Barbara Stanosz
CODE OF CONDUCT FOR NATURAL LANGUAGE

The aim of this paper is to draw attention to and characterize certain methodological changes that are taking place in modern linguistics. More specifically, I intend to survey those aspects of methodological evolution (or perhaps revolution?) in linguistics that are responsible for the progress it has made in defining itself vis-á-vis the formal-logical theory of language. Identifying the existing similarities and differences between the two disciplines with regard to their subject matters, methods and research objectives will perhaps contribute to removing at least some of the misunderstandings that arise in the 'professional' exchanges between linguists and logicians.

The present treatment of the methodological situation in linguistics is deliberately sketchy and fragmentary: it discusses only some of the current trends and schools of thought, emphasizing only the most fundamental assumptions. There are several reasons why a more detailed description would not be feasible. Firstly, the heterogeneous nature of the problems classified as falling in the scope of linguistics makes some approaches methodologically incomparable to one another. Secondly, some problems, concepts and research techniques associated with the different approaches are undergoing radical changes. And, thirdly, some concepts, positions and proposals that can be found in the literature are far from clear. The conclusions of this paper are tentative, as they rely on one of the many possible interpretations of the published material in linguistics.

Empirical Basis and Subject of Study

All the questions posed within the science of linguistics arise from particular human behaviors that we label as 'linguistic;' these behaviors also serve as the ultimate criterion of the cognitive value of linguistic claims. The knowledge offered by linguistics about the linguistic behaviors of a language community is framed in terms of a description of the LANGUAGE spoken by that community.

The empirical basis for the study of language consists of physical products of linguistic behaviors: namely, strings of noises. Linguistic description presupposes a particular way of segmenting those strings of noises and relies on some established principles for identifying certain segments as tokens of the same EXPRESSION. The description does not simply provide a characterization of empirically discovered expressions; it goes beyond the available empirical data, encompassing the set of expressions that MAY BE the products of linguistic behavior in a given social group.

The history of science reveals the following, arguably with exceptionless regularity: the researchers' professional self-knowledge lags behind their scientific achievements. This is especially true of the definitions that are offered of the subject matter of a given scientific discipline. They often remain vague, metaphorical, or, indeed, completely illusory long after the science in question has made significant theoretical progress. The concept of language, as a subject of linguistic study, suffered a similar fate: for a long time, its ontological and methodological status remained undetermined. It seems, however, that recent years have seen a breakthrough in this regard.

The inspiration came from de Saussure, and more specifically from his famous distinction between *langue* and *parole* (de Saussure 1916) — language 'in the proper sense' and speech. *Langue* is an abstraction, a system of relations (rules), form rather than substance — whereas *parole* is a physical realization of *langue*, which is both determined by *langue* and serves as evidence of its existence.

The distinction has become the subject of many interpretations and critical analyses (e.g. Zawadowski 1958). It is difficult to assess its contribution to the advances of 20th-century linguistics. It is clear, however, that research practice in linguistics has born out de Saussure's claim that the abstract *langue* constitutes the proper subject of linguistics. It also seems that recent construals of this distinction are bringing us closer to its complete and adequate explication.

What I have in mind are the concepts of competence and performance, employed by Chomsky and his followers (Chomsky 1965, ch. 1, §§1, 2.). Competence, defined loosely as 'the linguistic knowledge possessed by an

ideal speaker-hearer,' receives an exact characterization in the scientific work of Chomsky's school. Performance, or the products of the linguistic behaviors of actual speakers, is the psychologically constrained realization of this knowledge. The first and foremost task for linguistics is to construct a theory of competence, which is a fairly radical idealization of performance. Only with such a theory in place can one begin a scientific study of performance, which, together with psychology, would yield theories of actual human linguistic behavior.

An important component of the latest views on the nature of the subject matter of linguistics is the belief that the available empirical data and the inductive procedures known from the philosophy of science do not provide a sufficient basis for reconstructing any natural language. Acquisition of competence with respect to a particular language is only explicable on the assumption that the speaker possesses some a priori knowledge of possible languages, a knowledge which comes as part of a cognitive endowment characteristic of the human mind (Chomsky 1965, ch. I, §8; Chomsky 1969; Katz, Postal 1964, §5.5). Correspondingly, it is impossible to obtain an adequate description of any language whatsoever by way of a pure observation of facts, a classification of collected data or inductive generalization.

Thus, modern linguistics has parted with many past conceptions of its subject of study and scope of research. It first abandoned the vague psychologistic slogan of analyzing language as 'an instrument for communicating thoughts.' Then it rejected the program of conducting narrowly empirical research into 'actualized' language, or speech, with a view to discovering its governing patterns and customs. It now focuses on a product of abstraction: the body of ideal linguistic knowledge which is supposed to form a theoretical basis for language as an empirical phenomenon. It proposes to replace a faithful description of facts with theory. It presents itself as an explanatory discipline.

This being so, some traditional arguments for the claim that the subject matter of linguistics is essentially distinct from that of the formal-logical theory of language lose their bite. More specifically, the claim that the subject matter of linguistics is given in experience (as physical products of linguistic behavior), while the language studied by logicians is given theoretically (as a set of rules), was justified only on the assumption that *parole* constitutes the sole subject of linguistic study. After all, de Saussure's *langue* and Chomsky's competence are not physical objects but rather sets of rules! The argument purporting that natural languages studied by the linguist are subject to certain constructive restrictions, of which artificial

languages studied by logic are free, collapses for much the same reason: the restrictions in question, which hold for actual linguistic behaviors, derive solely from the physico-psychological causes of those behaviors. There is no evidence that the notion of ideal linguistic knowledge should come into conflict with the principle of unrestricted construction; on the contrary, the principle seems to be a necessary component of an adequate description of natural language (Bach 1964: 12-13). Finally, given that the fact that no natural language is strictly determined by its empirical realizations entails that observational and inductive procedures can be employed only at the first stage of language reconstruction, one can venture that the kind of cognitive faculty operative at subsequent stages of that process is essentially no different from the faculty engaged in creating the 'arbitrary' languages constructed and studied in logic — regardless of whether we shall regard it as a sort of innate 'intuition' or a product of empirical interactions with an extra-linguistic reality.

The Goals of the Theory

The psychological and neuro-physiological nature of the linguistic competence of a grown person remains outside the purview of linguistics. Nor is linguistics interested in the biological endowment by virtue of which man can achieve this competence. Instead, linguistics seeks to answer the following two questions:

(1) What does a competent language user 'know' about his or her language?

(2) What must one 'know' about languages in general in order to be able to learn any of them?

It goes without saying that neither of these questions is about the kind of meta-linguistic knowledge that could be verbalized by a competent language user; a sufficient criterion of knowledge possession, in the intended sense of the term, is proficiency of language use and the capacity to learn any language in the usual way, respectively.

An answer to the first question should explain several kinds of facts which reflect the linguistic competence of natural language users. Chief among those facts is the ability to produce and interpret an indefinite number of sentences, most of which are novel ones (such that the language user has never heard them before). A second important fact is the ability to detect ambiguous sentences and to establish their possible interpretations, a third — the ability to identify, and sometimes interpret, deviant sentences.

An answer to the second question must explain how linguistic competence is acquired on the basis of the limited linguistic data available in the course of learning a language. One consequence of the creative character of linguistic competence, supported by the facts described above, is that learning a language cannot be a simple matter of storing information and generalizing upon it. The extent to which the human mind processes the scarce linguistic information and selects from a number of admissible options a single correct extrapolation of the data indicating the existence of some 'task constraints,' a knowledge about possible languages.

A major goal of linguistics is to provide, for every natural language, a description of the linguistic competence of its users formulated in terms of a set of rules. Let us call such a set of rules the GRAMMAR of a language. Providing adequate grammars for a sufficient number of natural languages would enable us to take up another task of theoretical linguistics: the construction of a UNIVERSAL GRAMMAR, or rather a GENERAL THEORY OF GRAMMARS, which would represent our knowledge about possible languages and thereby explain the mysterious properties of human 'linguistic intuition.'

Linguistics has only recently begun to make conscious and potentially fruitful attempts at tackling the latter task.[1] Although one can treat the theoretical results of traditional grammar as a partial explication of the linguistic competence of the speakers of particular languages, given the notorious incompleteness of this explication, it is better to regard it as a by-product of an altogether different endeavor. As a matter of fact, even the most thorough grammars contain many gaps; indeed, they are sets of rules that serve merely as EXAMPLES and GUIDELINES for the reader as to how sentences of a given language should be constructed or understood. Given this heuristic nature of traditional grammars, one should classify them as practical rather than theoretical achievements — significant in the context of education rather than science.

A grammar that explicitly characterizes the relevant properties of a natural language without appeal to the intelligence of the reader is known as a GENERATIVE GRAMMAR. A generative grammar of a given language is a set of rules that pair each phonetically possible utterance with a STRUCTURAL DESCRIPTION, which determines all the constituents of the utterance together with all inter-constituent relations. Grammatical

---

[1]The publication of N. Chomsky's *Syntactic Structures* (1957) is usually regarded as the turning point, although many of its ideas have since lost their topicality or been modified.

sentences of the language correspond to a special set of such structural descriptions. The set of utterances defined by such descriptions is called the LANGUAGE GENERATED BY A GIVEN GRAMMAR.

Generative grammars grew out of the theoretical assumptions, methods and research techniques of the structuralist schools in linguistics. However, the generative grammars constructed within the conceptual framework of classical structuralism did not provide an explanation for the basic linguistic facts of sentence production and sentence comprehension. The explanatory power of those grammars was negligible because, loosely speaking, the structural descriptions of utterances that they offered in terms of the so-called phrase markers were not subtle enough. As a scheme for adequately representing theories of linguistic competence, grammars of this type (called PHRASE STRUCTURE GRAMMARS or CONSTITUENCY GRAMMARS) are contrasted with the so-called TRANSFORMATIONAL generative grammars.[2]

For over a decade now, the methodological assumptions of transformational grammars have been a topic of great interest for mathematicians, logicians, psychologists and philosophers of language as well as linguists. In the course of many discussions, the theory has undergone significant changes,[3] and nothing suggests that it has already taken its final shape. It seems, however, that one can regard its basic research objectives as fixed. We shall discuss them in a little more detail, focusing on the already mentioned task of explaining the interpretability of sentences.

In providing the sentences of a language with their structural descriptions, a transformational grammar should 'mimic' the language user in how he or she comprehends an utterance with respect to those of its aspects that do not presuppose reference to an extra-linguistic reality. Thus, for example, it should distinguish the grammatical from the deviant, as well as the unambiguous from the ambiguous, assigning the correct number of interpretations to the latter. It should identify synonymous sentences, capture differences in meaning between any pair of sentences and characterize those differences in terms of inter-sentence meaning relations (entailment, incompatibility, independence, etc.).

---

[2]For the relation between transformational grammar and phrase structure grammar see Szaumian 1966; Chomsky 1964.

[3]From our present perspective, the publication of Katz and Fodor's *The Structure of Semantic Theory* (1963) was the turning point in the development of this theory, supplementing Chomsky's phonological-syntactic conception with the semantic component.

All the proposals to date as to the internal structure of grammars satisfying these conditions divide the task of their fulfillment between the syntactic and the semantic component; also, all of them respect the principle that the meaning of a sentence is a function of the meaning of its constituents. From the general point of view we are assuming here, differences between those proposals appear irrelevant.[4] It is an important fact, however, which has many consequences for the relation between linguistics and the logical theory of language, that they abstract from those properties the acts of comprehension that associate linguistic expressions with extra-linguistic reality. The apparatus of transformational grammar does not distinguish between sentence meanings whose interpretation depends on situational context (though it takes into account linguistic context);[5] it does not even distinguish such sentences from other sentences. Above all, it does not capture those aspects of the phenomenon of synthetic (situation-independent) sentence comprehension that enable the language user to determine which sentences are true: it does not account for successful acts of naming objects in the world or for any other referential procedures. It does, however, specify the set of analytic truths for a language, as by assumption it accommodates all intra-linguistic meaning relations.

The idea of describing language in terms of transformational grammar clearly coincides at this point with the formal-logical notion of an uninterpreted language. The latter defines language L in terms of the set of its expressions, E, (with a special subset S of sentences) and the set of sentences, A, where $A \neq S$, that are true in all 'possible worlds' (models) of language L, i.e. in all the domains of reality that can be described using L. Thus, A is the set of analytic truths of L. One can represent language L as an ordered pair of the following form:

(i) $<E, A>$.

An interpretation transforms language L into an ordered triple:

(ii) $<E, A, M>$,

where M is 'the real world' (selected model) of language L, i.e. the domain of reality actually talked about in L. Model M fixes the denotation of the expressions of L and thereby specifies the set of synthetic truths for L.

---

[4]Cf. Katz and Postal's proposal (1964) and Weinrich's counterproposal (1966).

[5]Arguments for this construal were presented by Katz and Fodor (1964: 486-491).

Logicians investigate systems of both kinds, making certain standard assumptions about their structure and relations between their elements. Systems of type (i) are investigated within logical syntax; systems of type (ii) are investigated within logical semantics.

It follows that, on the meaning of the term 'semantics' adopted in logic, the theory of language constructed in terms of transformational grammar is programmatically asemantic. By refusing to investigate any relations between linguistic expressions and the things they are used to talk about, it assumes a purely syntactic perspective. Yet, in linguistics, the terms 'semantics' and 'syntax' are used differently than in logic. Moreover, linguists disagree over their precise meaning: the question of how to draw the distinction between syntax and semantics is far from resolved. It seems that one possible solution to this problem would be to classify investigations into set E as belonging to syntax and investigations into set A as belonging to semantics (qua elements of systems of type (i)). To incorporate the 'semantic component' mentioned above into a transformational grammar would be, simply, to decide to describe a language in terms of both set E and set A.[6]

The following question now suggests itself: To what extent can a theory of language based on these principles account for the basic facts of linguistic competence? In particular, what *are* the properties of the act of sentence comprehension (interpretation) which the theory allegedly sets out to explain? If the remarks made above are correct, one can assume that said properties coincide with those stipulated by the following definition of sentence comprehension, which is sometimes investigated in logic:

x comprehends sentence s of language $L \equiv_{df} x$ can identify every sentence entailed analytically by s in L.

The notion of sentence meaning associated with this definition is as follows:

The meaning of sentence s in language L is the set of all analytic entailments of s in L.

It is clear that these definitions are fairly good approximations of our intuitions. It is equally clear, however, that there are certain intuitions that they fail to capture. Indeed, a person may comprehend (in the sense defined above) the sentence 'Man has two eyes' (by identifying its entailments such

---

[6]One could modify slightly the terminology adopted above by calling systems of type (i) partially (verbally) interpreted and reserving the term 'uninterpreted language' for set E.

as 'Man has a pair of eyes,' 'Man has two visual sense organs,' 'The number of man's eyes is smaller than 3,' etc.) and yet have no idea whether it is true! One can sincerely assert 'It's not raining,' and know perfectly well what the meaning of this sentence is in English, while getting soaked to the bone in a thunderstorm!

If we accept the above definitions as good 'first approximations' of the intuitive notions of meaning and sentence comprehension, these somewhat paradoxical consequences force us to recognize that a gap exists between the explicatum and the explicandum. In order to fill this gap, we must take into account the referential function of linguistic expressions, which the linguistic theory under consideration programmatically ignores. This sometimes gives rise to the objection that linguistic theory does not describe the principal function of language, namely the role it plays in cognition.

The linguist can answer this criticism by pointing out that not only has he not set himself the task of describing the cognitive function of language, but also that such a description is beyond his expertise, as its fulfillment requires certain epistemological and ontological commitments which he is not qualified to make. The referential version of the theory of language ought to be developed within disciplines that have at their disposal a suitable conceptual apparatus and a repertoire of appropriate research methods. In particular, it is the formal theory of knowledge, i.e. logical semantics, that seems up for the task of conducting research into the properties of languages (including natural ones) construed as systems of type (ii).

The Structure of the Theory

The task of answering the second principal question of linguistics, formulated in the previous section, lies with the general theory of grammars. The goals of the theory include providing a method for selecting one from among the many grammars that are compatible with the appropriate corpus of empirical linguistic data (and structural descriptions assigned to them). Above all, however, the general theory of grammars ought to identify the class of POSSIBLE GRAMMARS of natural languages.

Despite continuing discussions over the issue of the internal structure of these grammars, the main methodological outline is quite uncontroversial. First of all, given their generative character, it is necessary to characterize the grammars in purely formal terms; in particular, grammatical rules must be formulated in a way that guarantees the possibility of their 'automatic' application, which is to say they are modeled on formalized deductive systems. (One consequence of this constraint is that the traditional semantic definitions

of categories such as 'noun,' 'verb,' 'adjective,' etc. have been abandoned in favor of enumerative specification). Furthermore, given the intended scope of grammatical description of natural languages, their grammars must include the syntactic, semantic and phonological components (the last one being necessitated by the type of physical realization characteristic of those languages). Finally, well motivated considerations support the view that grammatical rules belonging to each component are heterogeneous; for example, the syntactic component seems to combine so-called rewrite rules with transformation rules.

Ignoring the absence of a phonological component in the formal-logical scheme of the description of artificial languages (which are essentially non-spoken), it seems that the only distinguishing feature of grammars of natural languages is the structural heterogeneity of their rules. It also seems that this feature is exhaustively accounted for by the fact that expressions of artificial languages lack the syntactic and lexical ambiguity characteristic of many expressions of natural languages. In particular, introduction of transformation as well as rewrite rules into the syntactic component of grammar is motivated by the need to distinguish between superficially identical syntactic forms with different semantic interpretations. With regard to syntactically unambiguous expressions, rewrite rules and structural descriptions with which they are associated (phrase markers) carry a sufficient amount of information about the grammatical structures of sentences. On the other hand, the procedure of specifying the set of expressions of a given language by means of describing their structure via inductive definition, so commonly employed in logic, bears a close resemblance to the method of phrase structure grammars.

However, regardless of all the differences and similarities between the formal methods that are in fact being used, what has, in recent years, brought the logical and linguistic studies of language closer together (putting an end, in the words of Bar-Hillel, to their peaceful and uninspiring coexistence) is the idea of mechanizing the determination of syntactic structure. In order to realize this idea in linguistics, it is necessary to fully formalize grammars of natural languages — to reconstruct their internal CODE, as it were — which would open up the prospect of systematic research into the structural properties of those languages and their interrelations.

Bibliography

1. Bach, Emmon W. (1964) *An Introduction to Transformational Grammars.* New York: Holt, Rinehart and Winston, Inc.

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

3. Chomsky, Noam (1964) "The Logical Basis for Linguistic Theory." In *Proceedings of the 9ᵗʰ International Congress of Linguists, Cambridge, August 27-31, 1962*, Horace G. Lunt (ed.), 914-978. The Hague: Mouton.

4. Chomsky, Noam (1965) *Aspects of the Theory of Syntax*. Cambridge: MIT Press.

5. Chomsky, Noam (1969) "Form and Meaning in Natural Language." In: *Communication*, John Dale Roslansky (ed.). Amsterdam: North-Holland Publ. Comp.

6. Katz, Jerrold J. and Jerry A. Fodor (1963) "The Structure of Semantic Theory." *Language* 39: 170-210.

7. Katz, Jerrold J. and Paul M. Postal (1964) *An Integrated Theory of Linguistic Descriptions*. Cambridge: MIT Press.

8. de Saussure, Ferdinand (1916) *Cours de linguistique générale*. Lausanne — Paris: Payot.

9. Szaumian, Sebastian K. (1966) "Teoretyczne podstawy gramatyki transformacyjnej." in: *Z zagadnień językoznawstwa współczesnego*, Marian Jurkowski (ed.). Warszawa: PWN.

10. Weinreich, Uriel (1966) "Explorations in Semantic Theory." In *Current Trends in Linguistic*, vol. 3, Thomas A. Sebeok (ed.), 395-477. The Hague: Mouton.

11. Zawadowski, Leon (1958) "Odróżnienie języka od tzw. parole." *Sprawozdania Poznańskiego Towarzystwa Przyjaciół Nauk* 2.