The KANT Machine Translation System:  
A Brief Overview

Eric Nyberg  
Teruko Mitamura  
Center for Machine Translation  
Carnegie Mellon University

1. Introduction

KANT (Knowledge-Based, Accurate Natural Language Translation) is a software system which supports the creation of customer-specific applications for fast, accurate translation of controlled source texts in a document production environment without requiring postediting. This short introduction to KANT is intended to describe the KANT software, the kinds of translation tasks it can perform, and the steps necessary to build a KANT application in a particular customer domain.

2. The History of KANT

KANT is comprised of a set of software modules (parser, interpreter, mapper, generator) which work together to produce target language translations from controlled source text. These modules are the result of long-term research and development in practical machine translation at the Center for Machine Translation (CMT) at Carnegie Mellon University, located in Pittsburgh, PA. The KANT software grew out of extensions and refinements to earlier systems developed at the CMT, which include the CMT-SEMSYN system, a collaborative effort with the University of Stuttgart in the domain of doctor patient communications (Japanese and English source languages to Japanese, English and German target languages), and the KBMT-89 system, a funded project with IBM’s Tokyo Research Laboratory in the domain of PC installation manuals (Japanese and English to Japanese and English; cf. (Goodman and Nirenburg, 1991)).

3. The KANT Approach

KANT is a knowledge-based translation system. For each source language to be analyzed and each target language to be produced, the system makes use of specific lexicons, grammars, semantic rules, etc. to perform its task. The KANT software itself is language-independent — the same code modules are used regardless of the source and target languages. Extending the system to a new language involves writing a new lexicon, grammar, etc. for the language, but does not require any modification to KANT itself.

KANT is an interlingua-based translation system. For each sentence in the source language, KANT produces a semantic representation or "interlingua" expression. The interlingua is independent of the source and target languages, and is based on the set of concepts (objects, events,
properties) relevant to the domain of translation. The generation phase in KANT produces a target language sentence for each interlingua expression. Because interlingua is used as an intermediate representation, the analysis of the source language and the generation of the target language are independent of each other, which eliminates the need for bilingual dictionaries and transfer grammars for specific language pairs (see Figure 1).

![Diagram](image)

**Figure 1: Interlingua Translation**

4. The KANT Domains

KANT is specifically designed for multi-lingual document production from a single, controlled source language. KANT is intended for use in domains where technical information (such as documentation for machinery, electronics, software/hardware, etc.) is to be authored preferably at a central location and translated for dissemination world-wide. KANT achieves high accuracy (with no need for post-editing) precisely because it takes advantage of a controlled source language and a well-defined application domain.

The source language is not limited by the size of its vocabulary; rather, KANT places certain restrictions on the types of sentences and phrases that can be used, to eliminate unnecessary vagueness and ambiguity in the source text. This not only improves translatability, but also encourages uniform, concise text in the source language as well. Although many constructions are eliminated from the source language, the set of constructions supported by KANT is more than expressive enough for the authoring of technical information in a particular domain.

It should be clear that KANT is not intended for the translation of general, unrestricted text. By limiting KANT applications to those with a well-defined, controlled source language and particular technical domains, KANT can achieve a high degree of understanding of the input text, thus producing a much higher quality translation than all other "general purpose" commercial systems.

KANT is the first and only existing system that combines the knowledge-based, Interlingua technology with principled domain definition to achieve special-purpose translation with no postediting required.

5. Building KANT Applications

In order to develop a KANT application for a particular customer, a period of domain analysis and customization is necessary. The combination of products and document types that will be translated are analyzed to extract the vocabulary (for the KANT source lexicon), appropriate sentence structures (for the KANT source grammar), and domain concepts (for the KANT domain model) (see Figure 2). Target texts from previous translations are also analyzed to produce corresponding lexicons and grammars for each of the desired target languages.
Although much of the analysis of new domains is automated or semi-automated (using CMT corpus analysis tools), some engineering by CMT linguists is required. This customization process may require a year or so, unless the languages and domains have been previously analyzed.

6. Ongoing Applications

The Center for Machine Translation is currently building a large-scale KANT application for Caterpillar, Inc. When completed, the system will translate texts authored in a controlled technical English into eleven different languages, for the entire Caterpillar product line. The CMT has completed an initial application prototype for translation of controlled English to French. Present work is focussed on extending the system to the entire Caterpillar domain and the additional languages required by Caterpillar. At present, the CMT has developed knowledge sources for English, French, German, Japanese, and Spanish, with active development ongoing for several additional languages. For a short example of controlled English source and three target language KANT translations, see Figures 3-6.

The CMT is also in the process of negotiating contracts with two other international corporations with similar translation needs, in the domain of documentation for computer hardware/software.

7. Contacts

For further information concerning the KANT system and/or the Center for Machine Translation, please contact Jaime Carbonell, Teruko Mitamura or Eric Nyberg:

Jaime Carbonell
Director
Center for Machine Translation
5000 Forbes Ave,
Carnegie Mellon University
Pittsburgh, PA 15213
Phone: (412) 268-7279
FAX: (412) 268-6298
Internet: jgc@cs.cmu.edu

Teruko Mitamura
KANT Project Leader, Linguistics
Center for Machine Translation
5000 Forbes Ave,
Carnegie Mellon University
Pittsburgh, PA 15213
Phone: (412) 268-6596
FAX: (412) 268-6298
Internet: teruko@cs.cmu.edu

Eric Nyberg
KANT Project Leader, Software
Center for Machine Translation
5000 Forbes Ave,
Carnegie Mellon University
Pittsburgh, PA 15213
Phone: (412) 268-7281
FAX: (412) 268-6298
Internet: ehn@cs.cmu.edu
Figure 2: KANT: Knowledge-Based Natural Language Translation
8. Sample Source and Target Texts

Safety Warnings
Read the "General Installation Information" section of this manual. Then, follow the instructions in the "Safety Warnings" section.
In order to prevent a fire hazard, do not overload AC outlets.
In the following cases, TV sets can overheat:
  1. The ventilation slots are blocked.
  2. The TV set is placed in a built-in enclosure.

Periodically clean the ventilation slots with your vacuum cleaner.
If the TV set has been dropped, a shock hazard may exist. In this case, unplug the TV set. Then call your dealer.

Figure 3: Sample English Source Text Input to KANT

Conseils de sécurité
Consultez la section de ce manuel intitulée "Renseignements pour installation". Ensuite, se conformer aux instructions figurant à la section intitulée "Conseils de sécurité".
Afin d'éviter tout risque d'incendie, ne jamais surcharger les prises CA.
Dans les cas suivants, un téléviseur peut surchauffer:

  1. La grille de ventilation est bloquée.
  2. Le téléviseur est placé dans un coin renfoncé.

Dépoussiérer périodiquement la grille de ventilation à l'aide d'un aspirateur.
La chute du téléviseur peut provoquer un risque de choc électrique. En ce cas, débrancher le téléviseur. Ensuite faire appel au détaillant.

Figure 4: French Target Text Produced by KANT
Sicherheitsbestimmungen

1. Die Kühlsschlitze sind blockiert.
2. Das Fernsehgerät steht in einem Einbauschrank.


Figure 5: German Target Text Produced by KANT

「安全の注意」

このマニュアルの「一般設置情報」の章を読んで下さい。それから「安全の注意」の章の指示に従って下さい。

火災の危険を防ぐために、コンセントに電流の負荷をかけすぎないで下さい。次のような場合はテレビがオーバーヒートすることがあります。

1. 通気孔が詰まっている。
2. アイレーンが使用された場所にテレビがある。

定期的に通風口を掃除して下さい。

テレビを落とすと、電気ショックの危険があるかもしれませんが。その場合はテレビのプラグを抜いて下さい。それから販売店に連絡して下さい。

Figure 6: Japanese Target Text Produced by KANT

References


