



Dana Dannélls

With the rapid growth of the Semantic Web, an increasing share of the information available on the internet will be in a form designed for processing by machines, and not directly comprehensible to people. This creates a need for user interfaces capable of rendering this information in written or spoken language, and since we live in a multilingual world, in more than one language.

This thesis explores the optimal ways in which natural language generation techniques can be brought to bear upon the problem of processing a structured body of information in order to devise a coherent presentation of text content in multiple languages. We investigate through cross-linguistic studies how coreference is expressed in English, Swedish and Hebrew and suggest both discourse and coreference strategies to guide the generation of paragraph-sized descriptions about artworks from formal structured representations presented in the Semantic Web. We show how these strategies are incorporated into a multilingual generation application and demonstrate how it successfully produces coherent, well-formed descriptions in all three languages.

Dana Dannélls / Multilingual text generation from structured formal representations

Multilingual text generation from structured formal representations

Dana Dannélls



GÖTEBORGS
UNIVERSITET

ISBN 978-91-87850-48-6
ISSN 0347-948X



23•2012



Data linguistica