

FOUNDATIONS OF LANGUAGE



EMPIRICISM WITHOUT FACTS

Charles F. Hockett, *The State of the Art* (Janua Linguarum Nr. 73), Mouton, The Hague, 1968, 123 pp.

The State of the Art is a very strange book. Its title might lead one to believe that it is a contemporary document, detailing work on the frontiers of linguistics and drawing upon the vast amount of research that has been done in the past decade. Unfortunately, the title is misleading. It is very much a voice out of the past, doting on dead issues and nursing old misunderstandings. It is what Hockett considers to be a reply to Chomsky, and it is disappointing.*

One would have hoped that Hockett's reply when it came would be substantive; that it would have contained a mountain of examples taken from his own research and from his wide experience, showing in detail just why transformational grammar could not handle these examples and how the author proposed to handle them himself. Most of all, one would have hoped that it would come to grips with a reasonable proportion of the empirical research done in the past decade that seems to support transformational grammar. Why couldn't Hockett have taken representative examples and shown, case by case, how he could have handled them as well or better? If he were seriously interested in empirical linguistics, this is what he would have done. Unfortunately, this book lacks any substantive discussion of actual linguistic data (aside from some anecdotal examples in chapter 6, pp. 100-107). The result is a confused and inconclusive discussion of theory divorced from empirical content.

Perhaps it is unfair to take Hockett to task for not writing a book that lives up to the title. Perhaps a reviewer should simply point out that the title is much too pretentious for the book, and then proceed to review the book on its own terms. The trouble is that this is difficult. Bare as the book is of actual language data, there isn't all that much to review. The central weakness of the book is exactly this lack of empirical support for his claims.

In what is billed as a reply to Chomsky, Hockett makes every effort not to reply to Chomsky. Hockett implicitly distinguishes between Chomsky's

* This work was supported in part by grant NSF-1934 from the National Science Foundation to Harvard University.

theory of the structure of language and the methodological assumptions on which this theory is based. Hockett does not attack Chomsky's theory of language structure, which can be attacked only on empirical grounds, by calling on actual linguistic data. Rather Hockett attacks Chomsky's methodological assumptions and his views on the nature of language use and acquisition. Hockett refers to these views as 'Chomsky's theories'. By Chomsky's theory of language structure, I mean his claims that a grammar of a language has syntactic, phonological, and semantic components that relate meanings to phonetic outputs; that the syntactic component has a 'base' defined by context-free phrase structure rules, and has a set of cyclically-ordered transformations of a specified form mapping phrase-markers into phrase-markers; that phonological rules operate in terms of a universally specified set of distinctive features and that they are of a certain specified form; and so on. Further elaborations of the theory and an account of the empirical evidence supporting it can be found in the literature of the field.

Hockett isn't interested in discussing any of Chomsky's substantive claims about the structure of language. As he says in his preface,

Two topics that might be expected to loom large in a discussion of this sort in fact will not. Little will be said about syntactic transformations, and nothing about Chomsky-Hallean 'phonology'. The former are helping us to discover subtle facts about various languages, and are surely here to stay. The latter is, in my opinion, completely bankrupt, but is under adequate debate elsewhere (p. 3).

Having ruled syntax and phonology out of the discussion, Hockett has left no role at all for empirical linguistic data in his presentation. As might be expected, the discussion degenerates into speculations, anecdotes, and personal feelings.

What is particularly ironic about this is that Hockett is attempting to portray himself in this book, as he has elsewhere, as the defender of empiricism in linguistics and to put an onus on Chomsky as a purveyor of "metaphysics" with an "antiscientific bias" (p. 79). Oddly enough, Hockett feels free to indulge in such name-calling without having to bring in any empirical facts to back up his observations. By any objective criterion, it is Hockett who is cast as the metaphysician here.

Hockett is trying to discredit Chomsky's theory of the structure of language by asserting (not showing) that his methodological assumptions are wrong – especially the assumptions that rules of grammar exist (that language is 'well-defined') and that there is a distinction between performance and competence. He claims that if that distinction falls, then the whole edifice of generative grammar will collapse. Hockett seems to realize that there is

an empirical issue here, though he does not quite see what the real issue is.

Empirically, we might be led by our observations of speech to propose that the underlying system, the set of habits we call the language, is well defined. But Chomsky was not led to this conclusion empirically, nor, so far as I know, has anyone else proposed this as an empirical hypothesis. Rather, his opinion that a language is well-defined seems to come first in a logical if not a chronological sense. This assumption is hardly supported by observation, which, in addition to the discernible regularities of speech in any one community, reveals oddities and vagaries of all sorts. However, the assumption can be retained in the face of the evidence if one posits an obscure sort of 'underlying' system that *by definition* meets the requirements of the assumption, and then explains (or explains away) the vagaries of actual speech as due to the participation of other factors. But this step moves the underlying system completely out of the reach of the methods of empirical science. The notion thereby ceases to be a hypothesis, and becomes merely idle philosophical speculation (p. 66).

Hockett does not see that the performance-competence distinction is an empirical hypothesis. Hockett seems to think Chomsky would have the public believe in the distinction simply because he says it exists, and that he would have the public believe in transformations, base rules, cyclical rule ordering, variables, distinctive features, markedness, etc. simply because he says they exist. Nothing could be further from the truth. Chomsky is quite clear as to the empirical nature of all these issues. He makes the performance-competence distinction as a methodological assumption. Given that assumption, he defines the notion 'transformational grammar' and applies it to actual linguistic data. If transformational rules are discovered which turn out to explain deep facts about language, then there is every reason to believe whatever methodological assumptions are necessary in order to permit one to state such rules. The empirical question is whether those rules account for the facts that they are supposed to account for.

In short, the question as to whether Chomsky's methodological assumptions about the performance-competence distinction, language use, etc., are right or wrong is an empirical question in that it directly depends on the empirical question as to whether Chomsky's theory of the structure of language is right or wrong. To answer this question, we need to discuss in detail just how well that theory accounts for real linguistic data in syntax and phonology. But this is just what Hockett explicitly refuses to discuss. This refusal is particularly surprising in view of Hockett's charge that Chomsky's notion of a transformational grammar "is hardly supported by observation" (p. 66). From his very earliest work, Chomsky has taken pains to discuss the empirical evidence that supports his theory of language structure (see chapters 5 and 7 of *Syntactic Structures*). Since then, papers and books numbering in the hundreds have provided a vast amount of

empirical evidence for the existence of transformational rules, and this evidence in turn supports whatever methodological assumptions are necessary to allow one to formulate such rules precisely. The only way that Hockett could refute Chomsky's methodological assumptions is to refute, point by point, the empirical evidence supporting the notion 'transformational rule' in *all* of these works. This would be a vast undertaking, just as the accumulation of such evidence over the past decade has been a vast undertaking. In a book of this size, Hockett could at least have considered the evidence Chomsky gave in *Syntactic Structures*, which was an even smaller book. Instead, this self-styled defender of empiricism steadfastly refuses to consider the only empirical evidence that really bears on the question he is discussing. Like the metaphysician he accuses Chomsky of being, he indulges in definition-making and expects his audience to believe that this is a scientific activity.

A large part of Hockett's argument rests on a tenuous analogy between football scores and sentences of a natural language. In chapter 3 Hockett defines the concepts 'well-defined' and 'ill-defined'.

A *well-defined system* is any system (physical, conceptual, mathematical) that can be completely and exactly characterized by deterministic functions (p. 45).

All other systems are 'ill-defined'. Football scores, Hockett claims, fall into the latter category. The reason is that football has a time limit. Hockett reasons thus:

Even against no opposition, scoring requires time, and play is confined, at most, to slightly more than 60 minutes of time in. A score of 1,000,000 is obviously impossible. The highest score on record is 227. Could speed and skill be increased (and strength of opposition decreased) to squeeze this up to 228? Possibly. To 229? Perhaps. The fact that we can easily name an integer greater than any member of S_r does not mean that there is a precise maximum element in S_r . The set is neither computable nor noncomputable: it is *ill-defined* (p. 47).

Hockett contrasts football with baseball. Since baseball has no time limit, there is no bound on how high the score in a baseball game could go. Therefore, Hockett says, the set of baseball games is 'well-defined', since any non-negative integer could be a baseball score, even 1,000,000. It should be noted however that if, say, a ten-hour time limit were placed on baseball games, the set of possible baseball scores would then be 'ill-defined', just like the set of possible football scores. Although the structure of the game would remain the same, the imposition of a time limit would change a 'well-defined' set to an 'ill-defined' set. But in the case of baseball a time limit does not change the structure of the game. Thus we see that the set of scores defined by the structure of the game may be 'well-defined', while that defined by

game structure plus time limit can be 'ill-defined'. Hockett, however, does not note this fact, and this failure, as we shall see, leads him to make a false analogy between football scores and sentences.

In chapter 5, Hockett makes the great leap. He claims that a natural language, *in the sense that should concern linguists*, is like the set of football scores (or time-limited baseball scores) rather than like the set of (no-time-limit) baseball scores. Since people have finite life spans and limited endurance, they could not utter sentences of, say, a billion words in length. Yet there is no such thing as the longest sentence of English. Therefore, the set of physically possible English sentences is like the set of possible football scores or time-limited baseball scores: it is 'ill-defined'. Now Hockett makes the implicit, but never articulated, methodological assumption that the set of physically possible English sentences is what linguists should study. This is, of course, a matter of opinion. It does not follow from any mathematical or scientific principles. Hockett recognizes that this is a matter of opinion.

It is very easy to write down the beginning of a series of sentences each of which seems to be perfectly good English:

One.
 One and one.
 One and one and one.
 One and one and one and one.
 ...

Chomsky's views require us to believe that, if the first few of such a series (not necessarily the particular series displayed) are English, then so is the thousandth, or the billionth, or the n th for any positive integer n . This is empirically absurd. I claim [sic] that the millionth – or even the thousandth – term of this series is in fact not English, just as a million is not a possible football score. This does not mean that we can specify exactly which term of the series is the longest that is good English, any more than we can specify the longest possible football score or write down the formula for the longest possible methane-series hydrocarbon molecule. As one attempts a longer and longer sentence of the kind shown, or of the kind defined by any open-ended pattern, one encounters certain *flexible constraints*, that are, in my opinion [sic], *part of the language*, just as the time limits of a football game are part of football. Moreover it seems [sic] that *all* constraints in a language are of this more or less rubbery sort, yielding no definite boundary to the 'set of all possible sentences' of the language; and for just these reasons languages [sic] are ill-defined (pp. 60–61).

Hockett admits that it is only his opinion that physical limitations like time and endurance are 'part of the language', as linguists should study it. He builds this opinion into his definition of the term 'language'.

The linguist seeks theories which are generalizations from observations, and are about *speech*. These yield predictions and are corrected by subsequent observations.

The linguist is led to posit that observable regularities of actual speech are a matter of habits, resident in the users of language – rather than, say, a matter of automatic chemical response to impinging sunlight. He calls those habits ‘language’. This proposal is part of our theorizing about *speech*. It makes no sense to pretend that there can be a separate and distinct theory of *language* (pp. 65–66).

Here Hockett defines ‘language’ – what a ‘linguist’ studies – to suit his own opinion, that there is no distinction between *speech* and *language*, *parole* and *langue*, *performance* and *competence*. Hockett’s feelings do not necessarily reflect the real world, and there is no reason whatever to believe his *ex cathedra* statement that there can be no “separate and distinct theory of language”. This, as was mentioned above, is an empirical question. Since linguists presumably study the structure of language, Hockett seems to be building into his definition of ‘language’ the *claim* that physical limitations like time, life spans, endurance, etc. are part of the *structure* of language. Again, this is an empirical claim; it cannot be decided *a priori*, which is just what Hockett is attempting to do.

So far as I can pin down Hockett’s reasoning, there are just a few central arguments in chapter 5. I have tried to make them explicit.

ARGUMENT I:

- (1) The set of physically possible sentences of a natural language is ill-defined (BY DEFINITION OF ‘ILL-DEFINED’).
- (2) The notion ‘grammar’ makes sense only for well-defined sets (BY DEFINITION OF ‘GRAMMAR’ AND ‘WELL-DEFINED’).
- (3) The set of physically possible sentences of a natural language cannot be scientifically studied in terms of a related well-defined set of sentences (ASSUMPTION BASED ON HOCKETT’S OPINION).
- (4) Therefore, the set of physically possible sentences of a natural language cannot be scientifically studied in terms of the notion ‘grammar’.

Step 3 is the crucial step in this argument. If step 3 is based solely on Hockett’s opinion, then there is no reason to accept the argument. Hockett seems to sense this, and his appeal to the analogy between football scores and sentences of English seems to be an attempt to justify Step 3. Let us therefore consider what seems to be the logic of his analogy.

ARGUMENT II:

- (1) The time limits of football are essential to the structure of the game (UNDISPUTED ASSUMPTION).
- (2) The set of football scores is ‘ill-defined’ because of the game’s time limit (DEFINITION OF ‘ILL-DEFINED’).

- (3) If any system is 'ill-defined' because of time limits, then time limits are essential to the structure of the system (UNWARRANTED ASSUMPTION).
- (4) 'Language', or the set of physically possible sentences, is 'ill-defined' because of time limits (DEFINITION OF 'ILL-DEFINED').
- (5) Therefore, time limits are essential to the structure of 'language'.

Step 3 in this argument is not only unwarranted, but false. As was pointed out in the case of baseball, the imposition of a ten-hour time limit would turn the 'well-defined' set of baseball scores into an 'ill-defined' set, without changing the structure of baseball. Thus the analogy between football and natural language fails because of a false assumption.

Hockett's last basic argument is based entirely on definitions.

ARGUMENT III:

- (1) 'Language' is the set of all physically possible sentences (DEFINITION).
- (2) The set of all physically possible sentences is ill-defined (DEFINITION OF 'ILL-DEFINED').
- (3) 'Language' is 'ill-defined' (FROM 1 AND 2).
- (4) The notion 'grammar' only makes sense for 'well-defined' sets (DEFINITIONS OF 'GRAMMAR' AND 'WELL-DEFINED').
- (5) Therefore, the notion 'grammar' does not make sense for 'language'.

This argument is valid, but completely empty. Since neither Chomsky nor anyone else ever intended that the notion 'grammar' should be used in talking about the set of *physically* possible sentences, the argument is pointless. Since Hockett's definition of language is not the same as Chomsky's, the argument proves nothing. Hockett, however, is not consistent in his use of quotation marks in the discussion of his concept of 'language'. In general, he does not use quotation marks and speaks of 'language' as though his definition were the only possible one.

If languages [sic] are ill-defined, then the theory of computability and unsolvability ceases to have any relevance for linguistics. This knocks the props out from under current mathematical linguistics, at least in the form of algebraic grammar, whose basic assumption is that a language can be viewed as a well-defined subset of the set of all finite strings over a well-defined finite alphabet (p. 61).

Since the proper definition of 'language', that which is amenable to scientific study by linguists, is just what is at issue here, all of Hockett's argumentation turns out to prove nothing.

In chapter 6, Hockett turns his attention to the question of how people can say and understand new sentences that they have never heard before.

His answer is that old saw, analogy. "An individual's language, at a given moment, is a set of habits – that is, of analogies" (p. 93). The trouble is that this is a relatively empty idea. Hockett leaves the notion of analogy undefined. What is a possible analogy? Why do some analogies work, but not others? On p. 97 Hockett gives three examples of analogies, which are pretty much of the traditional sort.

(1) John shot the tiger : The tiger was shot by John :: The butcher weighed the meat : The meat was weighed by the butcher ::

This example is supposed to be self-evident, but it is anything but that. Clearly the analogy is not defined in terms of strings of words, but rather in terms of higher-level constituents. 'John' and 'the butcher' are considered the same by the analogy. Presumably the analogy, when stated formally, would mention a general constituent like noun phrase. But then why is it that some analogies fail?

The lightning bolt struck the boy who was rich: *X*

Solving for *X* with respect to the analogy of (1), we could get either 'The boy was struck by the lightning bolt who was rich' or 'The boy who was rich was struck by the lightning bolt'. Clearly the first analogy fails and the second succeeds. Why? Presumably the passive analogy must apply to the full noun phrase 'the boy who was rich', not to the constituent noun phrase 'the boy'. But this means that analogies must be sensitive to phrase structure, not just to strings of phrases, and must here choose the 'higher' noun phrase (as in Chomsky's A-over-A principle).

Presumably Hockett would have a question analogy, just as he has a passive analogy. It might work something like this.

John has shot the tiger : Has John shot the tiger? ::
 John was running : Was John running? ::

But what about passive questions? What about 'Was the tiger shot by John?' Presumably, one would get this sentence from 'John shot the tiger' by applying the passive analogy.

John shot the tiger : The tiger was shot by John

Then the question analogy would apply:

The tiger was shot by John : Was the tiger shot by John?

But if Hockett were to attempt to account for passive questions in this way, he would need *ordered analogies*! The passive analogy would have to apply before the question analogy.

To get this far, a theory of analogy would have to apply to phrase structure trees and would have to have ordered analogies. One could easily go on in this way, showing that a theory of analogy would have to have variables, exceptions, cyclical application of analogies, – in fact, all the theoretical devices that Chomsky sets up in *Syntactic Structures* and that have been shown since to be necessary to define the notion 'transformation'. Hockett is almost right when he says "Transformations are analogies" (p. 97); he should have said 'analogies are transformations'. So far as I can see, any attempt to define the notion 'analogy' precisely and to test it on real linguistic data (like that discussed in *Syntactic Structures* and elsewhere) would lead to a definition of analogy that is indistinguishable from that of transformation. If Hockett does not think that this is true, let him try to define analogy precisely and apply his definition to the range of data discussed in the literature on transformational grammar. Until he does this (at least for the tiny number of early works listed in his bibliography), he cannot be taken seriously.

One more point: 'analogy', according to the examples Hockett gives, would be recursive. He gives the following example (p. 97):

(2) John shot the tiger : I watched John shoot the tiger ::
 The butcher weighed the meat : I watched the butcher weigh the meat
 ...

Presumably the analogy would apply to its own output:

John watched the butcher weigh the meat : I watched
 John watch the butcher weigh the meat ::
 Max watched John watch the butcher weigh the meat :
 I watched Max watch John watch the butcher weigh the meat ::
 ...

This analogy, if iterated, would produce sentences arbitrarily long – even a billion words long. Presumably, the analogy can only produce sentences of English. But Hockett claims that sentences a billion words long cannot be sentences of English. If analogies are recursive, at what point can Hockett shut them off? How can he constrain them so that they produce just those sentences which are English and no more? If English is 'ill-defined' in Hockett's sense, he obviously cannot do it. Thus Hockett's theory of analogy seems to be subject to exactly the 'faults' that Hockett finds in transfor-

REVIEW ARTICLE

mational grammar. To the extent that Hockett makes himself explicit, he contradicts himself.

On the whole this is an empty book, of no empirical linguistic significance at all. Hockett argues not from linguistic data, but from definitions and opinions, and refuses to challenge the details of linguistic analyses that seem to support transformational grammar. This is especially damning since Hockett attacks as 'antiscientific' transformational grammarians who have based their conclusions on a great mass of empirical evidence. Given Hockett's aprioristic arguments, this book makes an odd empiricist manifesto.

Harvard University

GEORGE LAKOFF