

# Squinting at Dali's Lincoln: On How to Think about Language<sup>1</sup>

Peter W. Culicover  
The Ohio State University

## Concept and theory

Ernst Mayr notes in his book *This Is Biology* that “in biology, concepts play a far greater role in theory formation than do laws. The two major contributors to a new theory in the life sciences are the discovery of new facts (observations) and the development of new concepts” (62-3).

With this general perspective in mind, I address my remarks in this section of the conference labeled “Emerging Theories” on a related question, which is ‘How should we think about language?’ The idea is that in linguistics as in biology, the way that we think about language strongly determines the kinds of theories that we come up with. Focusing on syntax in particular, it can be argued that to a considerable extent contemporary mainstream theory is the product of a very particular way of conceptualizing language. If we shift to another perspective, which I will outline here, then the way in which we theorize about the phenomena is in many cases substantially different.

## Resolution

Before getting into the linguistic substance, let me give you an example from another domain that illustrates in an approximate way what I am talking about. Consider this painting by Salvador Dali.

---

<sup>1</sup> I am grateful to Farrell Ackermann, John Goldsmith and Anne Charity for conversations that suggested aspects of the structure of this paper, as well as ideas about what to put in it (and what to leave out). Much of the work reported on here has been done with Ray Jackendoff and Andrzej Nowak, and I am grateful to them for their friendship and stimulating collaboration. I take full responsibility for any shortcomings.



**Figure 1.** Dali's Lincoln

The title of this paper may help you figure out what this is a picture of – but for most American viewers, at least, familiarity with the subject is sufficient. This is a painting of Abraham Lincoln. It is possible to see the image of Abraham Lincoln more clearly by squinting, and the more you squint, the more it looks like Lincoln. Alternatively, if we make the image smaller, the image of Lincoln is also easier to see. In either case, that of squinting or resizing the image, we are able to see the image by eliminating detail, in fact, a considerable amount of detail that is not essential to the image of Lincoln.



**Figure 2.** Small Dali's Lincoln

There is **so much detail** in the Dali painting that does not conform to the image of Lincoln that even with a lot of squinting, or making the image very small, the image is at best suggestive of Lincoln. When there is a mass of detail, squinting may bring out the structure, while losing the detail. The general point is applicable to linguistics (and no doubt other sciences). Syntactic theories in the tradition of Mainstream Generative Grammar (MGG)<sup>2</sup>, that is, *Syntactic Structures* through the Minimalist Program, have progressed largely through the equivalent of large scale squinting. We argue in Chapters 2 and 3 of *Simpler Syntax* is that if there is a significant amount of important detail, as there is in the case of Dali's Lincoln, massive squinting gives a very distorted view of what the object that we are looking at really is.

*Simpler Syntax* suggests that linguistic squinting rests on three basic foundations:

### **Foundations of MGG**

- **Structure:** There is (abstract) structure.
- **Syntactocentrism/syntacticization:** All relevant structure is (narrowly) syntactic.
- **Uniformity:** The structure, and everything that we say in way of accounting for the structure, is maximally uniform.

The first point is uncontroversial, but can be satisfied in a number of ways. The second incorporates the idea that all important linguistic phenomena are ultimately to be represented and accounted for in narrow syntactic terms, that is, in terms of syntactic categories like NP and VP and the relationships between them. And the third has led to large scale theory development in the MGG tradition, by radically generalizing structures, derivations, and interface mappings. (Please see *Simpler Syntax* for many examples.)

### **On explanation in linguistics: beyond the core and the periphery**

It is interesting to observe that in its earlier incarnations MGG had the goal of accommodating within the theory the higher degree of detail found in the data. This accommodation is often called 'the theory of markedness'. Here's a quote from "On wh-movement" (Chomsky 1977), referring back to "Conditions" (Chomsky 1973) that captures the essential idea in discussing the role of constraints in a theory of grammar:

“[O]ne might construct a rule to ‘violate’ the A-over-A condition, but only at a cost: the rule would have to make explicit the relevant structures so that it can apply without failing under the condition. ‘The logic of this approach ... is essentially that of the theory of

---

<sup>2</sup> We use this term as a convenient way to refer to the approach found in syntactic theories from Chomsky 1957 through Chomsky 1994, which although varying in specifics over the years, share certain features (see Jackendoff 2002 Culicover and Jackendoff 2005.). The reader is welcome to substitute whatever term s/he prefers for this purpose.

markedness.’ That is, the conditions become an integral part of an evaluation measure, rather than imposing absolute prohibitions.”

On this view, which appears as early as *Aspects* (Chomsky 1965), there are many genuine possibilities in natural language that are complex in some sense; some are too complex to actually occur, while others are relatively rare.

Chomsky subsequently set aside markedness in favor of a more sharply etched approach, relegating the complexity to the ‘periphery’ and focusing on the ‘core’ part of the core-periphery pair. One way to deal with complexity is to perform a radical idealization, that of an ideal speaker-hearer in a homogeneous environment with instantaneous learning, choosing among a small set of parameter values on the basis of the linguistic experience. Just one measure of Chomsky’s extraordinary insight, in my view, is that he saw how the complexity would constitute a formidable if not impenetrable barrier to progress in research, and found a straightforward and effective way around the barrier. Ironically, perhaps, the consequence of this radical idealization has been not only a series of important insights into the structure of what we know when we know a language, but also a more refined appreciation of the true complexity of the phenomena that we are dealing with, but are perhaps unable to account for satisfactorily.

It is our contention in *Simpler Syntax* that the focus on a narrow (and increasingly narrower) set of phenomena to get at the latent structure rules out of consideration a vast amount of the detail that exists in natural language. In effect, we fail to see the images that make up Dali’s Lincoln, images that are real and cannot be ignored in favor of the grosser structure. In the case of language, the detail is effectively acquired by children and represented in the minds of adult speakers, and must be accounted for by a linguistic theory. The picture that *Simpler Syntax* presents is one in which the core is at the extreme end of a continuum of form-meaning relations, a point that is characterized by maximal generality and semantic transparency, e.g. phrase structure rules, robust filler-gap dependencies of the wh-movement variety, and so on.

But there is a vast array of less than general detail to account for along the continuum as well, from the individual word, to idioms, to semi-regular derivational morphology, to fixed expressions with open slots (like *to put one’s money where one’s mouth is*), to constructions that reflect the basic structures of a language but have special idiosyncratic properties (like *to Verb one’s way Path*), to what I have called “syntactic nuts”, which are semi-productive syntactic constructions (like Sluice-stranding: *She said she was talking with someone, but I can’t remember who with.*)

I argue in *Syntactic Nuts* that our intuitions as native speakers are as sharp and reliable about the periphery as they are about the core. Learners are capable of learning very specific details, small scale generalizations, idiosyncratic exceptions, and lexically restricted but semi-productive syntactic and

morphological constructions. Generalization, when it occurs, may go beyond the actual experience, but does not appear to be so radical as to induce maximal uniformity on the basis of limited experience. We need to have an architecture for language acquisition and processing and the representation of knowledge that captures this state of affairs.

### **The Dynamical Perspective on language**

The picture suggested by *Simpler Syntax*, *Syntactic Nuts* (Culicover 1999), and related work, viewed against the backdrop of contemporary syntactic theory, is that the architecture is that of a complex system in which there is a huge mass of detail, much of which is well-organized and reflects various degrees of generalization, and underlying mechanisms that are not seen directly, but whose effects are seen in the ways in which they impose organization on the whole thing.

My colleagues and I have attempted to capture this picture by modeling this system as a dynamical system that forms an internal representation of the language on the basis of each and every piece of linguistic experience, which we may take as pairings between sounds and Conceptual Structure (CS) representations. (See Culicover and Nowak 2003.) Even without prior knowledge about syntactic categories and syntactic structure, there is information in the primary linguistic experience of a learner that supports the formulation of provisional grammatical categories and generalized correspondences between strings and meanings. It is known, for example, that sequences of sounds have statistical properties that may in principle allow a learner to segment them into subsequences of units, and there is evidence that learners in fact can do this, even for arbitrary sound sequences (Saffran et al. 1996). At the same time, learners can segment the non-linguistic context into objects, properties, relations, actions, etc. and correlate these with sequences of sounds (see Siskind 2000 for a computational simulation).

The architecture that we envision is dynamical in two senses. One, which is not directly germane to the points I want to make here, is that a linguistic expression is represented in the mind as a trajectory through states that reflects the temporal organization of language. The second is that categories, structures, and rules are formed as a consequence of the organization and re-organization of the system, forming generalizations to the extent warranted by the evidence. The organizing mechanism does not exist in the data *per se*, but its hand can be seen in the way that individual pieces of data cluster with one another. This organizing mechanism, *Simpler Syntax* suggests, consists of three components:

#### **Organizing mechanism of *Simpler Syntax***

- Conceptual structure
- Grammatical functions (Subject and Object)
- Syntax (constituent structure and linear ordering) and morphosyntax

The dynamical perspective has implications for language acquisition, and for the broader social context in which language acquisition occurs. I will sketch out briefly what these are, and then consider the possible implications for linguistic theory.

### **Language acquisition**

For a number of years we have been developing a computational simulation of language acquisition, called *CAMiLLe*. The objective of this simulation is to explore the interactions between and the effects on the course and content of acquisition of three main factors in the theory of language acquisition: (i) the computational capacities of the learner and its 'prior knowledge' about language and grammar, (ii) the input to the learner, and (iii) the target of learning. The outcome of the simulation is an explicit representation of the knowledge of the simulated language learner that can be examined in detail, and compared with what human learners are presumed to know.

We have proceeded by making minimal assumptions about each of (i)-(iii). In the case of (i), prior knowledge, we assume that the learner has no knowledge of grammatical categories, linguistic structure, or grammatical principles. The learner has only the capacity to extract correspondences between form and meaning based on the statistical properties of the linguistic input, to form categories based on similarity of distribution, and to form limited generalizations. In the case of (ii), the input, we assume that the learner is presented only with pairs consisting of forms (phrases and sentences) and their corresponding meanings. In the case of (iii), the target of learning, we assume that it is not a grammar in the sense of MGG, but a set of form-meaning correspondences that is coextensive with the form-meaning correspondences computed by a suitable grammar for the language to be learned (e.g. the grammar that is in the head of a native speaker of the language). In other words, the learner simply has to construct a grammar that is compatible with what a native speaker knows, but not necessarily with any particular linguistic theory.

Clearly, these assumptions are in many respects too Draconian, and a realistic account of how language is acquired will have to elaborate (i)-(iii) in many ways. The objective of the simulation is not to demonstrate that the strongest form of these assumptions is correct, but to determine in exactly which ways they are too radical. Moreover, it is of some interest to discover how far a language learner can get, even given these very minimal assumptions.

The results thus far have been encouraging, and interesting, although not entirely conclusive. *Syntax Nuts* and *Simpler Syntax* have argued that what is to be learned includes a large set of correspondences ranging in generality from individual words to compositional phrase structures, as outlined earlier. Since there is no way 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 as the weight of the evidence warrants further generalization (Tomasello 2000). The

weight of the evidence is at least in part determined by statistical properties of the input (e.g. Newport and Aslin 2004).

### What *CAMiLLE* does

*CAMiLLE* is exposed to sets of form-meaning pairs, e.g.,

- (1) house = HOUSE(\$TYPE:BLDG)  
see the house ? = YNQ(SEE(\$EXP:YOU, \$THEME:HOUSE(\$TYPE:BLDG)))

I leave open the question of how a learner knows what meaning a given expression is to be paired with, and simply assume that this question has a resolution in developmental cognitive psychology; see for example Bloom 2000, Gleitman et al. 2005.

On the basis of each such pair, *CAMiLLE* attempts to formulate one or more correspondence rules. *CAMiLLE* proceeds from the assumption that strings of words and their corresponding meanings are organized according to heads and non-heads (dependents or adjuncts). Our experiments suggest to us that without this assumption, *CAMiLLE* cannot acquire even minimal linguistic knowledge.<sup>3</sup> Even with this assumption, since there is no overt connection between the individual words and the individual meanings, *CAMiLLE* is prone to making many bad rules. But *CAMiLLE* also will some make correct rules. For example, after having encountered the sentences in (2) –

- (2) ted is nice. = BE(\$THEME:TED, \$PRED:NICE)  
ted is small. = BE(\$THEME:TED, \$PRED:SMALL)

– *CAMiLLE* has enough information to guess that *Ted is* means either BE(\$THEME:TED), or TED, or BE. *CAMiLLE* keeps track of the evidence that supports each hypothesis, so that after enough experience, the diversity of exemplified correspondences continues to support the first hypothesis, but not the other two. At the same time, this experience provides evidence that *Ted* corresponds to CS TED and *is* corresponds to CS BE. The evidence is purely statistical; the rules that are not supported remain in the learner but gradually get pushed out by rules that are more strongly supported by the evidence.

If *CAMiLLE* finds that two rules have a similar form, then to the extent possible it forms a cluster (i.e. a mini-category) . For example, if *CAMiLLE* has strong evidence for the following two correspondences --

- (3) ted is <=> BE(\$THEME:TED)  
sally is <=> BE(\$THEME:SALLY)

---

<sup>3</sup> Jerry Feldman has suggested to us that it is possible that it may not be necessary to assume that there are heads, and that some less explicit knowledge about the nature of the correspondence will be sufficient for the learner to be able to identify the heads. We have not yet looked into this possibility.

Then *CAMILLE* will form a correspondence rule of the form

(4) [ted;sally] is <=> BE(\$THEME:[TED;SALLY])

Clearly, correspondences such as these are not equivalent to rules of grammar in the traditional sense. For one thing, they are much too specific – they do not mention categories but simply clusters of individual elements. For another, they provide information only about the linear order of elements, not structure. And they do not provide any phrasal information.

At the same time, it is at least plausible that what *CAMILLE* comes up with is comparable in some important respects to what an early language learner comes up with, prior to the point at which generalization and the formation of large-scale categories and correspondences kicks in, not to mention the recognition of discontinuous dependencies. We are entertaining the hypothesis (suggested by Tomasello 2003 and arrived at independently through our own preliminary experiments with *CAMILLE*) that first there is a pre-grammatical stage, which is modeled by *CAMILLE*, followed by a grammatical stage. In the pre-grammatical stage we expect to see *CAMILLE* treat everything as though it is a construction. In the grammatical stage, we expect to see those aspects of the language that are fully or almost fully regular to be reflected in dramatic generalizations, while those aspects of the language that retain some significant idiosyncrasy, e.g. constructions of the sort that we noted earlier, would be retained in their pre-grammatical form.

### **Constructed input**

Constructed input allows us to test *CAMILLE*'s ability to deal with a particular grammatical phenomenon. *CAMILLE* requires a certain amount of exposure to a grammatical phenomenon in order to form a reasonably informed hypothesis about it. A file of naturally occurring, transcribed speech to children from the CHILDES database (MacWhinney 1995) does not in general provide enough instances of a specific phenomenon,<sup>4</sup> and running *CAMILLE* on composites of files, while potentially useful, does not allow us to focus on specific grammatical phenomena. So we have constructed files by hand. An example of a constructed input file is given below as **Sample Input 1: word\_order-1.txt**.

---

<sup>4</sup> It is an empirical question whether for any given grammatical phenomenon, the naturally occurring data taken as a whole provides sufficient evidence for a learner. If it does not, then this is an argument (from poverty of the stimulus) for innateness. The sorts of things that we are interested in are those that are not universally found in languages of the world, and therefore we may pretty confidently presume that they are learned on the basis of evidence in the linguistic input to the learner.

### Sample Input 1: word\_order-1.txt

```
house = HOUSE($TYPE:BLDG)
see the house ? = YNQ(SEE($EXP:YOU,$THEME:HOUSE($TYPE:BLDG)))
mary = MARY($TYPE:PERSON)
here's mary = $POINT($THEME:MARY($TYPE:PERSON))
see mary ? = YNQ(SEE($EXP:YOU,$THEME:MARY($TYPE:PERSON)))
john = JOHN($TYPE:PERSON)
see john ? = YNQ(SEE($EXP:YOU,$THEME:JOHN($TYPE:PERSON)))
here 's a flower = $POINT($THEME:FLOWER($TYPE:PLANT))
see the flower = $IMP(SEE($EXP:YOU,$THEME:FLOWER($TYPE:PLANT)))
here's a boy = $POINT($THEME:BOY($TYPE:PERSON))
see the boy = $IMP(SEE($EXP:YOU,$THEME:BOY($TYPE:PERSON)))
horsie = HORSE($TYPE:ANIMAL)
see horsie ? = YNQ(SEE($EXP:YOU,$THEME:HORSE($TYPE:ANIMAL)))
a baby ! = BABY($TYPE:PERSON)
see the little baby ? = YNQ(SEE($EXP:YOU,$THEME:BABY($TYPE:PERSON)))
nice baby ! = BABY($TYPE:PERSON)
talk to the baby = $IMP(TALK($AGENT:YOU,$GOAL:BABY($TYPE:PERSON)))
talk to mary = $IMP(TALK($AGENT:YOU;$GOAL:MARY($TYPE:PERSON)))
i see mary = SEE($EXP:ME,$THEME:MARY($TYPE:PERSON))
i am talk ~ing to may = TALK($AGENT:$ME;$GOAL:MARY)
do you see john ? = YNQ(SEE($EXP:$YOU,$THEME:JOHN($TYPE:PERSON)))
```

The purpose of this particular file is to try to get *CAMILLE* to correlate individual words with their meanings, and to correlate position in the string with semantic role. (The roles used here are Theme, Exp(eriencer) and Agent.) The output after processing ten sentences consists of 103 rules, many of them overlapping, and many of them highly idiosyncratic but low in weight. Lack of space precludes listing all of the rules here, so I will show a few ‘correct’ rules (5) and a few ‘incorrect rules’ (6).

- (5) 5. [89] MARY ⇔ mary  
6. [82] JOHN ⇔ john  
16. [13] YNQ(\*NULL\*:SEE) ⇔ 1.see 3.?  
21. [10] \$IMP(\*NULL\*:SEE) ⇔ 1.see  
27. [9] SEE(\$THEME:[JOHN; MARY;]) ⇔ see+1->[john; mary;]  
95. [2] \$POINT(\$THEME:MARY) ⇔ 1.here's 2.mary
- (6) 1. [172] SEE(\$EXP:YOU) ⇔ 1.see  
23. [10] YNQ(\*NULL\*:SEE)\$=#3 ⇔ 1.see 2.the 3.?  
72. [2] BOY ⇔ 3.boy | a+1->boy see | see X boy  
103. [2] FLOWER ⇔ 1.here 2.'s 3.a 4.flower | 1.see 2.the 3.flower

The ‘correct’ rules and the ‘incorrect rules’ all reflect *CAMILLE*’s exposure to the data. For instance, (6.1) reflects the fact that *see* in initial position correlates highly with the meaning SEE(\$EXP:YOU). This is because there are a lot of sentences beginning with *see* (imperatives and questions) in which the subject is not expressed overtly. We may take this to be a very early stage of development, in which the learner has not yet determined that such sentences have a missing

subject; such a determination can be made when the learner recognizes that all sentences of English have subjects.

More strikingly, (6.23) shows that in the limited input data, *see the* correlates highly with the interrogative of *see*. This is an artifact of the particular dataset, and is not an error on *CAMILLE*'s part, but a correct hypothesis under the circumstances. Similarly, *CAMILLE* finds evidence to form correspondences between the meaning boy and *boy* in third position, *a boy*, and *see ... boy*. A more diverse set of experiences will disabuse *CAMILLE* of these errors, or at least it should if we have designed the simulation correctly. And it is possible, although difficult to determine experimentally, that actual learners may form such incorrect, yet fleeting, mistaken correlations in the early stages of learning.

Rule (5.16) reflects the fact that *see* is used as an interrogative (with '?' in third position in the sentence – an artifact of the input data). Similarly, (5.21) reflects the interrogative case. (5.27) indicates that *CAMILLE* has identified *John* and *Mary* as elements that have the same distribution (with respect to the Theme of *see*). This observation may, if we wish, form the basis for a generalization that *John* and *Mary* have the same distribution with respect to everything, although we will want to exercise caution in formulating the rule of generalization. Finally, (5.95) is a small construction, correlating *here's Mary* with pointing to *Mary*.

### Natural Input

We have begun to look at what *CAMILLE* does with natural linguistic input. There are approximately 775K English sentences spoken to children in the CHILDES database. Since the sentences in the CHILDES database do not for the most part have meanings associated with them, in order to use those sentences as input to *CAMILLE* it is necessary to provide them all with meanings. Doing so manually is prohibitively labor intensive.<sup>5</sup> Our approach has been to parse the sentences with a fast (but, unfortunately, inaccurate) parser (Mini-par), translate the output of the parser into rudimentary meanings, present the resulting sets of sentence/meaning pairs to *CAMILLE* and ask *CAMILLE* to figure out the correspondence rules.

The results thus far are somewhat inconclusive, but for reasons that do not necessarily reflect on *CAMILLE*. First, the 775K sentences in the CHILDES database are taken from many speakers, a number of dialects, and are spoken to children of widely varying ages. This means that there may be an overall lack of consistency that might interfere with *CAMILLE*'s ability to extract reasonably accurate generalizations. Second, and more seriously, the parses produced by Mini-par are often wildly mistaken; hence the meaning that is automatically generated from the parse is also wildly mistaken. To take just one example, Mini-

---

<sup>5</sup> If it takes one minute to construct each meaning, and a person does this eight hours a day, five days a week, it would take over six years to assign meanings to each of these sentences. Who knows how long it would take to correct the errors?

par treats *going to* as a directional, even when it is used in sentences like *Are you going to kiss me?* Conjunction, which is found very often in the CHILDES data, is also very problematic for Mini-par (and other parsers, for that matter). Thus, there are numerous errors that arise out of the misparsing of the input. Finally, the range of subject matters found in the CHILDES data is quite extensive, and a successful meaning assignment to a large number of sentences, even if they are correctly parsed, is a non-trivial task that we have yet to contemplate.

This being said, *CAMILLE* does produce some useful output when dealing with the naturally occurring input, and some of this output very definitely has the feel of emergent constructions. The following rules are a small sample of what *CAMILLE* came up with after processing approximately 19K sentences from the CHILDES database.

- (7) **1. [59152]** BE(\$THEME:[HE; IT; THAT; THIS; WHAT; WHERE; WHO;]) ⇔  
 [he; it; that; this; what; where; who;]+1->is
- 6. [38487]** [BED; BOOK; BUG; BUNNY; CHAIR; COOKIE; CRAYON; DUCK; IT; PICTURE; THAT; THIS;](\$REF:[\$DEF; \$INDEF;]) ⇔ [a; the;]  
 [?; bed; book; bug; bunny; cookie; duck; picture;]
- 15. [12036]** [KNOW; LIKE; SEE; THINK; WANT;](\$EXPERIENCER:YOU) ⇔  
 you [know; like; see; think; want;]
- 62. [1491]** IMP([GET; GIVE; LIKE; PUT; SAY; SEE;]) ⇔ 1.[get;  
 give; like; put; say; see;]
- 155. [372]** [WHAT; WHERE;](\$REF:\$WH) ⇔ 1.[what; where;] is |  
 [what; where;]+1->is
- 163. [353]** PLAY\_WITH ⇔ play+1->with
- 195. [266]** YNQ(\*NULL\*:WANT) ⇔ do X want
- 228. [227]** WANT ⇔ want+1->to
- 590. [43]** NEG(\*NULL\*:BE) ⇔ 2.is 3.not

Rule (7.1) characterizes a construction in which there is a pronominal theme of *be*. (7.6) shows that *a* and *the* correlate with the features \$DEF and \$INDEF on nominal concepts. This may ultimately form the basis for a more general rule expanding NP as Det-N, although getting there requires additional generalization. (7.15) correlates *you* ordered before a set of verbs expressing knowledge and perception with the experiencer role of these verbs. And so on for the rest of the rules shown. Each one indicates that *CAMILLE* has correctly extracted some correspondence, one that is specific to lexical items. This is how we characterized the most specific end of the correspondence spectrum. Hence *CAMILLE* appears to be capable, in principle at least, of carrying out the preliminary work of forming correspondence rules.

### Summary and prospects

There is clearly a lot more that can be said about what *CAMILLE* does, if only because even in its current form, it produces so much output. The massive output provided by *CAMILLE* is both a curse and a blessing. It is a curse, because it is so much to deal with, and not particularly easy to analyze. But it is a blessing, because what we are doing in creating *CAMILLE* is simulating what takes place

in the mind of a language learner. If it is fact true that early language learners begin by creating numerous constructions and only later generalize over and perhaps beyond them, then looking at *CAMILLE*'s output is like looking directly into the developing language faculty.

Of course it would be a serious mistake to claim that this program is anything more than a simulation, or that it is necessarily a correct simulation of how learning proceeds. The ultimate test will be whether *CAMILLE*, or a subsequent development of *CAMILLE*, is capable of producing a representation of the language learned that comprises in a satisfactory way a native speaker's knowledge of language (or at least, the form-meaning correspondences). Such a representation has to go beyond the actual experience. Moreover, it must capture generalizations that are formulated at a level of abstraction that goes well beyond what is available to *CAMILLE* at this point.

These issues are the focus of our current work with *CAMILLE*. Our immediate goal, besides improving the input to the simulation, is to provide *CAMILLE* with the capacity to generalize beyond individual or clustered correspondences (the outcome of the pre-grammatical stage) to correspondences in terms of general categories (the grammatical stage). We are also experimenting with various 'local' relations, such as Subject-Aux inversion, to show that *CAMILLE* can master them without elaborate knowledge of syntactic structure beyond linear order. Our experiments with wh-questions and topicalization are intended to show that *CAMILLE* can construct adequate local variants of these unbounded dependencies, which may serve well enough in the pre-grammatical stage.

Beyond this, it is clear that *CAMILLE* is not able to identify the locus of a 'gap' in a sentence. That is, *CAMILLE* cannot connect a 'moved' constituent with the corresponding canonical position. While it is likely that this capacity does not exist in early learning (see Tomasello 2000), it is something that *CAMILLE* needs to be able to do at some point in the course of development. We see no way for *CAMILLE* to discover that such connections exist unless *CAMILLE* is endowed with the capacity to determine that something is absent from a particular position. True generalizations (i.e. those that speakers really make use of to assign interpretations to sentences and to judge acceptability) that crucially rely on grammatical notions such as subject and object, or thematic hierarchies, are also beyond the scope of *CAMILLE*, and would have to be built in or facilitated by specific knowledge – we see no way for *CAMILLE* to discover them given just the primary linguistic data.

### **Language change and variation**

Let us now shift our perspective. We may think of a language as being the sum total of all of the systems in the heads of all of the speakers who communicate with one another. There is a certain organization imposed upon this sum total by the fact that parts of it are encapsulated in individual brains. While any and all knowledge within a single individual may form the basis for generalization and

extension, much more limited knowledge is shared between individuals because of the restrictiveness of the communication channel or channels available to them. In the absence of telepathy, there is necessarily a lack of homogeneity.

Moreover, since not everyone communicates with everyone else, the basic picture is one of a network of linguistic agents, each with enough in his/her head in common with the other agents that he/she comes into contact with to be able to communicate with them.

For any linguistic phenomenon that allows for variation, we can map the distribution of the variants over the agents in the network. I have written elsewhere about our computational simulations of language interaction in a network of agents, so I will not reproduce the specific examples and illustrations here.<sup>6</sup> We assume for the sake of the situation that a language consists of a finite bundle of features and possible values for each feature, and that language is acquired by a learner who adopts a set of feature values on the basis of interactions with speakers. In case of competition, the majority ‘wins’; that is, the learner acquires the feature value most represented among its interaction partners. It is possible to adjust the parameters of the model with respect to the number of interaction partners, the distance between interaction partners, the strength of influence of agents, the amount of evidence that a learner needs in order to acquire or change a feature value, the amount of noise present in the network, and so on.

To summarize briefly, the overall picture that emerges from this simulation is the following. Given the appropriate parameters of interaction in the network:

- i. Some languages die out entirely. In other words, all of the speakers of some language change their feature values to the extent that the particular combination of values that defines this language is no longer represented.
- ii. Properties of language correlate highly (either positively or negatively). This is due to the fact that a speaker that is influenced to acquire one feature value through interaction with a particular group of speakers is likely to be influenced to acquire other feature values through interaction with the same group of speakers. Hence the typical Greenbergian situation (Greenberg 1963), wherein with greater than chance frequency, if a language has property P then it has property Q, emerges as a consequence of the network interaction. Thus, we argue, such correlations are not in themselves of linguistic interest. What **is** of linguistic interest is if the same correlations appear over and over in geographically unrelated language groups.
- iii. Some languages have many speakers, some have very few, simply as a consequence of the evolution of the model. Demographic differences in themselves cannot be used to draw linguistic conclusions.

---

<sup>6</sup> See Culicover and Nowak 2002, Culicover and Nowak 2003, Culicover et al. 2003.

iv. But, if we introduce a **bias** into the model against a particular property or combination of properties, the languages that are at a disadvantage because of this bias either die out completely, and always do on every simulation, or are able to survive only in virtue of being isolated from languages that are not disadvantaged.

One upshot of the simulation model is that one must be cautious in drawing theoretical conclusions on the basis of the absence of certain features or feature combinations, especially if we are focusing on a relatively small geographical area, e.g. Italy and the South of France. Studies of ‘microparametric variation’ are particularly susceptible to the criticism that the absence of certain combinations of features is not linguistically interesting. But suppose we find that the same feature combinations do not exist in widely scattered regions, where the languages are very likely not descended from the same source and have had no contact. In such a case, we may hypothesize that there is at the very least a bias (if not an absolute prohibition) that renders certain feature combinations preferable to others as languages come into contact and compete for standing in learners’ brains in the social network.

Pursuing this idea, we may combine the perspective on learning and the perspective on variation just outlined with the *Simpler Syntax* view of the architecture of linguistic knowledge to try to put together an understanding of why languages vary from one another in the ways that they do, why certain things are very common and other things very rare or unattested. The key idea goes back to Chomsky’s view of markedness, outlined earlier, and is related to more recent notions of economy:

**Markedness:** That which is more complex is less economical in some precise measurable sense, which can be stated in terms of the correspondences that constituent knowledge of language. The more complex will be more difficult to learn, and there will be a bias against it in the competition in the social network. It may be attested, but if it is it will be relatively rare, and survive best where it is isolated.

### **Language processing and syntactic theory**

A this point let us see if we can tie all of these diverse threads together. The proposal from *Simpler Syntax* that the organizing mechanism that accounts for the appearance of structure when we squint at a natural language is the following organizing mechanism, given earlier.

#### **Organizing mechanism of *Simpler Syntax***

- Conceptual structure
- Grammatical functions (Subject and Object)
- Syntax (constituent structure and linear ordering) and morphosyntax.

*Simpler Syntax* proposes that all of these play a crucial role in account for the mapping between sound and meaning.

Suppose as a first approximation we take economy/markedness to be a measure of the transparency of the mapping between Syntax and Conceptual structure (CS), as suggested by Culicover and Nowak 2002. Assuming that economy is maximized wherever possible, what we see when we ‘squint’ at a natural language is the imposition of CS and the grammatical function on syntactic structure. This is not to say that syntactic structure is identical to or reducible to CS, but that there is a high value placed on making the mapping between them transparent.

But there are other considerations. There are statistical and distributional properties of linguistic forms that learners and speakers generalize, and in the extreme case these properties take on a life of their own. In other words, aspects of syntax are autonomous. This is a useful way to think of such things as the fact that every tensed sentence in English must have an overt subject, even if it is a dummy subject (e.g. *it is raining*), or the fact that the order in the English VP is

(8) V NP PP S

regardless of whether the VP is semantically transparent (*eat the cake*), a construction (*giggle down the hall*), or an opaque idiom (*kick the bucket*) (see *Simpler Syntax*: Chapter 1).

A counterpart to this view is that things that are less than maximally transparent can exist and even live robustly in a language if the evidence to the learner in support of them is extensive, and if there is no simpler competition from correspondences that serve the same function. So, actually explaining what actually occurs is not simply a matter of form, it is also a matter of function in a narrow sense (that of accounting for meaning), since it is the transparency with respect to CS of a particular locution that will determine its frequency of occurrence in a corpus of linguistic experience for learners, summed over the network of learners.

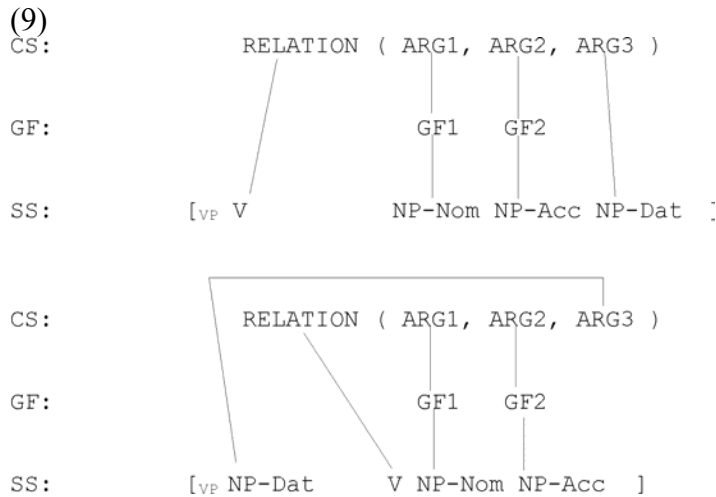
On this view, then, there is a range of variation, and also a way of measuring complexity of alternative ways of doing the same work. The key idea is that CS is fixed, and various ways of mapping to CS vary in transparency. A preliminary attempt to characterize this complexity in terms of transparency was suggested in Culicover and Nowak 2002. It can be formulated in terms of the *Simpler Syntax* architecture, in terms of the following:

- a triple consisting of syntactic structure, GF tier, and CS
- linkings from syntactic structure to GF
- linkings from GF to CS

- linkings from syntactic structure to CS

Can we be precise about the measure of complexity? Can we come up with a measure that is not circular? A reasonable approach is to use the central cases where we believe we understand the interaction between complexity and transparency of mapping well in order to formulate the preliminary metric, and then apply it to other cases and assess the extent to which it produces sensible conclusions.

Consider first scrambling. In this case, there is a direct linking between NPs, GFs and CS arguments, regardless of how things are ordered with respect to one another, since linear order is irrelevant to CS (at least the argument structure part of it). For example,



The case marking in the default case indicates the grammatical function regardless of where in the sentence the NP is with respect to the other arguments.

But in (10), an illustration of iterated A' movement adapted from Culicover and Nowak 2002, the situation is very different.



supposed to be, i.e. on the right branch of VP. Culicover and Nowak 2002 measure the relative complexity in terms of the fact that the more complex case involves less deformation of the CS relations, which we can measure in terms of the directness of the linking relations (shown with double arrows).

No doubt there are more sophisticated and more precise ways to measure the complexity, but as a first approximation, this appears promising. Other things being equal, transparency is a function of proximity. Deformation of a structure may have the consequence of making two related constituents non-proximate. I believe that this is one way to understand Chomsky's (1995) proposal in the Minimalist Program that overt movement is a cost. From our perspective, we understand overt movement as contributing to the computational cost of recovering the GF-structure and CS-structure from the surface tree. A measure of complexity that takes proximity into account along the lines suggested here will yield results similar to the locality constraint in Culicover and Wilkins 1984, Rizzi's (1990) Relativized Minimality, and the Minimal Link Condition of Chomsky 1995. At the same time, for relativized minimality effects, local domains might be sufficient to define proximity, essentially recapitulating the insight of subjacency. There are many open questions here, but I believe that the overall approach makes sense, particularly if viewed from the general perspective developed in this paper.

## **Summary**

What does this approach suggest about grammatical theory? A central feature of *Simpler Syntax* is that it takes as primitive the PF-SS-CS triple. The dynamical approach predicts that there will be a range of generality of these triples, from the most idiosyncratic (words and idioms) to the most general (phrase structure rules with compositional interpretations). This constructionalist view can be seen as a natural consequence of the dynamical approach, which hypothesizes that the mappings between the representations are built up and generalized on the basis of concrete experience.

At the same time, a significant aspect of *Simpler Syntax* is that it takes conceptual structure to be fundamental. To the extent that we are able to say something about the complexity of the mapping between syntactic structure and conceptual structure, we can understand grammatical universals in part in terms of the tendency of the less complex mappings to force out the more complex ones in competition in the social network, other things being equal. *Simpler Syntax* and the dynamical approach, taken together, offer the opportunity of genuine explanation of the true complexity of natural language, which we can approach with eyes wide open, not squinting.

## References

- Bloom, Paul. 2000. *How Children Learn the Meanings of Words*. Cambridge, MA: MIT Press.
- Chomsky, Noam. 1957. *Syntactic Structures*. The Hague: Mouton.
- \_\_\_\_\_. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- \_\_\_\_\_. 1973. Conditions on Transformations. In Steven Anderson and Paul Kiparsky (Eds.), *Festschrift for Morris Halle*. New York: Holt, Rinehart and Winston: 232-286.
- \_\_\_\_\_. 1977. On Wh Movement. In Peter W. Culicover, Wasow and Adrian Akmajian (Eds.), *Formal Syntax*. New York: Academic Press: 71-132.
- \_\_\_\_\_. 1994. Bare phrase structure. In *MITWPL Occasional Papers*. Cambridge, MA: MIT.
- \_\_\_\_\_. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.
- Culicover, Peter W. 1999. *Syntactic Nuts*. Oxford: Oxford University Press.
- \_\_\_\_\_, and Ray Jackendoff. 2005. *Simpler Syntax*. Oxford: Oxford University Press.
- \_\_\_\_\_, and Andrzej Nowak. 2002. Markedness, antisymmetry and the complexity of constructions. In Pierre Pica and Johan Rooryk (Eds.), *Language Variation Yearbook, Vol. 2*. Amsterdam: John Benjamins: 5-30.
- \_\_\_\_\_, and Andrzej Nowak. 2003. *Dynamical Grammar*. Oxford: Oxford University Press.
- \_\_\_\_\_, Andrzej Nowak, and Wojtech Borkowski. 2003. Linguistic theory, explanation and the dynamics of language. In John Moore and Maria Polinsky (Eds.), *The Nature of Explanation in Linguistic Theory*. Stanford, CA: CSLI Publications.
- \_\_\_\_\_, and Wendy Wilkins. 1984. *Locality in Linguistic Theory*. New York: Academic Press.
- Gleitman, Lila R, Kimberly Cassidy, Rebecca Nappa, Anna Papafragou, and John C. Trueswell. 2005. Hard Words. *Language Learning and Development*, 1(1): 23-64.

- Greenberg, Joseph. 1963. *Universals of Language*. Cambridge, MA: MIT Press.
- Jackendoff, Ray. 2002. *Foundations of Language*. Oxford: Oxford University Press.
- MacWhinney, Brian. 1995. *The CHILDES Project: Tools for Analyzing Talk*. Hillsdale, N.J.: L. Erlbaum.
- Newport, E.L., and R.N. Aslin. 2004. Learning at a distance: I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, 48: 127-162.
- Rizzi, Luigi. 1990. *Relativized Minimality*. Cambridge, MA: MIT Press.
- Saffran, Jennifer R., Richard N. Aslin, and Elissa L. Newport. 1996. Statistical learning by 8-month old infants. *Science*, 274: 1926-1928.
- Siskind, Jeffrey Mark. 2000. Learning word-to-meaning mappings. In Peter Broeder and Jaap Murre (Eds.), *Models of Language Acquisition: Inductive and Deductive Approaches*. Oxford: Oxford University Press: 121-154.
- Tomasello, Michael J. 2000. Do young children have adult syntax. *Cognition*, 74: 209-253.
- Tomasello, Michael. 2003. *Constructing a Language*. Cambridge, MA: Harvard University Press.