

Adaptive learning and concrete minimalism*

Peter W. Culicover, Center for Cognitive Science and Department of Linguistics, The Ohio State University

We consider the problem of triggering in the theory of language acquisition, called T-theory. T-theory incorporates a number of significant conceptual paradoxes and empirical puzzles. We suggest that a different type of theory, 'adaptation' or A-theory, avoids the paradoxes and puzzles. Interestingly and perhaps surprisingly, A-theory crucially incorporates core notions of markedness theory and shares a number of characteristics of the Minimalist Program.

1. T-theory

1.1 Triggering

On the standard, idealized view of language acquisition, the learner proceeds through a sequence of discrete states on the basis of various "triggering" experiences, where the structure of the data determines the form of the grammar. (See for example Wexler and Culicover 1980 and Gibson and Wexler 1993). In earlier manifestations of this view (e.g. Wexler and Culicover 1980), the learner moved from state to state by hypothesizing grammatical rules. In such a framework, knowledge is modified and added to over time as a consequence of experience. In more recent perspectives, knowledge states are arrived at through the setting of values for a set of universally available parameters (see for example the papers in Roeper and Williams 1987).

Let us call all theories of this general type "triggering" or "T"-theories, and the general approach, "T-theory". In T-theory, the language acquisition mechanism \mathcal{L} extracts a grammatical pattern from the linguistic input T , the potential trigger, checks this pattern against what is licensed by its existing grammar G_i , and if there is a mismatch, makes some kind of a change in the grammar, producing G_{i+1} . The set of possible states that the learner can enter into, the possible grammars, is specified in advance by Universal Grammar (UG). The basic idea of T-theory is that the learner has a very specific idea of what to expect.

I will suggest that T-theory, by its very nature, incorporates a number of significant conceptual paradoxes and raises fundamental empirical puzzles. Unfortunately, limitations of space preclude an extensive discussion of possible solutions to these problems and their consequences.

1. Puzzles

First, some puzzles about T-theory: Embodiment, Subjacency, Parameters and Development.

1.2.1 The embodiment puzzle

It is standard in linguistic theory to think of the grammar as a description of the knowledge that

* I wish to acknowledge the essential contributions made to this paper by Andrzej Nowak. Many of the ideas were stimulated by the work of Jeff Elman and Janet Fodor. I also thank Stephen Crain, Michael Gasser, Ray Jackendoff, Andreas Kathol, Robert Levine, Rita Manzini, Jordan Pollack, Carl Pollard, Craige Roberts, Jack Smith, Laurie Stowe, Deliang Wang, Frank Wijnen, and Jan-Wouter Zwart for suggestions, comments and conversations that have led to numerous improvements in content and presentation.

is somehow EMBODIED in the processing mechanisms in the speaker's mind/brain, which are used for speaking and understanding language.¹ The puzzle is, how does it happen that a real-time psychological mechanism for speaking and understanding, arrived at through a process of acquisition, can embody knowledge with properties such as current (or any) linguistic theory specifies? There is little if any independent direct evidence that competence exists within the mind/brain except the fact that the learner displays knowledge of the language (and by implication, the grammar). (i) As far as we know, for example, brain damage produces psycholinguistic, not linguistic, deficits.² (ii) In the domain of psycholinguistics, there is no evidence that beyond its content the formal grammar explains anything about language processing. (iii) The vocabulary of Universal Grammar is often used to express properties of the course of language acquisition (e.g. "the child sets parameter A in such-and-such a way") but this does not constitute independent evidence for the existence of parameters, particularly if there is no evidence that a parameter is ever set to the wrong value. In brief, there is little if any clear evidence that there exists a computational device in the mind/brain that corresponds in any interesting way to a formal grammar.

1.2.2 Subjacency and learnability

Another puzzle concerns the role of linguistic constraints, in particular Subjacency, but also ECP, Relativized Minimality, Manzini's Locality Principle, etc. It has been suggested that constraints such as Subjacency ensure learnability of language, by restricting the range of hypotheses available to the learner to a relatively small set of possibilities consistent with any linguistic experience. (See e.g. the quote from Chomsky 1965; see also Wexler and Culicover 1980.) These constraints form a part of UG. The constraints rule out the possibility that the learner will fail to converge over time to a "correct" hypothesis about the grammar of the language.

But recent formulations of linguistic theory (especially Principles and Parameters and the Minimalist Program) restrict the expressiveness of the theory to the point that learnability in this sense is no longer an issue. What is the explanatory role of these constraints, then?

On the other side of the coin, it appears that Subjacency is true of some languages, such as English. This fact at first appears to lend some empirical credibility to the proof of the efficacy of Subjacency from learnability, as argued by Wexler and Culicover (1980). But there are languages such as Italian, Swedish and Icelandic in which systematic violations of Subjacency are possible, in the sense that sentences which would be judged ungrammatical in English are judged grammatical in these languages. There is evidence that it does not even hold uniformly for English (Erteschik-Shir 1993; and elsewhere; Pollard and Sag 1994).

The Subjacency Puzzle is, If constraints are not relevant to learnability, where do they come from?

1.2.3 Parameters

Next let us consider the Parameters Puzzle. The vision of Principles and Parameters theory, as exemplified by Chomsky and Lasnik 1991, is that complex linguistic phenomena, expressed as grammaticality judgments, can be understood as the product of the interaction of relatively simple and **substantive** parameters with a small number of values, preferably two. (The point is made particularly clearly by Safir 1987.) By setting the parameter values one way or another, we hope to derive complex patterns of grammatical sentences, but only those particular patterns that can be expressed in terms of the precise parameters. But if it turns out that the observable differences are

¹ The term "embodiment", used in in Culicover (1993), is not familiar, but the concept is.

² There have been attempts to find evidence for the architecture of linguistic theory in the phenomena of language deficits, e.g. Grodzinsky (1990). For discussion and criticism, see Crain's (1992) review.

more or less arbitrary and unprincipled, or if everything imaginable is possible, we can continue to call the differences "parametric" but the term is without content. The puzzle is, if substantive parameters exist, why do we find so few good examples of them in the data? (Presumably the learner would have no trouble doing so, if they exist.)

1.2.4 Development

A last puzzle concerns the nature of initial and intermediate states. The assumption that there is a fully specified default state for the grammar either in part or in its entirety runs up against the obvious fact that at the outset young children are unable to speak or understand any language at all, and their linguistic ability improves gradually over time. It is true that there are some things about language that children do appear to know from the very earliest stages of development, which they could not have learned through experience (Crain 1991). But crucially the idealization assumes that all of X' theory is present from the outset, and would have the learner set parameters on the basis of single exposures. Even if we concede that at the earliest stage children lack the full elaboration of the functional categories in some way (e.g. Radford 1990; see Poeppel and Wexler 1993 for a contrasting view), it comes as somewhat of a surprise that children do not display essentially complete and well-elaborated phrase structure very early in their lives, on a par with the ability of other animals to walk immediately after birth, for example. I call this the Puzzle of Development.

One might say that there is a maturational component to development of phrase structure, one that is driven not by experience but by the basic biology (see Borer and Wexler 1987; Rizzi 1993). But until we understand the nature of the biology, maturation must be understood as our last and least preferred refuge after we have failed to account for development in any other way, a point that has been made by a number of people (see for example Bloom 1991).

1.3 Paradoxes

Let us now turn to the paradoxes, of which there are two, Parsing and Idioms.

The Parsing Paradox concerns the already mentioned idea that the learner creates mechanisms for production and comprehension that "embody" the knowledge of language expressed by the grammar. Presumably, in order to extract a grammatical pattern from a input sentence, the learner must already have a mechanism for comprehension (or at least for parsing), so that the input structure can be checked against the current grammatical hypothesis or trigger a revised hypothesis. Analysis of the input, whether or not it is compatible with the current state, is an essential component of learnability theory. The question is, where did this mechanism for input analysis come from?

One answer is that there is an innate functioning Universal Parser ρ_U , which seems to contradict the assumption that there is something to be learned, given that a parser is the embodiment of a learned grammar.³ The other is that there is not. Suppose that there is no Universal Parser, and suppose that at any stage i the parser ρ_i "embodies" the grammar G_i . Suppose that the grammar G_i incorporates an error, so that the datum T has a structure that conflicts with the structure licensed by G_i . Then the parser ρ_i cannot assign a structural representation to this datum. In order to explain learning in such a circumstance, T-theory might assume that there is some other (universal) mechanism, ρ_U that can extract from T the offending structure, analyze it, and use this analysis to map G_i into G_{i+1} as shown in (10). Once again we are back to a Universal Parser.

The implausibility of this solution has led some T-theories to assume that failure to analyze a structure leads to random revision of the current grammar. (See, for example, Hamburger and

³ The Universal Parser could be restricted just to the analysis of particular types of constructions, along the lines of Fodor's (1993) "designated triggers".

Wexler 1975, Wexler and Culicover 1980, Gibson and Wexler 1993). Under certain circumstances such (Markovian) revision of the grammar produces convergence to a correct grammar for the input data, in the limit. This approach obviously does not require a Universal Parser.

The Parsing Paradox is, then, the following. T-theory entails either that there is an fully functioning innate Universal Parser, or that language learning in the sense of grammar revision is Markovian. As far as I know there is no good empirical evidence in support of either alternative.

Another difficulty with T-theory concerns the status of isolated cases that conflict with the general pattern, a point that has been made by Fodor (1993). There are two types of isolated cases, mistakes and idioms. A mistake is a slip of the tongue, while an idiom is an isolated special case. In addition, there are subregularities that do not generalize to an entire class.

In T-theory, a single exemplar is sufficient in principle for establishing the pattern in the processing mechanisms. As Fodor has pointed out, the existence of idioms produces a paradox (and mistakes appear to produce the same paradox). Just as a single exemplar of the pattern to be learned triggers learning, a single exemplar of the idiom whose structure contradicts this pattern should trigger unlearning, or learning of the wrong grammar. While the proportion of "typical" to "idiomatic" patterns may be enormous, such a discrepancy in the statistical likelihood of encountering the various patterns plays no role in T-theory. Call this the Idiom Paradox.

2. A-theory

2.1 Concepts

What I would like to do now is to briefly explore a different way of thinking about the acquisition knowledge of language, one that may be called "adaptive" or "A-theory". The idea that A-theory in general is a more satisfactory way of accounting for the representation of language, and of cognition in general, is familiar from connectionist proposals, but the basic idea is independent of the particular implementation. For some precedents, see for example the papers in Sharkey (1992). Connectionist approaches are attractive from the perspective of language acquisition because they appear to approximate in a more or less natural way the growth of complex knowledge on the basis of experience. However, a standard criticism of some connectionist proposals regarding language is that they may show how a **language** as a set of sentences can be acquired by some device, but at best display "nothing more than a superficial fidelity to some first-order regularities of language" (Pinker and Prince 1989; see also Lachter and Bever 1989). What is needed is an account that accounts not just for sentence patterns but for knowledge of language.

2.2. A dynamical approach

I summarize one plausible approach. Suppose that the representation of linguistic expressions in Syntactic Space should be dynamic, in the sense that what corresponds to a linguistic expression is a trajectory with extent and direction. Understanding or producing a sentence involves traversing a particular path through the space. If a particular trajectory has been traversed, subsequent traversal of it requires less energy. Syntactic generalizations (such as phrase structure rules and transformations like Move α) correspond to regions of flow within this space; that is, trajectories that more or less parallel one another, that link the same regions. A flow is a trough in Syntactic Space. If such a flow exists, this does not mean that every trajectory within the region of flow has been traversed as a consequence of experience. But traversing a new trajectory within the region of flow requires relatively little energy, as does traversing a previously laid down trajectory. A sentence is judged grammatical if traversing its trajectory requires little energy. (In effect we have

the core of a solution here to the so-called Projection Problem (Peters 1972), but I cannot pursue the point here.)

It is natural to think about such a model in the following way: if every trajectory in a region were realized, the result would be equivalent in content to the linguist's (ideal) grammatical description of the knowledge of the native speaker. Such a configuration would be the result of exposing the learner to infinite experience over infinite time. T-theory applies well enough to this ideal conception of learning; however, such an account in principle cannot deal with any real aspects of language acquisition that are sensitive to the finiteness of the learner's experience.

2.3. Solutions

We now have enough of a picture to begin to see the conceptual basis for eliminating the Embodiment Puzzle, the Idiom Paradox, the Paradox of Subregularities, and the Parameters Puzzle. The embodiment of the grammar in the processing mechanisms is in fact achieved through the development of flows that together define the language represented in the model. The configuration of this space is a reflection of the learner's experience, it incorporates the learner's knowledge of language, and drives the learner's language processing; hence it answers in principle the question of how knowledge is embodied in processing. Of course, the trick is to provide the details.

Learning on this model does not proceed by triggering, but by accretion. A mistaken hypothesis, should it arise on the basis of particular experience, will not be sustained by future experience; the region of space that it occupies will eventually be absorbed by the "correct" cases. An idiom will be strongly supported by experience, but will not occupy a large region of space. A subregularity will occupy a well-defined region, its boundaries may be somewhat ill-defined, but it will be able to coexist peacefully and stably with other general patterns. In effect, the size of a region of this space corresponds to the statistical properties of the input. The Parameters Puzzle is eliminated, since there are no parameters; again, learning is through concrete experience.

There is no Parsing Paradox, since in A-theory, no analysis of the input and matching of the analysis with a grammar is performed. On this approach, syntactic categories do not exist except as clusterings of lexical items and flows in Syntactic Space. Specific patterns are laid down and categories form through the groupings of elements with similar properties. Representation of linguistic expressions in terms of trajectories provides a natural way of representing word order without extracting an explicit grammatical representation from the input.

3. Implications

3.1 Conceptual structure

A standard criticism of the A-theoretic approach to language has been that it may show how a LANGUAGE as a set of sentences can be acquired, but it does not explain why a GRAMMAR takes the form that it does. On the standard view in linguistic theory, the explanation resides in UG, of course, which A-theory lacks. With respect to the lexical categories, for example, the question naturally arises, if there is no UG with a universal inventory of categories, how does the learner figure out which words belong with which in the first place? The answer must be that those strings that exemplify genuine syntactic categories must correspond or be systematically related to elements in a representation that is external to language; I suggest Conceptual Structure (Jackendoff 1990). Assume that CS has representations for physical objects corresponding to nouns, for concrete actions corresponding to verbs, for predication and the subject-predicate relation, and for operator-variable binding structures. I hypothesize that the connection with CS is sufficient to bootstrap the

syntax (Pinker 1984; Lebeaux 1988; Gleitman 1990); once the sentential core is up and running, there can be significant divergences between CS and syntax.⁴ .

3.2. Constraints

On this approach constraints must come from one of two sources: either they correspond to regions of Syntactic Space that are relatively difficult to traverse, or they are reflections of processing limitations. I will appeal to a notion of relative complexity defined over syntactic structures. The basic idea is that the possibility of a movement chain in a simple structure does **not** generalize to more complex domains. The idea of generalization is a natural one if we think of linguistic knowledge as a grammar that contains something of the form Move α , but not if we think of it as a set of dense flows in Syntactic Space.

The idea, then, is that the regions of Syntactic Space corresponding to more complex extractions are attainable, but only on the basis of positive experience. Failure to experience more complex extractions leads to a state of affairs in which such extractions are judged to be ungrammatical, in the sense introduced earlier; the corresponding regions of Syntactic Space cannot be easily traversed. Locality constraints such as Subjacency follow, then, not from grammar learnability considerations but from the relative complexity and accessibility of structures.⁵

The picture suggested by the Subjacency Puzzle is that violations of Subjacency are possible, but less highly valued by the learning mechanism, other things being equal. Hence, if a learner is exposed only to examples that satisfy Subjacency, after sufficient experience, violations of Subjacency will be substantially below threshold (in fact, near zero) while extractions obeying Subjacency will be substantially above threshold. It will therefore appear to be the case that in the absence of positive evidence, the learner "knows" that Subjacency cannot be violated. But if the learner is exposed to violations of Subjacency in sufficient quantity, the learner will "know" that Subjacency can be violated, and, given that there is some type of independent complexity metric, that it is computationally more costly to do so.⁶ It is through this means that the A-theory eliminates the argument from the poverty of the stimulus so often cited as evidence for innate knowledge specific to language, without requiring negative evidence in exchange. What is wired into the system is not that Subjacency is a constraint, but a complexity metric that renders violation of Subjacency less highly valued. We have thus returned to the perspective of Chomsky (1964), where a more "marked" alternative is possible if there is specific evidence to support it. A-theory provides us with a natural way of capturing the markedness phenomenon.

This approach promises to account not only for Subjacency phenomena, but for a more extensive hierarchy of extraction possibilities. Suppose we take the subject/predicate split to be a fundamental feature of CS. Then the accessibility of extraction appears to be more or less as in (11), where the subject is immediately accessible, the arguments of the predicate next, and so on. Also, local extraction is more accessible than long extraction:

SUBJECT > OBJECT > PREPOSITIONAL OBJECT > ADJUNCT

⁴ As suggested by Culicover and Jackendoff (1995).

⁵ Such complexity can be formulated in terms of Complete Functional Complexes, at least as a first approximation; I will not pursue the point here. Notice also that this approach suggests a possible account for the fact that linguists who work on Subjacency phenomena in English find certain extractions to be better over time, as they encounter them more frequently. We call this 'life-long learning'.

⁶ We can account in this way for the phenomenon of 'lifelong learning' (thanks to Carl Pollard for the term) whereby linguists who are native speakers of English find that the more often they produce and judge Subjacency violations, the weaker the violations become.

LOCAL > FROM SENTENTIAL COMPLEMENT > FROM SENTENTIAL ADJUNCT (Subjacency)

This formulation appears more or less consistent with the work of Keenan and Comrie (1977) on NP accessibility. Moreover, such a hierarchy recalls the notion of o-command used in the analysis of binding in HPSG (Pollard and Sag 1994), and may be seen as providing a possible foundation for such hierarchical generalizations.

3. Concrete minimalism

What now of the distinction between grammar and parser (or competence and performance) that is so fundamental to T-theory? In T-theory, the grammar is a characterization of the possible structures of a language; the grammar is translated into mechanisms for speaking and understanding that implement the knowledge expressed in the grammar. In A-theory, the grammar is embodied in the performance mechanism that resides in Syntactic Space, through the particular flows that are established.

If the foregoing is more or less on the right track, it offers the following picture of the relationship between linguistic theory and what is in the native speaker's head. Linguistic theory, as the home of Universal Grammar, is responsible for markedness, that is, for the complexity metric stated over possible linguistic structures. Semantics provides a universal characterization of core phrase structures, including variable binding constructions. Variation beyond the core is in principle unbounded, subject to the constraint of interpretability. The relative complexity of a given construction on the complexity hierarchy will determine the likelihood that it will occur with sufficient frequency to make it an actual grammatical construction of some language.

Furthermore, on this approach syntactic theory itself has little if any content. Syntax is indeed "minimalist" with a vengeance. None of the specific formulations of GB theory and its extensions (Chomsky 1981; 1986) that rely on a type of Uniformity principle are particularly plausible, unless of course we find some motivation within the dynamical systems approach for such a uniformity principle. Effect we have arrived, albeit by a very different route, to many of the positions adopted in the Minimalist Program of Chomsky (1995). A fuller discussion of the similarities and differences must be left to another place.

4. Conclusion

If the conclusions that we have drawn are more or less correct, a single theory will account for competence, performance, and acquisition. But it does not follow that we should stop doing syntax or formulating syntactic theory. Syntactic theorizing remains as the means by which we try to understand what it is that the language system is capable of doing. The importance of formal syntax on this view is not so much that it provides the solutions, it is that it provides the problems, many of which exist independently of particular theories. In seeking to understand how language is represented in the mind/brain, I expect that we will find that the lasting value of theories such as GB Theory, Principles and Parameters, the Minimalist Program, LFG, GPSG, HPSG and so on, rests to a large extent on how they have made clear precisely what the term "language" refers to, and what it is that needs to be explained.

References

- Bloom, P. (1991). Subjectless sentences in child language. *Linguistic Inquiry* 21, 491-504.
- Borer, H. and K. Wexler (1987). The maturation of syntax. In: T. Roeper and E. Williams (Eds.) *Parameter Setting*. Dordrecht: D. Reidel.
- Chomsky, N. (1964). *Current Issues in Linguistic Theory*. The Hague: Mouton.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, Mass.: MIT Press.
- Chomsky, N. (1981). *Lectures on Government and Binding*. Dordrecht, Holland: Foris Publications.
- Chomsky, N. (1986). *Knowledge of Language: Its Nature, Origin, and Use*. New York: Praeger.
- Chomsky, N. (1995). *The Minimalist Program*. Cambridge, Mass.: MIT Press.
- Chomsky, N. and H. Lasnik (1991). Principles and Parameters Theory. In: J. Jacobs, A. von Stechow, and T. Vennemann (Eds.) *Syntax: An International Handbook of Contemporary Research*. Berlin: Walter de Gruyter.
- Crain, S. (1991). Language acquisition in the absence of experience. *Brain and Behavioral Sciences* 14, 597-650.
- Culicover, P. W. and R. Jackendoff (1995). *Something else for the Binding Theory*. *Linguistic Inquiry* 26.
- Erteschik-Shir, N. (1993). *The dynamics of focus structure*. ms. Tel Aviv: Ben Gurion University.
- Fodor, J. D. (1993). How to obey the Subset Principle: binding and locality. to appear. In: B. Lust, G. Hermon, and J. Kornfilt (Eds.) *Syntactic Theory and First Language Acquisition, Vol 2: Binding, Dependencies and Learnability*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Gibson, E. and K. Wexler (1994). Triggers. *Linguistic Inquiry* 25, 407-453.
- Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition* 1, 3-55.
- Hamburger, H. and K. Wexler (1975). A Mathematical Theory of Learning Transformational Grammar. *Journal of Mathematical Psychology* 12, 137-177.
- Jackendoff, R. (1990). *Semantic Structures*. Cambridge, Mass.: MIT Press.
- Keenan, E. and B. Comrie (1977). Noun phrase accessibility and universal grammar. *Linguistic Inquiry* 8, 63-99.
- Lachter, J. and T. Bever (1988). The relation between linguistic structure and associative theories of language learning - a constructive critique of some connectionist learning models. *Cognition* 28, 195-247.
- Lebeaux, D. (1988). *Language Acquisition and the Form of the Grammar*. unpublished doctoral dissertation, Amherst, Mass.: University of Massachusetts.
- Peters, P.S. (1972). The projection problem: how is a grammar to be selected? In: P.S. Peters (Ed.) *Goals of Linguistic Theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Pinker, S. (1984). *Language Learnability and Language Development*. Cambridge, Mass: Harvard University Press.
- Pinker, S. and A. Prince (1989). On language and connectionism: analysis of a parallel distributed processing model of language acquisition. In: S. Pinker and J. Mehler (Eds.) *Connections and Symbols*. Cambridge, Mass: Bradford Books.
- Poeppl, D. and K. Wexler (1993). The full competence hypothesis of clause structure in early German. *Language* 69, 1-33.
- Pollard, C. and I. Sag (1994). *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press.
- Radford, A. (1990). *Syntactic Theory and the Acquisition of English Syntax*. Oxford: Basil Blackwell.
- Rizzi, L. (1993). Early null subjects and root null subjects. In: T. Hoekstra and B. Schwartz (Eds.) *Language Acquisition Studies in Generative Grammar*. John Benjamins.
- Roeper, T. and E. Williams (1987). *Parameter Setting*. Dordrecht: D. Reidel.
- Safir, K. (1987). Comments on Wexler and Manzini. In: T. Roeper and E. Williams (Eds.) *Parameter Setting*. Dordrecht: D. Reidel.
- Sharkey, N. E. (Ed.) (1992). *Connectionist Natural Language Processing*. Dordrecht: Kluwer

Academic Publishers.
Wexler, K. and P. W. Culicover (1980). *Formal Principles of Language Acquisition*. Cambridge, Mass.: MIT Press.

Peter W. Culicover
Center for Cognitive Science and Department of Linguistics
The Ohio State University
208 Ohio Stadium East
1961 Tuttle Park Place
Columbus, OH 43210-1102
USA

culicover.1@osu.edu