ISA OR NOT ISA: THE INTERLINGUAL DILEMMA FOR MACHINE TRANSLATION

FLORENCE REEDER
The MITRE Corporation[1] / George Mason University
freeder@mitre.org

ABSTRACT
Developing a system that accurately produces a good translation between human languages is the goal of Machine Translation (MT) systems. This problem requires a sophisticated lexicon which associates words with "meanings". This paper looks at MT as a bizarre problem (Kercel, 1999). First we describe MT and typical solutions while identifying the notion of an interlingual, or language-independent, representation. We present challenges and issues in developing this representation, particularly the notion of language-independence. Finally, we recommend a potentially more effective approach to developing lexical components based on the bizarre system model.

INTRODUCTION

Natural language processing (NLP) systems represent the ultimate in bizarre systems. Designing systems to understand, utilize or transform the languages we speak into actions, information and other languages is a complex and difficult job. Machine translation (MT) is a type of NLP system that can be thought of as bizarre. In this paper, we describe the field of MT and look at the "holy grail" of the field – developing a language-independent representation, or interlingua. We illustrate with a few current interlingual representations. Afterwards, we show how MT can be considered a bizarre system process and discuss a new model of interlingua based on the bizarre system model.

MACHINE TRANSLATION

Machine translation systems translate between human, natural languages. This section describes Machine Translation (MT) by first showing its origin. A brief depiction of the processing involved is followed by a demonstration of the difficulties in translation which help to characterize it as a bizarre process. Weaver (1947) describes the translation process as a kind of noisy channel decryption process: the French document is really an English document coded in French.[2] For instance, the literal translation of **omelette du fromage** is **omelette with cheese**. So far, the process is simple enough. Yet, as anyone fluent in more than one language knows, the process of translating between

---

[1] The views expressed in this paper are those of the author and do not reflect the policy of the MITRE Corporation.
[2] "When I look at an article in Russian, I say: 'This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode.'" (W. Weaver, March 1947)

languages is more than the substitution of one word for another. Consider the example **Parlez vous Français?** Literally translated, it is **Speak you-formal French? –** understandable, but not quite good English. More analysis of the source or better generation of the target is necessary for quality translation.

Traditionally, translation systems have been composed of several components or processing layers. The process is frequently described in terms of a pyramid (Vauquois, 1968), **Fig. 1**. To show the processing, a system begins at the left side and analyzes the source language – the analysis is more complex as we ascend the pyramid. At the transfer point in any given system, it changes the results of the analysis into a format suitable for the generation of the target language. The system then descends the pyramid into the target language. Interlingua represents a pure translation process where analysis yields a language-independent representation from which generation can be done. The components which address these different levels include: a bilingual lexicon (or list of words), a grammar of the source and target languages, a set of transfer rules which translate structures between the source and target representations.

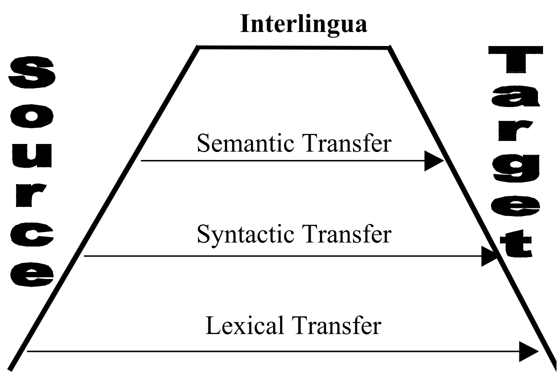

Figure 1 The Machine Translation Pyramid

To demonstrate the difficulties inherent in translation and why it resists traditional processing, consider the word **bank**. It can mean a financial institution, a side of a river, the act of counting on an event, or driving around a corner. Translating this word into foreign languages is not straightforward, as other languages have different words corresponding to the different meanings of **bank**. Because of this many-to-many mapping of words, word for word substitution is insufficient, therefore, a lexical transfer system translates poorly.

The next level of representation, the syntactic level, provides some disambiguation. For instance, determining if **bank** is a noun or verb contributes sufficient evidence to cut the number of readings in half. The grammatical categorization of words is accomplished through stochastic models of language or through rule-based analysis. The problem of accurate translation remains as we still have multiple meanings for the nominal **bank** and the verb **bank**.

Further up the pyramid, the semantic level, information about possible meanings of words contributes to the translation process. To continue with the banking example, it is possible to differentiate **bank** by looking at neighboring words (a, b). Semantic disambiguation can be addressed through probabilistic translation models utilizing n-grams (word pairs or triplets). Yet this is also not optimal as neighboring words do not necessarily provide sufficient information to assist translation in as in (c). In all three examples (a-c), the words **bank** and **left** occur in the sentence, yet the meanings are very different.

(a) First National Bank is on the left.

(b) The Left Bank is the site of the first national parade.

(c) I left the bank yesterday.

Another way to try to pick the correct word sense is to develop a large ontology of concepts and meanings to support the translation process.

As we have moved up the translation pyramid, the necessary amount of target and source background knowledge has increased as has the preparation cost. Also, we still do not have sufficient representations for the transformation of language at the conceptual level. To say **I am banking on the horse coming in** implies a concept which literal translation may not be able to account for. In this case, building an interlingual representation is an appropriate choice.

INTERLINGUAL REPRESENTATIONS

An interlingua is a "language independent" representation. While it represents long-term cost savings in translation[3] systems, it also demands much more complex up-front knowledge acquisition and engineering. Even more problematic is the notion of what it means to be a true interlingual representation. Consider the following possible interlingual representations shown in **Table 1**.[4] Each of these is an approach to interlingua, but all of them are language-dependent. The degree to which they are language dependent reflects how they fit into their domain.

<table>
<tr>
<td><code>COMPANY: [NAME: N [text]</code>
<code>HEADQUARTRS: H [office]</code>
<code>SUBSIDIARIES: S [set: company]</code>
<code>EMPLOYEES: E [set: human]</code>
<code>SALES: V [currency] …]</code>
(Farwell & Helmreich, 2000)</td>
<td><code>A:give-information+price+room</code>
<code>(room-type=double,</code>
<code>price=(quantity=150,</code>
<code>currency=dollar,</code>
<code>per-unit=night))</code>
(Levin, et al., 2000)</td>
</tr>
<tr>
<td colspan="2"><code>Activity: [ACTION]</code>
<code>Participants : [agent: AGENT] [objects: OBJ1, OBJ2]</code>
<code>Applic_cond : [reachable (OBJ1)] [have(AGENT, OBJ2]</code>
<code>Preparatory_spec: [get (AGENT, OBJ1)]</code>
<code>Termination_cond: [contact (OBJ1, OBJ2)]</code>
<code>Post_assertions: [contact (OBJ1, OBJ2)]</code>
<code>Path, duration, motion, force</code>
<code>Manner : [MANNER]</code>
(Kipper & Palmer, 2000)</td>
</tr>
</table>

Table 1 : Interlingual Representation Examples

[3] If M is the number of source languages and N is the number of target languages, one only needs to develop M+N instead of M*N systems.

[4] Note that each of these supports a different notion or function.

Kipper and Palmer (2000), for instance, developed an interlingua for providing a multilingual interface to a simulation. Their interlingua reflects the needs of the physical environment and the planning system which operates in it. While reasonably language independent (if one grants that the primitives are language independent), the amount of work necessary for this very limited domain was substantial. Levin et al (2000) measure the notion of language-independence when they test for inter-coder agreement.[5] While they had interlingual coders in different countries, all of them were English speakers and they related difficulty in getting non-English speakers to code the same information accurately. Having described the processing necessary for MT and illustrated the difficulty inherent in it, we will now argue that MT is a bizarre process that requires a new notion of interlingua to effectively reach the next level of processing.

MT AS A BIZARRE PROCESS

Past approaches to MT system development have relied on traditional knowledge representation ideas, based on notions that language is reducible for processing. Therefore, they break down a paragraph into sentences, sentences into phrases, phrases into syntactic parts, syntactic parts into words and words into meaning. Each of these can be accomplished through a rule-based or statistically-based method which is based on the notion that words are independent. Yet, it is our argument that this is not a realistic view for MT systems. The sentiment is not new in the MT community.[6] We will now show how MT systems rightly fit into the category of bizarre systems, an argument for a new approach to representing information necessary to build them.

MT qualifies as bizarre system (Kercel, 1999) because of the partially entangled behavior of words. For instance, meanings can be compositional: **cheese fries** literally interpreted is either **fries with cheese on them** or **cheese is a substance that can be fried**. At the same time, we have many wordings that are not compositional, defying a logical representation. The saying **He bought the farm** has a non-compositional, or idiomatic, meaning that is different from the sum of the parts. Yet, there is a reasonable and rational causality to idiomatic language usage. The phrase **thumb-rule** stems from the days where the **Rule of Thumb** determined the size of a stick with which a man could beat his wife.[7] So the entanglement of words is one reason for considering MT under the bizarre system model.

If MT can be viewed in the bizarre model, then we can apply aspects of bizarre systems to improve MT systems – such as a record of the past; prediction of the future; inferential elements; causal elements in our ontology; and anticipatory behavior. A record of the past can be gleaned from corpora as in

---

[5] There is a standing joke in the MT community, that an interlingua is just English with upper case letters, e.g., **dog** ⇔ DOG ⇔ **chien**.

[6] "… as to the problem of mechanical translation, I frankly am afraid the boundaries of words in different languages are too vague … to make any quasimechanical translation scheme very hopeful." (N. Weiner, April 1947)

[7] No greater than the width of his thumb.

statistical language models. Ontological representations with causal elements under the bizarre model should contribute sufficient evidence of relationships to support language independence. To achieve the next level of capability, we require a new way of looking at computing translations because we are modeling a process that is bizarre. The question, then, becomes how to utilize the best of the traditional models and combine it with a bizarre scheme? Or should we even try this? Part of the answer may be found through evidential reasoning models[8] and psycholinguistic research. The end goal is a representation which supports the combination of multiple pieces of evidence that contribute to the "meaning" of a concept in a language independent way.

EVIDENTIAL LEXICONS – A BIZARRE MODEL?

Because many biological systems are bizarre, we look to psycholinguistics for ideas in modeling the bizarre nature of MT. Psycholinguistic models suggest why we select the words and grammatical constructs we use. Language is one of our primary means of acquiring information – through talk, through reading, through words. Language is not the form of the representation, but words and the combination of these words are part of our knowledge structure. The notion that we remember a gist of what was said in place of the exact words means that there is something else going on. Professional translators show this often when they translate sentences according to general meaning instead of for exact words.

Connectionists (e.g., Winograd & Flores, 1986; Cummins, 1989) model this associations of words with neural networks. Taking from this tradition, Jurafsky (1996) treats lexical organization with probabilistic or evidential models. Words, and therefore some aspects of knowledge, are organized into associated structures in the mind.[9] The following experiment describes that this association exists and that it is probabilistic in nature:

> "As expected, people pressed the button faster when recognizing *ant*, which is related to *bug,* than when recognizing *sew*, which is unrelated. Surprisingly, people were just as primed to recognize the word *spy*, which is, of course, related to *bug*, but only to the meaning that makes no sense in the context." (Pinker, 1994)

Our recognition and usage are sensitive to stronger associations, so a word is more easily recognized if it is related to a previously stated word. New information is learned by incorporating it into the existing network or strengthening or renewing current connections. This knowledge structure can then be processed via following the links and connections. The links have strengths which indicate their associative force. Koestler (1964) is one of the first to show the biological basis for this, indicating these are useful phenomena to incorporate into a lexical model.

In describing this possible model, we realize that space does not permit us to develop the idea fully. If we assume our lexical model to be evidential with

[8] Schum (1994) describes many of these.

[9] Note that Chomsky () is the first to see this relationship of pre-wired language structures and it is from Chomsky that Pinker derives much of his work.

both traditional and bizarre data, the core evidence for a translation of a word comes from dictionaries. These are the evidential equivalent of jurisprudence representing the work of people whose job it is to know how words are used and translated. Yet these dictionaries contain ambiguity in both meaning and usage of words. Therefore, we would augment a dictionary entry with additional information which is learned from multiple evidence sources both traditional and bizarre: associated words (n-gram measures); part of speech measures; syntactic measures; morphological information; ontology information. In selecting a translation, the evidence is combined to find the most likely translation – based on a number of sources.

In a basic representation, the evidence could be combined using Bayesian networks. This follows the Jurafsky model from psycholinguistics. We have not divorced ourselves from the need to use language to label nodes in the network. We are, however, more language neutral than previous models because of the ability to combine evidence sources. For instance, the Japanese distinguish the different forms of rice (cooked, uncooked or in the field). One could easily envision a Japanese lexicon where the form of the rice has a great enough weight to allow for the correct word to be chosen. More advanced models could rely on other evidential models (Schum, 1994), although this is an area for further exploration.

CONCLUSIONS

We have described machine translation as a bizarre system especially in the area of interlingual representation. We have very briefly presented a possible computational solution through evidential reasoning. We would like to thank our reviewers for their astute comments.

REFERENCES

Cummins, R., 1989., *Meaning and Mental Representation*, MIT Press.

Farwell, D. and Helmreich, S., 2000, An Interlingual-based Approach to Reference Resolution, In *Proceedings of the Workshop for Applied Interlinguas: Practical Applications of Interlingual Approaches to NLP*, ANLP/NAACL-2000, Seattle Washington.

Kercel, S., 1999, Embedded Intelligence: Bizarre Systems for the Steel Industry, Presentation to AISE – Sensors Meeting.

Kipper, K. and Palmer, M., 2000, Representation of Actions as an Interlingua, In *Proceedings of the Workshop for Applied Interlinguas: Practical Applications of Interlingual Approaches to NLP*, ANLP/NAACL-2000, Seattle Washington.

Koestler, A, 1964, *The Act of Creation*, The Penguin Group.

Levin, L., Gates, D., Lavie, A., Pianesi, F., Wallace, D., Watanabe, T. and Woszczyna, M., 2000, Evaluation of a Practical Interlingua for Task-Oriented Dialogue, In *Proceedings of the Workshop for Applied Interlinguas: Practical Applications of Interlingual Approaches to NLP*. ANLP/NAACL-2000, Seattle Washington.

Pinker, S., 1994, *The Language Instinct—How the Mind Creates Language,* Harper Perennial.

Schum, D. A., 1994, *Evidential Foundations of Probabilistic Reasoning.,* John Wiley & Sons, Inc.

Vauquois, B., 1968, A survey of formal grammars and algorithms for recognition and transformation in machine translation, *IFIP Congress-68* (Edinburgh), pp. 254-260.

Weaver, W., 1947, Letter to Norbert Wiener, March 4, 1947, (Rockefeller Foundation Archives).

Weiner, N., 1947, Letter to Warren Weaver, April 30, 1947, (Rockefeller Foundation Archives).

Winograd, T., and Flores, F. , 1986, *Understanding Computers and Cognition*, Addison-Wesley.