

MT Workshop

Benoit Lavoie
CoGenTex, Inc.
840 Hanshaw Road
Ithaca, NY
USA, 14850
benoit@cogentex.com

Richard Kittredge
CoGenTex, Inc.
840 Hanshaw Road
Ithaca, NY
USA, 14850
richard@cogentex.com

Tanya Korelsky
CoGenTex, Inc.
840 Hanshaw Road
Ithaca, NY
USA, 14850
tanya@cogentex.com

Owen Rambow *
ATT Labs-Research, B233
180 Park Ave, PO Box 971
Florham Park, NJ
USA, 07932
rambow@research.att.com

Abstract

MT Workshop is a development environment for machine translation (MT). The environment consists of a declarative language for specifying transformation rules (conversion, transfer and realisation) of syntactic structures (such as the output from a syntactic parser) and a graphical interface for interactive verification of the processing of the specified rules.

1 Introduction

In this demonstration paper we present MT Workshop, a development environment for machine translation (MT). MT Workshop uses a declarative language for specifying syntactic dependency structures (Mel'cuk, 1988) and their transformations for conversion, transfer and realisation. MT Workshop includes a graphical interface which allows the developer to validate and verify the specified transformations of dependency structures. MT Workshop was developed using a general framework for developing NLG and MT applications (Lavoie et al., 2000). Features that distinguish MT Workshop from other MT systems include its declarative lexico-structural transformation language and its interactive development environment.

2 Input Representations and their Processing

Figure 1 illustrates the scope of MT Workshop. The input representations are parsed syntactic structures (PSyntSs) in a source language which

are converted into deep-syntactic structures (DSyntSs) in order to be either realized directly in the source language or transferred to corresponding deep-syntactic structures in a target language and then realized in the target language.

All syntactic structures used in MT Workshop (PSyntSs and DSyntSs) are represented as syntactic dependency tree structures defined in a formalism based on RealPro notation (Lavoie and Rambow, 1997). The main characteristics of a syntactic dependency tree structure are:

- A dependency tree is unordered;
- All the nodes in the tree correspond to lexemes (i.e., lexical heads);
- Labelled arcs indicate the syntactic dependency relationships between the lexemes.

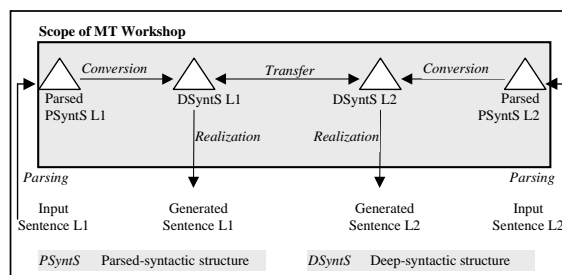


Figure 1: Scope of MT Workshop

Dependency structures used in MT Workshop and lexico-structural transformations during conversion, transfer or realisation are described in detail in (Lavoie et al., 2000).

3 Graphical Interface

The graphical interface of MT Workshop can be used for loading, editing and validating the syntactic dependency tree structures and for

* The work performed on MT Workshop by this co-author was done while at CoGenTex, Inc.

displaying the intermediate results of the processing.

Figure 2 illustrates MT Workshop interface for Korean-English translation with the four following structures (from top to bottom):

1. A parsed Korean syntactic structure (a reformatted output of an off-the-shelf Korean parser (Yoon et al, 1997)).
2. A Korean deep-syntactic structure, a conversion from the parsed Korean syntactic structure (1). The conversion process eliminates the discrepancies that might exist between the feature system used by the parser and the one used in transfer rules.
3. An English deep-syntactic structure, a result of transfer from the Korean deep-syntactic structure (2).
4. An English surface sentence, a realisation of the English deep-syntactic structure (3).

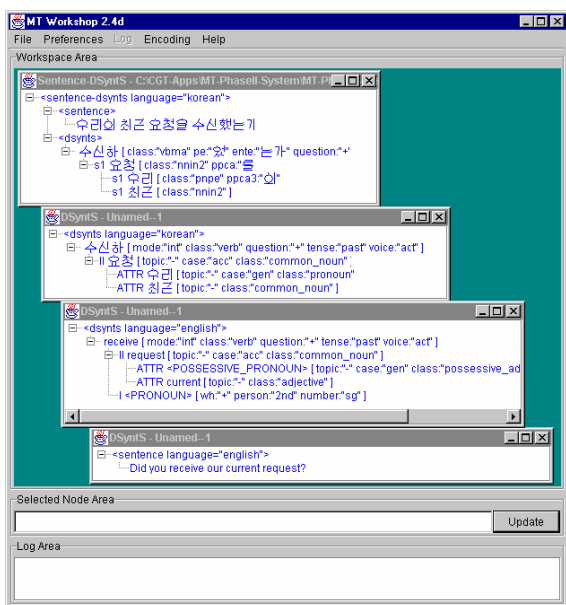


Figure 2: MT Workshop Interface

The dependency structure elements shown in Figure 2 (lexemes, features or relationships) are directly editable which allows the developer to interactively test possible variations of the transformations rules.

4 Validation with XML Grammars

MT Workshop uses customizable XML grammars for validating various types of syntactic dependency structures (such as Korean PSyntS, Korean DSyntS, English DSyntS or

other). Violations of XML grammar rules in the input or output structures are shown by highlighting the errors in the displayed structures.

Status

MT Workshop is currently implemented in Java and is being used in a project for bi-directional Korean-English translation in a military domain. Some components of MT Workshop were originally implemented in C++ for a small MT application in the military weather and battlefield domain for English-Arabic and English-French translations (Palmer et al., 1998).

Acknowledgements

The first C++ implementation of MT Workshop was sponsored by SBIR DAAL01-97-C-0016 from the Army Research Laboratory (ARL). The Java implementation, the graphical interface and XML enhancements were developed under SBIR DAAD17-99-C-0008 from ARL. We are thankful to Chunhye Han, Meesook Kim, Nari Kim, and Eon-Suk Ko for their comments and evaluation of MT Workshop. We are also thankful to Ted Caldwell, Daryl McCullough, Alexis Nasr and Mike White for their comments on and criticism of the reported work

References

Lavoie, B., Kittredge, R., Korelsky, T. and Rambow, O. (2000) A Framework for MT and Multilingual NLG Systems Based on Uniform Lexico-Structural Processing. To Appear in *Proceedings of the Fifth Conference on Applied Natural Language Processing*, Seattle, Washington.

Lavoie, B. and Rambow, O. (1997) A Fast and Portable Realizer for Text Generation Systems. In *Proceedings of the Fifth Conference on Applied Natural Language Processing*, Washington, DC.

Mel'cuk, I. (1988) *Dependency Syntax*. State University of New York Press, Albany, NY.

Palmer, M., Rambow, O. and Nasr, A. (1998) Rapid Prototyping of Domain-Specific Machine Translation Systems. In *Proceedings of the Third Conference on Machine Translation in the Americas (AMTA-98)*, PA, USA.

Yoon, J., Kim, S., and Song, M. (1997) New parsing method using global association table. In *Proceedings of the 5th International Workshop on Parsing Technology*.