

Time to eat peaches:

Language-specific Information in Interlingua

Stephen Helmreich

Computing Research Laboratory

New Mexico State University

shehnrei@crl.nmsu.edu

This presentation continues the discussion, begun at the pre-workshop, on language-specific information in an interlingual approach to machine translation. At that workshop I argued that language-specific information must be present in the IL representation, since any bit of information about the source language text could be relevant for translation and therefore must be present in the IL.

In a paper in this workshop, David Farwell shows that this argument fails (if I may paraphrase his position) because it is not the language-specific data that is relevant, but rather the information that is communicated through that data, and that this information can be expressed in IL terms, without reference to the language-specific information. That is, what is important for translation is not (for example) that a source language sentence used a passive-construction, but rather what that construction communicated (different focus, de-emphasis on agent, etc.)

However, even given the validity of this argument. I argue here that, ultimately, language-specific information will still need to be available in the IL representation. First, though it may be possible to represent in IL in a non-language-specific way all the information that is expressed by the form of the language-specific source language text, it is compactly represented in the a language-specific way. Thus it seems likely that it would not be feasible to actually do this processing of making non-semantic language-specific information available in the IL.

Second, there is the problem of “retranslation,” where the source language text represents already a translation from the target language, or where there already exists translations into the target language. In such cases, the machine translation system should be able to take into consideration this information. Here, for example is a “clipping” (undated) from the *New Yorker*:

[From "What the Maid Saw," by Yasutaka Tsutsui, translated from the Japanese by Adam Kabat]

Katsumi had come across the image of the peach in a poem by an American poet.

The years are taking their toll.

It's time to roll up the cuffs of my pants—

Time to part my hair from the back—

Time to eat peaches.
I put on my white flannel pants and walked on the beach.
I heard mermaids singing to each other—
Mermaids who would never sing out to me.

Here is the original T. S. Eliot (from *The Love Song of J. Alfred Prufrock*):

I grow old ... I grow old...
I shall wear the bottoms of my trousers rolled.

Shall I part my hair behind? Do I dare to eat a peach?
I shall wear white flannel trousers, and walk upon the beach.
I have heard the mermaids singing, each to each.

I do not think that they will sing to me.

I argue that in order to get a retranslation to turn out as the original, there must be language-specific information (namely the original form) available in the IL.

One of the advantages of an IL method of translation is that there is a level of representation of the source text which can be used as a basis for inferencing. Both the “world knowledge” and the source text are stored in logical forms that support that inferencing. But, ultimately, this data base of world knowledge will contain information that is language specific. In order to reason adequately, it must have access to to the language specific form of the source language text.