DOCUMENT RESUME

ED 267 604 FL 015 534

AUTHOR Paillet, Jean-Pierre
TITLE Linguistics and High Technology.
PUB DATE 85
NOTE 11p.; In: Carleton Papers in Applied Language Studies. Volume II, 1985. Aviva Freedman, Ed. (FL 015 530).
PUB TYPE Information Analyses (070) -- Viewpoints (120)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Applied Linguistics; *Artificial Intelligence; *Biology; Futures (of Society); *Information Science; Interdisciplinary Approach; *Semiotics; *Technological Advancement; Trend Analysis
IDENTIFIERS *Bioengineering

ABSTRACT

There are two main areas of activity hidden behind the catch-phrase "high technology": the domain of computers and communication, and biological engineering. Because both deal with patterns of organization and information, it would seem that linguistics could contribute to their advancement. The main concept in information systems is distinctiveness, with its attendant notion of code. The main activity in information technology is the designing, implementation, and checking of devices capable of effecting specified operations on codes while preserving certain types of distinctiveness. Similarly, in biological engineering, the primary concern is with genetic codes expressed in the systems of the cell. The type of awareness linguists have developed about language, or psychologists and ethnologists about other sense systems, is needed to bring scholars in technology a wider, contextual perspective on what they are trying to do, especially since the objects they are dealing with are increasingly linked to their contexts. (MSE)

Linguistics and High Technology

JEAN-PIERRE PAILLET

Department of Linguistics
Carleton University

What is "High Technology"?

The phrase is used everywhere to wave a hand vaguely in the general direction of Kanata, Silicon Valley or those even more mysterious places where they manipulate genes....

There is also the image of an activity whose products are going to rob us of jobs, soak us with radiation, turn us into video game junkies, and/or breed various kinds of biological monsters.

In short, Hitech is the new name for everything we do not understand about the applications of the more recent scientific developments. There are two main areas of activity hidden behind the catch-phrase.

The first one is the domain of computers and communications. The tremendous changes which have happened there in the past fifteen years are a direct spin-off from the space program in the U.S.A. On the one hand, spacecraft required light, compact and very robust control and communication systems. The money and manpower available made possible the start of the "chip" industry, which has now acquired many other markets (once the products were there, it was easy to see how to use them in novel ways) and is self-sustaining. On the other, communication satellites have made practical many types of operations on information which were previously too expensive or too cumbersome.

For the public at large, the central element of all these activities is not so much the array of technical developments which has made all this possible, but the intrusive presence in everyday life of objects whose nature is not *material*, but *informational*. The concept of information as a legitimate mode of being was introduced only in the forties, and already it forces itself on an unprepared general public!

There is nothing new, of course, in what information machines do. We have been doing it for ages, with other means. What is new is the rate at which we can do it, which forces our attention, by inducing a massive shift in the social and economic parameters of our activities. Consider the analogy of the steam engine. For millenia before, people had been spinning and weaving: nothing new at that level. But if a steam engine can run a whole textile mill, we have to start thinking differently about textile production, use and management. We must also start thinking consciously about energy management in the production of goods, in transportation, home and office practices, etc. We all know that for a long time we have not done so, because the attention of the public was focused on the visible objects of the industrial revolution rather than on its ecological aspects.

Similarly, while chips and comsats are impressive an... ous, the real mastery of the new technology is to be found in ...spread understanding of its systematic implications. This requires t... the public at large learn to think consciously in terms of information objects, their properties and their ecology. We shall examine below what Linguistics can contribute in this matter.

However, we must first look at the other main area of high technology: biological engineering. In many respects, bio-engineering is now at the same stage that computers and other information machines were in the fifties. Then, computers were very expensive, hand-made (or nearly so), cumbersome, and generally on the fringe of things. They were featured in comic movies as the source or focus of various catastrophies, where their high priests and attendants were ridiculed. We have a similar attitude now towards bio-engineering labs, except that, in keeping with the spirit of the times, we have turned from comedy to horror stories. Sooner or later, some economic incentive is going to boost bio-engineering work to the same level as *informatique*—in fact, it is already happening. It would be nice if we were not caught short again in our degree of understanding.

The objects that bio-engineering deals with are of a different nature again. Specifically, even as information objects consist of patterns of organisation of matter and energy, those "new" objects consist of patterns of organisation of information. No term has yet been consecrated to designate such objects, and I have proposed the word *sense*, for various reasons, at least one of which should become clear somewhere in this paper.

Sense objects too have their own specific properties and their special ecology. Somehow we must become familiar with them, in a conscious way, before we are overwhelmed and bewildered again. Linguistics can help here too.

1. Some properties of Language

It is trivial to tell a linguist that language is complex, less so to try to see into the nature of its complexity. I believe that it is precisely such an insight that makes linguists useful to the task of giving people mastery over the implications of High Technology.

I see the main sources of complexity in language to be the following:

— Language is self-referring; while this property is not immediately relevant to our present "Hitech", the time will come when it will be, inasmuch as the properties which make self-reference possible are even now being developed for information machines.

— Language involves information systems *and* sense systems, as well as "interfaces" between these and systems of a different nature, energy and matter on the one hand, and on the other the vast uncharted domain vaguely called *meaning*.

The next two sections give a brief sketch of the aspects of the linguist's experience which are directly relevant to an understanding of the issues of High Technology.

2. Linguistics and information

The main concept in information systems is *distinctiveness*, with the attendant notion of a *code*. A code is essentially a collection of distinct objects, with rules for combining them and conventions as to which combinations shall be considered equivalent. The main activity in information technology is the designing, implementing and checking of devices (material, *hardware*, or informational, *software*) capable of effecting specified operations on codes while preserving certain types of distinctiveness.

All this area of thinking ought to be well known to linguists, since the "discovery" of the phoneme. The object of phonology, increasingly well understood, is to study the special codes called *phonological systems*, which serve the purpose of representing in a stable and distinct way the entities of interest in a human language. The principles of phonology are of course independent of the particular implementation of the code, although most of the attention has been given to the "natural" implementation by means of vocal sounds. Over the years, many people have tried to apply information theory to phonology, but most of the results have been trivial or disappointing. The main reason for this is that the kind of systems that information theorists deal with, like most mathematical systems, are extremely regular, and unencumbered with such considerations as the wide variability of the apparatus which is to implement the code in "real life". More recently, the multiplication of uses of computing machines in widely different environments has made such considerations very important. The linguist has a wealth of experience, however informal, with the way in which the evolution of the human species has coped with the problem of implementing regular codes in *very* irregular environments. Starting with Martinet's *Economie des changements phonétiques*, linguists have devised models of the interaction of the requirements of codes and their ecological context; there is some understanding of the conditions of stability for a code, of the dynamics of its evolution, etc. In this way, historical linguistics and linguistic typology are directly relevant to software and hardware design—if only the specialists involved were to know of each other's concerns.

Similarly, the design of error-resistant codes could benefit greatly from what sociolinguists know of language communication across social and dialectal clines. It goes without saying that benefits would flow both ways: communication engineers have done a lot of work on performance degradation of communication systems in adverse conditions, and the concept e.g., of "graceful degradation of performance" would be very useful to teachers of second languages, to evaluate the teaching of sound systems.

Another area where the linguist's experience is relevant is that of abstract syntax. Characteristically, the syntax of computer languages has so far been simplistic. There was no particular reason to make it more elaborate as long as the only users were specialists who could invest the effort necessary to "think like a computer." Now, on the other hand, "user-friendliness" is *de rigueur*, and that means that our machines must take our preferences into consideration, both for input/output (the "human language interface") and for the structure of the tasks we want to give them (more on this under 3.)

The work of Chomsky has made the technical half of the world aware of linguistic syntax, to the extent that this domain is where interaction between linguistics and technology is the most extensive and the most fruitful. Bypassing the problem of speech recognition, many have developed systems which can interact with the user in a decently flexible form of written English (and other languages too). Now is the time for linguists to jump in, inform themselves and contribute to the refinement of these still-crude systems. There are, of course, many benefits to be gained for linguistics in such an interaction, if only by finding out how and why such systems work so well—or so poorly. Researchers like Z S Harris, M Gross, T Winograd have already voiced and/or published many interesting and seminal ideas on this subject.

Another form of research which should be of interest to linguists is the attempt at *designing* a human language for specific purposes. For instance, the LOGLAN Institute, expanding on the work of its founder. J. Cooke Brown, is developing Loglan into a full-fledged language. What distinguishes this attempt from previous designs is that the authors are conscious of semantic and cultural implications (more on this under 3.), and that they are seeking a delicate balance between the objective of total unambiguity at the purely formal level (Loglan grammar can be automatically compiled using the *yacc* compiler compiler on Unix) and human constraints on syntax.

The seemingly chaotic nature of human syntax is rooted in the same constraints as that of human phonology. In the case of phonology the "boundary conditions" imposed by the anatomy and physiology of the vocal tract and breathing apparatus of humans are obvious. In the case of syntax, one might think that such constraints are irrelevant. But an increasing number of studies show that there are iconic connections between syntactic patterns and patterns of human cognition. A fairly clear, although by no means simple, case is that of the subject/agent/topic complex. In most, if not all languages, one noun group (or whatever entity plays such a role in the language) receives a privileged syntactic treatment. The interpretation of this observation is now occupying many linguists the world over, with considerations such as new and old information, the requirements of human syntactic processors, the universality of the semantics of agentivity, and whatnot. Other linguists (and the same ones too) consider matters such as natural person hierarchies, and their implications as to the morphology and syntax of verbs. At the present time, computers can routinely decode such sentences as *The technician burned the circuit* and *The circuit was burned by the technician*, and correctly extract from both the same knowledge about the state of the circuit and the world at large. On the other hand, they cannot tell why one would use one sentence rather than the other, nor correctly select one in appropriate circumstances. Nor can they decide whether to mark a question by preposing the tense or by appending *eh?*. It is up to linguists to study the boundary conditions on syntax and to tell the computers, which on the whole are ready to comply. And this brings us to sense, etc.

3. Linguistics and sense

The main concept in sense systems has not been given a name. I propose to call it *semiosis*, and hope to justify it, albeit sketchily, together with the use of the word "sense."

The customary use of the word "sense," as in "the five senses" is intimately connected with unconscious, one is tempted to say "instinctive," pattern recognition in the biologically established modes. This can be extended, and has been, to other modes, as in "infrared sensor, remote sensing, pressure-sensitive switch." The point of pattern recognition is to assign the raw input signal—an information object—to preestablished or dynamically adaptive categories of stimuli. What constitutes such categories is that they are differentially correlated with objects of another class, in some way or other.

For instance, Gestalt psychologists have thrown light on the phenomenon of *closure*, in which a portion of a configuration of (e.g., visual) information is *recognized* as the whole configuration. One can at present but speculate on the foundation for the establishment of perceptual categories; however, much work in psychotherapy suggests that survival requirements must have a strong influence. In other words, in this model, the operation of perception consists in assembling information objects into patterns which are *correlated with something else*, for instance biological imperatives.

Linguists are familiar with such structures; for instance, morphemes are understood to be the correlation of certain kinds of configurations of phonemes with something else, which distributionalists, with virtuous rigour, refused to call meaning. And indeed language does not carry or create meaning, human (and possibly other) beings do. What language does is offer a bunch of ready made, though evolving, correlations between information patterns and whatever the users want to make of them. These correlations have been called *signs*, and linguists (ethnologists also, and more recently semioticians) have bravely muddled their way through them, trying to understand their properties, interactions, conditions of existence and stability, in other words their ecology.

The most insightful definition of a sign I know is that of C.S. Pierce: though obscure at first glance, it pins down the essential dynamics of the process whereby a sign (or, as I like to say, a sense object) arises. This process, which, borrowing an articulate Greek structure, he calls *semiosis*—i.e., *semeion*, 'sign' + *-sis*, 'the process of'—involves a form of interaction between two objects which generates a similar relation between one of them and a third. In other words, if some configuration correlates sucessfully with something else, this very correlation tends to give rise to others like it: in language, such events are called *effets de sens*, derived meanings, connotations, and the like.

Pierce's definition is purely phenomenological. there is a lot of research and *reflective inquiry* to be done in order to work out the details of such processes, and surely linguists have some headstart.

Pierce's definition also entails a recursive character for semiosis, and ultimately the possibility of self-reference. This was exploited by Godel in the proof of his famous theorem, and is slowly finding its way into computer systems, bootstrapping them from the status of information machines to that of sense machines. This is the area of Artificial Intelligence (AI), which, not surprisingly in the context of this discussion, has as its subdomains Pattern recognition (i.e., the reproducing of human sensory functions), Language processing (self-explanatory), Problem-solving (i.e., seeking acceptable or feasible transformations of sense patterns into a form matching a stated goal, starting from a description of a given situation). Following the lead of Japan, many countries are committing resources to the development of this field, in an effort which has been compared to the Moon program of the sixties. It would seem desirable to make as much use as possible of the experience of linguists, psychologists and ethnologists in such an endeavour, in order to avoid reinventing cognitive wheels.

Biologists and bio-engineers, from the start, have to deal with sense entities. When the nature of the genetic code was revealed to the world, many speculated on the similarity, or lack of it, between the genetic code and language. They sensed (pun intentional) that the way language uses sounds could tell us something about the way in which the cellular machinery uses sequences of certain special chemical radicals.

Generally speaking, chemical species are intriguing to linguists. Some use chemical metaphors, the most common revolving around the notion of valence and bonding. Not much has come out of these speculations, however, because we have not focused much on the underlying dynamics of the perceived similarities.

As biologists progressively worked out the details of how the genetic code is expressed in the systems of the cell, we have begun to see how the notion of sense can link linguistic research and biology. In this case, the connection is more at the fundamental level than in technology. In cellular systems, we can directly observe the properties of sense objects. Messenger RNA works very much like a recoder, transforming the information while preserving specific features of distinctiveness. Ribosomes are like morphemes, recognizing specific sequences of codons at one end, while matching specific aminoacids at the other: molecular signs! A fascinating question, of interest to linguists and biologists alike, is whether the correlations between codons and aminoacids are arbitrary. It is not clear even how to ask this question in a precisely answerable way, but there it is in biology, after a long history in linguistics. φυσει or θεσει ? i.e., by evolving or by positing? It should not be difficult to see that a serious answer to this question is crucial to our handling of biological engineering....

Another property of cellular systems, which may throw light on the previous question, is that biological molecules are often self-assembling

On selectional criteria alone, this makes sense (it does, doesn't it?); but how wonderfully lucky that it is possible at all! (On the other hand, if it were not, we would not be here to speculate on it). Do sense systems in language show the same property? This has been suggested several times (e.g. Hockett's notion of *deep structure*, in which the senses of morphemes contract relations which are not otherwise signaled in the overt structure) and is the very basis of R. Schank's approach to language understanding.

As a final example of the investigation of sense systems, we note that cellular systems exhibit a form of self-reference, embodied—literally—in the immune system. This organisation is engaged in the production both of signatures identifying the organism—ultimately the genome out of which the organism arose—and of sensors to distinguish between legitimate components of the system and aliens. In a similar way languages have evolved special sense units to police the language itself. Very little systematic work has been done, however, on the ecology of these elements (e.g. Hockett's, again, considerations on editing, or Harris's metalinguistic operators).

4. In sum

It is characteristic of the beginning of a change, as we noted earlier, that people notice small, isolated aspects without necessarily perceiving the major trend. I have tried to take the emphasis of these comments away from the immediate aspects of technology, towards an understanding of their longer-term significance. The trend I see is that, having acquired a remarkable mastery over matter and energy systems (we are still struggling to acquire the corresponding wisdom), we are now turning to other kinds of systems, those of information and sense. The task at hand is to understand their dynamics, in order to make the right kinds of decisions both at the level of practical implementation and, more importantly, at the level of social and ethical concerns. A feature of these new systems which distinguishes them from previous kinds of objects of technology is that they increasingly resemble mental objects, the study of which has heretofore been the province of "social scientists" and humanists. The nature of the new technology is best summarized in the area of AI mentioned above. And within AI, the subdomain of Language processing gathers all the significant elements found in the other subdomains. It may be that solutions can be found separately to each individual practical problem of AI, or any other area of High Technology. If so, we would still find ourselves in a state of general disorientation. The type of awareness linguists have developed about language, or ethnologists about other sense systems, is needed to bring technologists a wider, contextual perspective on what they are trying to do. This is central to the work at hand, since, to borrow a phrase from Bateson, the objects we want to operate on are increasingly "creatures of context."

Beyond the technology of information and sense, we can predict in the distance a technology of increasingly mental objects. D. Hofstadter's columns in *Scientific American* contain many speculations on this topic, from the nature of musical creativity to the ecology of ideas. In the latter case, elementary ideas, or *memes*, are viewed in analogy with viruses, which can invade a population of organisms, in this case the minds of a human group, and use their internal machinery to reproduce and propagate. This may lead to a technology of propaganda, of course, but also to mental prophylaxis...But nil novi. Science fiction writers have long ago broached this topic. One may only hope that by then semioticians will have developed a sound basis of understanding to guide the practical efforts of the device makers.

REFERENCES

Bateson, Gregory

1975 *Steps to an Ecology of Mind.* New York: Ballantine Books.

1979 *Mind and Nature.* New York: Bantam Books.

Brown, J. C.

1972 *Loglan 1: A Logical Language.* Palm Springs: The Loglan Institute Inc.

Chomsky, Noam

1956 Three Models for the Description of Language, *IRE Transactions on Information Processing,* IT-2, 113-124

Dawkins, Richard

1976 *The Selfish Gene.* Oxford: The University Press.

Godel, Kurt.

1962 *On Formally Undecidable Propositions,* translation of 1931 German original, New York: Basic Books.

Halliday, M. A. K.

1961 Categories in the Theory of Grammar, WORD 17, 241-292

Harris, Zellig

1968 *Mathematical Structures of Language. Interscience Tracts in Pure and Applied Mathematics 21* Toronto, Interscience publication

1970 The Two Systems of Language, *Papers in Structural and Transformational Linguistics,* ed. Z. S. Harris Dordrecht: Reidel

Hockett, Charles F

1958 *A Course in Modern Linguistics.* New York: MacMillan

1968 *The State of the Art.* New York: The Humanities Press

Hofstadter Dougla
1981 *Godel Escher Bach An Eternal Golden Braid* New York Random House

1983 Matemagical Themas, *Scientific American*

Lamb Sidney
1966 *An Outline of Stratificational Grammar* Washington The Georgetown University Press

Paillet Jean Pierre
1975 Problemes de Notation pour l'etude du contenu linguistique *Languages* 35 27-69

Paillet Jean-Pierre and A Dugas
1982 *Approaches to Syntax* Philadelphia John Benjamins

Winograd T
1972 *Understanding Natural Language* New York Academic Press

Language as a Cognitive Process Reading, Mass Addison-Wesley

Winston Patrick Henry
1984 *Artificial Intelligence* Reading Mass Addison-Wesley