

Translation Between Representation Languages

Jeffrey Van Baalen
Computer Science Department
P.O. Box 3682
University of Wyoming
Laramie, WY 82071
jvb@uwyo.edu

Abstract

A capability for translating between representation languages is critical for effective knowledge base reuse. We describe a translation technology for knowledge representation languages based on the use of an *interlingua* for communicating knowledge. The interlingua-based translation process consists of three major steps: (1) translation from the source language into a subset of the interlingua, (2) translation between subsets of the interlingua, and (3) translation from a subset of the interlingua into the target language. The first translation step into the interlingua can typically be specified in the form of a grammar that describes how each top-level form in the source language translates into the interlingua. We observe that in cases where the source language does not have a declarative semantics, such a grammar is also a specification of a declarative semantics for the language. We describe a methodology for building translators that is currently under development. A "translator shell" based on this methodology is also under development. The shell has been used to build translators for multiple representation languages and those translators have successfully translated non-trivial knowledge bases.

1. Introduction

Acquiring and representing knowledge is the key to building large and powerful AI systems. Unfortunately, knowledge base construction is difficult and time consuming. The development of most systems requires a new knowledge base to be constructed from scratch. As a result, most systems remain small to medium in size. The cost of this duplication of effort has been high and will become prohibitive as attempts are made to build larger systems. A promising approach to removing this barrier to the building of large scale AI systems is to develop techniques for encoding knowledge in a reusable form so that large portions of a knowledge base for a given application can be *assembled* from knowledge repositories and other systems.

For encoded knowledge to be incorporated into a system's knowledge base or interchanged among interoperating systems, the knowledge must either be represented in the receiving system's representation language or be translatable in some practical way into that language. Since an important means of achieving efficiency in application systems is to use specialized representation languages that directly support the knowledge processing requirements of the application, we cannot expect a standard knowledge representation language to emerge that would be used generally in application systems. Thus, we are confronted with a *heterogeneous language problem* whose solution requires a capability for translating encoded knowledge among specialized representation languages.

We are addressing the heterogeneous language problem by developing a translation technology for knowledge representation languages based on the use of an *interlingua* for communicating knowledge among systems. Given such an interlingua, a sending system would translate knowledge from its application-specific representation into the interlingua for communication purposes and a receiving system would translate knowledge from the interlingua into its application-specific representation before use. In addition, the interlingua could be the language in which libraries would provide reusable knowledge bases. An interlingua eases the translation problem in that to communicate knowledge to and from N languages without an interlingua, one must write $(N-1)^2$ translators into and out of the languages. With an interlingua, one need only write $2*N$ translators into and out of the interlingua.

We consider in this paper the problem of translating declarative knowledge among representation languages using an interlingua with the following properties:

- A formally defined declarative semantics;
- Sufficient expressive power to represent any theory that is representable in the languages for which translators are to be built.

In practice, one cannot expect any given interlingua to have sufficient expressive power to support usable representations of *any* theory that is representable in *any* language. However, an interlingua with the expressive power of first-order logic, such as the Knowledge Interchange Format (KIF) being developed in the ARPA Knowledge Sharing Effort [Genesereth & Fikes 92], can provide that support for a broad spectrum of theories and languages. For our purposes in this paper, we will assume an interlingua and a set of languages for which the properties listed above hold.

The interlingua-based translation process can be thought of as consisting of three major steps:

- Translation from the source language into a subset of the interlingua;
- Translation between subsets of the interlingua; and
- Translation from a subset of the interlingua into the target language.

Since the interlingua is assumed to be at least as expressive as the source language, the first translation step into the interlingua can typically be specified in the form of a grammar that describes how each *top-level form* (e.g., sentence, definition, rule) in the source language translates into the interlingua. Our methodology includes techniques for specifying such grammars so that they are *reversible*, i.e., they can be used not only to translate into the interlingua, but also to translate out of a subset of the interlingua. If one has such a reversible grammar for the target language, then step 2 involves translating from the subset of the interlingua produced by the source language grammar to the subset of the interlingua that is translated (i.e., recognized) by the reverse of the target language grammar. For any given top-level form F_s in the source subset, translation step 2 involves determining a top-level form F_t in the target subset such that F_s is logically equivalent to F_t . Thus, formally, step 2 requires hypothesizing an equivalent form in the target subset and then proving the equivalence.

We have developed the following in support steps 1 and 3:

- A formal description of the translation process into and out of an interlingua;
- A method for determining whether a given grammar in fact specifies how to construct a translation for every top level form in a given source language; and

- A method for determining whether a given grammar is reversible so that it can be used to translate both into and out of an interlingua.

These languages and methods have been incorporated into a "translator shell" system that provides facilities for specifying interlingua-based translation using KIF as the interlingua. The system has been used to build translators for multiple representation languages and those translators have successfully translated non-trivial knowledge bases. Among the systems built so far are a bi-directional CLASSIC [Borgida, et al 89] to KIF translator and a LOOM [MacGregor 91] to KIF translator[Fikes, et al 91].

2. Interlingua-Based Translations and Semantics

We consider here *equivalence preserving translations* [Buvac and Fikes 93] in which the translation of an axiomatization of a logical theory is an axiomatization of an equivalent logical theory. To make such a requirement on translators meaningful, a declarative semantics including logical entailment needs to be formally specified for both the source and target languages. We are assuming such a declarative semantics for the interlingua. In cases where a language does not have such a declarative semantics, specifying a translation of that language into the interlingua provides a declarative semantics for the language. Thus, another advantage of using an interlingua is that it offers a relatively easy way to specify a semantics for new representation languages. This use of an interlingua for specifying the semantics of representation languages may turn out to be at least as important as its role in facilitating translation among representation languages. This method of semantics specification is based on the following definition:

Definition 2.1 (interlingua-based semantics): Let L be a language, L_i be an interlingua language with a formally defined declarative semantics, $TRANS_{L,L_i}$ be a binary relation between top-level forms of L and top-level forms of L_i , and BT_L be a set of top-level forms in L_i . The pair $\langle TRANS_{L,L_i}, BT_L \rangle$ is called an *L_i -based semantics* for L when for every set T_L of top-level forms in L , there is a set T_{L_i} of top-level forms in L_i such that

$$\begin{aligned} \forall s_1 \in T_L \exists s_2 \in T_{L_i} TRANS_{L,L_i}(s_1, s_2) \\ \forall s_2 \in T_{L_i} \exists s_1 \in T_L TRANS_{L,L_i}(s_1, s_2) \end{aligned}$$

and the theory of $T_{L_i} \cup BT_L$ is equivalent to the theory represented by T_L .

Hence, $TRANS_{L,L_i}$ specifies translations of top-level forms in L to top-level forms in L_i . Roughly speaking, BT_L is the set of axioms that are included in the semantics of L expressed in L_i . For example, a device modeling language might have a vocabulary of measures (e.g., INCH, FOOT) and include in its semantics the axioms that relate those measures.

If $\langle TRANS_{L,L_i}, BT_L \rangle$ is being used to define the semantics of L , then "the theory represented by T_L " is equivalent to "the theory of $T_{L_i} \cup BT_L$ " *by definition*. If L has an independently defined semantics, then the equivalence of the two theories is a requirement on the definition of $TRANS_{L,L_i}$.

TRANS is defined as a relation rather than a function because we allow there to be more than one translation of a top-level form in L so long as it does not matter which translation is picked. Thus, TRANS can be viewed as a function into equivalence classes of interlingua top-level forms. Note also that TRANS defines what it means for two sentences in L to be equivalent, namely that their translations are equivalent sentences in L_i .

An additional advantage of the interlingua-based approach to semantics is that if such a semantics is given in a machine executable form, it can be used to automatically translate a new language into the interlingua. Hence, with a single effort, one can give both a semantics for a new language and a procedure for translating it into the interlingua.

In our language translation methodology one specifies the semantics of a new representation language using a special kind of definite clause grammar [Pereira & Warren 80] that we call a *definite clause translation grammar* (DCTG). This grammar can be used to translate top-level forms in the new language into an interlingua. A DCTG is a set of Horn clauses that has a distinguished binary predicate symbol TRANS such that if s_1 is a top-level form in the new language and s_2 is a top-level form in the interlingua, TRANS(s_1, s_2) follows from the grammar just in case s_2 is a translation of s_1 .

We provide a formal technique for showing that such a grammar is a translator, i.e., that for every sentence in the new representation language, the grammar produces a sentence in the interlingua. We also provide a technique for showing that such a grammar is reversible. Both of these techniques have the feature that when a grammar does not have the desired property, they pinpoint locations in the grammar that require repair in order to obtain the property.

3. Translating Between Subsets of the Interlingua

Normally, step 2, translating between subsets of the interlingua, is far more difficult than steps 1 and 3: for each sentence in the source subset of the interlingua we must find an equivalent sentence in target subset, if possible. What makes this difficult is that some sentences have no equivalent sentences in the target subset, while others have such sentences but they are difficult to find.

Our approach to this problem is to treat the target subset of KIF as a *pseudo-canonical* form for KIF and to construct a rewrite system that transforms KIF sentences into this pseudo-canonical form. This use of rewrite systems differs from the standard use [Dershowitz & Jouannaud 90]. Normally one develops a set of rewrite rules from a system of equations that specify equivalences between terms in a language. The goal is to develop a set of directed rules from which it is possible to infer that two terms are equivalent whenever it was possible to infer this from the original undirected equations. An additional goal is to construct rule sets with the following properties: first, given any term t , every possible rewrite sequence from t should end in the same term t' . Second, when two terms are equivalent, rewrite sequences from those terms should end with the same t' . When a set of rules has these properties, we say that every term in the language has a *canonical form* and that the language itself has a *canonical form*.

One can think of the problem of translating into a target subset of KIF as the problem of finding a set of rewrite rules making the target subset a canonical form. Unfortunately, a translator developer does not have a set of equations specifying all the equivalences between terms in KIF and, furthermore, no techniques are known for developing a set of rewrite rules for a particular canonical form. Therefore, we have relaxed some of the requirements on rule sets and call the target subset of KIF a pseudo-canonical form. We provide special rewrite mechanisms that allow a translator to search for rewrite sequences that will lead to sentences in pseudo-canonical form.

4. Status

The KIF-CLASSIC translator was completed in the first three months of the project. In early October 1992, a series of tests of the KIF-CLASSIC translator. The first test translated a "toy" knowledge base from CLASSIC to KIF and then back again. This translation was completely successful, i.e., all of the KIF version of the knowledge base was translated back into CLASSIC. Some of the translations were different than the original CLASSIC statements, however, the resulting knowledge base was equivalent to the original in the sense that CLASSIC did all the same inferences from the translated version as from the original version.

The second test translated into CLASSIC a toy knowledge base that was originally written in KIF. This knowledge base contained knowledge that was appropriate for representation in CLASSIC, however, it was developed by someone who has never used CLASSIC and, hence, the knowledge did not conform to the idioms of the CLASSIC language. Consequently, this KIF knowledge base had a considerably less constrained form and constituted a much more rigorous test of the KIF-CLASSIC translator, requiring it to do many reformulations of the knowledge base in order to get it into a translatable form. Remarkably, this test was also 100% successful in the sense that every statement in the KIF knowledge base was translated into one or more CLASSIC statements.

Having had this much success, it was decided to try a test involving translation from one specialized representation language to another, through KIF. In particular, we translated the ROME Planning Initiative knowledge base from LOOM to KIF using a LOOM-KIF translator developed by Ramesh Patil at USC ISI. Then the KIF-CLASSIC translator was used to translate the result into CLASSIC. One would not expect the translation from KIF-CLASSIC to be 100% successful since LOOM is a strictly more expressive language than CLASSIC.

The first several runs of the KIF-CLASSIC translator translated only around 50% of the KIF knowledge base. However, the translator is designed to flag untranslatable statements and allow the user to assist in their translation. Inspection of the untranslated statements showed that many of them were not correct translations of the LOOM knowledge base into KIF. When these difficulties in the LOOM-KIF translator were repaired, there remained approximately 20% of the KIF version of this knowledge base that the KIF-CLASSIC translator could not translate. Analysis has shown that there is no translation into CLASSIC for this 20% of the KIF knowledge base.

Hence, the KIF-CLASSIC translator succeeded in translating a real LOOM knowledge base into CLASSIC. Every KIF statement generated by the LOOM-KIF translator that was representable in CLASSIC was translated by the KIF-CLASSIC translator. The KIF-CLASSIC translator's ability to flag untranslatable statements proved useful in several ways including debugging the LOOM-KIF translator.

The above tests represent success in all of the milestones planned for this year as well as partially meeting the second milestone planned for next year. Because of this early success, additional unplanned tasks were initiated this year: the development of an EXPRESS to KIF translator and the development of a LOOM-KIF translator. The EXPRESS to KIF translator is currently 95% complete and the LOOM-KIF translator is currently approximately 80% complete.

6. Summary

We have described a methodology for translating knowledge representation languages based on the use of an *interlingua* for communicating knowledge. The interlingua-based

translation process can be thought of as consisting of three major steps: (1) translation from the source language into a subset of the interlingua, (2) translation between subsets of the interlingua, and (3) translation from a subset of the interlingua into the target language. The methodology advocates that the first translation step into the interlingua be specified by a grammar consisting of a set of Horn clauses (called Definite Clause Translation Grammars) that constructively implements a translation predicate relating top-level forms in a source language to their translations in an interlingua. We observed that in cases where the source language does not have a declarative semantics, specifying a translation of that language into the interlingua provides a declarative semantics for the language. Thus, another advantage of using an interlingua is that it offers a relatively easy way to specify a semantics for new representation languages.

A developer of a specialized representation language that desires to build a translator from the specialized language to an interlingua first writes a DCTG G that is an interlingua-based semantics for the language. The developer then uses the methods we have provided to show that G constructs a translation in the interlingua for any top-level form in the specialized language and therefore that G is a translator from the specialized language to the interlingua. The developer then again uses the methods we have provided to show that G also is a translator out of the interlingua in that it constructs a top-level form in the specialized language as a translation for any top-level form in the subset of the interlingua that could be produced by G when it is being used as a translator from the specialized language. Such a reverse translator provides a first approximation of a translator from the interlingua to the specialized language. We provide techniques for augmenting the capability of this first approximation translator. The subset of KIF handled by the reverse grammar is treated as a pseudo-canonical form and the translator developer constructs a rewrite system to transform sentences into this pseudo-canonical form. We provide various methods for assisting with the construction of such a rewrite system.

These languages and methods have been incorporated into a "translator shell" system that provides facilities for specifying interlingua-based translation using KIF as the interlingua. The system has been used to build translators for multiple representation languages and those translators have successfully translated non-trivial knowledge bases.

7. References

- [Borgida, et al 89] A. Borgida, R.J. Brachman, D.L. McGuinness, and L.A. Resnick; "CLASSIC: A Structural Data Model for Objects"; Proceedings of the 1989 ACM SIGMOD International Conference on Management of Data; Portland, Oregon; May-June, 1989.
- [Buvac and Fikes 93] S. Buvac and R. Fikes; "Semantics of Translation"; Proceedings of the workshop on Knowledge Sharing and Information Interchange at the 13th International Joint Conference on Artificial Intelligence; Chambery, France; August 1993.
- [Dershowitz & Jouannaud 90] N. Dershowitz and J.-P. Jouannaud; "Rewrite Systems"; in The Handbook of Theoretical Computer Science, J. van Leeuwen (ed), MIT Press; pp. 245-320, 1990.
- [Fikes, et al 91] R. Fikes, M. Cutkosky, T. Gruber, and J. Van Baalen; "Knowledge Sharing Technology Project Overview"; Stanford University Report KSL 91-71, 1991.

- [Genesereth & Fikes 92] M. Genesereth and R. Fikes; "Knowledge Interchange Format Version 3.0 Reference Manual"; Logic Group Report Logic-92-1; Computer Science Department, Stanford University; June 92.
- [MacGregor 91] R. MacGregor; LOOM Users Manual, USC/Information Sciences Institute; Technical Report Working Paper ISI/WP-22; 1990.
- [Pereira & Warren 80] F. C. N. Pereira and D. H. D. Warren; "Definite Clause Grammars for Language Analysis -- A Survey of the Formalism and a Comparison with Augmented Transition Networks, *Artificial Intelligence*, 13, pp. 231-278, 1980.

ACKNOWLEDGEMENTS

Portions of this paper were derived from a paper written jointly by the author and Richard Fikes. Ramesh Patil assisted with the LOOM to KIF grammar. Sasa Buvac and John Cowles participated in helpful discussions. This work is supported by AFOSR under contract number F49620-92-J-0434, by the Advanced Research Projects Agency, ARPA Order 8607, monitored by NASA Ames Research Center under grant NAG 2-581, and by NASA Ames Research Center under grant NCC 2-537.