Machine Translation of Spatial Expressions: Defining the Relation between an Interlingua and a Knowledge Representation System*

Bonnie J. Dorr and **Clare R. Voss**
Department of Computer Science
A.V. Williams Building
University of Maryland
College Park, MD 20742
{bonnie,voss}@cs.umd.edu

Abstract: In this paper we present one aspect of our research on machine translation (MT): defining the relation between the *interlingua* (IL) and a knowledge representation (KR) within an MT system. Our interest lies in the translation of natural language (NL) sentences where the "message" contains a spatial relation — in particular, where the sentence conveys information about the location or path of physical entities in the real, physical world. We explore several arguments for clarifying the source of constraints on the particular IL structures needed to translate these sentences. This paper develops one approach to defining these constraints and building an MT system where the IL structures designed to satisfy these constraints may be tested. In this way, we have begun to address one of the basic issues in MT research, providing independent justification for the IL itself.

Keywords: Natural Language Processing, Knowledge Representation, Machine Translation, Lexical Knowledge, Spatial Knowledge

1 Introduction

In this paper we present one aspect of our research on machine translation (MT): defining the relation between the *interlingua* (IL) and a knowledge representation (KR) system within an MT system called LEXITRAN. Our interest lies in the translation of natural language (NL) sentences where the "message" contains a spatial relation — in particular, where the sentence conveys information about the location or path of physical entities in the real, physical world.

We will be looking at sentences such as:

(1) Die Kirche liegt im Süden der Stadt

which may have either of the following interpretations:

(2) (i) The church lies in the south of the city
(ii) The church lies to the south of the city

Here the location of a **church** is ambiguous with respect to the **city**: the church may lie in the southern part of the city, within the city limits, or it may lie south of the city. The need to translate such sentences accurately presents a clear case of where general as well as specific real world knowledge should assist in eliminating inappropriate translations.

For example, had the sentence above been about a **mountain** lying *im Süden* of the **city**, the MT system should be able to use the *default* knowledge in a KR system that mountains typically are physical entities distinct and external to cities, to produce only the second, "outside the city" translation of the sentence.[1] The MT system should also be able, when the information is available, to take advantage of specific facts to override the default reasoning. For example, if the KR system contains a fact, or is able to infer from other facts, that the particular mountain named in the sentence is in a city, the MT system should only produce the first (*i.e.*, "inside the city") translation of the sentence.[2]

What is intriguing about this translation is that the ambiguity concerns such a conceptually clear distinction (*i.e.*, lying inside of, *vs.* outside of, a geographical region), yet this conceptual distinction is not "lexicalized", *i.e.*, it is not readily noticeable in the words of the sentence. This observation has led us to ask how the encoding of spatial relations — such as being lexicalized or not — should result in different formal representations for these relations in the components of an MT system, and what the interdependencies of these encodings should be.

In producing the correct translation of a sentence, an MT system may need to have access to information about a spatial relation that is only logically implicit, *i.e.*, not lexicalized — as was the case with the mountain/city sentence above. We will argue in this paper that, for a particular sentence, its logically implicit relations should be kept distinct from its lexicalized relations. In the sections to follow, we will explain how this position is reflected in our system design by maintaining the main components — the syntax, the interlingua, and the KR system — as separate modules.

In this context we address the question of what the relation between the interlingua (IL) and KR means. In general terms, our discussion will focus on the specific ways a KR system can assist the MT system in filtering out incorrect translations. In particular, of all the IL

*This research has been partially supported by the National Science Foundation under grant IRI-9120788, by the DARPA Basic Research Program under grant N00014-92-J-1929, by the Army Research Office under contract DAAL03-91-C-0034 through Batelle Corporation, and by the Army Research Institute under contract MDA-903-92-R-0035 through Microelectronics and Design, Inc.

[1] We are assuming a non-interactive MT system here, not a system that has recourse to asking a person on-line during the translation process which of the possible meanings was intended or is most likely.

[2] One of our reviewers noted, for example, that towns like Edinburgh have mountains in them.

structures built during the analysis phase, where each IL structure represents a distinct interpretation of the input sentence, we ask how the KR system eliminates those interpretations that are incorrect or highly unlikely before the generation phase begins.

The issues we will be examining are:

- What primitives in the domain of spatial relations must be in both the IL and KR components?
- What structures are passed between the IL and KR?
- What relations will the KR need to infer, *i.e.*, that are not in the IL?

The system design of LEXITRAN reflects two research issues. First, we wish to capture the insights of Dorr (1993) and those of Nirenburg *et al.* (1992) in the same model because they are complementary: Dorr has streamlined the syntax-IL mapping and Nirenburg *et al.* (1992) has demonstrated the advantages of including a taxonomic, or ontological, knowledge base (KB) in a MT system. No MT system currently exists that combines these two approaches or is able to make the claims of both Dorr and Nirenburg *et al.*[3]

Second, MT theory has not yet defined the issues surrounding how IL and KR formalisms should be related, either in terms of primitives, structures, or overall MT system computational issues, such as efficiency. In the development of grammatical theory, for example, the "points of contact" between the syntax and the real world knowledge[4] have been addressed in natural language processing (NLP) systems (*e.g.* Winograd (1973), and others in Grosz, Sparck-Jones, and Webber (1987)). However, with respect to a theory of the IL, these issues are more complex because no consensus exists yet on the criteria for evaluating ILs.

It is this second concern for defining the relation between the IL and the KR components of a MT system that we focus on in this paper.

2 Background

This section first describes our system, focusing on certain issues relevant to defining an interlingua, and then introduces the formalism we are using as an interlingua for our system.

2.1 From UNITRAN to LEXITRAN

In translating from the input, *i.e.*, a source language (SL) sentence, to the output, *i.e.*, one or more target language (TL) sentences, an IL-based MT system proceeds through two phases:[5]

- An *analysis phase*: the SL sentence is translated into the IL formalism.
- A *generation phase*: the IL structures are translated into TL sentences.

MT system vary widely with respect to the processing components that are used in these two phases and the manner in which these components exchange intermediate representations during the translation. A syntactic processor is required both for analysis of the SL sentence and synthesis of the final TL sentence. Also in both phases, an IL processor is needed to compose the IL representation (during the interpretation of the SL input) and to decompose the IL representation (during the production of the TL output).

In our current system, LEXITRAN, we have adopted the syntactic and IL processors from UNITRAN (see Dorr (1987, 1990, 1993)). However, our system differs from the UNITRAN design in two ways. First, in terms of the translation steps, LEXITRAN has an intermediary "filter" phase between the analysis of the SL sentence and the generation of the TL sentences. Second, in terms of system components, LEXITRAN has a component containing a KR system, which is separate from the syntactic and IL processors.

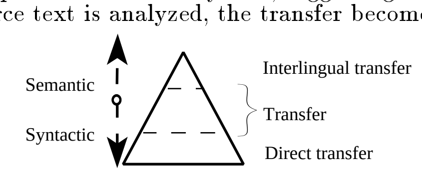
The filter phase makes use of a KR component and an MT system IL-KR interface program. During this phase, each structure that is output by the semantic analysis/composition phase is passed separately via the interface program to the KR component.[6] This component in turn filters out, or discards, those structures containing spatial relations that are incompatible with the system's facts; the resulting representations comprise the interlingua. The entire translation process is illustrated in Dorr and Voss (1993).

We should note that an MT system that includes a filter phase, by taking the extra step of having the KR system check its interpretations of a SL sentence (the IL set produced during the analysis phase), tackles two significant problems efficiently: (1) the MT system may be scaled up in terms of the number of natural languages it handles, without requiring changes to the KR system which is isolated from the syntax; and (2) the MT system

[3]For an introduction to the various approaches used in MT systems, see Hutchins and Somers (1992), chapter 4. Reference is made to other IL-based MT systems in section 2.2 below.

[4]For example, the notion of selectional restrictions — such as the requirement that the verb *sleep* have an animate subject to explain the anomaly of *The ideas are sleeping* — hinges on one's definition of terms that are taxonomic (or ontological) and syntactic — "animate" and "subject" in the case of *sleep*.

[5]Another type of MT system is the transfer-based model. For examples of this approach, see Abeillé, Schabes, and Joshi (1990), Alonso (1990), Arnold and Sadler (1990), Boitet (1987), Colmerauer *et al.* (1971), Kaplan *et al.* (1989), McCord (1989), van Noord *et al.* (1990), Thurmair (1990), among others. Simplifying somewhat, the following pyramid diagram is often used to illustrate a range of levels at which transfer is possible in an MT system, suggesting that as more of the source text is analyzed, the transfer becomes simpler.



See Hutchins and Somers (1992), chapter 6, for a critical discussion of different MT models.

[6]For the KR component we are using PARKA (Spector, Hendler, and Evett (1990) and Spector *et al.* (1992)), a frame-based KR system which was designed to provide a principled approach to multiple inheritance and the representation of part-whole relations. (Also, see Woods and Schmolze (1992) for an overview of the KL-ONE family of KR systems that PARKA belongs to.) Since at this time our KR needs are quite narrow, we have opted to use this adequate and readily available resource.

may be scaled up in terms of the number of words in a language,[7] without requiring changes to the syntactic component, similarly because that component is isolated from the KR.

2.2 Defining an Interlingua

As mentioned in section 1, the field of MT research lacks a consensus on what evaluation criteria should be applied to IL formalisms. Other IL-based MT systems have drawn on a variety of semantic formalisms as the basis for their IL. For example, the project Rosetta (Appelo and Landsbergen (1986)) uses an IL based on M-grammar, a representation derived from Montague grammar (Dowty, Wall, and Peters (1981)). Barnett *et al.* (1991) at MCC have taken Discourse Representation Theory (DRT) (Heim (1982), Kamp (1984)) as their starting point for an IL. In both these cases, the original representations were developed as a theoretical formalism and then were adapted for an MT system.

This raises the question of how the IL in an MT system relates to a theory of semantics. The semantic or IL structures of MT systems must meet two different types of criteria. First, IL structures must map somehow to KR structures and we must justify what differentiates the representations in the IL and KR components since neither is constrained by perceptual data. Defining a IL-KR mapping is a precondition to building an MT system that can take advantage of the KR capabilities to filter out incorrect translations. Second, the IL structures in an MT system must map to and from syntactic structures of all the languages in the system — not just one language.

Our approach in LEXITRAN has been to assume that the "languages" of the IL and KR systems share many of the same predicates, but are not identical. Instead, the IL predicates are a proper subset of those in the KR system because we wish to allow, in principle, for KR concepts that are not needed for language-to-language translations. This avoids the problem of trying to represent a "full" meaning for each word in a sentence being translated.

Another advantage to making this distinction is that we wish to have the predicates of the IL system be driven by the demands of the syntax-to-IL mapping, rather than by the KR system. This design consideration protects our system from becoming unnecessarily brittle as the KR system grows or changes with the domain of translation. It also reflects our bias toward maintaining the advantages of assumptions made by Dorr (1993) over those of Nirenburg *et al.* (1992) when the two have different consequences for LEXITRAN.

2.3 Lexical Conceptual Structures

LEXITRAN bases its IL formalism on the theory of semantic structures developed by Jackendoff (1983, 1990). The representation he developed, referred to as *lexical conceptual structure* (LCS) and later *conceptual structure* (CS), is defined at the word level. That is, for each word, there exists one or more CSs that defines its meaning as a structure. For the meaning of a sentence, simplifying somewhat, the CSs for the words in that sentence are composed into one CS. The resulting CS then represents the meaning of the sentence.

When a word has multiple meanings, it has, for each of those meanings, a separate CS associated with it in the lexicon. This occurs, for example, in the case of the English word *under* which is 'overloaded' and can convey several distinct interpretations. For these same interpretations German uses its word *unter* and then relies on the grammatical mechanism of case markings and an additional word to make further distinctions. Consider the translation of the English sentence *The mouse ran under the table* to its three German equivalents:

(3) (i) Die Maus ist unter dem Tisch gelaufen
'The mouse ran (about in the area) under the table'
(ii) Die Maus ist unter den Tisch gelaufen
'The mouse ran (to a place somewhere) under the table'
(iii) Die Maus ist unter dem Tisch durch gelaufen
'The mouse ran (past a place somewhere) under the table'

The English preposition *under* together with the verb *run* conveys ambiguously three possible spatial relations, *i.e.* three different paths that the mouse may take. In German, two of these paths are distinguished from the third by explicit case markings: the accusative and the dative cases show up on the determiner *den* and *dem* of the noun *Tisch*, and distinguish between the path having an endpoint (as when the mouse stops under the table) and the path being open-ended (as when the mouse continues to move either past or about under the table).

The mechanism of a verbal prefix is then also available in German for conveying additional information about the path.[8] Here the prefix *durch* is needed to convey that the path is not only under the table, but that it also continues 'past' or beyond being under the table.

Note that these two ways of explicitly distinguishing the path types — namely, the presence of the word *durch* and the different case marking options in the German translations — give us evidence that we do indeed need to have enough information in our CSs for the English word *under* to differentiate among these path types. Without the path details being stored in the CSs, the information needed to generate the German translations correctly would not appear in the IL and hence would be lost in the analysis phase.

Now consider a change of **window** to **door** in the sentence: *The mouse ran under the door*. This change does not affect the IL composition process in the translation. We would expect however that a KR system would be able to filter out 2 of the 3 interpretations — namely, those corresponding to running to a place under the door and those for running about in a place under the door. These should follow from the assumption that a typical door is upright on its hinges and so has inadequate space for a mouse to run 'to a place under' or 'about in an area under' and yet still be understood as having run *under the door*.

3 Analysis

The aim of the last section was to provide a brief introduction to LEXITRAN and the issues of defining an

[7] We assume here that the syntactic categories of those words have already been included in the syntax component.

[8] The grammatical status of these prefixes is subject to debate within linguistic theory (van Riemsdijk (1990)).

interlingua as a level of representation that is distinct from the syntax and the KR components of the system. Now we will examine more carefully the domain of the sentences being translated and the evidence for the representations in the different components.

3.1 Spatial Domain and Spatial Predicates

We have been using the phrase "spatial relation" to refer to the relative positions of objects in 3-dimensional physical space. Thus, when referring to the "spatial relation" of a cup being on a table, we are locating one object, a cup, in terms of the top surface of another object, a table. This phrase is meant to capture a conceptual level of representing such relations. We could describe spatial relations in a mathematical notation, such as with Cartesian coordinates. However the symmetry of mathematical formalizations for spatial relations does not extend to the natural language expressions of spatial relations (Talmy (1978)).

In contrast to "spatial relation" we use term "spatial expression" to mean the linguistic surface structures that express spatial relations.[9] Not all natural languages have or use the same set of linguistic forms to convey the location or path of motion of objects in physical space. For example, the spatial relation expressed in a preposition in English may appear as a verbal prefix in a Russian translation, or as a postfix on the head noun of an NP in a language such as Korean. Or the equivalent of the English preposition may not actually appear as a distinct surface element in a French translation, but instead be incorporated into the meaning of a verb.

In order to identify more narrowly the parts of a spatial expression that we will be discussing, we will use the term *spatial predicate*. A *predicate* is a structure (composed of a predicate-relation and arguments) that exists at the IL or semantic level of representation: it is a theoretical construct.

Modifying the categories of Talmy (1985) and adapting work on PLACEs and PATHs by Jackendoff (1983, 1990) to our MT framework, we identify the following components within spatial predicates:

T1: The type of spatial predicate being conveyed, one of two high-level characterizations we will be examining: a PLACE or a PATH. *Example: he stood on the boat* contains a PLACE, whereas *the cargo was loaded onto the boat* contains a PATH.

T2: The "target" object or event, the item being located. *Example: the boy* in the phrase *the boy is on the boat*.

T3: One or more "reference" objects, items whose locations are known. *Example: the boat* in the phrase *the boy is on the boat*.

T4: The spatial operator, one of a few high-level characterizations, including a LOCATION or a DIRECTION. *Example:* a direction such as *south*, *left*, *down*, or *away*; a location in a physical relation such as *against* or a geometric configuration such as *around*

T5: A "perspective" location or frame of reference. *Example:* German *hin/her* distinction; English *here/there* distinction; also the distinction between *come* and *go* as in *he came into the room* and *he went into the room*.

[9] Many other similar, less inclusive terms exist in the literature. (See Dorr and Voss (1993).)

In a simple sentence such as *the cup is on the table*, the spatial predicate T1, corresponds to a PLACE, meaning that location on the table where the cup is. T2, the target or object being located, and T3, the reference object, correspond to the phrases *the cup* and *the table*, respectively. The spatial operator T4, corresponds to the word *on*. This sentence conveys a spatial relation that is independent of the viewer's perspective and so T5 has no value.

The mapping from T1–T5 to the parts of a sentence is not always one-to-one however. Here are a few examples where the mapping is not so obvious:

Spatial Component	*(a) He lifted the box*	*(b) He went ashore*	*(c) He fell down*
T1	PATH	PATH	PATH
T2	box	he	he
T3	—	shore	ground
T4	up	to	down
T5	—	(the sea)	—

In (a), T4 corresponds to a lexically implicit value (the word *up* does not appear in the sentence). Similarly, in (b) and (c) there are other non-explicit values for the components, T3, T4, and T5.

Currently our IL structures contain spatial predicates corresponding to relations for the T1, T3, and T4 components; we have not yet implemented the T5 relation and have chosen to treat the T2 part of the predicate as an "external argument" (*i.e.*, it is outside the IL spatial predicate constituent structure).[10]

We should note here that the components in our predicates will need to be refined as we develop a richer model of spatial expressions. For example, some languages make fine-grained distinctions with respect to distances in their frames of reference (our T5). We have not dealt with the structure of measurement and quantity, so we have not formalized phrases like *under many tables*. And, in order to extend our work to an intersentential, or discourse level of analysis, our predicates may need additional components for tracking spatial focus (Maybury (1991)).

Dorr and Voss (1993) present a discussion of the changes made to the CSs that were adapted from Jackendoff's framework for LEXITRAN. One critical argument made there concerns the need for within-language synonymy tests, as well as cross-linguistic evaluations, in an iterative approach to developing the lexical-semantics for PATH and PLACE predicates. The results of such tests provide the first step in establishing evidence for the particular structures being hypothesized as IL predicates. The set of structures developed in this way can then be tested across languages. Furthermore, as noted in Dorr and Voss (1993), since there is a finite set of "lexicalization classes" that enumerate where spatial predicates may appear in the spatial expressions of a language, research can proceed by testing structures that fall into each of the relevant lexicalization classes.

3.2 Evidence for Encodings

In order to talk about the encoding of spatial relations in the various parts of the MT system and examine the

[10] This is analogous to the syntactic treatment of a sentence subject which is generally considered to be external to the verb phrase.

role of the KR system in filtering out incorrect interpretations during the translation process, we need to clarify which encodings appear in which part of the MT system.

The following terms are used to classify the encoding of spatial relations on the basis of the "evidence" we have for them:

- *lexically explicit:* a spatial relation encoded explicitly in a word.
- *lexically implicit:* a spatial relation encoded implicitly, or internal to the structure representing the meaning of a word.
- *logically inferable:* a spatial relation logically inferred from lexically explicit or implicit relations, but not itself part of the structure representing the meaning of a word.

In the first two cases, the relation appears in the lexical entry for the relevant word; in the third case, the relation does not appear in the lexical entry.

An example of the first case is the direction SOUTH as an abstract concept, which is lexically explicit in the word *south*.[11] An example of the second case is the direction UP as a lexically implicit component of the word *lift*. The implicit presence of this constituent is apparent in tests for synonymy: *he lifted the baby*, *he lifted the baby up*, and *he lifted up the baby*. Finally, as an example of this last category, the direction FROM is logically inferable in the sentence *John arrived home*, where the lexically implicit relation PATH contains the explicit PLACE from *home*, and where we can infer logically that in a PATH ending at home, there was also a DIRECTION from which the arriver, *John*, came.

The definition of these categories is tied to the way we have modularized LEXITRAN into components. In the chart below, the X's mark which types of encoding of a spatial relation may appear in which of the components in our MT system.

Spatial Relation	*Component of LEXITRAN*		
	Syntax	IL	KR
lexically explicit	X	X	X
lexically implicit		X	X
logically inferable[12]			X

Following up on the examples above, the relation SOUTH in *south* will be represented at all levels in LEXITRAN, whereas UP in *lift* will only be represented at the IL and KR levels, and FROM in *John arrived home* will only be represented at the KR level (as the result of inferencing).

We can readily see that the Syntax-IL mapping requires tracking which components in the spatial predicates (at the IL level) appear in the surface SL and TL sentences and where in the sentence syntax they will be positioned. The IL-KR mapping involves no such transformation of structures. Instead, the IL structures are passed to the KR component for the checking of its spatial predicates; thus, the term *spatial predicate* extends to these structures once transferred into the KR component as well. However, one must not confuse the spatial relations that were inferred in the KR system from those brought in by the IL representation.

To clarify this last point, consider the following English sentences:

(4) (i) He took the book to Tanya's table
(ii) He took the book from Florence's floor

If the sentences are translated into German, the *take-to* component of the first sentence translates to *bringen* whereas the *take-from* component in the second sentence translates to *nehmen*. In both sentences there is an implicit PATH relation where a book moves from one location to another. The FROM direction is logically inferable in the first sentence but lexically explicit in the second sentence. The situation is reversed with a TO direction: the TO is lexically explicit in the first sentence, but only logically inferable in the second sentence. If our IL representation of the first sentence were to include the FROM relation — and similarly if our IL representation of the second sentence were to include the TO relation — then at the point in translation where the system must generate a German sentence, we would have lost the information from the lexicalization and could no longer use it to select between the two German verbs.

This last example and the chart above help illustrate the double set of justifications that are required in a theory of the interlingua. In one direction the syntax-IL mapping provides one set of constraints on the IL, and in the other direction, the IL-KR mapping provides another set. Currently no theory of the interlingua defines these constraints and addresses the criteria to be used in evaluating them.

4 Results and Discussion

If we consider the status of the sentence *Daniel drove to the south of Colorado*, we quickly determine that the phrase *the south of Colorado* is ambiguous. One interpretation of this sentence is that Daniel drove to southern Colorado. That is, the phrase *the south of Colorado* refers to the region inside of Colorado that is considered its south. The IL structure for this part-to-whole relation, where the "part" is the meaning of the entire phrase and the "whole" is Colorado, is viewed as a "place-place" relation by the KR component which checks for a part-whole interpretation when it encounters two "place" predicates in an IL structure.

Now consider the following examples:

(5) (i) Maria drove to the south of Florida
(ii) The skipper navigated to the south of Gibraltar.

In the first case, *the south of Florida* refers to a part-whole spatial relation, with an "inside of" Florida interpretation. In the second case, *the south of Gibraltar* does not mean a part-whole relation — it refers to a region "outside of" Gibraltar. This distinction is captured in the translation into Spanish:

(6) (i) Maria manejó hacia el sur de la Florida
(ii) El capitán navejó al sur de Gibraltar

In other words, what appears as a conceptual distinction must be detected in the MT system in order to appropriately select the proper translation into Spanish.

Syntactically the English sentences are identical. At the IL level they are ambiguous. The IL processor will

[11] The words in capital letters refer to the spatial relation, the abstract term.

[12] The logically inferrable relations can be broken out into the "logically explicit" facts explicitly encoded in the KB of the KR system and the "logically implicit" facts that are derived from other facts and inference rules in the system.

create the "inside of" as well as the "outside of" IL interpretation for both of these sentences since it has no knowledge of which interpretation makes sense for which sentence. In other words, it does not check the T1–T4 values in predicates against real world knowledge. Rather it is the KR processor that performs this checking in the filter phase of the translation: it must allow for the part-whole IL structure for the first sentence and discard that structure for the second.

The KR processor makes use of the information in the IL structure that, in this case, was contributed by the verbs. Sentences (5)(i) and (6)(i) contain the equivalent of "go by land-vehicle" in its IL structure, whereas sentence (5)(ii) and (6)(ii) contain the equivalent of "go by water-vehicle" in its structure.[13] The KR, using inference rules that disallow a "go by X-vehicle" event composed with a path not on X (X would be "land" or "water" here), would rule out the two anomalous cases that concern us: (1) the IL interpretation of "driving to the south of Florida" as going by car to the outside of Florida, and (2) the IL interpretation of "navigating to the south of Gibraltar" as going by boat to the inside of Gibraltar. This result — enabling the KR to filter out anomalous IL interpretations by virtue of IL structures where it can readily identify the arguments within spatial predicates — also extends to other internal-external distinctions among spatial entities.

Our approach has been to assume that, in general, the syntactic properties of phrases reflect the underlying predicate-argument structural meaning of the words that head those phrases. Since we can "see" and test properties of phrasal structure, but not those of semantic structure, we must take advantage of what information we can glean from phrasal structures. The idea here is to use the differences in word meaning that correlate with syntactic distribution patterns to refine the hypotheses we have for the meaning structure — rather than developing lexical semantic structures solely based on our intuitions.[14]

We have described our approach to translating spatial expressions in an IL-based MT system and presented several arguments for the next steps in developing a theory of the *interlingua*. Such a theory must specify what can count as a constraint on the IL structures and thus provide independent justification for the particular structures being used. Our approach combines promising syntactic and semantic aspects of existing translation systems; we see this as the most appropriate framework for addressing some of the tough issues in MT, including the development of criteria for evaluating IL representations.

[13] See Dorr (1993) for a more complete discussion of verbs' CS structures.

[14] Two related issues should be addressed here. First, there is the notion that current lexical semantics basically draws on the well-known linguistic work on case (*e.g.* Fillmore (1968)) and so seems uninteresting. The second issue, is that the current computational work done in lexical semantics basically does not go beyond the insights achieved in the 70s work in AI (*e.g.* Schank (1973)), and so again seems uninteresting. However, we argue that this earlier work does not place any constraints on (1) what can be represented as a predicate, (2) the number of arguments, (3) the obligatory or optional nature of those arguments, or (4) the definitions of what constitutes a valid argument, each theory must provide independent justification for its hypothesized structures. It is at the lower level of *how* the ideas about case are integrated with the rest of modern linguistic theory that the current research challenges exist.

References

Abeillé, A., Schabes, Y., and Joshi, A. K. 1990. Using Lexicalized Tags for Machine Translation. In Proceedings of Thirteenth International Conference on Computational Linguistics, 1–6. Helsinki, Finland.

Appelo, L. and Landsbergen, J. 1986. The Machine Translation Project Rosetta. First International Conference on State of the Art in Machine Translation. Saarbruecken, Germany.

Barnett, J., Mani, I., Rich, E., Aone, C., Knight, K., and Martinez, J. C. 1991. Capturing Language-specific Semantic Distinctions in Interlingua-Based MT. In Proceedings of the Machine Translation Summit, 25–32. Washington, DC.

Dorr, Bonnie J. 1987. UNITRAN: An Interlingual Approach to Machine Translation. In Proceedings of the Sixth Conference of the American Association of Artificial Intelligence, 534–539. Seattle, Washington.

Dorr, Bonnie J. 1990. Solving Thematic Divergences in Machine Translation. In Proceedings of the 28th Annual Conference of the Association for Computational Linguistics, 127–134. University of Pittsburgh, Pittsburgh, PA.

Dorr, Bonnie J. 1993. *Machine Translation: A View from the Lexicon.* Cambridge, MA: MIT Press.

Dorr, Bonnie J. and Voss, Clare. 1993. Constraints on the Space of MT Divergences, Working Notes for the AAAI 1993 Spring Symposium, Building Lexicons for Machine Translation, Technical Report SS-93-02, Stanford University, CA.

Dowty, D. R., Wall, R. E. and Peters, S. 1981. *Introduction to Montague Semantics.* Dordrecht, Holland: Reidel.

Fillmore, C. 1968. The Case for Case. In *Universals in Linguistic Theory,* Bach, E. and R. Harms, eds. New York, NY: Holt, Rinehart, and Winston, 1-88.

Grosz, B., Sparck-Jones, K., and Webber, B. eds. 1987. *Readings in Natural Language Processing.* Los Altos, CA: Morgan Kaufmann.

Heim, I. 1982. The Semantics of Definite and Indefinite Noun Phrases. Ph.D. diss., Dept. of Linguistics, University of Massachusetts, Amherst, MA.

Hutchins, J. W. and Somers,H. L. 1992. *An Introduction to Machine Translation.* London, England: Academic Press.

Jackendoff, Ray S. 1983. *Semantics and Cognition.* Cambridge, MA: MIT Press.

Jackendoff, Ray S. 1990. *Semantic Structures.* Cambridge, MA: MIT Press.

Kamp, H. 1984. A Theory of Truth and Semantic Representation. In *Formal Methods in the Study of Language.* Dordrecht, Holland: Foris Publications.

Maybury, M. T. 1991. Topical, Temporal, and Spatial Constraints on Linguistic Realization. *Computational Intelligence* 7(4): 266–275.

Nirenburg, S., Carbonell, J., Tomita, M., and Goodman, K. 1992. *Machine Translation: A Knowledge-Based Approach.* San Mateo, CA: Morgan Kaufmann.

Riemsdijk, H. van 1990. Functional Prepositions. In *Unity in Diversity,* Pinkster, H. and Genee, I., eds. Dordrecht, Holland: Foris Publications, 229-241.

Schank, R. 1973. Identification of Conceptualizations Underlying Natural Language. In *Computer Models of Thought and Language,* R. Schank and K. Colby, eds. San Francisco, CA: W. H. Freeman, 187-247.

Spector, L., Hendler, J. and Evett, M. 1990. Knowledge Representation in PARKA, UMIACS TR 90-23, CS TR 2410, Dept. of Computer Science, University of Maryland, College Park, MD.

Spector, L., Anderson, W., Hendler, J., Kettler, B., Schwartzman, E., Woods, C. and Evett, M. 1992. Knowledge Representation in PARKA - Part 2: Experiments, Analysis, and Enhancements, CS-TR-2837, UMIACS-TR-92-16, Dept. of Computer Science, University of Maryland, College Park, MD.

Talmy, L. 1978. Figure and Ground in Complex Sentences. *Universals of Human Language*, Vol. 4: Syntax J. H. Greenberg, C. Ferguson, and E. Moravcsik, eds. Palo Alto, CA: Stanford University Press, 625–49.

Talmy, L. 1985. Lexicalization Patterns: Semantic Structure in Lexical Forms. *Grammatical Categories and the Lexicon* Timothy Shopen, ed. Cambridge, England: University Press, 57–149.

Winograd, T. 1973. A Procedural Model of Language Understanding. In *Computer Models of Thought and Language,* R. Schank and K. Colby, eds. San Francisco, CA: W. H. Freeman.

Woods, W.A. and Schmolze, J. A. 1992. The KL-ONE Family. *Computers and Mathematics with Applications,* 23(2-5): 133-177.