# LACUS FORUM XXXIV

*Speech and Beyond*

EKU

# REDUCTIONISM IN GENERATIVE LINGUISTICS

Patrick J. Duffley
*Université Laval*

> [A]ll scientifically meaningful statements are translatable into physical terms—that is, into statements about movements which can be observed and described in coordinates of space and time. (Bloomfield 1970[1936]:375)

> Anthropologists and philosophers have found themselves forced to invent pseudo-linguistic 'mental' entities such as 'ideas' or 'concepts', in place of the obvious and empirically discoverable morphemes and larger grammatical forms of a language. (Hockett 1958:139)

THESE TWO STATEMENTS, made by well-known linguists twenty-five years apart, are emblematic of a certain conception of what it means to be scientific. They reflect the view that all scientific statements must be based upon physically observable categories. This view is expressed forcefully by Victor Yngve (2006:267), who proposes what he calls a hard-science linguistics which rejects the study of the purportedly non-existent object called 'language' in favour of the study of "real people and other parts of the real physical world". In this paper I wish to show how this sort of view has led generative linguistic theories into a cul-de-sac. In passing a few comments will be made on other approaches which are based on similar presuppositions.

The restriction of science to the physical domain referred to in the paragraph above reflects a legitimate concern to eliminate subjective judgements from science. Yngve (p. 268) thus rejects introspective observations and feelings "because they are subject to observer bias and cannot be verifiably reproduced by others." One can only agree with him that a scientific statement must be based on observable data and consequently be testable by other researchers, as an untestable assertion could very well correspond to a purely subjective and personal opinion with no connection to reality. Recent reflections in the philosophy of science have suggested, however, that subjectivity cannot be completely eliminated from science since any observation necessarily requires an observer. Along with observability, another essential condition for data to be scientifically admissible has been proposed—that of intersubjectivity:

> [T]he arguments that we use in empirical science are expressed in intersubjective languages and must include well-established references to empirical facts, so that anyone can examine whether the arguments and the empirical proofs are valid (Artigas 2000:231).

**Reprinted from LACUS Forum 34: Speech and Beyond, ed. by Patricia Sutcliffe, Lois Stanford, and Arle Lommel. Houston, TX: LACUS. 2008**

Scientific data must not only be observable in some way, but all observers must perceive the same thing and be able to communicate their observations in such a way that other observers can check whether these observations, as well as the arguments based upon them, are valid. Intersubjectivity also plays a central role in natural language, where one sees the distinctly human ability to take another's perspective at work in everyday cognitive processing, which, like science, allows us to learn about the world through other people (cf. Verhagen 2005).

I wish to argue in this paper that meaning—an essential aspect of language which must be excluded from consideration if one adopts the positivist stance exemplified by the quotations from Bloomfield and Hockett above—is intersubjectively observable, and consequently can be admitted as scientific data. More importantly, I will argue that in order to adequately deal with its object, linguistics must take meaning into account. A good part of my argument will consist in showing the impasse to which the exclusion of semantics leads in the understanding of human language. But before doing that I would like to briefly comment on a position which includes semantics but treats meaning as a physical object.

Lamb (2006) in a paper entitled "Being realistic, being scientific" defends the legitimacy of treating language as a valid object of scientific study, his argument being that the individual linguistic system as it exists in a particular speaker is "a concrete observable physical object" (p. 208). This object corresponds to a cortical network in the brain of the individual which, for each word, associates a sub-network representing the spoken form of the word to another sub-network "representing information pertaining to that word" (p. 207), i.e., its meaning. This position raises several problems. First of all, one wonders how a "concrete physical object", such as the meaning of a word like *tomorrow* is purported to be, can evoke something which has no concrete physical existence: how could we ever have a concept of the future if meaning was merely a cortical network? Secondly, by situating meaning on the level of our neurons, this view detaches it from human experience, for it implies that if one could stimulate in the right way the neurons of someone who had never seen a dog, one could cause him to know the meaning of the word *dog*. But how could such a neurological stimulation be meaningful to someone if that person cannot relate it to any experience? Thus while neural networks should certainly be of interest to the linguist, they are not sufficient to account for the properties of natural-language meaning and so one cannot simply identify meaning with such networks.

Let us return now to the discussion of what happens when meaning is excluded from the study of language. One of the things that happens is the reduction of grammar to mere distribution. Thus Hockett, who defines morphemes as "the smallest individually meaningful elements in the utterances of a language" (1958:123), sees the grammar of a language as a description of "(1) the morphemes used in the language, and (2) the arrangements in which these morphemes occur" (p. 129). Meaningful units such as morphemes are dealt with as much as possible in the same terms as meaningless units of sound, or phonemes. Just as the recognition of a phoneme is based on complementary distribution and phonetic similarity, so the definition of a morpheme is based on complementary distribution and semantic similarity, but the recourse to semantics is limited to answering the question 'Is the meaning the same or is it different?'. This allows one, for example, to classify the forms



Reprinted from LACUS Forum 34: Speech and Beyond, ed. by
Patricia Sutcliffe, Lois Stanford, and Arle Lommel. Houston, TX: LACUS. 2008

[ɪz], [s] and [z] in the English words *bridges*, *lids* and *bits* as one morpheme, since they are each found in different phonetic contexts (complementary distribution) but express the same meaning.

As long as one's goal is merely to identify the minimal semantic units of a language, this method is fairly adequate. However, when it comes to accounting for the way these units are combined and used, its deficiencies quickly become apparent. In this view of language, syntax has to do merely with the arrangements in which morphemes occur. Since the focus is restricted to the observable sign, the essence of syntax concerns how linguistic signs are grouped together into constituents, gradually building up from basic blocks of morphemes to the complex level of the sentence. So, for instance, in *The ducks flew away*, the morphemes 'duck' and 'plural' combine to form the noun *ducks*; the latter combines with the article *the* to form the noun phrase *the ducks*; on the side of the predicate, the morphemes 'fly' and 'past' combine to form the verb *flew*, which combines with the adverb *away* to form the verb phrase *flew away*; and the latter combines with the noun phrase *the ducks* to form the sentence. A sentence is consequently treated as a sequence of morphemes which are grouped together according to a hierarchical structure. To describe the syntax of a language is to describe in as simple and general a fashion as possible all the ways in which the morpheme and phrase units of a language can be arranged.

Here another positivist axiom must be mentioned concerning the nature of the object of linguistics, language itself. Since all that is directly observable of language is the production of certain phonetic, graphic or gestural sequences, a language is defined by Bloomfield as "the totality of utterances that can be made in it" (1970 [1936]:129–30). This corresponds to Chomsky's (1978[1957]:13) famous definition of a language as "a set (finite or infinite) of sentences, each finite in length and constructed out of a finite set of elements." Such a view of language reduces it to a set of observable objects.

Now a set can be treated as a mathematical object describable by means of rules which allow one to generate it: with the simple instruction 'n × 2' (where *n* is a whole number) one can generate the entire set of even numbers. If a number does not conform to this rule, it is not a member of the set. The goal of linguistics according to Chomsky becomes then "to separate the *grammatical* sequences which are sentences of L [language] from the *ungrammatical* sequences which are not sentences of L and to study the structure of the grammatical sequences" (*ibid*). The grammar of a language accordingly is "a device that generates all of the grammatical sequences of L and none of the ungrammatical ones" (*ibid*).

One form that such a grammar can take is that of a series of rewrite rules, such as:

1. Sentence → NP + VP
2. NP → T + N
3. T → *the*
4. N → *man*, *ball*, etc.
5. VP → Verb + NP
6. Verb → *hit*, *took*, etc.



Reprinted from LACUS Forum 34: Speech and Beyond, ed. by
Patricia Sutcliffe, Lois Stanford, and Arle Lommel. Houston, TX: LACUS. 2008

This corresponds to a phrase structure grammar, which is one possible way of describing the arrangements of morphemes in a language. However, if one adds to such a grammar another type of rule, which treats some arrangements as derived from others, one can avoid the reduplication of certain rules and achieve a more elegant and economical description. (*ibid*:44). Thus, to give a very simplified example, deriving the passive *John is frightened by sincerity* from *Sincerity frightens John* allows one to exclude **John is admired by sincerity* from the set of grammatical sentences of English based on the ungrammaticality of **Sincerity admires John*.

Mainstream Generative Grammar has never called into question the basic goal of developing a computational procedure for generating all of the grammatical sentences of a language and excluding the ungrammatical sequences. In minimalist theory, its most recent formulation, a sequence is generated by the syntax through the gradual assembling of lexical units to form a structure which is spelled out and submitted to the perceptual-auditory and conceptual-intentional interfaces for interpretation. If it meets the legibility conditions at both interfaces, it is acceptable; if not, the derivation is said to "crash" (Seuren 2004:33). Consequently, a generative grammar is primarily a distributional grammar. It treats sentences as arrangements of morphemes, and seeks to formulate configurational rules which could generate the right set of positional arrangements. The starting-point for the operation of the rules has varied—in minimalism it begins with the selection of a certain number of lexical items from the lexicon. But the nature of the grammar has not changed, and meaning is kept out of the picture as much as possible, the essence of grammar being taken to be syntax, i.e., position or configuration.

Since language obviously serves the purpose of communication, semantics must be incorporated in some way, however. In the generative model this is done as late as possible, at the peripheral level of the interface between the syntax and the conceptual-intentional system. Minimalism does include 'semantic' features along with phonological and syntactic ones in the composition of lexical items (cf. Hornstein, Nunes & Grohmann 2005:291), but semantic features are treated as merely "interpretable", i.e., they are assigned an interpretation only once they reach the interface level with Logical Form. This entails that semantics is interpretative and propositional, and syntax is autonomous from meaning. There are very important problems with the way in which this model deals with semantics and its relation to syntax. Seuren (2004:16) has aptly characterized minimalism as proposing a "random-generator" model of syntax, whereby the latter acts as an "unguided sentence generator, randomly selecting items from the lexicon and 'merging' these into proper syntactic structures". He criticizes this model for not respecting the chain of causality observed in the real use of language, which always begins with a thought that the speaker wishes to communicate which guides the whole process of sentence formation. Seuren also points out (p. 161) that it is "absurd" to postulate that "a randomly generated sentence structure should be taken to pass an instruction to the cognitive system of the same organism for the sentence to be interpreted." On the contrary, the speaker must assemble lexical items in an appropriate way so as to express the cognitive content he wishes to communicate.

Regarding the reduction of meaning to propositional semantics, two factors converge to make this the natural option for a generative grammar. The first of these has to do with



Reprinted from LACUS Forum 34: Speech and Beyond, ed. by
Patricia Sutcliffe, Lois Stanford, and Arle Lommel. Houston, TX: LACUS. 2008

the starting premise that meaning is not treatable scientifically, and consequently must be excluded from the analysis as much as possible. We have seen how this led to treating morphemes, defined as the smallest *meaningful* units of language, as if they had no meaning, i.e., a distributional approach to syntax. Here we see that the inevitable return to meaning occurs only once the syntax has operated, i.e., on the level of the sentence. Since the question of truth and falsity arises on the level of the sentence, and not on the level of the word or morpheme, it is natural in a generative model for semantics to have recourse to the categories of propositional logic. Moreover, since truth conditions are defined in terms of how the world must be in order for the sentence to be true of it, this creates the illusion of being able to treat meaning as something empirically observable, which is reassuring to the scientific positivist.

Delaying the recourse to meaning to the sentence level is the source of many formidable problems, however. Doing this assumes that sentences, or better, the sequences of morphemes that are grammatical in a language, have meaning in the abstract, an assumption which is also shared by logic. Real utterances, however, cannot be determined to be true or false without taking into account the situation in which they are uttered and the intention of the speaker who produced them. If someone says *It is raining*, one must avert to the meteorological conditions at the time and place of utterance, or perhaps to some other place that the speaker has in mind and to which he wishes to refer. Even a statement such as *Water is composed of hydrogen and oxygen*, which might appear to be true outside of any particular situation and independently of the speaker's intention, would not be true applied to ordinary tap water if the speaker's intention were to convey the message that tap water is composed exclusively of hydrogen and oxygen, as the liquid that flows out of our taps also contains chlorine, fluoride, bacteria, etc. The utterance is an ephemeral unit assembled on-line for the purposes of communicating a particular cognitive content. To treat such a unit as having a stable permanent meaning is to misrepresent its nature. It is to seek the existence of a stable relation between sound and meaning on a level at which such a relation does not exist.

This has grave consequences for the understanding of language, as it severs the essential bond upon which the latter is based. Once cut loose from any stable relation to the linguistic sign, semantics is almost completely unconstrained and can float off into the realm of possible worlds. Since any sentence can correspond to an infinity of situations in which it could be said, with pragmatic factors and possible variation of speaker intentions complexifying the picture even more, one wonders how the hearer could ever guess the meaning the speaker was trying to convey if this meaning corresponded to the exact state of the world being referred to. Even something as apparently simple as the sequence *The circle is inside the square* can refer to a vast multitude of different real-world situations: Is the circle 100% inside the square, or only 99.9%, 99.8%...; Where is the circle inside the square?; Are there other objects inside the square?; etc. In this view, the semantics of a language becomes a relation between an infinite set of sentences, on the one hand, and an infinite set of infinite sets of possible worlds in which these sentences would be true. Defining a language as an infinite set of sentences makes it into an imaginary infinite; defining the meaning of a sentence as an infinite set of possible worlds makes it, too, into an imaginary infinite. It is



**Reprinted from LACUS Forum 34: Speech and Beyond, ed. by Patricia Sutcliffe, Lois Stanford, and Arle Lommel. Houston, TX: LACUS. 2008**

paradoxical that the starting premise that linguistic science must limit itself to the realm of the physically observable should lead to such a proliferation of physically unobservable entities. Moreover, if this corresponded to what language and meaning were, a language would be fundamentally unlearnable, as no child could ever acquire both an infinite set of morpheme-sequences and an infinite set of possible worlds constituting their meaning. The only way out of this impasse would seem to be to appeal to some form of powerful generative capacity, as generative grammar does for syntax. The multifarious diversity of the universe might be supposed to account for the infinity of the set of meanings (but that would still not solve the problem of talk about things that do not really exist.) Recourse to an autonomous syntactic generator, however, renders mysterious any connection between form and meaning, as it operates independently of both, as if the cognitive content which the speaker wishes to express had nothing to do with the forms that he uses to communicate this content and their arrangement.

Given the impasse to which the exclusion of meaning leads in linguistics, it would seem legitimate to call this postulate into question. The exclusion of semantic content as a *persona non grata* from the realm of scientific data has been shown to render human language incomprehensible. The question arises at this point as to whether, if meaning is an integral part of language, linguistics can claim to be a science. To answer this question I would like to return to the notion of intersubjectivity as the condition for data to be amenable to scientific investigation. If this is so, it can be argued that meaning in human language constitutes a paradigm case of intersubjectivity, and so, in principle, should be a possible object of science. The essence of intersubjectivity is that all competent observers must agree on the nature of the data used to confirm or disprove a scientific hypothesis. I would suggest that the essential property of human language is that all speakers basically agree on the meanings of the words and morphemes of the language used in their speech community. This is the very condition for a language to be able to function as an instrument for communication between speakers. Of course, misunderstandings are possible (cf. Tannen 1990), and for some concepts there can be divergences among speakers such as that between the chemist's and the ordinary person's conception of water. However the chemist can still understand what the ordinary person means by the word, and so communication between the two is possible. Consequently, meaning is perfectly admissible as scientific data in the multitudinous cases where native speakers agree that a given utterance in a given situation conveys a given message. Whether linguistics will be able to explain why this is so for specific utterances is not guaranteed. Nevertheless, at least the data are there to provide a legitimate object of scientific enquiry.

## REFERENCES

Artigas, Mariano. 2000. *The mind of the universe*. Philadelphia: Templeton.

Bloomfield, Leonard. 1970. *A Leonard Bloomfield anthology,* ed. by C.F. Hockett. Bloomington: Indiana University Press.

Chomsky, Noam. 1978 [1957]. *Syntactic structures*. The Hague: Mouton.

**Reprinted from LACUS Forum 34: Speech and Beyond, ed. by Patricia Sutcliffe, Lois Stanford, and Arle Lommel. Houston, TX: LACUS. 2008**

HOCKETT, CHARLES F. 1958. *A course in modern linguistics*. New York: Macmillan.
HORNSTEIN, NORBERT, JAIRO NUNES & KLEANTHES K. GROHMANN. 2005. *Understanding minimalism*. Cambridge: Cambridge University Press.
LAMB, SYDNEY. 2006. Being realistic, being scientific. *LACUS forum* 32:201–9.
SEUREN, PIETER A.M. 2004. *Chomsky's minimalism*. Oxford: Oxford University Press.
TANNEN, DEBORAH. 1990. *You just don't understand*. New York: Ballantine Books.
VERHAGEN, ARIE. 2005. *Constructions of intersubjectivity*. Oxford: Oxford University Press.
YNGVE, VICTOR H. 2006. Formalizing the observer in hard-science linguistics. *LACUS forum* 32:267–76.



Reprinted from LACUS Forum 34: Speech and Beyond, ed. by
Patricia Sutcliffe, Lois Stanford, and Arle Lommel. Houston, TX: LACUS. 2008