Compositional Translation Revisited

Jan Landsbergen

Utrecht Institute of Linguistics OTS
Trans 10, 3512 JK Utrecht, The Netherlands
landsbrn@let.uu.nl

Compositional Translation is an approach to (the linguistic aspects of) Machine Translation, which has been developed in the Rosetta-project, in the eighties. It is extensively described in Rosetta (1994). Recent work by Huijsen (1998) gives occasion to consider the underlying ideas once again, in a wider context.

The compositional approach is based on the following principle: *Two expressions are translation-equivalent if they can be built up from parts which are translation-equivalent by means of translation-equivalent rules.* The principle suggests a system that operates as follows: (i) first an analysis module must find out how a sentence of the source language (SL) can be built up from its smallest translatable parts by means of composition rules, (ii) then a transfer module relates these parts to translation-equivalent parts of the target language (TL) and the composition rules of SL to translation-equivalent rules of TL, (iii) finally a TL generation module applies the translated composition rules recursively to the translated parts.

If we elaborate the notion of translation-equivalence and require it to comprise logical equivalence, the composition rules must have a well-defined truth-conditional semantics and corresponding rules of SL and TL must be logically equivalent. The question is: how do we arrive at an adequate set of such corresponding composition rules? Taking into account that a translation system should include information about the grammars of SL and TL anyway, it lies at hand to describe the languages by means of compositional grammars (grammars of which the rules have a well-defined truth-conditional semantics) and to relate the composition rules needed for translation to the compositional grammar rules.

This idea has been worked out in the Rosetta-system in a specific way. The compositional grammars of SL and TL must be *isomorphic*: basic ex-

pressions of SL must correspond to one or more translation-equivalent basic expressions of TL and each rule of SL must correspond to one or more translation-equivalent rules of TL. Sentences are considered translations of each other if they have isomorphic derivation trees, i.e. if they can be derived by means of corresponding rules from corresponding basic expressions. So, here the grammatical analysis is identical to the compositional analysis on behalf of translation. An attractive property of this method is that a translated sentence can be guaranteed to be syntactically correct and to have the same truth-conditional meaning as the original (if the grammars are correct and correctly related).

The isomorphic grammar method has been criticized because it necessitates the grammars of SL and TL to be attuned to each other. It was expected that writing mutually dependent isomorphic grammars would be more difficult than writing grammars for the individual languages and that attuning would lead to 'pollution' of the grammars. In actual practice this was not really a problem, thanks to the power of the grammar formalism (M-grammar) that was used in Rosetta. The rules of M-grammars operate on constituent trees; their applicability conditions relate not only to the syntactic category but also to the syntactic structure of their arguments; application of the rules may include operations like deletion, permutation, insertion, as long as certain conditions, such as rule reversibility, are satisfied. So, translation-equivalent rules can have quite different effects on the surface structure. Thanks to this and other features of M-grammars the required attuning could be achieved rather painless. However, the power of the grammars caused other problems. One of these problems related to *translation completeness*, the property that each correctly analyzed SL sentence will indeed be translated. Unfortunately, due to the complexity of the M-grammar rules, completeness of the Rosetta-systems could not be proven.

Recently it has been investigated by Huijsen (1998) how the original idea of compositional translation can be combined with the use of simple concatenative grammars, in particular context-free grammars. For some applications, such as translation of controlled languages, a simple grammar formalism is feasible and preferable. But then the technique of isomorphic grammars is too restrictive. While in M-grammars there is a clear distinction between derivation structure and constituent structure, in context-free grammars these are similar. The use of isomorphic context-free grammars would then imply that only sentences with similar constituent structures could be translation-equivalent. Apparently, the use of simple concatenative grammar formalisms forces us to complicate the definition of the composi-

tional translation relation, to allow a form of 'structural transfer'.

Elaborating the algebraic foundations of the Rosetta framework, Huijsen developed a translation method, called *polynomial translation*, where the translation relation is defined between combinations of rules, polynomials, rather than between individual rules. This enlarges the power of the translation relation, while maintaining the advantages of isomorphic translation such as guaranteed meaning preservation and syntactic correctness. Moreover, for polynomial translation between context-free languages necessary and sufficient conditions can be formulated such that translation completeness can be proven. This is important for the intended application, translation of controlled language, where the users are the authors of the texts to be translated, who need not be familiar with the target language. They have to make a special effort in order to formulate their texts within the limits of a restricted grammar and should then be rewarded by a reliable automatic translation into one or more target languages.

Because of its logical foundations, the isomorphic grammar method has sometimes been compared with translation methods which use a logical language as an interlingua. The latter approach seems to have the advantage that attuning of the grammars is not necessary. This is not quite correct: translation completeness in such systems can only be achieved with a certain degree of attuning in order to guarantee that translations relate to identical (not just equivalent) logical expressions. But a more important weakness of translation via a logical interlingua is the loss of all translation-relevant information that cannot be expressed in logic. The isomophic grammar method provides the possibility to transfer this type of information: the system developer can select from the pairs of logically equivalent SL and TL rules those that are also similar in other, non-logical, aspects. Only those rules are considered translation-equivalent. In a polynomial translation system this idea can be exploited much further. The system developer can now select from the many pairs of logically equivalent rule *combinations* those pairs that agree optimally in other translation-relevant aspects.

Huijsen, W.O., Completeness of Compositional Translation, Doctoral dissertation, Utrecht Institute of Linguistics OTS, 1998.

Rosetta, M.T., Compositional Translation, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1994.