SEMANTIC ENGINEERING[1]

JERZY POGONOWSKI[2]

Department of Applied Logic
Adam Mickiewicz University
www.logic.amu.edu.pl

I. Introductory remarks

The title of this paper might suggest that I'm going to talk about manipulation with meaning in natural language. As an example of such a manipulation one could take the following, rather non-standard use of the passive voice:

> One of our missiles flying yesterday night on its peaceful mission in the south-east region of the country was hit by Boeing 747 with 269 passengers on board, all of them, including children, suspected of espionage on behalf of an enemy. None of our citizens was harmed but the missile was totally destroyed.

It is likely that, when you hear of „semantic engineering", then you automatically think of artificially created human languages (e.g. Esperanto) or computer languages, pidgins, slangs, persuasion systems (propaganda, commercials), Lewis Carroll's poetry, some uses of euphemisms and idioms, etc. The term „semantic engineering" has to be understood in this paper not as manipulation with particular meanings but rather as manipulation with the whole semiotic systems. Those systems are supposed to be possible human languages. The main goal of semantic engineering is to establish a borderline between those semiotic systems which could be treated as possible human languages (including all natural languages, existing presently as well as in the past) and those which could not. In a certain sense, then, one may think of semantic engineering as a method of thought-experiments in linguistics which could appear useful in the investigation of language universals. Given a fixed bundle of semantic conditions one can try to construct a language satisfying those conditions. Whatever is the case, i.e. whether such a language can be constructed or not, one is usually able to draw some general conclusions concerning the investigated semantic properties. Thus, for instance, if one can prove that there can be no human language with a given property F, then this means that the property *non-*F is universal.

[1]Published in: *Semiotic Theory and Practice*, Mouton de Gruyter, Berlin New York Amsterdam, 1988, 899–907.

[2]The work on this paper has been supported by the Alexander von Humboldt-Foundation. I am very grateful to the Foundation and would like to thank them as well as Prof. Theo Vennemann (Ludwig-Maximilians-University in Munich) for providing me with excellent conditions for my research work. Finally, my thanks go to Gaby Hollmann for carefully typing the manuscript.

Our main thesis is that the property of being a possible human language depends essentially on the assumed linguistic theory. I'm going to discuss two aspects of the concept of a possible human language which could be called global and local, respectively. The notions of ideal typology and preference relations introduced by Theo Vennemann (cf. Vennemann 1982a, 1982b, 1983) are used in the next section for imposing some global algebraic structure into the space of all possible languages. The last section of the paper deals with a few examples of semantic engineering.

II. Typologies, preferences and possible languages

Given a class of linguistic entities to be typed, say G (where elements of G are possible human languages, language varieties, idiolects, etc. — depending essentially on our domain of interest), and a general linguistic theory, say T, Vennemann's ideal types for G based on T are properties of elements of G expressible in T. More exactly, those ideal types are formulas with one free variable, ranging over elements of G, whose non-logical constants are determined by the general linguistic theory T. The formulas in question should also be contingent, i.e. neither tautological nor contradictory. Further, Vennemann's ideal typologies for G based on T are those sets A of logically non-equivalent ideal types which are partially counterinstantiated, i.e. for which there is L in G and t in A such that L does not have the property expressed by t.

Elaborated examples of ideal typologies in the above sense can be found in Vennemann, 1982a (e.g. a formal reconstruction of the typological ideas of Schlegel and Humboldt).

Observe that Vennemann's ideal types correspond only to properties (or, in extensional terms, to classes) of languages. Consequently, it seems difficult — if possible at all — to describe in terms of those ideal types e.g. typological *relations* between languages or such phenomena as for instance incorporation or polysynthesis. Furthermore, it is in many cases easier to *compare* two languages with respect to a given parameter than to provide for a rigid classification of languages (or even an association of particular languages with ideal types); for example, it is easier to decide that one language is more synthetical than the other one than to define explicitly the class of all synthetic languages. For those reasons, I would like to suggest that ideal types should be understood in a more general way.

Let G, as before, be the class of all possible languages and assume that T_1, T_2 is a pair of theories such that:

- T_1 is a general linguistic theory describing internal properties of languages from G,
- T_2 is a theory in which we can talk about elements of G — i.e. we have variables ranging over elements of G, predicates (constructed with the help of T_1) corresponding to relations between elements of G and a suitable set of axioms characterizing those predicates.

Let Rel be the family of relations about which we can talk in T_2 (e.g. translation, typological classes and relations, diachronic relations, etc.). Our object of investigation is thus the relational structure (G, Rel).

Let Fml_n be the set of all formulas from the language of T_2 with exactly n free variables and let Tp_n be the family of all classes of T_2-equivalent formulas from Fml_n. It is well known that Tp_n is a Boolean algebra for all $n \geqslant 0$. Let us call its elements *ideal n-types*. One can easily see that all Vennemann's ideal types are those elements of Fml_1 which are contingent and which, in addition, do not contain any bound variables. Hence his ideal types are contained in elements of Tp_1.

A non-empty subset A of Tp_n is called an *ideal n-typology*, if:

- (i) A does not contain either zero or unit element of the algebra Tp_n,
- (ii) A is partially counterinstantiated, i.e. there are $L_0, \ldots, L_{n-1}$ in G and t in A such that the n-tuple $(L_0, \ldots, L_{n-1})$ does not have the property expressed by t.

The above slight modification of Vennemann's original idea enables us to apply a uniform formal framework to practically all typological constructions. It introduces simultaneously an algebraic structure into the space of all ideal n-types. This structure can be also used in the characterization of the elements of G, i.e. possible languages.

For any L from G let $U(L)$ be the set of all ideal 1-types of formulas realized (in the model-theoretic sense) by L in the structure (G, Rel). As the set of all such formulas is a maximal consistent set, it follows that $U(L)$ is a principal ultrafilter of the algebra Tp_1. Intuitively, the ultrafilter $U(L)$ corresponds to the family of *all* properties of a possible language L which are expressible in the theory T_2. Obviously, if A is any ideal 1-typology, then a given language L instantiates exactly those ideal 1-types from A which belong to the ultrafilter $U(L)$. Similar constructions can be introduced in all algebras Tp_n.

Several further structures in the algebra Tp_n are of linguistic relevance. Thus, for instance, the canonical ordering of Tp_n, as well as the Boolean operations in Tp_n, can be used in the investigation of mutual relationships between ideal n-typologies, operations on them, etc. Some special role play the atoms of Tp_n, i.e. the minimal non-zero elements of Tp_n. If the algebra Tp_n has any atoms at all (which depends on the theory Tp_2), then some, or even all, of its elements can be represented as Boolean combinations of the atoms. Consequently, ideal n-types can be in such a case represented as combinations of „atomic" (maximally simple) ideal n-types.

A method of comparing possible languages from G with each other or with ideal 1-types can be obtained by taking into account the quotient algebras $Tp_1/U(L)$ for any L from G. Intuitively, elements of any such algebra are sets of ideal 1-types. Two ideal 1-types are put together into an element of $Tp_1/U(L)$ iff their Boolean equivalence belongs to the ultrafilter $U(L)$, i.e. if the properties expressed by them are equivalent „from the point of view" of the possible language L. Again, similar constructions can be introduced in the case of all algebras Tp_n.

Another sort of global structure in the space G of all possible languages may be obtained by taking into account *preference relations*. The predicate „is preferred to" arises in theorems of the following kind (cf. Vennemann 1982b, 1983):

> For all languages L_1, L_2 from G, if L_1 has the property a and L_2 has the property b, then L_1 is preferred to L_2, assuming that for every property c, different from a and b, L_1 has c iff L_2 has c (i.e. assuming „everything else being equal").

As a and b are the only free variables in this formulation, one may think of preference relations as relations between linguistic properties (e.g. those expressible in a theory like T_2): property a is preferred to property b, everything else being equal. Examples of preference laws are:

- (i) Having only affixation for the symbolization of derived categories is preferred to having mutation, either alone or in addition to affixation, everything else being equal.
- (ii) Having agglutination and vowel harmony is preferred to having agglutination and no vowel harmony.
- (iii) Having unidirectional serialization is preferred to not having unidirectional serialization.

Thus, preference laws tell us what is usual and what is rare in the languages of the world. Theories of linguistic preference can be also used in the description of diachronic language changes as well as in the characterization of the differences between existing and possible (but only imaginary) human languages. In order to achieve this goal, however, one should provide for a formal axiomatic characterization of the predicate „is preferred to" — only after that it could be possible to draw a borderline between existing (i.e. preferred) languages and the hypothetical ones. A few examples of possible meaning postulates for the predicate aPb (to be read: the property a is preferred to the property b, everything else being equal) are listed below:

- 1. if aPb and bPc, then aPc
- 2. if aPb, then not bPa
- 3. if aPb and c implies a, then cPb
- 4. if aPb and b implies c, then aPc
- 5. if a implies b (in T_2), then not aPb.

(Further meaning postulates — of the form $aPnon$-a — should depend on the underlying theory T_2).

It is also possible to give a precise formulation of the „local" character of preference laws (i.e. to replace the condition „everything else being equal" by suitable explicit formal conditions). The calculus of preference relations could be also augmented by considering the structures occurring in the algebras Tp_n described previously.

III. A few examples of semantic engineering

By local aspects of semantic engineering I understand the construction of (fragments of) hypothetical human languages, e.g. in order to test linguistic universals or to discuss the possible relationships between grammatical structures and ontological representations.

Following this idea, I've invented, together with one of my friends in Poznań (Poland), a language, called Ningueño (from Spanish „ninguno" — nobody) — cf. Majewicz and Pogonowski 1983. Ningueño is a polysynthetic language (in Boas' sense of the term). We were trying to put as much syntax of Ningueño into its morphology as we could. Thus, each Ningueño word is an 11-place morphological complex; those places are occupied by exponents of separate grammatical categories, e.g.:

- 3 modes of assertion (expressing speaker's own experience, experience reported by others and hypothetical experience, respectively),
- 12 number/person categories marked by suitable confixes,
- category of localization determined on the basis of speaker's accessibility to the fragments of the external world,
- highly developed systems of aspect and modality, and
- categories of focus and circumstantial, whose status, though at first sight comparable to that of lexical information in the SAE languages, is nevertheless a little bit different: the morphs expressing those categories are mutually interchangeable as well as interchangeable with morphs of other categories.

The properties of the categories of focus and circumstantial, together with the fact that the aspectual system of Ningueño is a rather sophisticated one, suggest the hypothesis that Ningueño semantics is „eventistic" rather than „reistic". Fundamental components of the denotations of Ningueño expressions are individual events; such ontological categories as e.g. processes, states, objects and their properties, etc. may be conceived of as (set-theoretical) constructs over the class of all individual events.

Let us now give a few examples of further properties which can be interesting when creating a hypothetical language. They are divided into three groups, corresponding to the following three assumptions about fundamental components of all possible human languages:

- **A. Plane of content.** One can distinguish a bundle of semantic dimensions in any language (e.g. lexical, inflectional, derivational, syntactic, pragmatic, etc.). Each of these dimensions consists of *semantic parameters*. They correspond to sorts of information conveyed by linguistic units (cf. tense, person, number as examples of inflectional parameters) and have a suitable number of values (elementary meanings). Semantic dimensions and parameters are specific for particular languages. There exists a correspondence between semantic parameters of a given natural language and ontological representations presupposed by that language.

- **B. Plane of expression.** Elements of the plane of expression are connected by a network of relationships — *cohesions* of several sorts, serving as a basis for the combinability of simpler units into more complex ones (e.g. combinability of morphs within words, words within phrases and sentences, etc.). Those cohesions are governed by compositionality rules which describe the dependencies between the meaning of a complex expression (i.e. bundle of values of semantic parameters conveyed by that expression) and the meanings of its parts.

- **C. Connections between the two planes.** Each language has its specific set of *modes of expression*, i.e. ways of connecting elements of the above two planes (as e.g. ordering, auxiliaries, affixation, mutation, reduplication, etc.). Those modes are present in the compositionality rules mentioned in B.

We think that all linguists unanimously accept the above, very general in fact, assumptions about possible natural languages. Any model of language which satisfies the mentioned conditions can still have a certain „degree of freedom" in the interpretation of the concepts listed in A.–C. Ideally, such a model should be general enough in order to embrace all possible languages but it should be also possible to impose some additional conditions on it in order to characterize existing languages. Still in other words, one should be able to prove metatheorems of the following form: if our model has a given property F, then it also has a property H. Ultimately, one should be able to prepare a list of properties which, from the point of view of the applications of the model, are most adequate.

Let us point to a few, unusual at first sight, properties which could play some role in the metatheorems of the above kind:

ad. A. ***Languages with a fixed bundle of semantic parameters.*** It is interesting to see how ontological representations proposed as semantical models for particular languages depend on that information which is expressed in a regular way in those languages (cf. the Ningueño example, cf. also the well-known remarks concerning American Indian languages). One can think, in this respect, of simple examples — „nearly" instantiated by the existing languages — as e.g. languages without derivation, with inflection but without derivation (negation of Greenberg's Universal 29), languages which have only lexical (or only grammatical) morphs, as well as of more sophisticated examples — e.g. languages with only one grammatical category or languages without negation.

ad. B. ***Languages with/without particular kinds of cohesion.*** It is not difficult to imagine a language which is totally incorporative, i.e. whose sentences contain always one word only (in fact many polysynthetic languages nearly instantiate this ideal type). In such languages, syntax is simply included in inflection. But one can go even further and think of languages with syntax included in derivation.

Further examples of imaginary languages with/without special sorts of cohesion are: languages without hypotaxis (or without parataxis) or — a difficult case, as it seems — completely idiomatic languages, i.e. languages in which all syntactically complex expressions are idioms. This last case is a sophisticated example of an „ideal" Newspeak with its doublethink all over the place.

ad. C. ***Languages with special conditions imposed on modes of expression.*** An example of a condition of this sort has been provided in the case of Ningueño, where negation is expressed in a rather unusual way — by a reduplication of the corresponding morph. One can think of numerous further examples of this kind. One can also investigate the cases where such conditions influence linguistic theories used for the description of the languages in question — just think of a regularly suppletive language, i.e. a language with a suppletive form in every inflectional paradigm (or, even worse, with irregular distribution of suppletion over its inflectional paradigms). One could bet that such a language would be a nightmare for those who believe in generative morphonology (with its nice formal „underlying structures").

REFERENCES

- Majewicz, Alfred F. and Jerzy Pogonowski. 1983. „Kilka obserwacji na temat struktury języka ningueño i konsekwencje dla teorii lingwistycznej" (Some remarks concerning the structure of Ningueño and the consequences for linguistic theory), *Working Papers of the Institute of Linguistics*, Adam Mickiewicz University, Poznań.

- Vennemann, Theo. 1982a. „What is a linguistic typology?" *Studies in Linguistics*, **2**, 3–44.

- Vennemann, Theo. 1982b. „Universals, preferences, typologies. Definitions and delimitations." Paper presented at the 15th Annual Meeting of the Societas Linguistica Europea, Athens.

- Vennemann, Theo. 1983. „Causality in language change. Theories of linguistic preferences as a basis for linguistic explanations." *Folia Linguistica Historica*, **VI/1**, 5–26.