[*MT Summit III*, Washington, DC, USA, 1-4 July 1991]

Capturing Language-Specific Semantic Distinctions in Interlingua-Based MT

James Barnett, Inderjeet Mani, Elaine Rich,
Chinatsu Aone, Kevin Knight, Juan Carlos Martinez

Microelectronics and Computer Technology Corporation (MCC)
3500 West Balcones Center Drive
Austin, Texas 78759

Abstract

We describe an interlingua-based approach to machine translation, in which a DRS representation of the source text is used as the interlingua representation. A target DRS is then created and used to construct the target text. We describe several advantages of this level of representation. We also argue that problems of translation mismatch and divergence should properly bo viewed not as translation problems per *se* but rather as generation problems, although the source text can be used to guide the target generator. The system we have built relics exclusively on monolingual linguistic descriptions that are also, for the most part, bidirectional.

1 System Goals

We describe an approach to machine translation that is motivated by the following goals. First, we want to maximize the ability of the system to produce correct and natural sounding translations even when translation mismatches and divergences occur. Second, we seek to minimize the cost of adding new languages to the system. Although the second of our two goals is clear, we should say a bit more about what we mean by the first. There are many cases in which the natural translation of one language into another has a very different form than the form in the original. These cases can be divided into two categories:

- Translation divergences
- Translation mismatches

Translation divergences [Dorr, 90] arise when the same information is conveyed in the source and target texts, but the structures of the sentences are different. An example of this is the English/French pair, “I swam across the river” and “J'ai traversé la rivière à la nage” [Mounin 63].

Translation mismatches [Kameyama, et. al, 91] occur when there are actually differences in the information that is conveyed. If the source text is vague or ambiguous in ways that are not allowed in the target language, then the translation process must *add information* by first making the best possible guess about the intent of the source text and then rendering that intent into the target text. This problem comes up, for example, in translating from English into Japanese, where it is necessary to add politeness information. If, on the other hand, the target language allows vagueness or ambiguity that is not. allowed in the source, it may be necessary, in creating a natural translation, to *throw away some information* that is present in the source text and not render it completely into the target text. An example of this occurs in translating from English into Japanese, since it is often necessary to throw away number information that is mandatory in English and not in Japanese. Although syntactic differences between languages cause many cases of translation mismatch, they are not responsible for all of them. Lexical differences between languages can also cause both kinds of translation mismatch. For example, the English word “fish” has two translations into Spanish: “pez” and “pescado”, depending on whether the fish is still in its natural state or caught and suitable as food. So to go from English to Spanish, it is necessary to add information. To go from Spanish to English, we must throw information away,

For an MT system to be able to add this kind of information, it must be possible to derive the required information from the discourse context and a model of the domain that is being discussed. The MT architecture that we describe makes this possible by creating a meaning-based representation of the content of the source text.

2 Two Issues that Overlap

Translation divergences and translation mismatches share a common problem. When they occur, straightforward transfer from source structures to target ones fails to produce the desired translation. But when we look more closely at where the problems are and what kinds of solutions might exist, we see that there are actually two separate issues that happen to overlap:

- Generation of text that sounds natural and says what needs to be said.
- Use of the source text form as a guide to generating matching target forms.

In transfer-based MT systems, these two issues are treated in the same way, using transfer rules that map from specific structures (at one or more levels) in the source language to the desired structures in the target. But transfer systems have two important, deficiencies from the point of view of the goals we described in Section 1. The first is that, since they require separate rules for each language pair, the cost of adding each new language to the system is high. The second is that they cannot deal effectively with many cases of translation mismatch in which it. is necessary to add information. They are limited by the fact that they cannot reason about, the subject matter, and so cannot derive information that is not present linguistically in the source text. For example, although a transfer system could use local type checking to choose the correct translation for "fish" in the sentence, “John ate fish for dinner”, most transfer systems cannot distinguish between the two sentences, “There were five fish in the lake” and “There were five fish in the casserole” in a way that would enable them to choose “pez” in the first case and “pescado” in the second.

The only way to achieve the stated goals, then, is to move toward an architecture that exploits a meaning-based interlingual form. Analysis of the source text creates the interlingua expressions; then generation from the interlingua creates the target text. We will say more about the form of the interlingua that we are using in Section 4. But the details of the interlingua are not important from the point of view of the two issues that we mentioned above. As soon as we commit to the use of a semantic interlingua, then we accept that one component of the resulting MT system is a generation system that takes as its input a description of some message to be conveyed, along with an indication of which language the message is to be conveyed in. This system must first decide exactly what to say (since there may be several things that are equivalent in illocutionary force) and then decide how to say it.

The key insight here is that this is exactly the same problem that must be solved in any system that uses an NL generation system to create linguistic representations of expressions in a meaning language. The fact that the expressions in the meaning language came from linguistic expressions in a source language is irrelevant except for a few special cases in which the form of the source language expressions can provide help in making generation decisions. This insight is not profound. In fact it is obvious. But it is usually overlooked,

Most of the problems that have been described in the MT literature as problems of translation divergence and mismatch are really not problems of translation at all; they are primarily problems of generation that have been finessed in many non-MT-based generation efforts by a careful design of the meaning representation language that guarantees that it maps straightforwardly to the appropriate linguistic structures in the one language for which the generation system has been designed to work. The real difference between generation in the context of MT and generation in most monolingual contexts is that MT forces us to find a systematic solution to the problem of choosing what to say and how to say it. It has this effect because the alternative, namely choosing a meaning representation that is already very close to the desired linguistic form, is not available since such a representation is necessarily different for different languages.

To make this clearer, let's consider a typical case of translation divergence (taken from [Dorr, 90])- The English sentence, “John usually goes home” is naturally translated into Spanish as “Juan suele ir a casa” (literally, “John tends to go home.”). In English, the main event is a going; in Spanish it is a tending. In a good semantic representation system, it can be either of these, both of them, or perhaps some third thing. Suppose that the origin of the fact that these two sentences are conveying were not source text in English or Spanish. Suppose instead that a reasoning program had concluded it based on an analysis of a database of individual events. Now we want a generation program to apply to this derived fact and render it in English (or Spanish, or whatever). If the semantic representation focuses on going, the job of English generation is relatively straightforward, while generation in Spanish will require more work. If the semantic representation focuses on the habituality, then generation into Spanish will be fairly direct and English will be more work. But in either case, the task is not to translate from one language to another but to translate from an internal semantic representation to a natural linguistic expression in a particular language. And a big part of that task is to take a complex meaning representation expression and break it apart into pieces whose size, focus, emphasis, and so forth correspond to a natural form of expression in the chosen language.

This is a difficult problem, particularly because it often happens that a language will offer several different ways to say the same thing. For example, in English one can say, “John tends to go home.” In addition, any meaning representation language that is powerful enough to represent the content of any substantial corpus of texts probably allows for many (provably) equivalent ways of saying the same thing. So a general solution to this problem requires a way to move from one set of assertions in the representation language to other sets that are equivalent, a way to compute linguistic renderings for all of them, and a way to rank those renderings so that the most natural one, in a particular context, can be chosen. Much work remains to be done on this. But by characterizing it as a generation problem rather than a translation one, it is easier to see what kinds of knowledge must be applied to solve it.

There are, however, some cases in which generation should be affected by the fact that it is being done as part of translation. We will return to this issue in Section 6.

3 System Architecture

In this section, we describe the architecture of an MT system that is based on the interlingual framework we have just described. In many ways, this system is similar to other interlingua-based systems, such as KBMT [Goodman and Nirenburg, 89]. It exploits no language-pair-specific transfer rules. In contrast to some other interlingua-based systems (e.g., [Uchida and Zhu, 89]), it is designed to be used with an interlingua that is as independent as possible of the set of languages that it will be used to support.

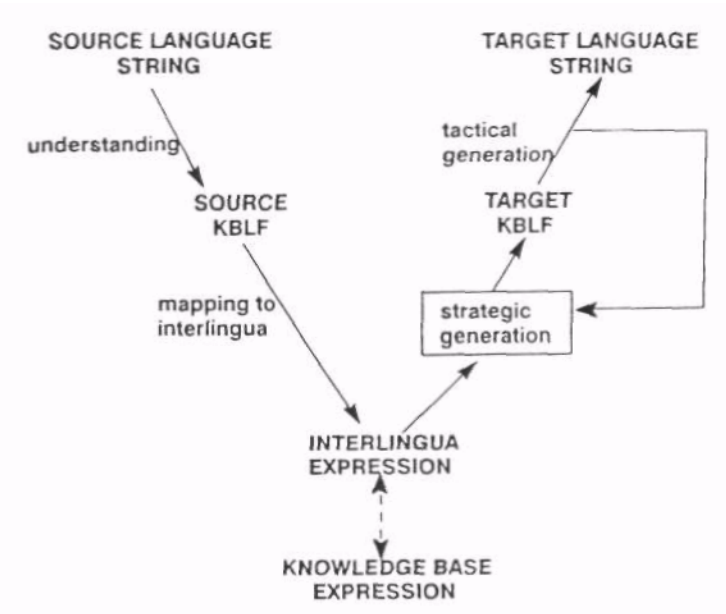

Figure 1: An Interlingua-Based Architecture for MT

A schematic description of the system is shown in Figure 1. All of the representations in the figure, except the source and target language strings, are described in terms that are drawn from a knowledge base (KB) that describes the domain(s) of discourse. So the system relies heavily on dictionaries, for both source and target languages, that define mappings between words and knowledge base objects. In addition to providing a common set of terms that enable meanings to be defined, this backend knowledge base is important because it provides the ability to reason about meanings and thus the ability to add to the target text information that was omitted from the source. We will assume that all the KB-based representations can be treated as sets of logical assertions (although they can of course be implemented in a variety of ways, including the frame-based system [Crawford, 90] that we are using).

To translate a sentence, this system must do the following things:

- Map the source sentence into an internal representation of what was said. (Ideally, this entire process would happen for units larger than sentences, but for now, sentences are processed one at at time, although a discourse model is built as the text progresses and can be used as necessary during the processing of individual sentences.) We call this the source *DRS;* it is isomorphic to the Discourse Representation Structures described in [Kamp, 84] and [Heim, 82], except that its terms are taken from the backend knowledge base rather than from the words of the source language.
- Map the source DRS into the interlingua. Because we have not been able to discover any additional information that can be contributed at this stage, this mapping is currently the identity. So the interlingua representation is equivalent to the source DRS, both in form and in content. Thus it contains assertions corresponding to exactly what was said in the source.
- Map the interlingua expression to a target DRS. At this point, decisions about what to say in the target text must be made. Some assertions in the interlingua may be dropped. Some new assertions may be added. Some groups of assertions may be replaced by others that are equivalent with respect to the KB but more appropriate as a basis for a natural sounding text in the target language.
- Map the target DRS into a target string. Unfortunately, it is often not possible to enforce a clean separation between these last two (strategic and tactical) generation steps, so it may be necessary for them to interact and to inform each other, as shown in by the loop in the figure.

We have implemented an English/Spanish MT system using this architecture, with the KBNL system [Barnett et al, 90], [Barnett et al, 91a] as its linguistic basis. KBNL has three main components: an understanding system, Lucy [Wittenburg and Barnett, 88], [Rich and Luperfoy, 88], which performs the initial mapping from source text to source DRS; a generation system, Koko [Barnett and Mani, 90], which performs the interlingua to target string mapping; and a lexical acquisition component, Luke [Wroblewski and Rich, 88], [Knight et al, 89], which assists in building the dictionaries that define the required mappings between linguistic structures and knowledge base objects. These systems exploit a generic KB interface [Barnett et al, 91c], so they can run on any KB that contains the necessary domain knowledge.

The understanding half of the process we have described here has already received a great deal of attention in both the natural language and the MT literatures. We will not say anything more about it here except to mention that a major design goal for Lucy has been to build it on a KB interface that enables Lucy's performance to improve as the power of the underlying KB grows.

The generation half is much less well-understood, but is, unfortunately, in the general case outside the scope of this paper. We will, however, in Section 6 talk about specific ways in which being embedded in an MT system affects the generation process.

4 The Significance of the DRS Level of Representation

The DRS level of representation has several important properties.

First, we note that the source DRS, the interlingua expression, and the target DRS are all encoded in the same representation language. This language is closely tied to the language of the backend knowledge base. In particular, the distinctions that are made in that KB are precisely the distinctions that will be able to be made in the DRSs and the interlingua. Ideally the KB will be designed with a broad view in mind of the class of ways that various languages carve up the world. Of course this is never completely possible, but this approach still works, even if the KB has been designed with only one language clearly in mind. In that case, the mappings between that language and the KB will be straightforward. When a new language is added, there are two choices. Either new objects can be added to the KB (along with the necessary relationships between them and the old entities) and the new language can be mapped to them, or the KB can be left alone and more complex mappings, that define the new language in terms of the old KB, can be provided. But in either case, it will not be necessary to change the lexicon for the initial language. The necessary relationships are defined in terms of objects in the KB rather than between pairs of linguistic entities, with the advantage that the power of the KB reasoning system is available and can be used in particular discourse contexts as necessary.

The language of the DRSs and the interlingua is not, however, identical to the language of the backend KB. In fact, the DRS and interlingua representations are, in general, both more and less vague than the corresponding KB expression would be. They are less vague because they make commitments about the specific objects that are being explicitly mentioned, the order in which the mentionings occur, and the explicit assertions that are made (as distinct from all the other assertions that are entailed, in the KB, by the ones that are explicitly mentioned). The DRS level is more vague than the KB level in some other ways though. Temporal relationships are described linguistically (i.e., with respect to speech time and reference time), but there is no absolute time frame. Furthermore, no attempt is made to decompose linguistic quantifiers, such as "most" into primitives that can be supported by a general reasoning system. In the rare cases where it is necessary to create KB expressions from these linguistic level objects in order to enable the KB to reason about their meaning, approximate translations are available. But the generation process can still begin at the linguistic (DRS) level and does not have to undo the *ad hoc* transformations. In this respect, generation within MT differs from generation that begins with a goal that was derived from a nonlinguistic process. It gets to start a bit ahead.

A second important point is that the common DRS level makes language-pair specific rules unnecessary in this system. Since the interlingua expression is the same as the source DRS, there is a sense in which the first step in generation corresponds to a transfer from the source DRS to the target. But, with the exception of the cases that will be described below, all of which can be handled by looking at solely monolingual knowledge sources, this process is pure generation. It does not care what language the DRS expressions came from. So our goal of minimizing the cost of adding new languages to the MT system by avoiding the need for language-pair-specific knowledge sources has been achieved.

Finally, we should point out that the use of parallel DRS structures for both source and target languages makes it possible to use reversible linguistic descriptions that can support both understanding and generation. Our system is based on such representations [Barnett and Mam, 90], [Aone, 91], But it is important to note that the use of reversible linguistic descriptions alone is inadequate for MT since both understanding and generation rely on a set of preference rules in addition to the rules that define the set of legal mappings, and preferences are not in general symmetric [van Noord, 90]. See [Barnett and Mani, 9l] for a discussion of this issue.

```
x:  (Fish x)
    (measure  x NaturalUnit 5)
y: [(Plate y)
    (*F measure y NaturalUnit 1)]
    (Equal y previous-referant2)
z: [(State z)
    (*F precedes  z  Now)]
    (supported-by  z x y)
```

Figure 2: DRS for "There were five fish on the plate."

Figure 2 shows the DRS that corresponds to the En glish sentence, "There were five fish on the plate". The marker *F indicates that the assertion it is attached to was derived from a syntactic distinction that is forced in the source language (English). Adding these markers is one of the two changes we have made to our grammars to enable them to support MT. The importance of this marker will be described in Section 6.

The second change is what produced the brackets around the first two assertions on z. As we will see below, it helps the target generator to be able to tell when a group of assertions came from a single lexical item in the source text. So such groups are bracketed.

But notice that both of these changes have been made to the grammars of individual languages. They do not require any language-pair-specific information,

English: "I like Mary."

```
x:  (speaker x)
y:  (Mary y)
z: [(liking-state z)
    (*F overlaps z Now)]
    (agent z x)
    (object-of-liking z y)
```

Spanish: "Maria me gusta a mi"

```
x:   (speaker x)
y:   (Mary y)
z: [(pleasing-state z)
    (*F overlaps z Now)]
    (theme z y)
    (object-of-pleasing z x)
```

Figure 3: English and Spanish DRSs

Figure 3 shows the DRSs for the English sentence, "I like Mary", and the corresponding Spanish sentence, "Maria me gusta a mi", assuming that the backend

KB provides explicit representations for both liking and pleasing. If it allows only one, then the necessary mappings to it must appear in the lexicons and the two DRSs will then look alike. But notice that if that is the case (for example, suppose only liking is represented), then what is stored in the lexicon for the other language is not a translation into the first language; it is simply a mapping into a KB representation that is not particularly natural. So if the target language were changed, it would still not be necessary to change the source lexicon. All that would be required is to specify the mapping between the KB form and the new target. Assuming the two distinct representations however, the English generator (with the help of a KB that knows the required relationships) must be able to take a description of a pleasing event and decide to describe it as a liking, while the Spanish generator must be able to do the reverse.[1]

5 Markedness and Lexical Differences

When we consider the problem of lexical differences between languages, the concept of *markedness* is important. Consider the Spanish words "pez" and "pescado", There is no English word corresponding to "pescado", which refers to caught fish or fish as food. Furthermore, it is infelicitous in Spanish to use "pez" to refer to caught fish (just as it is infelicitous to use "cow" instead of "beef" in English when referring to the flesh of the animal). We therefore say that "pescado" is the *marked* term, while "pez" is unmarked[2]. In general, the marked member of a pair indicates the presence of some distinguishing property and the marked term should be used whenever that property is known to be present. The unmarked term carries no information about the presence or absence of the property and is the default choice.

Not all hyponymous pairs of words correspond to marked/unmarked distinctions. For example, Volkswagens are a subclass of automobiles, but the use of "automobile" to refer to a Volkswagen is perfectly acceptable (while the use of "cow" to refer to beef is not). In the absence of markedness distinctions, lexical choice is relatively free, but markedness introduces a kind of 'lexical forcing' into the system. The Spanish lexicon forces us to use "pescado" where it is appropriate, just as English forces us to use "beef".

As we have emphasized throughout this paper, these problems are not particular to MT and must be solved in monolingual generation systems. However, the problems are brought to our attention more forcefully in a MT context, particularly when one language contains a marked/unmarked pair that is missing in the other. We will consider this problem in more detail in the next section.

[1] This has not yet been implemented in the KBNL system.

[2] The marked/unmarked distinction that we ate exploiting here is analogous to the more traditional one that is used in morphology [Jakobson, 66], It is also related to the concept of blocking, cf. [Andrews, 90].

6 Explicit Influence of the Source on Target Generation

The generation problem can be described very generally as: "Given a communication goal (in the case of MT, the interlingua expression), find the best rendering of it. in the target language." This view suggests three kinds of generation problems: (1) deciding when not to give an explicit rendering of some assertion(s) in the communication goal; (2) deciding when to add assertions to the original goal and render them explicitly; and (3) how to make the stylistic choices that are necessary when the target grammar and lexicon license more than one rendering of the communication goal.

In the specific case in which the communication goal was generated by understanding a source language text (rather than for example, as the result of some problem solving program), it turns out that the source text, as well as the source language itself (as described in its grammar and lexicon) can be used to provide some advice to the generator in dealing with all three of these kinds of problems, the first two of which correspond to the problem of translation mismatch and the last of which often occurs in cases of translation divergence.

First consider the case in which it is necessary to add information. This can happen both when the syntax of the target language forces some distinction to be made that was not made in the source (and so the necessary facts are probably missing from the interlingua expression) and when the target lexicon contains marked forms that disable the use of the corresponding unmarked forms whenever the more specific marked forms are applicable. The source language text cannot help here since what it can contribute is assumed already to be in the interlingua expression. But when the problem is that the target lexicon forces information to be added, it is sometimes possible to save some work by looking at the source lexicon.

This happens when the generation system is looking for a lexical item to cover a set of assertions, and it finds one that cannot be used if some more specialized marked form can be used. For example, suppose that a Japanese generator tries to find a word for just the assertion

```
(rice-substance x)
```

It will find the unmarked form "kome", which is both a specific word meaning "raw rice" as well as the unmarked form that applies to rice when it is not known (or relevant) whether or not it is cooked. But "kome" cannot be used if, from context, it is clear that the rice is cooked. (In this case, the word "gohan" must be used.) So, in general, the generator must check all the marked forms and see if any of their additional requirements can be derived from the discourse context. If they can, then the corresponding form is used. If not, the unmarked form is chosen. But this procedure can be simplified when generation is being done as part of translation. Before checking to see whether any of the marked forms can be shown to be applicable, we check to see whether the same marked forms were available in the source language. If they were available but they were not used, then the speaker chose not to make the distinction, and so we can bypass the step of trying to apply them for the target language.

Next consider the case in which the generator should drop assertions that are present in the source DRS (and thus in the interlingua). This happens whenever two conditions both occur:

- The assertions are not necessary to convey the intended meaning, and
- the assertions are not forced by the target grammar or lexicon.

Unnecessary assertions can show up in any semantic representation, but in the specific case of MT they are usually there because they were forced by the source grammar or lexicon. For example, number assertions show up in English, while status distinctions appear in Japanese. As we showed in Figure 2, we anticipate this problem in the case of syntactically forced assertions by using the source grammar to mark them as forced. In the simplest scheme, the target generator would drop all such assertions unless they are also forced by the target grammar. This is of course too simplistic, because it is possible that a distinction that happens to be forced is also important, but deciding this requires reasoning about, the current discourse context and is not yet implemented.

Unnecessary assertions can be forced not just by the grammar of the source language but also by its lexicon, For example, the Japanese word "gohan" will introduce the assertions:

```
[(rice-substance x)
(cooked x)]
```

Yet in most contexts, the fact that the rice is cooked should be dropped by the English generator. In the case of syntactically forced assertions, we anticipate this problem in the source grammar and mark assertions as forced in the interlingua. It does not make sense to do this for lexically forced assertions, however, because, unlike syntactic ones, there is not a small, fixed set of forced distinctions. Every language carves up the world differently in its lexicon. There is no guarantee that assertions that originated from a single lexical item in the source will be grouped together to form a single lexical item in the target. (In fact, exactly what happens in the case of translation divergence is that they are chopped up differently.) So it does not make sense to look ahead and try to decide during source analysis which assertions will turn out to be superfluous during target generation. Too much depends on the target language itself. Instead, the generator must appeal to its general strategy for deciding what to say. In particular, it must choose an appropriate referring expression for each object in the current context. But it can make use of the source lexicon as one source of information.

Specifically, when it is given a set of assertions about a particular object and asked to generate a referring expression, it will first try to find a single lexical item. If it finds a word that matches but is too general, it must decide whether to drop the additional facts or generate modifiers to account for them. At this step, one factor it can consider is that if the additional assertions were derived from the same lexical item as the more general fact (as in the case of the fact that the rice is cooked), then it is more likely that they are present because of the structure of the source language. Further, if they were derived from a lexical item that was marked along a dimension that corresponds to the extra assertions, it is even more likely that they are unnecessary in the target.

Finally, we consider the case in which the generator has several options to choose from. In general, this problem is very difficult. It is often necessary to augment the lexicon with phrasal templates that describe the preferred ways of expressing concepts that have many allowable renderings [Jacobs, 85], [Hovy, 88]. But even here, it. is possible that the form of the source text, as well as its lexicon, can help.

When the generation system is given a set of assertions in the meaning language and it is asked to find a linguistic realization of them, the first thing it will usually try is to find a single lexical item that covers them. For example, given the set. of assertions

```
(bovine x)
(young x)
```

an English generator should normally produce the word "calf". But suppose that these assertions were derived from the predicate object of the Spanish sentence, "El becerro es la cria de la vaca"(The calf is a young cow). We do not want to generate the English sentence, "The calf is a calf." The general rule here is that we block the use of a single lexical item to cover multiple assertions if the assertions arose from multiple lexical items in the source text and the source text had a single lexical item that could have been used hut was not. This last check is important since, for example, we do not want to block translating the English expression, "cooked rice" into the single Japanese word "gohan".

In other cases, it may be possible to extract from the source text stylistic measures, in addition to content. Then those measures can be used to guide the target generator. So if the source text contained, for example, words that are marked in the lexicon as being informal, technical, formal, or whatever, then the words that are correspondingly marked in the target lexicon can be selected. We have not yet dealt with this issue, but see [DiMarco and Hirst, 90] for a discussion of the use of stylistic preferences in translation.

In [Barnett et al, 91b] we examine these all of issues in more detail and we describe an algorithm for lexical choice in MT that began as a general-purpose procedure for NL generation from a meaning representation (based on the algorithm described in [Calder et al, 89]). But onto it have been superimposed the particular places, such as the ones we have just described, where facts that can be derived from linguistic properties of the source text can be used to provide guidance to the generator.

7 Current Status

Lucy, Koko, and Luke were originally implemented for English. Lucy can map a wide range of English sentences into the appropriate DRS form. Koko is capable of tactical generation from a DRS, but, except in a few very simple cases, it does not yet do the earlier strategic step in which decisions about what to say are made. We have built a grammar of Spanish, so both Lucy and

Koko can now work in Spanish. Our treatment of forced and unforced propositions is somewhat incomplete, but sufficient to handle distinctions such as *"pez"/ "pescado"*. Our acquisition tool Luke, used to associate words with KB concepts, has also been extended to Spanish. The entire system is written in Common Lisp and runs on Symbolics machines and Sun workstations. It is being used in our text retrieval system (Scan) to enable multilingual text retrieval.

8 Summary

In this paper we have described our approach to the construction of a reversible, interlingua-based MT system that exploits a DRS level of representation that serves as both the interlingua and the description of the literal content of both the source and target texts. This approach is based on the observation that many of the problems that, arise in MT systems (besides the source text ambiguity problem that has been thoroughly discussed elsewhere) can best be thought of not as transfer (i.e., mismatch or divergence) problems but as issues in NL generation. They must be solved using powerful generation techniques that include the ability to assign preferences to the alternative ways of expressing a set of assertions in the meaning representation language. There are, however, some ways in which the source text, as well as the source language, can be used effectively to influence this process. But, importantly, all of them can be done by appeal to the monolingual descriptions of the text and the language (i.e., the lexicon and the grammar), so no language-pair specific transfer rules are required.

References

[Andrews, 90] A. Andrews, "Unification and Morphological Blocking", *Natural Language and Linguistic Theory, 8(4),* November 1990, pp. 507-558.

[Aone, 91] C. Aone, "Classifiers and an Interlingua for Multilingual Machine Translation", *Proceedings of the International Conference on Current Issues in Computational Linguistics,* Penang, Malaysia, 1991.

[Barnett et al, 90] J. Barnett, K. Knight, I. Mani, and E. Rich, "Knowledge and Natural Language Processing", *Communications of the ACM,* August, 1990.

[Barnett et al, 91a] J. Barnett, D. D'Souza, K. Knight, I. Mani, P. Martin, E. Rich, C. Aone, J. Blevins, W, Bohrer, S. Luper-Foy, J.C. Martinez, and D. Newman, "Knowledge-Based Natural Language Processing: the KBNL System", MCC Technical Report ACT-NL-123-91, 1991.

[Barnett et al, 91b] J. Barnett, I. Mani, P. Martin, and E. Rich, "Reversible Machine Translation: What to Do When the Languages Don't Line Up", *Proceedings of the ACL Workshop on Reversible Grammars* in *Natural Language Processing,* Berkeley, 1991.

[Barnett et al, 91c] J. Barnett, E. Rich, and D. Wroblewski, "A Functional Interface to a Knowledge Base for Use by a Natural Language Processing System", MCC Technical Report ACT-NL-OI9-91, 1991.

[Barnett and Mani, 90] J. Barnett and I. Mani, "Using Bidirectional Semantic Rules for Generation", *Proceedings of the Fifth International Workshop on Natural Language Generation,* pp.47-53, Dawson, Pa., 3-6 June, 1990.

[Barnett and Mani, 9l] J. Barnett and I. Mani, "Shared Preferences", *Proceedings of the ACL Workshop on Reversible Grammars in Natural Language Processing,* Berkeley, 1991.

[Bresnan, 82] J. Bresnan, "The Mental Representation of Grammatical Relations", MIT Press, 1982.

[Calder et al, 89] J. Calder, M. Reape, and H. Zeevat, "An Algorithm for Generation in Unification Categorial Grammar", *Proceedings of the 4th Conference of the European Chapter of the ACL,* pp. 233-240, Manchester, 10-12 April, 1989.

[Crawford, 90] J. Crawford, "Access-Limited Logic-A Language for Knowledge Representation", Ph.D. Thesis, The University of Texas as Austin, 1990.

[DiMarco and Hirst, 90] C. DiMarco and G. Hirst, "Accounting for Style in Machine Translation", in *Proceedings of the Third International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Language,* Austin, Texas, 1990.

[Dorr, 90] B. Dorr, "Solving Thematic Divergences in Machine Translation", *Proceedings of the 28th Annual Meeting of the ACL,* Pittsburgh, 1990.

[Goodman and Nirenburg, 89] K. Goodman and S. Nirenburg, "KBMT-89 Project Report", Center for Machine Translation, Carnegie Mellon University, 1989.

[Heim, 82] I. Heim, "The Semantics of Definite and Indefinite Noun Phrases", *University of Massachusetts,* Ph.D. Dissertation, 1982.

[Hovy, 88] E. H. Hovy, "Generating Language with a Phrasal Lexicon", in D. D. McDonald and L. Bolc, eds., *Natural Language Generation Systems,* Springer-Verlag, New York, 1988.

[Jacobs, 85] P. S. Jacobs, "PHRED: a generator for natural language interfaces", *Computational Linguistics, 11(4), 1985.*

[Jakobson, 66] R. Jakobson, "Zur Struktur des Russischen Verbums", in Hamp, Householder, and Austerlitz, eds,, *Readings in Linguistics, II,* Chicago: The University of Chicago Press, 1966.

[Kameyama, et. al, 9l] M. Kameyama, R. Ochitani, and S. Peters, "Resolving Translation Mismatches with Information Flow", *Proceedings of the 29th Annual Meeting of the ACL,* Berkeley, 1991.

[Kamp, 84] H. Kamp, "A Theory of Truth and Semantic Representation", in M. Groenendijk, J. Janssen, and M. Stokhoff, eds., *Formal Methods in the Study of Language,* Dordrecht: Foris, 1984.

[Knight et al, 89] K. Knight, E. Rich, and D. Wroblewski, “Integrating Language Acquisition and Knowledge Acquisition”, *Proceedings of Workshop on Automated Tools for Lexical Acquisition, IJCAI-89,* 1989.

[Mounin 63] G. Mounin, *Les Problèmes Théoriques de la Traduction,* Gallimard, 1963.

[Rich and Luperfoy, 88] E. Rich and S. Luperfoy, “An Architecture for Anaphora Resolution”, *Proceedings of the Second Conference on Applied Natural Language Processing,* Austin, 9-12 February, 1988,

[Uchida and Zhu, 89] H. Uchida and M. Zhu, “An Interlingua for Multilingual Machine Translation”, *Shizengengoshori, 1989. 5. 19.,* Japan, 1989.

[van Noord, 90] G. Van Noord, “Reversible Unification-Based Machine Translation”, in *Proceedings of COLING-90,* Helsinki, 1990.

(Wittenburg and Barnett, 88] K. Wittenburg and J. Barnett, “Canonical Representation in NLP System Design: A Critical Evaluation”, *Proceedings of the Second ACL Conference on Applied Natural Language Processing,* 1988.

[Wroblewski and Rich, 88] D. Wroblewski and E. Rich, “LUKE: An Experiment in the Early Integration of Natural Language Processing”, *Proceedings of the Second ACL Conference on Applied Natural Language Processing,* 1988.