

# On Representing Language-specific Information in Interlingua

David Farwell

New Mexico State University  
Computing Research Laboratory  
david@crl.nmsu.edu

## Introduction

In this paper I wish to defend the position that linguistic information about the source text need not, indeed should not, be included in the interlingual representation of that text. After reviewing a few preliminary assumptions, I will present four cases in which it appears that the representation of language particular information related to the source text is inescapable. In each case, I attempt to show that the actual language particular information, per se, is not central to the authors intent and, therefore, need not be represented. While no doubt there are further cases that will not be discussed, it is my hope that they would be addressed along lines similar to those set forth here.

## Preliminaries

An interlingual representation system must represent that information about a (arbitrary) source language (SL) communication which is necessary for producing an appropriate corresponding target language (TL) communication. The key point is that the representation is of a communication (a sequence of communicative acts) in an arbitrary language and that it serves as a basis for producing an appropriate corresponding communication (sequence of communicative acts) in an arbitrary language. Beyond this there are no further restrictions on what an interlingua must represent.

Given the above requirement, it is unreasonable to assume that any interlingual representation system whose goal is to represent crosslinguistic correspondences of form will be successful. That is to say, it would be a mistake to represent, say, the fact that a given SL expression takes the form of a passive construction under the assumption that there will be some particular corresponding construction, say, a reflexive or a middle voice or a passive, etc., in an arbitrarily selected target language. This is what I understand as being suggested by the “crosslinguistic ontological approach” described in the call for papers (Helmreich, 1997). The reason that such an approach is doomed is that with language we can apply the principle of form follows function and that it is the latter that underlies translation. Since there are many functions associated with any given form, there are, therefore, unpredictably many correspondences between any two languages, let alone among arbitrarily many languages.

It is likely rather that an interlingual representation system will have to represent who said what to whom along with how, where, when, and why it was said. That is, it will have to represent the

(sequence of) communicative acts performed by the SL speaker in such a way as to serve as a point of departure for performing a corresponding (sequence of) communicative acts using the TL.

## **Language-specific Information**

The question of whether or not, or to what extent, “language specific” information must be represented by an interlingual system hinges in part in what we define as language specific information. For the purpose of this discussion, it is any aspect of the linguistic system that is peculiar to that language. Clearly, this includes the language’s phonology (e.g., the 20 or so vocalics of English, the schwa-ing of unstressed vowels, and so on), morphology (e.g., in English, V+ing can act as N, V+s indicates 3rd singular subject present tense, and so on), lexicon (e.g., in English “watch” may be a V referring to various particular types of activity or states of affairs or a N referring to various types of objects or situations, and so on), syntax (e.g., in English, Subject precedes Predicate which, in turn, precedes Objects in unmarked finite clauses, pronominal datives precede nominal direct objects but follow pronominal direct objects, and so on), semantics (e.g., in English, “watch” under one of its nominal meanings is used to refer to “portable personal clock” but not other sorts of clocks, and so on) and pragmatics (e.g., in English, “hello” is used on meeting another while “good-bye” on leaving another’s company, “hello” is more formal while “hi” is more casual, and so on). These are all facts about English and the question is whether such information need to be represented by an interlingual system and, if so, what information needs to be represented.

In what follows, I will attempt to defend the position that such information does not need to be represented by the interlingual system since it is not (normally) part of what the SL speaker intends to communicate, why the speaker wishes to communicate it, nor is it instrumental to how the speaker is attempting to communicate it. The discussion will focus on four cases which, at least at first blush, appear to require the representation of language specific information.

### **Case 1: Metalinguistic discussions**

The first case involves the translation of texts in which the SL author talks about the source language. For instance, suppose the SL speaker says:

*"Ostrich" is used to refer to a kind of large, long-legged, flightless bird.*

For the sake of concreteness, we can translate this into Spanish:

*Se utiliza "ostrich" para referir a un tipo de pájaro grande, de patas largas y que no puede volar.*

Here the form of the English word *Ostrich* must be represented in the interlingua although probably not as a form per se but rather as a description of that form. If, for instance, the TL had been Russian or Arabic, it is quite likely that the translation of "*ostrich*" would be the Cyrillic or Arabic transliteration of the English word. What is key here as far as the author of the SL text is concerned is not that “ostrich” has the particular form that it has but rather that form is typically used to refer to the particular type of thing it is used to refer to.

Note that yet another alternative translation might have been:

*Se utiliza "avestruz" para referir a un tipo de pájaro grande, de patas largas y que no puede volar.*

Here the SL form itself has been translated. This might be done if the SL author was not actually interested in communicating what the English form is that is used to refer to ostriches so much as in showing that the person who uttered the original sentence knew which form was used to refer to which type of animal. The point here is that even communicative acts about the language of communication are not necessarily primarily about the language. Thus, it is the SL author's INTENTION that needs translating rather than the SL author's words per se.

## **Case 2: Phonological effect**

The second case involves the translation of texts that are exploiting the sounds of the source language to some end. Such would be the case, for instance, with tongue twisters which are designed to trip speakers up as they pronounce the text. Suppose you need to translate into Spanish:

*She sells sea shells down by the sea shore ...*

The tongue twister has semantic content but communicating that content to the addressee is not the central intent of the speaker in uttering the expression. If that were the case then something along the lines of:

*Vende ella conchas a lado del mar ...*

would be an appropriate translation. Normally, however, it is not.

This tongue twister is crucially dependent on the alternation of the consonants "s" and "sh". But in attempting to translate it into Spanish, this particular alternation cannot be used because Spanish does not have the latter phoneme in its inventory (although in some dialects it is a variant of "ch" and in others it is a variant of "s").

Does this mean the tongue twister cannot be translated into Spanish? Perhaps so. But in the end it is unlikely that either the specific phoneme alternation nor the specific semantic content of the tongue twister are crucial. I expect that, in most cases, some TL tongue twister such as (Bravo-Villasante, 1976, p. 106):

*Si den sierras asierran cien cipreses, seiscientas sierras asierran siescientos cipreses*

(If a hundred saws saw a hundred cypresses, six hundred saws saw six hundred cypresses)

will be a reasonable translation even though the phonemic alternation, in Castilian, is between "s" and "th" and the tongue twister is about sawing cypresses rather than selling sea shells. The central information that needs to be represented then is that the SL speaker wishes to show his or her capacity to utter successfully a tongue twister, a text which is intended to cause the reader to mispronounce it.

### Case 3: Word play-double entendre

The third case concerns texts in which the SL author intentionally exploits some multiple function of a given form. These are difficult for most translators since they often hinge on truly idiosyncratic properties of the SL. Consider Hamlet's chat with a gravedigger upon returning from England (Shakespeare, 1600, Act 5, Scene 1).

*HAMLET:* .... *Whose  
grave's this, sirrah?*

*First Clown:* *Mine, sir. ....*

*HAMLET:* *I think it be thine, indeed; for thou liest in't.*

A brief dialog ensues in which this pun on “lie” (to occupy vs. to prevaricate) is repeated several times.

The question is whether it is possible to represent this in a language non-specific way. In translating this exchange into, say, Spanish, the translator will be faced with some tough choices. The best situation is that in the target language there is a pair of homonyms meaning “to occupy” and “to prevaricate”. The interlingual representation would need only to represent the fact the utterance is intentionally ambiguous, that the two ambiguous readings are:

BECAUSE YOU ARE (LOCATED) IN IT  
BECAUSE YOU PREVARICATE IN IT

and, possibly, that the ambiguity arises from the multiple meanings of the predicate. However, in general, it cannot be expected that such a pair of homonyms will be found in the TL or, for that matter, that there is a single phrasal construction in the TL that can be used to express that desired ambiguity. Certainly this is the case for Spanish.

One option, then, is to attempt to translate the complex semantics of the expression by “unpacking” the ambiguity. This will most likely take the form of a conjoined TL expression such as:

*... como tanto se tiende Vd. como miente adentro (de ella)*

(... since you both stretch out and lie in it)

The key here would be to attempt to keep the dialog as clever and entertaining as possible assuming that is the central goal of the playwright.

Yet another approach would be to select one or the other of the two readings as central to the author's intent at that point in the dialog and to translate that reading. For instance, since the dialog continues to focus on who is to be buried in the grave, the translator might reasonably focus on the “occupy” reading as central and translate the expression as:

*... como se tiende Vd. adentro (de ella)*

(... since you are stretching out in it)

Note, that in conceding that there may not be an expression in the target language which can be used to fulfill all the functions that some original source language expression was used to fulfill, I am not conceding that information about the source language must be represented in the interlingua. I am conceding rather that information may be gained or lost in the translation. But this is a position with which I can live since it is probably true of every translation that has ever been done.

#### **Case 4: Word play**

The following case was first described at this workshop last year by Steve Helmreich (Helmreich, 1996). It involves a Spanish translation of a Tom and Jerry cartoon in which Tom approaches a chicken on a playground and asks:

-- *What are you doing?*

The chicken replies:

-- *I'm gonna cross the playground.*

Tom then asks why and the chicken responds:

-- *To get to the other slide!*

The joke, of course, depends on recognizing its likeness to the well known joke:

-- *Why did the chicken cross the road?*

-- *To get to the other side.*

which is invited by the fact that *slide* rhymes with *side* rhyme. If it did not, this joke would not even be cute, let alone funny.

The Spanish translation provided by the newspaper (*El Paso Times*, 11 Mar 1994?) was:

-- *¿Qué estás haciendo?*

-- *Voy a cruzar el patio.*

-- *¿Porqué?*

-- *Para ir al otro tobogán.*

Unfortunately, it is highly unlikely this cartoon can be felicitously translated into Spanish (or any other language) as is. The first prerequisite for success is knowledge of the standard “chicken crossed the road” joke, which would be surprising for any particular TL speech community (it may not even be known by all English speaking groups). The second prerequisite is that playgrounds both exist and serve essentially the same purpose (areas for children to play) and physical layout (jungle gyms, seesaws, slides) in the TL speech community. The third prerequisite is that

there is an appropriate correlate for “slide” such that it rhymes with the correlate for “side” in the TL. That all these prerequisites obtain is highly unlikely.

What aspects of the input then does the interlingua need to represent? First, it must represent the fact the authors goal is to entertain (make laugh) the addressee by telling a joke. Second, it should represent that the joke is a “play on” an old standard (joke) based on rhyme (sound similarity). Third, it might be pointed out that the old standard is funny (if indeed it is funny) because the punch line is a statement of the obvious. That is, it is funny because the addressee is expecting the non-obvious and clever but is fooled by being provided with the obvious and unclever. Beyond this, of course, we would expect to represent the semantics of the SL text (even though it may well turn out that this will not play an immediate role in determining the semantics of the TL text). We might also represent the semantics of the old standard to allow an analogy, and relevant expectations, to be set up. Finally, it might even be useful to represent the phonologies (or graphologies) of the rhyming words (even though it is highly unlikely these will play a role in determining the text).

## Conclusion

I have presented four cases in which it appears at first blush that one must represent language particular information in the interlingua. I have attempted to show in each case how that language particular information need not, indeed should not, be represented. Rather, what needs to be represented is the goal of the author in exploiting the particular form of the source language that were exploited and the way the author exploited them. This is the appropriate information for formulating a corresponding act using the TL. In the end, language particular information is just that, language particular. It cannot be used in formulating the TL act because it is simply not part of the TL communicative system.

## References

- Shakespeare, W. 1600. *Hamlet*. Electronic version at: The Complete Works of William Shakespeare, <http://the-tech.mit.edu/Shakespeare/works.html>.
- Bravo-Villasante, C. 1976. *Una, Dola, Tela, Católa: el Libro del Folklore Infantil*. Valladolid: Editorial Miñon.
- Helmreich, S., 1997. Workshop on Interlinguas at MT-Summit '97. Electronic communication, 1 Jul 1997  
Electronic version at: <http://crl.nmsu.edu/Events/FWOI/call.html>
- Helmreich, S. 1996. What Is an Interlingua and What Information should it Contain? Paper presented at the AMTA SIG-IL Pre-workshop on Interlinguas and Interlingual Approaches to Machine Translation, Montreal, Canada, October 1, 1996. Electronic version at: <http://crl.nmsu.edu/Events/FWOI/PreWorkshop/preworkshop.papers.html>.