

The Rationale Of The Idioglossary Technique

VICTOR A. OSWALD, JR.
University of California, Los Angeles

Upon reviewing recent linguistic publications in the field of Machine Translation, one cannot fail to be struck by the extent to which the principle of translation via idioglossaries, which met with wide acceptance at the first conference on Mechanical Translation (Massachusetts Institute of Technology, June 17-20, 1952), has either fallen into abeyance or is being applied unconsciously rather than with a full awareness of its implications. To cite specific instances, the most elaborate recent attempt to resolve lexical problems, Miss Masterson's chunk-lattice-thesaurus procedure¹, is riddled with needless complications introduced by the failure to isolate scientific discourse from the tangle of general language; and Mr. Gode's procedure of translation via Interlingua², which is to date the most wide-sweeping venture toward a solution of problems of word-order and syntax, incorporates, but does not acknowledge, translation by the use of an idioglossary. Evidently the proponents of the idioglossary principle³, instead of concentrating upon a demonstration of how to compile and apply a specific idioglossary, should have elaborated at greater length upon the rationale of the procedure.

It ought to be taken as axiomatic that each and every realm of scientific discourse requires for the expression of ideas peculiar to itself a vocabulary which is, at least in part, peculiar to itself. To be precise, the substantives which "name" specific objects or established concepts in a particular realm of discourse, together with the verbs that "describe" the operations carried out with these

¹ Margaret Masterson, "The Potentialities of a Mechanical Thesaurus," Cambridge Language Research Unit, 1956 (typescript).

² Alexander Gode, "Signal System in Interlingua," *Mechanical Translation*, Vol. II, No. 3 (December, 1955), pp. 55-60.

³ Victor A. Oswald, Jr. and Richard H. Lawson, "An Idioglossary for Mechanical Translation," *Modern Language Forum*, pp. 1-11.

objects or these concepts, constitute a unique jargon intelligible only to the initiated. This is true not only of scientific discourse, but of the communication of trades and handicrafts and hobbies as well. Let us descend for a moment from the austere domain of the sciences and examine a brief, but graphic, sample on the level of the *lingua populi*,

Not many years ago a gifted cartoonist, the late H. T. Webster, capitalized on the existence of popular “private” languages by running off a series of cartoons entitled “They Don’t Talk Our Language.” A hypothetical reconstruction of such a cartoon strip might run somewhat as follows (in all of the panels a husband and wife are represented as sitting at ease in the living-room after dinner).

Panel One

Husband: I checked the rig in the lab this afternoon. The new electrostatic tweeter brings the frequency way up.
Wife: But doesn’t it get out of phase with the old woofer?

Panel Two

Husband: Well, I think we can compensate by putting in a new pre-amp. I’m much more worried about a tendency in the arm toward lateral inertia.
Wife: I thought something was wrong with the tracking, but I assumed it was vertical inertia. Maybe a hydraulic arm would be best, after all.

And so for at least two more panels. The last panel carries the punch line.

Last Panel

Wife: Oh, by the way, May Johnson called just before you came home. She and Phil want us to come over tonight.
Husband: Heavens, no. They’d want to go over the slides they took last week with that new camera of theirs, and I wouldn’t understand a word all night. *They don’t talk our language,*

Please note that this specimen of “hi-fi” jargon displays two quite different kinds of impediments to intelligibility. One is the

use of coined words like “tweeter” and “woofer”, which simply mean nothing outside of hi-fi context. The other—and the more formidable impediment to intelligibility—is the use of what we may call analogical conversions, the adaptation of established words to new and more or less inscrutable uses. The “arm” (a conversion) is, of course, the “pick-up (another conversion) arm” which holds the “cartridge” (a conversion), and the cartridge in turn has in it the “stylus” (a conversion) whose “tracking” (a conversion), that is, its undeviating conformation to the “grooves” (a conversion) of the “disc” (a conversion), is vital to the attainment of “high fidelity” (a conversion).

Fortunately, for those who hope to see mechanical translation realized, the processes of coinage and analogical conversion are used by men everywhere in the world to create the unique terminology of any particular “mystery,” with the happy result that the vital terms of the jargon of, say, a German-speaking biologist, will have unique equivalents in the jargon of an English-speaking biologist. *An idioglossary is simply a bilingual list of such unique equivalents within a given realm of scientific discourse.*

Coinages, whether those of recent origin or those of long-established usage, are likely to have unique equivalents not only, say, between a jargon of English and a jargon of German, but even within the entire scope of these two languages. To roam about for a moment in the territory with which I am most familiar, English linguistic terms of recent origin, such as “phoneme,” “morpheme,” “taxeme,” or “sememe” will turn up in any foreign language with a unique equivalent or no equivalent at all. It would appear to be an unconscionable waste of machine time—no matter how rapidly machine analysis may proceed—to put terms like these through anything so complicated as a chunk-lattice-thesaurus process. On the other hand, analogical conversions probably cannot be latticed because they would frequently prove inextricable from the maze of general language. To put the case on a very homely level again, it would take more ingenious programming than I can imagine to extract from the jargon of plumbing the fact that in this context the foreign language equivalent for “snake” would have to be a word naming a long section of flexible metal used to dislodge foreign matter from a drain pipe; or that “plumber’s friend” would require a word designating a rubber suction cup with a wooden handle. If these examples seem unfair or extreme, consider the case of “set” in modern psychology, or “crotchet” in musicology, or “stop” and “click” and “stress” in phonetics. How

is a lattice going to separate “articulate” as used in anatomy from “articulate” as used in linguistics? One can only feel torn between admiration for the flexibility of the human mind and despair at the thought of trying to reproduce such flexibility in any mechanical device. It will be much easier to operate with a rigid comparison of terms which have unique equivalents.

One cannot but disagree with Miss Masterson that the chunk-lattice-thesaurus process “*feels* like a model of what we ourselves do when we translate”⁴. It must rather be said that what we do, when we ourselves translate, at least when we translate scientific discourse, is to run the foreign language input through a sorting mechanism in our personal translating machine, a process in which we refer all terms from any special realm of discourse to a special repository of target language output equivalents. We do not scan the context for clues to help us in resolving lexical ambiguities. In any special realm of discourse there are no lexical ambiguities. It is for this reason that Mr. Gode can say that, in translating “atrial fibrillation and flutter” into French: “There are but four equations involved which yield the French words “auriculaire, fibrillation, et, flutter.”⁵ Mr. Gode knows in advance that this example is from an “actual medical text,” and he automatically turns to his personal, built-in medical idioglossary for his French equivalents. If he had not done so, he would have found that “atrial” has not one, but at least three possible equivalents: in the domain of archeology “atrial” would require a translation meaning “pertaining to the central main room of an ancient Roman private house”; in architecture it would require an equivalent meaning “pertaining to a courtyard, usually surrounded by colonnades, in front of early Christian or medieval churches.” Only in his medical idioglossary can Mr. Gode find a one-to-one equivalent, meaning “pertaining to an auricle of the heart,” i.e., French “auriculaire.”

I have focused my comments on the work of Miss Masterson and Mr. Gode, not because I do not respect them, but on the contrary, because I think they have made, from the standpoint of linguistics, the most original propositions of recent years. Mr. Gode’s translation via Interlingua appears capable of solving, at one stroke, many of the tiresomely complicated problems of syntactic resolution. His

⁴ *Op. cit.*, p. 9

⁵ *Op. cit.*, p. 56

procedure represents, to be sure, pre-editing on a massive scale; but one cannot fail to be impressed with the fact that it also represents the first demonstration of how mechanical translation might operate on a multilingual scale. If his process is to be mechanized, however, he will have to provide the machine with the idioglossaries he himself uses. Miss Masterson's chunk-lattice-thesaurus operation can be used to save time in programming access to idioglossaries, but it cannot dispense with them. When she chose an Italian article on botany to carry out her demonstration with the "chunk" PIANT-, she was actually working with an idioglossary that was only in part cluttered up with general language terms and idioms.

I do not wish to belabor my point much further, but it might be well to point out that another recent proposition of considerable ingenuity, Mr. Reifler's "Mechanical Determination of the Constituents of German Substantive Compounds,"⁶ will likewise fail to work unless it is used with an idioglossary. I am persuaded that he can mechanically separate the components of compounds like "Säuge-tier" and "Säuge-tier-blume," and that for "Säuge-tier" he can eventually arrive at the English equivalent, "mammal"; but where, save in a *botanical idioglossary*, is one to find the information that a "Säuge-tier-blume" is neither a flower that feeds like a mammal, such as the Venus flytrap, nor some such incredible monstrosity as a flower that suckles its young, but simply a plant that depends upon mammals to carry out the process of pollenization? Or that a "Vogel-blume" is a plant with an analogous dependency on birds, and not a plant that looks like a bird, such as the flower we call "bird of paradise," or, perchance, a flower that flies like a bird?

The existence of a private jargon for every domain of scientific discourse and the need for idioglossaries to translate the jargon of one language into the jargon of another has hitherto seemed to me to be so obvious as not to require blunt and homely demonstration. Perhaps Mr. Lawson and I are to blame for not having made our point more forcefully or more colorfully. In 1951-53 we were intent upon finding out how to construct an idioglossary of no more than 5,000 terms in a foreign language (German) together with the same number of equivalent terms in a target language (English)—the hypothetical limit of a

⁶ Erwin Reifler, "Mechanical Determination of the Constituents of German Substantive Compounds," *Mechanical Translation*, Vol. II, No. 1, (July, 1955), pp. 3-15.

magnetic drum which could serve as a “mechanical dictionary.” To assure optimal precision within such a limited scope we compiled our own idioglossary—to the best of my knowledge, the first idioglossary ever compiled—from articles on brain surgery. By undertaking our task as we did we may have made it appear more formidable than it really is, and we certainly appear to have failed to make the point that the use of idioglossaries provides the only immediately practical solution for the major lexical problems of machine translation.

We did, at least, demonstrate the effectiveness of translation via an idioglossary, and we also made the important discovery that, along with their inevitable private jargon, scientists use a very limited range of the general language at their disposal.

In view of recent engineering developments that have provided programming devices of wider scope and storage devices with more rapid access time and vastly greater capacity than anyone would have dared hypothesize five years ago, there is no reason why idioglossaries cannot now be compiled from dictionaries for such large but essentially self-contained fields as Biology, Chemistry, Physics, or Medicine. The process of mechanical translation would then proceed in no more than four steps. First, the input material would be scanned as a segment from a special field of scientific discourse and the output material held “in storage.” This step would isolate those items of the input (primarily nouns and verbs, but also a certain proportion of adverbs and adjectives) which are peculiar to the particular jargon in question. Second, the input would be scanned for “general language” items of the same categories, a not very formidable proceeding in view of the fact that scientists characteristically restrict themselves to a rather narrow range of expression. Thirdly, the input material would be scanned for language constants: articles, pronouns, prepositions, and conjunctions which are used with the same frequency and in the same way in *any* segment of language. Finally, the input material would be submitted to whatever programming might be required to resolve the syntactical patterns of the input language into those of the output language.

I do not wish to sound more sanguine than I really am. The small-scale experiment with syntactical rearrangement that Fletcher and I carried out in 1950-51⁷, and the small-scale experiment that

⁷ Victor A. Oswald, Jr. and Stuart L. Fletcher, “Proposals for the Mechanical Resolution of German Syntax Patterns,” *Modern Language Forum*, Vol. XXXVI, No. 3-4 (Sept.-Dec., 1951), pp. 1-24.

Lawson and I performed with our brain-surgery idioglossary in 1952-53, appear to indicate that the maximal equivalence to be attained mechanically—and at that a very crude equivalence—could not be higher than about 80 per cent. I cannot pretend to know what results would be obtained from large-scale operations. I can only reiterate that the procedure outlined above appears to be the only wholly mechanical procedure likely to attain results of any kind. It is time for someone to produce the facilities to arrange for an investigation on a scale large enough to provide definitive results.