that go beyond the semantic, distributional and grammatical criteria. An approach to tying syntactic categories to semantic criteria would be an acquisitional one that establishes the core syntactic categories on the basis of the restricted semantic space available to the early learner, and then uses distributional criteria to generalize them. (See for example Grimshaw 1981, Macnamara 1982, Pinker 1984, and Anward 2000 for a range of proposals.) *CAMILLE*'s problem appears to be that it is limited in how far it is able to generalize; this limitation may well be a principled one.

We have demonstrated through simulation of language acquisition that a language learner must be endowed with certain architectural features that are specific to language. Some of these are standardly assumed to be universal features of natural language within linguistics. While our computational simulation does not constitute proof that the standard linguistic view is correct, it provides additional motivation for it.

Other features are specific to the problem of constructing a computational learner, and are not standardly assumed to be part of the language faculty. However, in our view, it is impossible to envision the language faculty without taking into consideration what kinds of operations it must perform on real data in order to arrive at an adequate representation of the sound/meaning correspondence. It may be, as we have argued, that some of these operations have nothing to do *per se* with the content of grammatical knowledge, e.g. specifics of the structure of phrases or constraints on rules of grammar. Nevertheless they do have to do with the architecture of the language faculty. We have in mind in particular the assumption that the learner is constructing a parser with certain characteristics for the input that it is presented with, on the basis of which the sound/meaning correspondences can be hypothesized and evaluated. In our view, the characteristics of this parser are very much a part of the language faculty. In fact, we would argue that the way in which this parser and the correspondence rules together constitute the learner's grammatical knowledge; see Culicover and Nowak 2002. We envision the parser as an idealization of the device that exists in the mind of the native speaker that performs the sound/meaning mapping in real time. It is an idealization because it is not subject to memory limitations, frequency effects, lexical structure, and similar factors that determine the actual behavior of speakers in producing and comprehending language.

References

- Anward, J. (2000). 'A Dynamic Model of Part-of-Speech Differentiation,' in P.M. Vogel, and B. Comrie, (eds.), *Approaches to the typology of word classes*. Berlin: Mouton de Gruyter. 3-46.
- Baker, M.C. (2003). *Lexical Categories: Verbs, Nouns and Adjectives*. Cambridge: Cambridge University Press.
- Berwick, R., and Niyogi, P. (1996). 'Learning from Triggers,' *Linguistic Inquiry*, 27: 605-622. .
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Culicover, P.W. (1999). *Syntactic Nuts*. Oxford: Oxford University Press.
- , and Jackendoff, R. (2005). *Simpler Syntax*. Oxford: Oxford University Press.
- , and Nowak, A. (2002). 'Markedness, Antisymmetry and the Complexity of Constructions,' in P. Pica, and J. Rooryk, (eds.), *Language Variation Yearbook, Vol. 2*. Amsterdam: John Benjamins. 5-30.
- , and Nowak, A. (2003). *Dynamical Grammar*. Oxford: Oxford University Press.
- Everett, D.I. (2005). 'Cultural Constraints on Grammar and Cognition in Piraha: Another Look at the Design Features of Human Language,' *Current Anthropology*, 46. .
- Fodor, J.D. (1998). 'Unambiguous Triggers,' *Linguistic Inquiry*, 29 (1): 1-36. .
- Gazdar, G., Klein, E., Pullum, G., and Sag, I. (1985). *Generalized Phrase Structure Grammar*. Cambridge: Harvard University Press.

- Gibson, E., and Wexler, K. (1994). 'Triggers,' *Linguistic Inquiry*, 25: 407-453. .
- Goldberg, A.E. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.
- Grimshaw, J. (1981). 'Form, Function, and the Language Acquisition Device,' in C. Baker, and J. McCarthy, (eds.), *The logical problem of language acquisition*. Cambridge, MA: MIT Press.
- Hale, K., and Keyser, S.J. (2002). *Prolegomena to a Theory of Argument Structure*. Cambridge, MA: MIT Press.
- Halpern, A., and Zwicky, A. (eds.) (1996). *Approaching Second: Second Position Clitics and Related Phenomena*. Stanford, CA: CLSI Publications.
- Harman, G.H. (1963). 'Generative Grammars Without Transformational Rules: A Defense of Phrase Structure,' *Language*, 39: 597-616. .
- Hauser, M., Chomsky, N., and Fitch, T. (2002). 'The Faculty of Language: What is It, Who Has It, and How Did It Evolve?' *Science*, 298: 1569-1579. .
- Jackendoff, R. (1990). *Semantic Structures*. Cambridge, MA.: MIT Press.
- . (1997). 'Twistin' the Night Away,' *Language*, 73: 534-559. .
- . (2002). *Foundations of Language*. Oxford: Oxford University Press.
- Levin, B., and Rappaport Hovav, M. (1995). *Unaccusativity: At the Syntax-Lexical Semantics Interface*. Cambridge, MA: MIT Press.
- Macnamara, J. (1982). *Names for Things: A Study of Human Learning*. Cambridge, MA: MIT Press.
- MacWhinney, B. (1995). *The CHILDES Project : Tools for Analyzing Talk*. Hillsdale, N.J.: L. Erlbaum.
- Pinker, S. (1984). *Language Learnability and Language Development*. Cambridge, Mass: Harvard University Press.
- , and Jackendoff, R. (2005). 'The Faculty of Language: What's Special about it?,' *Cognition*, 95, 201-236.
- Pollard, C., and Sag, I. (1994). *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press.
- , and Sag, Ivan. (1992). 'Anaphors in English and the Scope of Binding Theory,' *Linguistic Inquiry*, 23: 261-304. .
- Prince, E.F. (1987). 'Topicalization and Left-Dislocation: A Functional Analysis,' in S.J. White, and V. Teller, (eds.), *Discourses in Reading and Linguistics*. New York;: New York Acad. of Sciences. 213-225.
- Prince, E. (1998). 'On the Limits of Syntax, with Reference to Left-Dislocation and Topicalization.,' in P.W. Culicover, and L. McNally, (eds.), *The limits of syntax. Syntax and semantics. Vol. 29*. New York: Academic Press. 281-302.
- Quartz, S.R., and Sejnowski, T.J. (1997). 'The Neural Basis of Cognitive Development: A Constructionist Manifesto.,' *Behavioural and Brain Sciences*, 20 (4): 537 -596.
- Ross, J. R. (1967). 'Constraints on Variables in Syntax,' unpublished doctoral dissertation, MIT, Cambridge, MA.
- Tomasello, M. (2003). *Constructing a Language*. Cambridge, MA: Harvard University Press.
- Wexler, K., and Culicover, P.W. (1980). *Formal Principles of Language Acquisition*. Cambridge, MA: MIT Press.
- Yuret, D. (1998). 'Discovery of Linguistic Relations Using Attraction,' unpublished doctoral dissertation, MIT, Cambridge, MA.