ABSTRACT

THE TRANSLATOR'S ASSISTANT:
A MULTILINGUAL NATURAL LANGUAGE GENERATOR
BASED ON LINGUISTIC UNIVERSALS, TYPOLOGIES,
AND PRIMITIVES

Tod Allman, PhD.

The University of Texas at Arlington, 2010

The Translator's Assistant (TTA) is a multilingual natural language generator (NLG) designed to produce initial drafts of translations of texts in a wide variety of target languages. The four primary components of every NLG system of this type are 1) the ontology, 2) the semantic representations, 3) the transfer grammar, and 4) the synthesizing grammar. This system's ontology was developed using the foundational principles of Natural Semantic Metalanguage theory. TTA's semantic representations are comprised of a controlled, English influenced metalanguage augmented by a feature system which was designed to accommodate a very wide variety of target languages. TTA's transfer grammar incorporates many concepts from Functional-Typological grammar as well as Role and Reference grammar. The synthesizing grammar is intentionally very generic, but it most closely resembles the transformational-generative model. The meaning-based theory of translation underlies the TTA system.

The fundamental question that this research proposes to answer is as follows: if the semantic representations contain sufficient information, and if the grammar possesses sufficient capabilities, then is TTA able to generate drafts of sufficient quality that they improve the productivity of experienced mother-tongue translators? To answer this question, software was developed that allows a linguist to build a lexicon and grammar for a particular target language. Then semantic representations were developed for

one hundred and five chapters of text. Four unrelated languages were chosen to test the system, and a partial lexicon and grammar were developed for each test language: English, Korean, Jula, and Kewa. Three chapters of text were generated in Jula, four chapters of text were generated in Kewa, fifty chapters were generated in Korean, and all one hundred and five chapters were generated in English. Then extensive experiments were performed to determine the degree to which the computer generated drafts improve the productivity of experienced mother-tongue translators. Those experiments indicate that when experienced mother-tongue translators use the rough drafts generated by TTA, their productivity is typically quadrupled without any loss of quality.