Issues in Generating Text from Interlingua Representations

Stephan Busemann

DFKI GmbH
Stuhlsatzenhausweg 3
D-66123 Saarbrücken
`busemann@dfki.de`

Abstract. Multi-lingual generation starts from non-linguistic content representations for generating texts in different languages that are equivalent in meaning. In contrast, cross-lingual generation is based on a language-neutral content representation which is the result of a linguistic analysis process. Non-linguistic representations do not reflect the structure of the text. Quite differently, language-neutral representations express functor-argument relationships and other semantic properties found by the underlying analysis process. These differences imply diverse generation tasks. In this contribution, we relate multi-lingual to cross-lingual generation and discuss emergent problems for the definition of an interlingua.

1   Introduction

In this contribution, we relate multi-lingual to cross-lingual generation and discuss emerging problems for the definition of an interlingua. Multi-lingual generation starts from non-linguistic content representations for generating texts in different languages that are equivalent in meaning. The generation of weather forecasts or environmental reports are typical examples. In contrast, cross-lingual generation is based on a language-neutral content representation which is the result of a linguistic analysis process. Generation for machine translation is a most prominent example.

Non-linguistic representations do not specify linguistic semantics nor do they reflect the structure of the text to be generated. In contrast, language-neutral representations express functor-argument relationships and other semantic properties found by the underlying analysis process. These differences imply diverse generation tasks.

However, there are also commonalities. In both cases, generation is the mapping of some semantic representation onto linguistic strings. We may assume a single generation process that uses different separately defined language specific knowledge sources. In both cases, we may view the underlying representation as an interlingua, since it attempts to cross the language barrier by providing content descriptions independently of the target language.

An instance of each type of tasks has been implemented using the generation system TG/2 (Busemann, 1996), quickly overviewed in Section 2. The usage of the same framework allows us to relate the tasks to each other (Section 3) and to gain insights