Squinting at Lincoln in Dalivision[1]

Peter W. Culicover

The Ohio State University

1. Concept and theory

Ernst Mayr notes in his book *This Is Biology* (Mayr 1997) 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 in mind, I would like to focus my remarks in this section 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[2] 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.

[1] I am grateful to Farrell Ackermann, John Goldsmith, Ray Jackendoff 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.

[2] 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 2002a Culicover and Jackendoff 2005,). The reader is welcome to substitute whatever term s/he prefers for this purpose.

2. 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.

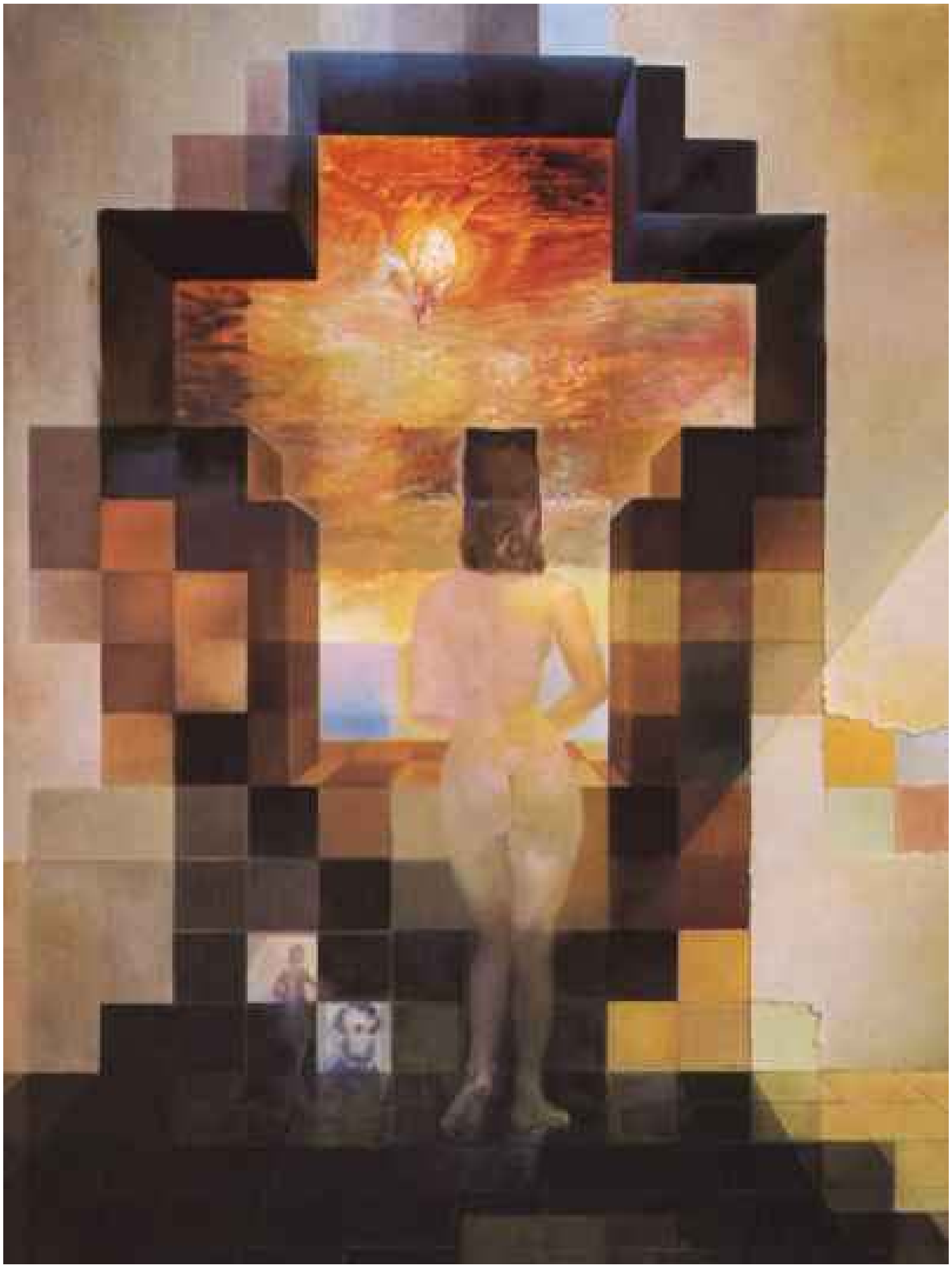

Figure 1. Lincoln in Dalivision by Salvador Dali

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 picture 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.

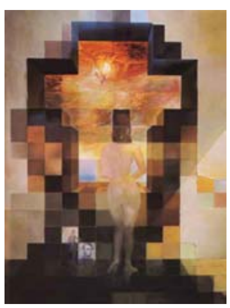

Figure 2. Small Lincoln in Dalivision by Salvador Dali

But, note that 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. Moreover, much of the detail that is lost is also quite important, and doesn't really have to do with see

Lincoln in the image. I will return to the these points and suggest that Dali is telling us something very important about how to think about language (and probably most other things).

Along similar lines, we can control the amount of information in an image that we create ourselves.

Figure 3. Pixelated Culicover's Culicover (Culicover in Culicovision?)

If you squint, the pixelated image approximates more closely the original, which is here.

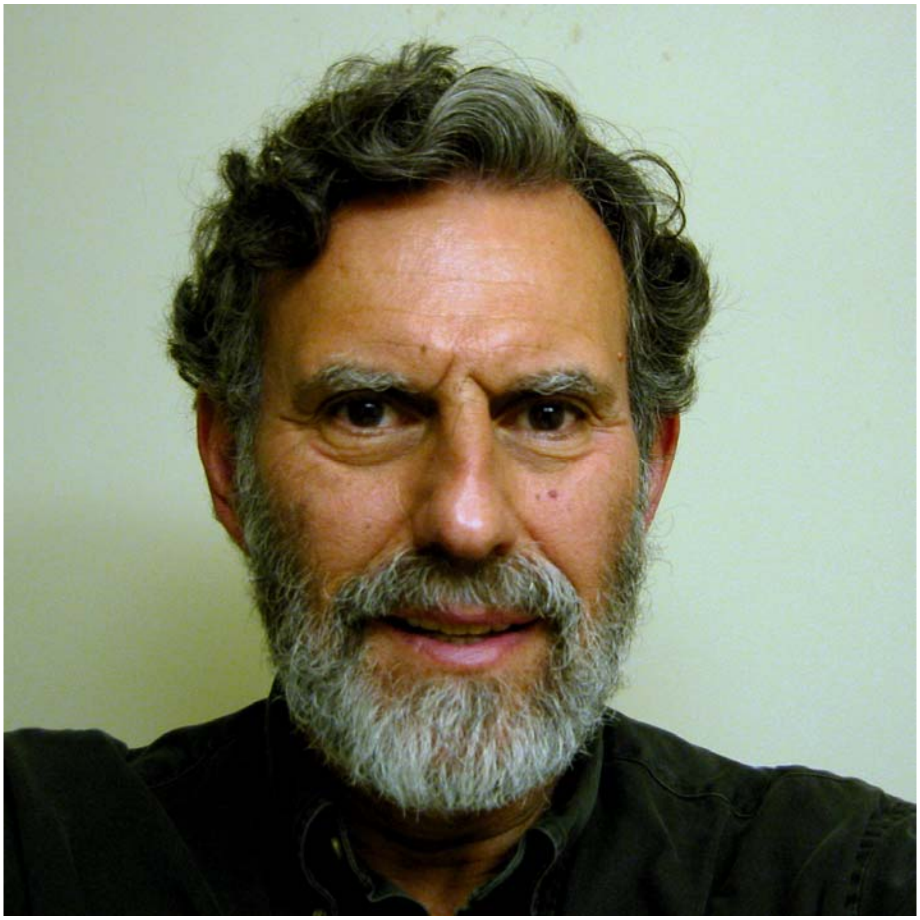

Figure 4. Original Culicover's Culicover

And see what happens when we make the pixilated image very small – you lose a lot of detailed information about the individual pixels.

Figure 5. Very small pixilated Culicover's Culicover

The point here is that when there is structure that plays a role in organizing a mass of detail, squinting or restricting the empirical scope of the phenomenon brings out the structure, while losing the detail. The point is applicable to linguistics. Syntactic theories in the tradition of Mainstream Generative Grammar (MGG), that is, *Syntactic Structures* through the Minimalist Program, have progressed largely through the equivalent of large scale squinting. The argument that Jackendoff and I make in Chapters 2 and 3 of *Simpler Syntax* is that if there is a significant amount of important, highly organized detail, as there is in the case of Dali's Lincoln in Dalivision, large scale squinting gives a very imperfect and inaccurate interpretation of what the object that we are looking at really is.

3. 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 linguistic architecture is that of a complex system in which there is a huge mass of detail, much of which is well-organized at various degrees of generality, as well as abstract structure that is not seen directly, but whose effects are seen in the ways in which it imposes 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). These tasks do not depend on the prior existence of categories, just the ability to categorize and correlate.

The architecture that we envision is dynamical in two senses. One is that a linguistic expression is represented in the mind/brain as a trajectory through states that reflects the temporal organization of language. The structure of a language is implicit in the organization and clustering of these trajectories. The second is that the categories, structures, constructions and rules are formed as a consequence of the organization and re-organization of the system, a process through which generalizations emerge to the extent warranted by the evidence. On this view, the organizing mechanism does not exist in the data *per se*, or explicitly in the representation of the knowledge, but its hand can be seen in the way that individual pieces of data cluster with one another – this is the structure that we see when we squint. For syntax, this organizing mechanism, *Simpler Syntax* suggests, consists of three components:

Organizing mechanism of *Simpler Syntax*

- Conceptual structure
- Grammatical functions (Subject and Object)
- Syntactic structure (constituent structure and linear ordering) and morphosyntactic marking

The dynamical perspective has implications for language acquisition, and for the broader social context in which language acquisition occurs. In my remaining time I will sketch out very briefly what these are, and then finish up by discussing what this means for linguistic theory.

4. 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 prior knowledge of particular 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 (including phrases and sentences) and their corresponding meanings. In the case of (iii), the target of learning, we assume that the target of learning is not a grammar in the sense of Mainstream Generative Grammar, but a set of form-meaning correspondence rules that comprises the form-meaning correspondences of the language to be learned. 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 strong, 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 strong. 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 compositionally interpreted phrase structure. 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).

4.1. What CAMiLLE does

Let me be a bit more specific. ***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)))
```

We 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 ultimately has a resolution in developmental cognitive psychology; see for example Bloom 2000 regarding word meanings. The problem of how learners know what expressions mean before they know the language is a fundamental and difficult one, but even if it is solved (which we assume for our simulation), the problem of language learning is far from trivial.

On the basis of each such sound/meaning 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 have suggested to us that without this assumption, ***CAMiLLe*** cannot acquire even minimal linguistic knowledge. But 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, since ***CAMiLLE*** doesn't know a prior what the head is. But ***CAMiLLE*** also will 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 `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 correspondences that are not supported remain, but gradually get pushed out by those 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)`

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 with respect to other individual elements. For another, they provide information only about the relative 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 everything treated as though it is a construction. In effect, the learner is establishing, testing, refining and eliminating **templates** of sound/meaning correspondences. These are not rules in the familiar sense of general, structure dependent correspondences.

These come, we suggest, in the grammatical stage. In this stage, we expect to see those aspects of the language that are fully regular or almost fully regular to be reflected in significant 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 template form. The key is that the learner must make the step from larger and larger sets of individual items that participate in correspondences to general categories. Whether this step can be taken by the learner on its own, and on what basis, or whether it needs to be pre-wired, are for us open questions.

4.2. *Constructed input*

It is particularly useful to explore ***CAMiLLE***'s performance using constructed rather than naturally occurring 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. Naturally occurring, transcribed speech to children from the CHILDES database (MacWhinney 1995) in general does not provide enough instances of a specific phenomenon,[3] and running ***CAMiLLE*** on composites of files, while potentially useful (see §**Error! Reference source not found.**) does not allow us to focus on specific grammatical phenomena. So we have constructed files by hand. An example of part of a constructed input file is given below as **Sample Input 1: word_order-1.txt.**

[3] It is of course 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 a good argument (from poverty of the stimulus) for innateness. However, the sorts of things that we are interested in are those that are not universally found in languages of the world, and therefore other things being equal we may 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 of one simulation, after processing fifty sentences, consists of 57 rules, many of them overlapping, and many of them highly idiosyncratic but low in weight. I cannot list all of the rules here, so I will simply show a few 'correct' rules (5) and a few 'incorrect rules' (6).

(5)

1.[180] SEE ⇔ see

4.[68] MARY ⇔ mary

5.[68] JOHN ⇔ john

8.[31] $YNQ(*NULL*:SEE) ⇔ see ?

17.[12] $POINT ⇔ here's a

20.[11] SEE($THEME:[JOHN;MARY;]#1) ⇔ see+1-> [john;mary;]

47.[4] $POINT($THEME:MARY) ⇔ here's+1->mary

53.[3] $IMP(*NULL*:SEE) ⇔ see

(6)

3.[83] SEE($EXP:YOU) ⇔ see

16.[12] SEE($EXP:YOU) ⇔ see+1->the

65.[2] SEE($THEME:BOY) ⇔ see+2->boy

The 'correct' rules and the 'incorrect rules' all reflect ***CAMiLLE***'s exposure to the data. For instance, (6.1) reflects the fact that *see* correlates highly with the meaning SEE. (6..8) reflects the fact that in a question about *see*, there is a question marker. And the imperative with *see* corresponds as well to *see* alone. We may take this last correspondence to reflect 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.

Rule (5.20) 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, (5.47) is a small construction, correlating *here's Mary* with pointing to Mary. And (5.17) is a small construction built around *here's a*.

Let's look at some mistakes, as well (and there are many). (6.23) shows that in the limited input data, *see the* correlates highly with the cases where YOU is the experience 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* two words to the right of *see*, as in 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.

4.3. 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 language faculty.

We have also applied ***CAMiLLE*** to naturally occurring data consisting of sentences spoken to children in the CHILDES database. The main difficulty here is assigning a meaning to every sentence. This has to be done automatically, because the sheer number of sentences makes doing this by hand impractical. On the other hand, the automatic assignment of meaning has proven to be considerably less than precise, so errors are introduced into the meanings which affects the quality of the correspondences that are formed. We have gotten some suggestive results out of the exercise, such as the following rules (from 54K sentences of the CHILDES corpus, spoken to children). At this point we are attributing the errors to the imperfect meanings.

(7) a. **[13]** [LIKE; WANT;](\$EXPERIENCER:YOU) ⇔ you [like; want;]

b. **[24193]** [DO; FINISH; GO; LOOK1; MEAN; NEED; PUT; SAY; TELL; THROW; TIE; USE;](\$AGENT:YOU) ⇔ you [.; ?; finished; look; mean; need; put; say; tell; throw; tie; to;]

c. **[12485]** [IT; STICK; THAT; THERE; THIS; TOY; WHAT;](\$TYPE:THING) ⇔ [it; stick; that; there; this; toys; what;] .

d. **[7]** KNOW(\$EXPERIENCER:YOU) ⇔ you+1->know

e. **[3]** SAY(\$SAID:WORD) ⇔say+1->the magic+1->word

f. [**3**] BE(\$THEME:WHERE) ⇔ 1.where's 2.the

Examples a, b and c show the beginnings of category formation, and examples d, e, and f show the emergence of limited constructions.

It would be an overstatement at this point to claim that ***CAMiLLE*** is anything more than a simulation, or that it is necessarily a correct simulation of how learning proceeds. The ultimate test will be whether or not a further version of ***CAMiLLE*** is capable of producing a representation of the learned language 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 abstractness 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.

```
(8) [51] $YNQ([BE; LIKE;]) ⇔ 1.[does; is;]
    [23] $YNQ([LIKE; PLAY;]) ⇔  1.does  3.[like; play;]
```

Our experiments with wh-questions and topicalization have shown that ***CAMiLLe*** can construct adequate local variants of these unbounded dependencies, which may serve well enough in the pre-grammatical stage. Inputs like (9)

```
(9)    who eat-past all my vegetables =
$WHQ(EAT($AGENT:WHO($REF:$WH),$THEME:VEGETABLES($POSSESSOR:I,$QUANT:ALL),$TIME:PAST))
```

lead to rules like (10.

```
(10)   [5] EAT($THEME:WHAT)  ⇔  what+1->do
    [18] [EAT; SPIT;]($TIME:PAST)  ⇔[eat; spit;] ~past
```

And if we modify the input and explicitly tell ***CAMiLLe*** where the traces are –

```
(11)    what do-pres that vaccuum do trace =
$WHQ(DO($AGENT:VACCUUM,$THEME:WHAT($REF:$WH),$TIME:PRES))
```

-- we get rules like (12).

(12) **[251]** `WHAT($REF:$WH) ⇔ what  trace`

But we have not yet seen ***CAMiLLe*** form any hypotheses linking a wh in displaced position to the thematic role assigned to a trace.

Beyond this, it is clear that ***CAMiLLE*** is not able to identify on its own 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 may have to be built in – at this point we see no way for ***CAMiLLE*** to discover them given just primary linguistic data.

5. Language change and variation

Let us now shift our perspective a bit. One way to think of a language is as what exists in the idealized head of the idealized native speaker. Another way to think of a language is as the sum total of what is in the heads of all of the speakers who can communicate with one another. There is a certain organization imposed upon this sum total by the fact that its component parts are acquired by and 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. All that learners have access to is what other people say, which reflects what is in their heads but is distinct from it. In the absence of telepathy, there is necessarily a lack of homogeneity.

Moreover, since not everyone communicates with everyone else, the basic picture that we arrive at 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. But each has somewhat different experiences.

For any linguistic phenomenon that allows for variation, we can map the distribution of the variants over the agents in the network. For example, suppose for the sake of illustration that there are three dimensions of variation, each with two values. They might have a pattern of distribution as follows.

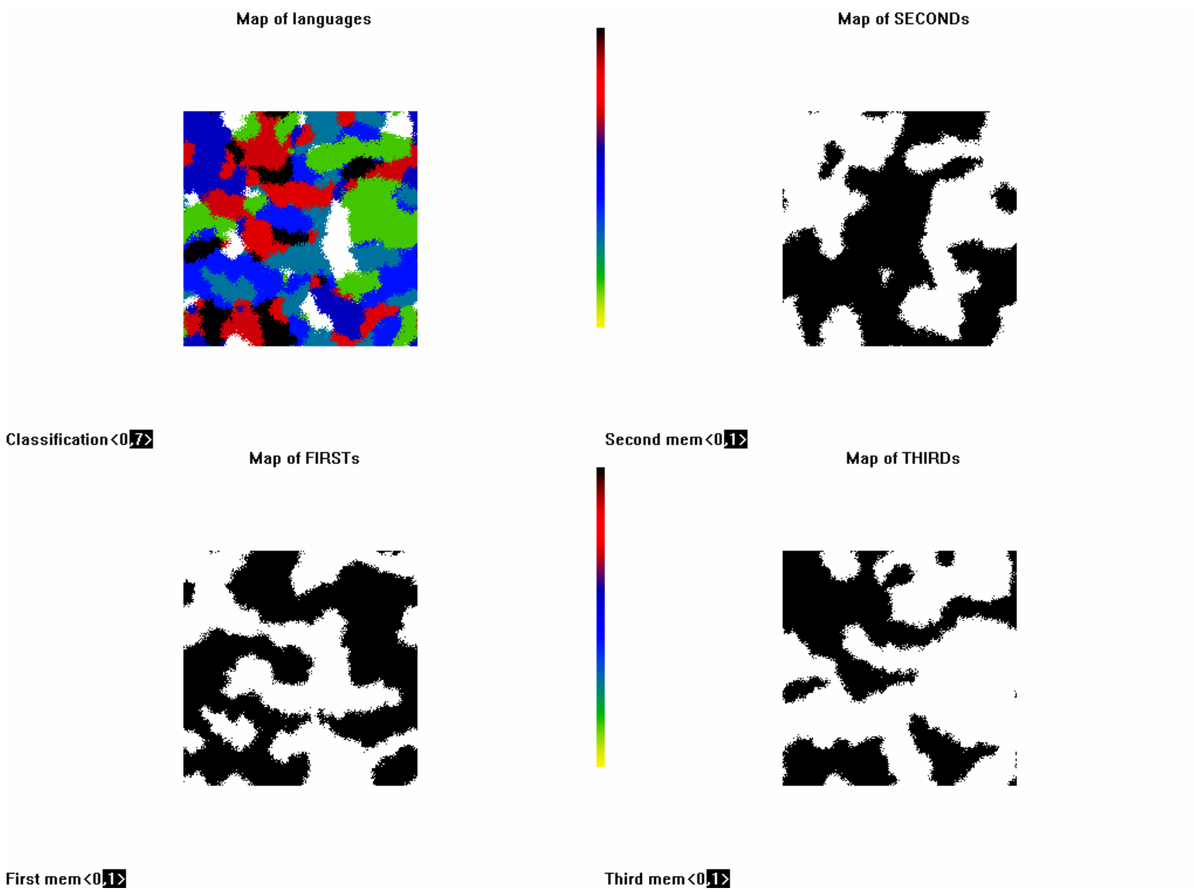

Figure 6. Features and languages

The three black and white arrays show the distribution of the values of the three parameters; the colored array in the upper left shows the combinations that exist.

Recall that earlier I characterized an individual's linguistic knowledge in terms of trajectories. The stability and strength of a particular piece of knowledge can be expressed in terms of the strength of the corresponding trajectory; the generality of this knowledge can be expressed in terms of the breadth of the trajectory, i.e. how much it covers. The most general trajectories will be those that cover all words in a category. Such trajectories correspond, for example, to rules of phrase structure.

Along the same lines, we can track variation in terms of the strength and breadth of similar trajectories across agents. Moreover, multiple variants may exist in a single agent, allowing for socially conditioned variation and in the extreme case bilingualism.

From this perspective, a language is a particular combination of the possible variants. The set of possible languages is then the set of logically possible combinations, however they are to be described. Gaps in the set of logically possible combinations of the possible variants, especially in a very limited geographical region, are very likely the result of the fact that agents will tend to adopt the variants of the agents that they interact with. That is, there will not be an even distribution of all of the possible combinations across the population.

This point is illustrated by a simulation. The initial state of affairs is shown in Figure 7. The composite of the features values for the three binary features is given in the upper left. This simulation assumes that at the outset there is an even distribution of the possible combinations across the population, so that all eight languages are attested and more or less evenly populated.

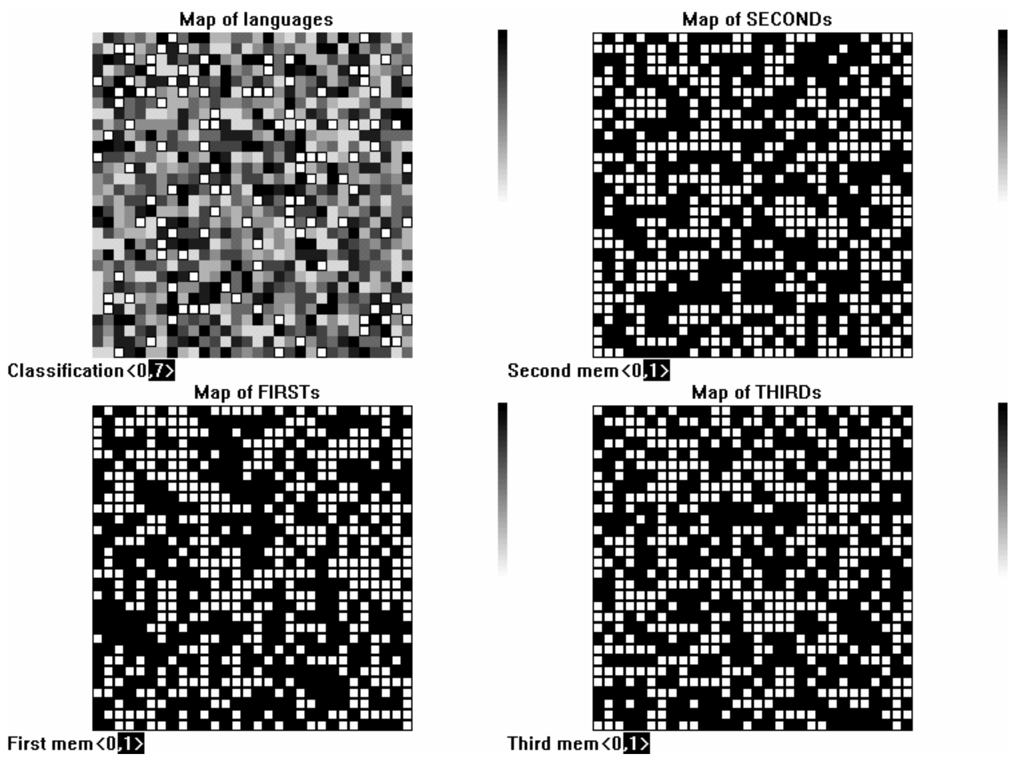

Figure 7. Feature values before clustering

At each step of the simulation, agents influence one another with respect to each feature of variation. An agent will change the value of a feature if more of its neighbors have the other value (where what actually constitutes ‘more’ can be made explicit and even weighted by distance, strength of agents, resistance to change, and other factors or ‘parameters’ of the simulation model).

After 69 simulation steps in which the agents influence one another and possibly change, the distribution of languages and features is as in Figure 8.

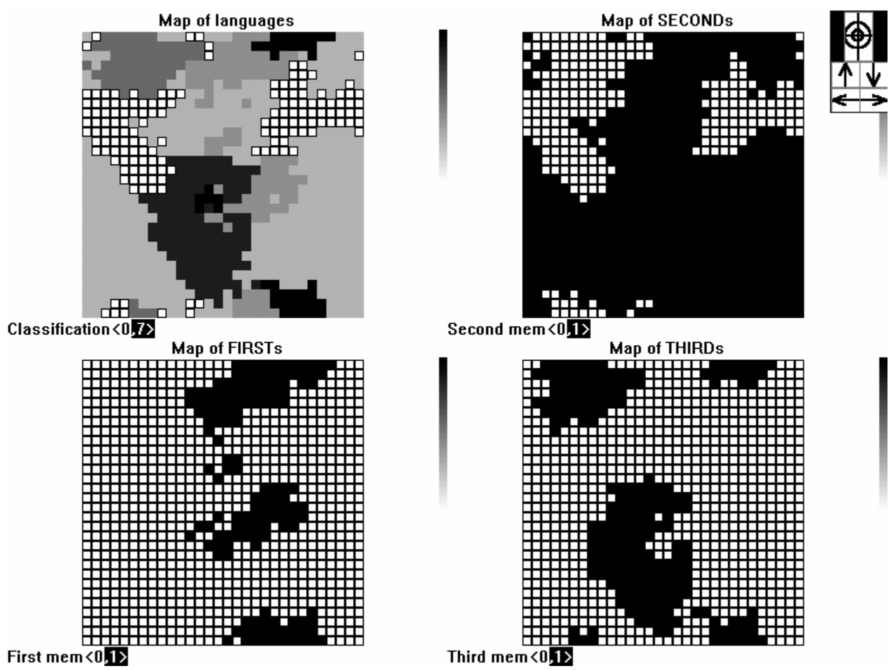

Figure 8. Feature values after clustering

This display shows how each agent has changed its values for each feature over the course of the simulation. As a consequent of changes, some combinations of feature values increase their distribution over the population, while others decrease. The histogram in Figure 9 shows the population levels of the eight languages at this point.

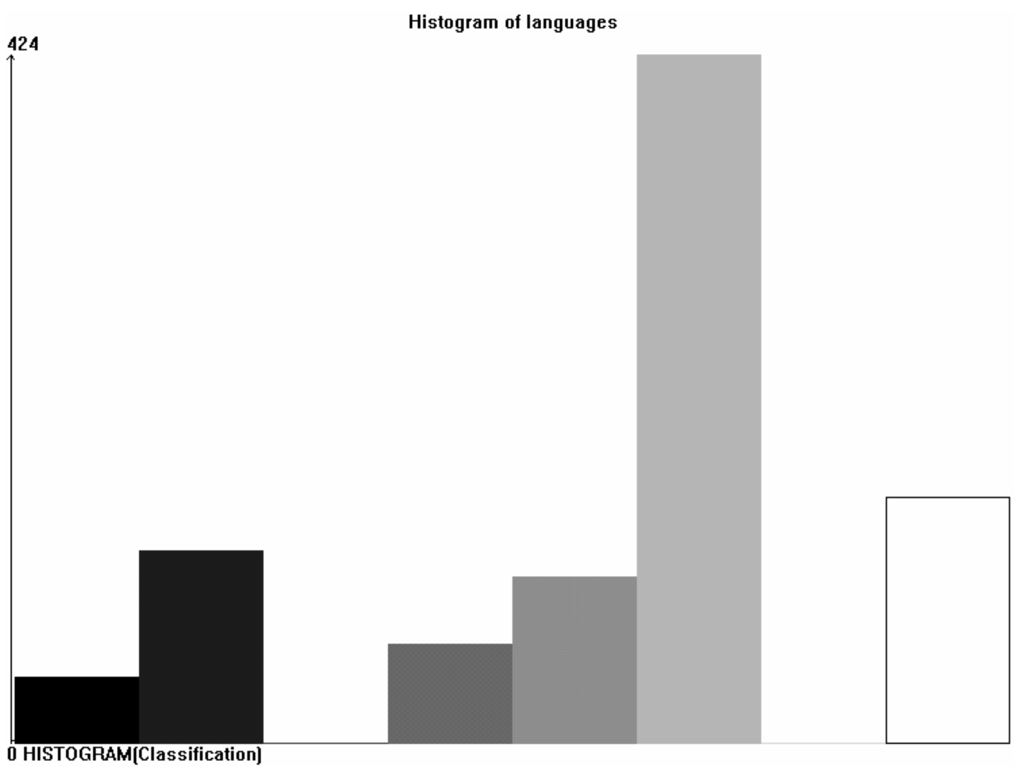

Figure 9. Number of languages of each type

Notice, crucially, that languages 3 and 7 are dead at this point, i.e. there are no agents with the relevant combination of feature values.

The loss of languages illustrated in this particular instance of the simulation is not unique to the starting point shown in Figure 7. It is a consequence of the particular assumptions made in the simulation about how individuals interact in the network. Running the same simulation under the same parameters of interaction yields a different pattern of feature values and languages each time, but the basic results are the same – some combinations survive, and some die out.

The appearance of gaps in the set of languages is an extreme case of a more general property of the simulation model, which is that feature values correlate. The value of a feature for an individual is determined by the interactions between that individual and its interaction partners, over time all of the feature values of an individual tends to become more highly correlated with those of its interaction partners. In other words, the model evolves over time so that there is a tendency for one feature value to appear on a agent when another appears on the same agent. However, the correlation is not 100%, so that the opposite pattern may also occur, although less frequently. An illustration of the history of correlation in a particular simulation is given in Figure 10.

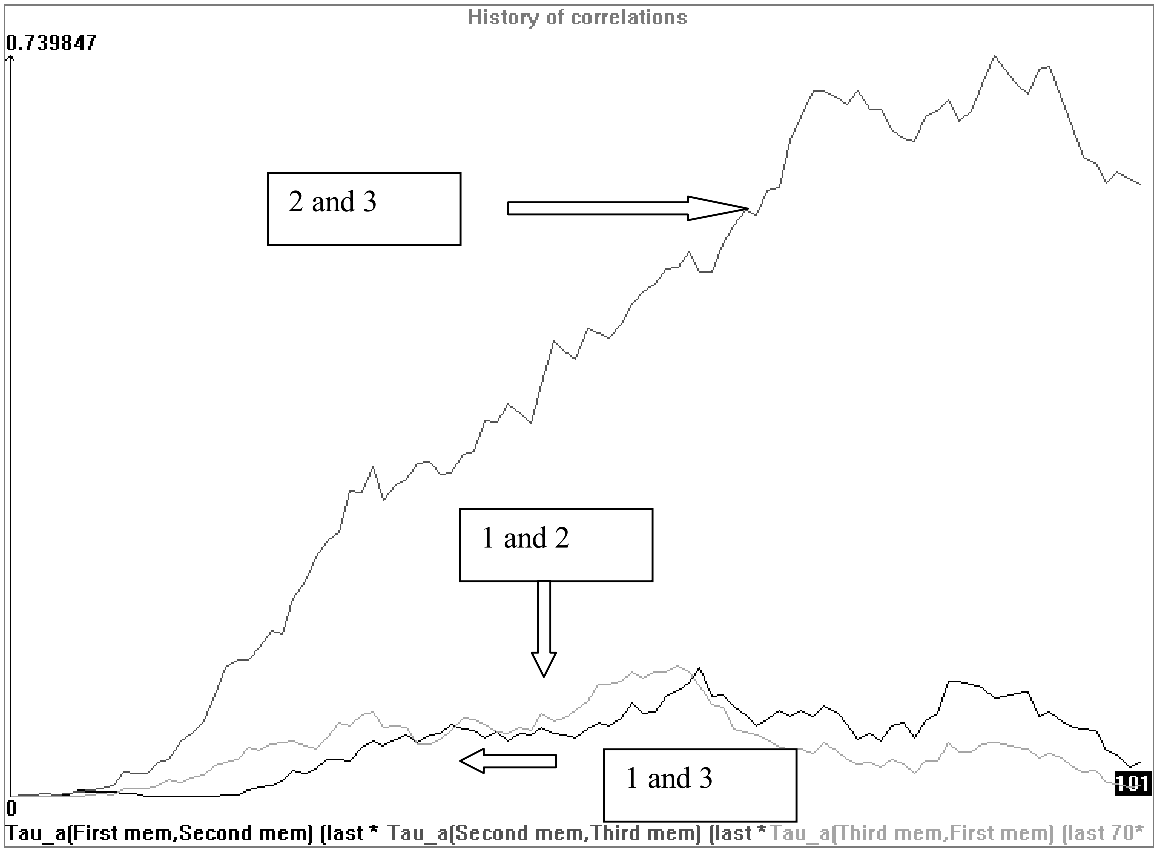

Figure 10. Feature clustering

As can be seen, features F2 and F3 correlate increasingly closely over the course of the simulation, while features F1-F3 and F1-F2 do not.

It is instructive to observe the kind of pattern that is associated with such correlation. In Figure 11 we see the space for F2 and F3 after 100 steps.

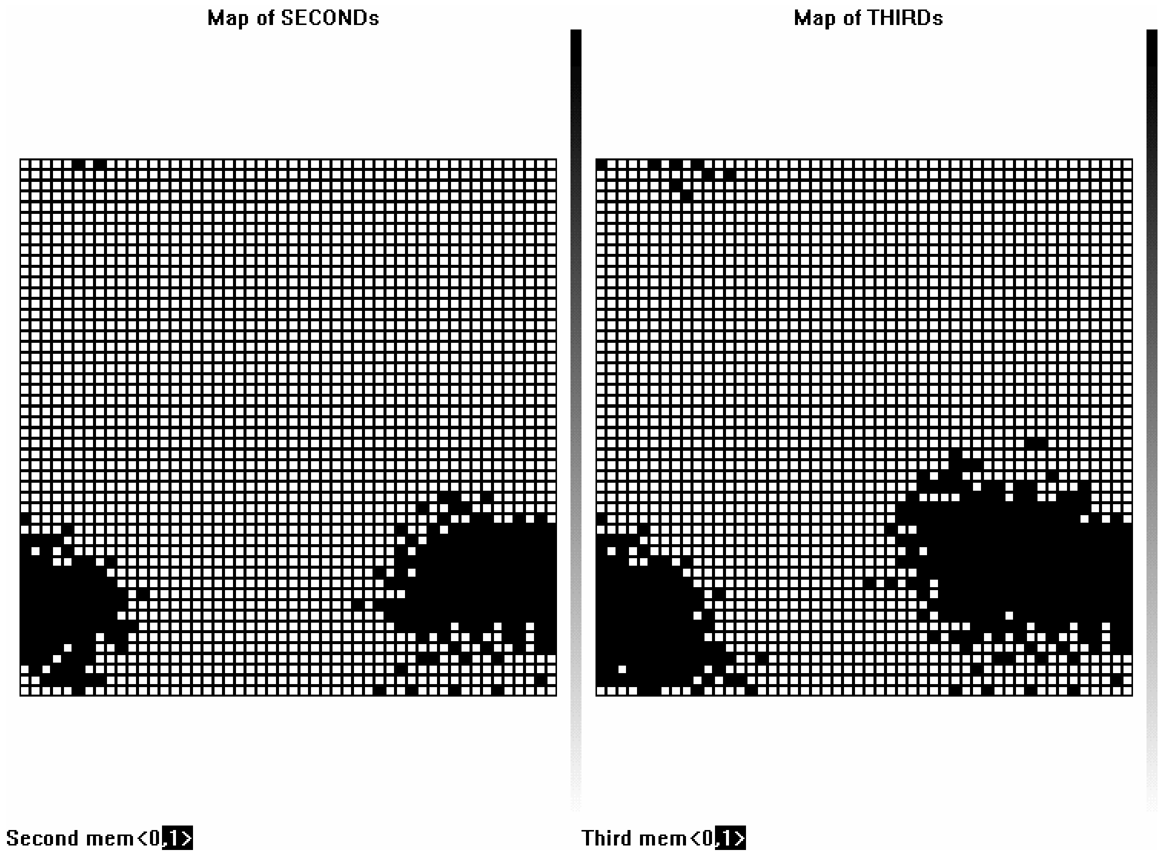

Figure 11. Features 2 and 3 after 100 steps

The visual evidence reveals the tendency for F2 and F3 to have the same value, even through there are exceptions.

Such tendencies may have a linguistic interpretation along the lines of some Greenbergian universals Greenberg 1963. Consider one chosen more or less at random, for example, Universal 4: "With overwhelmingly greater than chance frequency, languages with normal SOV order are postpositional." Let F2 be [+/-postpositional], and let F3 be [+/-OV]. Figure 11 can be understood as an illustration of the situation in which most SOV languages are postpositional, although some are not; we illustrate in Figure 12.

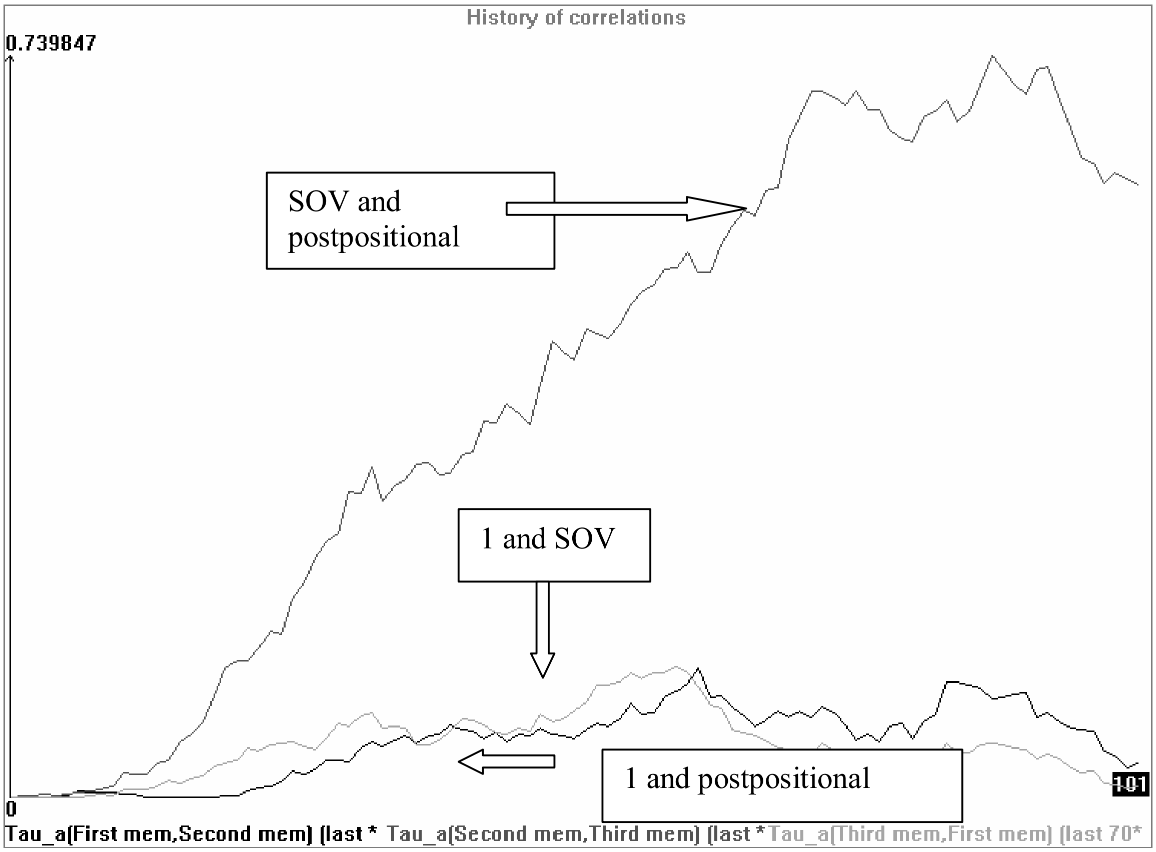

Figure 12. Development of a universal through clustering

The upshot of this simulation model is that one must therefore be cautious in drawing theoretical conclusions on the basis of the absence of certain feature combinations in a relatively small geographical area, e.g. Italy and the South of France. Cases of 'microparametric variation' are especially vulnerable to the claim that the gaps are not linguistically interesting, but arise from the clustering that necessarily takes place in the social network. 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 that 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.

6. Implications and summary

Pursuing this last 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 original view of Markedness (see for example Chomsky 1977) and is related to more recent notions of economy:

> *Syntactic 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. It 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 it will be relatively rare, and where it exists, it will have the status of an exception, an idiosyncrasy, or a lexically restricted subregularity.

The central feature of *Simpler Syntax* is that it takes as primitive the mappings between CS, GF and SS (which is linked to PF). 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 is a natural consequence of the dynamical approach that I have outlined, 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, in the sense that it is universal. 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 to force out the more complex in competition in the social network, other things being equal.

I believe that we can achieve considerable understanding of how language works by tying together the dynamical perspective on language acquisition, the notion that the features of language reside not just in individual brains but in the social network that these brains are part of, and the idea that relative complexity explains the extent to which various ways of accomplishing this mapping occur in languages. On this view, *Simpler Syntax* and the dynamical perspective offer the opportunity of genuine explanation for what we actually find in natural languages, a reality which we can approach with eyes wide open, not squinting.

References

Bloom, P. (2000). *How Children Learn the Meanings of Words.* Cambridge, MA: MIT Press.

Chomsky, N. (1957). *Syntactic Structures.* The Hague: Mouton.

-----. (1977). 'On Wh Movement,' in P.W. Culicover, Wasow, and A. Akmajian, (eds.), *Formal Syntax.* New York: Academic Press. 71-132.

-----. (1994). 'Bare Phrase Structure,' in *MITWPL Occasional Papers.* Cambridge, MA: MIT.

Culicover, P.W. (1999). *Syntactic Nuts.* Oxford: Oxford University Press.

-----, and Jackendoff, R. (2005). *Syntax Made Simpler.* Oxford: Oxford University Press.

-----, and Nowak, A. (2003). *Dynamical Grammar.* Oxford: Oxford University Press.

Greenberg, J. (1963). *Universals of Language.* Cambridge, MA: MIT Press.

Jackendoff, R. (2002a). *Foundations of Language.* Oxford: Oxford University Press.

MacWhinney, B. (1995). *The CHILDES Project : Tools for Analyzing Talk.* Hillsdale, N.J.: L. Erlbaum.

Mayr, E. (1997). *This is Biology.* Cambridge: MA: Harvard University Press.

Newport, E., and Aslin, R. (2004). 'Learning at a Distance: I. Statistical Learning of Non-Adjacent Dependencies,' *Cognitive Psychology,* 48: 127-162. .

Saffran, J., Aslin, R., and Newport, E. (1996). 'Statistical Learning by 8-Month Old Infants,' *Science*, 274: 1926-1928. .

Siskind, J.M. (2000). 'Learning Word-to-Meaning Mappings,' in P. Broeder, and J. Murre, (eds.), *Models of Language Acquisition: Inductive and Deductive Approaches.* Oxford: Oxford University Press. 121-154.

Tomasello, M.J. (2000). 'Do Young Children Have Adult Syntax,' *Cognition*, 74: 209-253. .

Tomasello, M. (2003). *Constructing a Language.* Cambridge, MA: Harvard University Press.