What a Universal Semantic Interlingua Can Do

Pieter A.M. Seuren

Department of General Linguistics and Dialectology
Nijmegen University
The Netherlands

In 1960, Bar Hillel presented an argument showing the non-feasibility of machine translation on grounds of lexical ambiguities. I will now present a further and more sophisticated argument based on lexical and non-lexical semantic distinctions.

Sentences do not wear their meanings on their faces: they are grammatical transforms of meaning representations. Some dialects of English, for example, have sentences like:

(1) I never stole no money

which does not mean what it literally says: there never was a time that I did not steal some money, i.e. I have always stolen money, or I am an inveterate thief. It means instead there never was a time that I stole any money, i.e. I am an honest person. It is the grammar of this English dialect that is responsible for this: this grammar contains a rule prescribing copying of the negation for any subsequent existential quantifier.

Or take the Akuapem (West Africa, Akan) sentence:

(2) Me-de aburow mi-gu msu-m

I-poured corn I-go water-into

But this sentence does not mean that the speaker went into the water as he poured the corn. It means that the corn went into the water as a result of being poured into it (to be washed): I poured corn into the water. Here it is the grammar of Akuapem that prescribes copying of the subject (i.c. I) with any following (serial) verb after the rule-governed deletion of the original semantic subject (corn).

It is not known why language likes to hide her semantic face, but it does. For that reason (most) linguists distinguish two levels of sentence analysis: surface structure (SS) and semantic analysis (SA) (also sometimes called Logical Form or LF). SAs are represented in the language of Predicate Calculus, which seems suitable for SA-representation.

The only serious attempt at a formal generative mapping from SAs to SSs is Semantic Syntax (SeSyn; Seuren 1996). (In Chomsky grammar small-scale and largely unsuccessful attempts have been made to map SSs onto SAs.) SeSyn now has a formal generative SA-SS-mapping (i.e. a syntax) for English, French, Dutch, German and Turkish. These grammars cover the central constructions in these languages to a degree of precision not achieved so far. The rule systems that do the job are extremely compact and highly uniform across languages.

For an adequate syntactic treatment it has proved necessary to assume SAs that are, though semantically adequate, not identical for all languages (universal): SAs are to some extent still language-specific. The question is therefore: Is it possible to construct a semantic representation system that is common to all languages and may serve as a bartering counter for formal (automatič) translation procedures, a Universal Semantic Interlingua (USI)? The answer to this question is, for the moment, negative. Translation is a deep process that requires full cognitive integration. And even if cognition as a whole could be modelled as software, there would be no place for a USI. The reason is that each language makes a large number of idiosyncratic semantic distinctions, which depend on cognitive decisions. To capture these in a USI would require the incorporation of cognition as a whole.

The idiosyncrasies are lexical and non-lexical. Lexical idiosyncrasies are omnipresent. They depend on culture as much as on chance. For example, the English verb pluck applies primarily to the removal of feathers from fowl. Then its use is extended to pick (fruit, flowers, weeds), snatch, pick up (courage), swindle, pinch, twang (strings of musical instrument). It is unlikely that any other language has a precise equivalent for all

these uses. A choice must therefore be made to get a proper translation equivalent in a given context. This choice can only be mediated through full cognitive comprehension of the text at issue. (Probability measures based on surrounding words may help but are unreliable.) This was, essentially, Bar-Hillel's 1960 argument.

Non-lexically, one finds, for example, important distinctions in the meanings and uses of verbal tenses. The English perfect, for example, expresses (roughly) (a) existential quantification over past events, with the implication that the set of events can still be augmented (there have been one or more occasions such that ...); (b) a resultative meaning relevant to the present situation (the car has been stolen), and (c), with stative verbs and progressives, continuation from the past into the present (we have known for a long time that ...). The English perfect tense is unfit for reference to a definite closed past: !*Yesterday I have fallen.

The Dutch perfect is syntactically identical to its English counterpart, but differs semantically roughly as follows. A context where a set of past events can still be augmented is allowed but not required. It is thus perfectly natural to say Gisteren ben ik gevallen (lit. yesterday I have fallen). The resultative meaning (b) is also valid in Dutch. Meaning (c), however, does not exist in Dutch, which, like most languages, uses the simple present for such cases. But then the Dutch perfect tense has a meaning unknown in English: it is typically used to take distance from a narrative and to present a proposition as a comment on a series of past events or a past situation. Typically, at the end of a story, an English-speaking child will ask Did it really happen? whereas a Dutch-speaking child will say Is het echt gebeurd? (has it really happened?). The Dutch child does this because it takes distance from the narrative and is asking for comment on it.

Other than Dutch, English or Russian, Malay is hardly interested in temporal information. Usually there is no tense indicator: the context has to supply the information. If necessary, futuricity can be expressed (akan), or resultative past (sudah). But there is no tense system.

The means that, apart from the rough propositional content, each language presents the speaker with a semantic questionnaire that must be filled in before a sentence can be formulated. The answers are obtained from cognition, and often a speaker has a choice between this or that mode of presentation. This is why a formal translation procedure must incorporate the totality of integrated cognitive comprehension.

Even so, there is room for further integration of language-specific SAs. It proves possible, for restricted groups of languages (e.g. the European languages) to set up certain semantic categories which are formally convertible into the various language-specific SAs, which then simply feed the grammars so that the corresponding sentences are generated. An example is provided by counterfactuals, with or without the possibility operator. English I would go, for example, is a present counterfactual, and I would have gone is a past counterfactual. With the possibility operator: I could go and I could have gone. But note that the German translation of the last sentence is Ich hätte gehen können, which has a very different grammatical structure. Likewise for the French and Dutch equivalents.

It will be shown how a common semantic representation can be converted into English, French, German and Dutch SAs, producing the correct grammatical forms for the expression of these four semantic categories. Other examples will be discussed as well, e.g. the well-known French Il vient de partir, translated into English as He has just left. Or English He likes to swim, which corresponds to the Dutch Hij zwemt graag (i.e. he swims gladly).

It thus seems possible and fruitful to consider the construction of a Deeper Semantic Analysis (DSA) which is not a USI but helps in many cases where existing treatments have so far failed (or have achieved success only in an ad hoc way). Though it has been shown that DSA construction is useful in relation to groups of languages, it is not known what the limits of this method are.

References

1. Bar-Hillel, Yehoshua 1960: The present state of automatic translation of languages. Advances in Computers 1:92-16
2. Schachter, Paul 1974: A non-transformational account of serial verbs. Studies in African Linguistics, Supplemen 5:253-270
3. Seuren, Pieter A.M. 1996: Semantic Syntax. Blackwell Oxford.