Standardization of the Generation Process in a Multilingual Environment

Jesús Cardeñosa, Carolina Gallardo and Edmundo Tovar

Universidad Politécnica de Madrid, 28660 Madrid.
{carde,carolina,edmundo}@opera.dia.fi.upm.es

Abstract. Natural language generation has received less attention within the field of Natural language processing than natural language understanding. One possible reason for this could be the lack of standardization of the inputs to generation systems. This fact makes the systematic planning of the process of developing generation systems to become difficult. The authors propose the use of the UNL (Universal Networking Language) as a possible standard for the normalization of inputs to generation processes.

1   Introduction

In natural language processing (from now on NLP) two areas can be differentiated: analysis and generation. However, one has not received the same attention as the other from the scientific community, that is why generation can be considered as the "poor brother" of the NLP. The reason for this minor development is the different nature of the input to the analysis and generation systems. The input to the analysis systems is always natural language, whose casuistic and phenomenology are known; while in a generation system, the output is always known, but not what it is going to generate from [1].

The input to a generation system varies depending on whether it is monolingual generation (dialogue systems) or a multilingual system (mainly machine translation systems). In dialogue systems it is difficult to establish appropriate characteristics common to all inputs, because "the problem" of generation is usually solved with solutions *ad hoc*, depending on the application and the system language. In machine translation systems, there are also many differences in the inputs to the generation subcomponents, conditioned by the nature of system architecture (transfer, interlingua, etc.), the kind of grammars being used (declaratives vs. procedural) [2], or the number of languages in the system.

This difference in the input to the generators makes a systematic planning of their development process impossible (main cause of the minor development of generation compared to analysis). It is necessary then, that the input to the "generator" can be supported with an appropriate model of contents representation, separated from the format or language that ensures a standard process for the development of generation systems.

In this article we propose the UNL as a possible standard for the generation inputs. To achieve this, in section 2 we will introduce the main generation architectures. Sec-