3-23-2001

Toward a Psychological Analysis of the Sentence from the Work of Lashley, Chomsky, Wundt, Polanyi, and Skousen's AML

Bruce L. Brown

Follow this and additional works at: https://scholarsarchive.byu.edu/dlls

BYU ScholarsArchive Citation

Brown, Bruce L. (2001) "Toward a Psychological Analysis of the Sentence from the Work of Lashley, Chomsky, Wundt, Polanyi, and Skousen's AML," *Deseret Language and Linguistic Society Symposium*: Vol. 27 : Iss. 1 , Article 3.
Available at: https://scholarsarchive.byu.edu/dlls/vol27/iss1/3

Toward a Psychological Analysis of the Sentence from the Work of Lashley, Chomsky, Wundt, Polanyi, and Skousen's AML

Bruce L. Brown

Lashley (1951) and Chomsky (1957) clearly demonstrated the inadequacy of "left-right" associationistic models in accounting for language and other kinds of holistically patterned behavior. Both argued persuasively that the kind of holistic dependencies among elements that characterize language syntax cannot be explained through behavioristic S-R connections. Chomsky (1957, 18–25) began his attack on behavioristic theories of language by demonstrating the inadequacy of Markov processes (a precise embodiment of S-R chaining theory) in accounting for patterned sequences of behavior. In particular, he showed that the kinds of holistic dependencies among elements that characterize syntactic structures in language could not be accounted for with left-right associationistic models, but rather, would require a top-down hierarchical approach.

In his influential paper on the problem of serial order in behavior, Lashley (1951) made a similar case for the necessity of hierarchical explanation, but from a neurological point of view. The lines of his argument were quite different from Chomsky's. Chomsky's argument was essentially formal and based upon artificial models of logical mechanisms. Lashley's argument was neurological, but also conversational and straightforward. He first reviewed a variety of anecdotal observations concerning language and then asked what kind of neurological organization would be necessary to account for them.

He pointed out that a given set of phonemes in spoken words (or of letters in typed words) can occur in a number of combinations, such as the reverse combinations *right* and *tire* (p. 115). Lashley then made the very obvious point that "the order must therefore be imposed upon the motor elements by some organization other than direct associative connections between them" (1951, 115). He further argued that words stand in relation to sentences as letters do in relation to words, and that words also have no intrinsic temporal valence as implied by the associative chaining models. Drawing upon an analysis of the language translation process, he argued that this syntactic order is also not to be attributed to the thought process—the same thought can be expressed with quite different temporal structures in different languages. Translators translate holistic thoughts, not word by word. As he summarized: "the mechanism which determines the serial activation of the motor units is relatively independent, both of the motor units and (also) of the thought structure." Lashley (p. 115) argued that language is not the only example of this kind of syntactically structured behavior, that a multitude of skilled behaviors in man and other animals display this kind of implicit hierarchical structure and cannot be explained in terms of associative connections among the elements.

Wundt (1912, Chap. 7, "Die Satzfügung") reasoned from a very different perspective. His primary task was to explain the formation

of sentences. He reasoned that any expla-
nation of the sentence that focuses only
upon its surface structure would obviously
be inadequate. He characterized the
sentence as "both a simultaneous and a
sequential structure" (see p. 21 of
Blumenthal). It is simultaneous because at
any given moment it is present in con-
sciousness as a totality even as the individ-
ual words are spoken. We focus upon the
whole of what we are saying even as the
words flow forth in a habitual way that is
not introspectible to us. As he said:

> The sentence, however, is not an
> image running with precision
> through consciousness where each
> single word or single sound appears
> only momentarily while the preced-
> ing and following elements are lost
> from consciousness. Rather, it stands
> as a whole at the cognitive level while
> it is being spoken. If this should ever
> not be the case, we would irrevocably
> lose the thread of speech. (quoted in
> Blumenthal 1970, 21)

Like Chomsky, Wundt held that any
explanation of the sentence that focuses
only upon its surface structure would
be obviously inadequate. But unlike
Chomsky's position, both Wundt's
account and that of Lashley left open the
question of whether the psychology
of the sentence requires one to posit
the literal existence of syntactical rules
in the human psyche.

Clearly a strong case can be made
for an explanation of patterned serial
behavior that does not attribute it to
associative connections among the ele-
ments. However, we cannot consider that
demonstration to be equivalent to making
the case for rule-based explanations.
Some have taken it this way. In particu-
lar, the Chomskian approach put phrase
structure rewrite rules and transforma-
tional rules in center stage and imbued
them with ontological status, thus open-
ing the way for a new era of mentalism in
the behavioral sciences. The new artificial

intelligence (AI) brand of cognitive
psychology further built upon this
unbridled mechanistic mentalism, much
to the detriment of a truly cognitive
approach to explanation. The excesses of
the AI movement were at least as out-
rageous as those of the behaviorists a
decade or two earlier. The behaviorists
insisted on mechanistic explanations, but
also on the law of parsimony. Neo-
cognitivists seem to be willing to sacrifice
parsimony as long as a computer
metaphor is satisfied, to guarantee mech-
anistic explanation.

But parsimony still makes sense.
There is no reason to create complex,
burdensome explanations if simpler ones
will suffice. Polanyi's characterization of
the nature of skills led the way for us
here. He began his discussion of the
psychology of skills (1962) with the
trenchant statement:

> I shall take as my clue for this
> investigation the well-known fact
> that *the aim of a skilful performance is
> achieved by the observance of a set of
> rules which are not known as such to
> the person following them.* (49)

He then went on to offer explanations
of the physical principles underlying
swimming and riding a bicycle, but with
the caveat that one certainly would not
have to understand those explanations to
perform either of these skills. Either of
these skills is acquired tacitly through
trial and error, or through apprenticeship,
but without explicit awareness of the
principles involved. This approach to
explaining the acquisition of skills
(including linguistic skill) is consonant
with influential theories of perception,
such as the "transactionalism" of
Ames (1946) and Kilpatrick (1961)
and J. J. Gibson's theory (1966) of
"direct perception."

A full explanation of the principles
involved in any of these skills is probably
beyond our present scientific capability.
However, Polanyi (1962) offered the

following as a first approximation of the physical principles involved in riding a bicycle:

> Again, from my interrogations of physicists, engineers and bicycle manufacturers, I have come to the conclusion that the principle by which the cyclist keeps his balance is not generally known. The rule observed by the cyclist is this. When he starts falling to the right he turns the handle-bars to the right, so that the course of the bicycle is deflected along a curve towards the right. This results in a centrifugal force pushing the cyclist to the left and offsets the gravitational force dragging him down to the right. This manoeuvre presently throws the cyclist out of balance to the left, which he counteracts by turning the handlebars to the left; and so he continues to keep himself in balance by winding along a series of appropriate curvatures. A simple analysis shows that for a given angle of unbalance the curvature of each winding is inversely proportional to the square of the speed at which the cyclist is proceeding. But does this tell us exactly how to ride a bicycle? No. You obviously cannot adjust the curvature of your bicycle's path in proportion to the ratio of your unbalance over the square of your speed; and if you could you would fall off the machine, for there are a number of other factors to be taken into account in practice which are left out in the formulation of this rule. Rules of art can be useful, but they do not determine the practice of an art; they are maxims, which can serve as a guide to an art only if they can be integrated into the practical knowledge of the art. They cannot replace this knowledge. (49–50)

Obviously, being able to explain bicycle riding at this high level of abstraction and physical decomposition is not a prerequisite to performing the skill. There are many six- and seven-year-old children who have mastered the skill of riding a bicycle, but the explanation given above would probably mean very little to any of them. Nor would it make sense to hypothesize the existence of "rules" of this kind in their heads, a kind of inborn, unconscious, unintrospectible BRAD ("bicycle riding acquisition device"). There are many ways to explain what one is doing, some explicitly physical (such as the foregoing), some metaphorical, some complex, and some simple, but probably none of these levels of explanation fully captures or exhausts what is actually going on.

I remember hearing Chomsky say in a talk at McGill University in 1967 that the logical capability implicit in the linguistic performance of a typical three-year old is more complex than the principles of calculus. At the time I found that statement preposterous. With thirty-five more years of experience the statement now seems obvious and correct. The two sentences "They are easy to please" and "They are eager to please" at first seem alike in structure, and their surface structure is similar. However, an impersonal transformation shows that they are very different in deep structure: "It is easy to please them," but not "It is eager to please them." Polanyi would explain this in terms of the contrast between *tacit* knowledge and *explicit* knowledge. We have a tacit apprehension of linguistic principles of great depth and subtlety, but we do not have explicit knowledge of the principles involved. Chomsky's subtle and complex linguistic rules could be viewed in this framework as being an explicit spelling out of the logic underlying what every person can do linguistically without taking thought, without being able to introspect. T. G. R. Bower (1977) and his colleagues have shown that Piaget's (1954) developmental stages for children are much too conservative. Infants and young children have a tacit

mastery of various cognitive tasks long before they can give proper explicit accounts, and Piaget made the mistake of basing his stages on what children would say, what they could explain.

One of the major approaches to language and cognition to come forth in the past thirty years is the work of the Parallel Distributed Processing (PDP) Group (Rumelhart, McClelland, and the PDP Research Group 1986), so-called "neural nets" or "connectionist models." The connectionist models capitalize on this "levels of explanation" approach, with the proposal that fairly simple associationistic mechanisms can be modeled on a computer to create close approximations to behavior that appears to be rule-governed. Chandler (1995) summarizes their major achievement: "They have shown that rule-like regularities can emerge from the massed interaction of relatively simple processes operating on homogeneous networks of information even though those networks contain and refer to no explicit representations of those rules" (234–35). The strategy is an ingenious one, and it has won for D. O. Hebb's neurological behavioristic associationism (on which PDP is based) a new hearing within contemporary cognitive psychology.

Skousen's (1989) analogical modeling of language (AML) also accounts for seemingly rule-governed behavior without recourse to explicitly represented rules. The approach is based upon a very simple principle of "natural statistics": to *minimize the number of disagreements* (Skousen 1992). In the same way that the complexities of hypothesized internalized linguistic rules can be avoided with this approach, the complexities of statistical decision theories can also be avoided. That is, there is no need to posit that the learner acquires some kind of "probabilistic rule" for dealing with linguistic categorization. Rather, his performance can be accounted for by the simple proposition that he samples from his own

stored linguistic experiences using this one basic principle. Close approximations to actual performance can be achieved by adjusting the level of "imperfect memory." It is intriguing how such a simple hypothesized process can create complex behavior that could be explained at the highest level in terms of a complex and subtle rule system of the kind Chomsky has described.

Both the connectionist models (PDP) and AML are what Skousen (1995, 227) referred to as "procedural" as contrasted with rule approaches, which are "declarative." As procedural models, both AML and PDP avoid the major conceptual problems encountered in rule-based models. Skousen (1995) identified at least three such problems: rule-governed approaches cannot deal with "leakage" across category boundries; they are not robust in dealing with missing information or ill-formed context in the way that actual speakers are; and they are pushed to revert to a competence/performance distinction to account in an ad hoc way for failures of the model to deal with real, dynamic aspects of language.

AML has a number of features to recommend it over other available procedural language models. One is its explicit incorporation of episodic memory into the learning process. Another is its potential to account for more general perceptual processes beyond language. Both Skousen (1995) and Chandler (1995) have pointed to a number of failings of the connectionist models that AML seems to overcome. For one, connectionist models, once trained, are deterministic and cannot handle probability matching. Furthermore, connectionist network training can often require an inordinately long time even for simple behaviors, can get stuck in local minima, and even when trained cannot adjust to learn new input but rather collapses into predicting nonsense (the so-called "catastrophe problem"). AML, on the other hand, is particularly good at probability matching

in a way that corresponds to actual human language learners. Also, no training is necessary, there are no local minima, and it adjusts well to new input, even contradictory input.

There are particular problems yet to be solved in the application of AML. One of the biggest problems is computational. With commonly used computational methods, each variable that is added essentially doubles the processing time as well as the memory requirements of the computer. Also, the notable successes of AML have been in the more well-defined areas of phonetics/ phonology, orthography, and morphology. Application to more abstract and difficult areas of semantics and syntax has yet to be demonstrated. However, initial work with syntax looks promising. Lonsdale (2001), for example, has found some success in translating from French to English using *analogical cloning,* following the method of Jones (1996).

The *probability matching* aspect of analogical modeling is particularly interesting to psychologists in that it foreshadows the possibility of higher level theoretical integration with other established principles of human and animal behavior. A case in point is the well known *matching law* of Richard Herrnstein (1961) whereby probabilities of response are found to match probabilities of reinforcement. There are probably many linguistic examples of probability matching of this kind. Tucker and his colleagues (1968), as one example, have documented a linguistic probability matching in native French speakers with respect to the categorization of grammatical gender of "artificial" French words. They found a close match between the gender selection probabilities for various invented words and the gender probabilities for words with the same endings in *Petit Larousse.* Skousen (1995) recognized this capacity of AML to deal with the ubiquitous phenomenon of probability matching as one of the many

advantages of AML over neural networks. AML can be seen as a sophisticated extension of associationistic principles, one that makes them capable of accounting for seemingly rule-governed behavior.

Given the arguments for the superiority of the AML approach to the modeling of human linguistic behavior, it could be argued that this paper has come full circle back to the associationistic approach criticized by Lashley and Chomsky. However, this is not just a case of "rocks break scissors, scissors cut paper, and paper covers rocks." A better metaphor would be an upward spiral, where the associationism implied in analogical modeling represents a much higher level of sophistication than the simple left-right associationistic chain theory that still falls vulnerable to the Lashley/Chomsky critique. Nor does it mean that with the continued ascension of analogical modeling we would expect to witness the demise of rule-governed approaches. In the concluding paragraphs of his fundamental work on analogical modeling, Skousen discussed the place of rule approaches:

> Despite the many arguments, both empirical and conceptual, in favor of an analogical approach to the description of language, there is a place for rule approaches too. An optimal rule description serves as a kind of metalanguage that efficiently describes past behavior and allows us to talk about that behavior. Whenever we attempt to summarize behavior or to discover relationships in data, our viewpoint is structuralist. But if we wish to predict language behavior rather than just describe it, we must abandon rule approaches. Rule descriptions have great difficulty in explaining actual language usage. (1989, 139)

Skousen went on to compare language rules with Boyle's Law and

Charles's Law as general physical laws that are only approximations to the real behavior of gasses. They are fairly accurate in accounting for gas molecules acting in the aggregate under most conditions, yet they have no real existence except in the minds of scientists. He made this comparison with linguistic rules:

> In no literal sense can it be said that individual gas molecules follow these laws. In a similar way, linguistic rules are meta-descriptive devices that exist only in the minds of linguists. Speakers do not appear to use rules in perceiving and producing language. Moreover, linguistic rules can only explain language behavior for ideal situations. As in physics, an atomistic approach seems to be a more promising method for predicting language behavior. (1989, 140)

This is reminiscent of Polanyi's characterization of a skillful performance as being achieved by the observance of a set of rules which are not known as such to the person following them. Linguistic behavior can be described in a general way by rules, but an analogical modeling approach is probably much closer to the actual psychological processes involved and accounts better for actual linguistic behavior (performance). Skousen's illuminating comments on the place of rules and analog constitute a fitting conclusion to his first published book on analogical modeling. They are also, perhaps, a promising prelude to the construction of a serious account of the psychology of the sentence, that mysterious process by which our holistic thoughts are automatically converted into a string of words.

REFERENCES

Ames, Adelbert, Jr. 1946. Binocular vision as affected by relations between uniocular stimulus-patterns in commonplace environments. *American Journal of Psychology* 59: 333–57.

Blumenthal, Arthur L. 1970. *Language and psychology: Historical aspects of psycholinguistics.* New York: John Wiley and Sons.

Bower, T. G. R. 1977. *A primer of infant development.* San Francisco: W. H. Freeman.

Chandler, Steve. 1995. Non-declarative linguistics: Some neuropsychological perspectives. *Rivista di Linguistica* 7, no. 2: 233–47.

Chomsky, Noam. 1957. *Syntactic structures.* The Hague: Mouton.

Gibson, James J. 1966. *The senses considered as perceptual systems.* Boston: Houghton Mifflin.

Herrnstein, Richard J. 1961. Relative and absolute strength of response as a function of frequency of reinforcement. *Journal of the Experimental Analysis of Behavior* 4: 267–72.

Jones, Daniel. 1996. *Analogical natural language processing.* London: UCL Press.

Kilpatrick, F. P. 1961. *Explorations in transactional psychology.* New York: New York University Press.

Lashley, Karl S. 1951. The problem of serial order in behavior. In *Cerebral mechanisms in behavior,* ed. L. A. Jeffress. New York: John Wiley and Sons. 112–36.

Lonsdale, Deryle W. 2001. Recent advances in analogical cloning. Presentation at the Deseret Language and Linguistics Society 2001 Symposium. Brigham Young University.

Piaget, Jean. 1954. *Origins of intelligence.* New York: Basic Books (original French edition, 1936).

Polanyi, Michael. 1962. *Personal knowledge: Towards a post-critical philosophy.* New York: Harper and Row.

Rumelhart, David E., James L. McClelland, and the PDP Research Group. 1986. *Foundations.* Vol. 1 of *Parallel distributed processing: Explorations in the microstructure of cognition.* Cambridge, MA: MIT Press.

Skousen, Royal. 1989. *Analogical modeling of language.* Dordrecht: Kluwer Academic Publishers.

Skousen, Royal. 1992. *Analogy and structure.* Dordrecht: Kluwer Academic Publishers.

Skousen, Royal. 1995. Analogy: A non-rule alternative to neural networks. *Rivista di Linguistica* 7, no. 2: 213–31.

Tucker, G. Richard, Wallace E. Lambert, Andre Rigault, and Norman Segalowitz. 1968. A psychological investigation of French speakers' skill with grammatical gender. *Journal of Verbal Learning and Verbal Behavior* 7: 312–16.

Wundt, Wilhelm M. 1912. *Die sprache.* Book 2, vol. 1 of the *Volkerpsychologie* series. Leipzig: Engelmann.