Integration of Natural Language Processing Techniques in the ICALL systems field: The treatment of incorrect knowledge

Montse Maritxalar, Arantza Diaz de Ilarraza
Lengoaia eta Sistema Informatikoak Saila
University of the Basque Country, 649 Postakutxa, 20080 San Sebastian, Spain
E-Mail: maritxalar@si.ehu.es

The aim of this poster is twofold. Firstly, the treatment of the incorrect knowledge is studied. Our objective is to detect that incorrect knowledge and the use of this knowledge for diagnosis. Secondly, the necessity of the integration of natural language techniques in the world of ICALL systems is claimed.

Since basque (an agglutinative language) is the second language we are considering, a brief description of basque morphology is presented in the poster. A general idea of XUXEN, a Spelling/Checker Corrector for basque based on Two-level Morphology which will be used in the system, is given as well.

Three different kinds of errors are differentiated: **lapses** or occasional errors, **superficial errors** which include real systematic surface errors and **deep errors** which represent the cause/s of the superficial errors. As the type of errors we have examined are restricted to the word-level, only morphonological and morphosyntactical errors have been checked. Next, we are giving a classification of superficial errors which has been abstracted from the study of a corpus of upper-intermediate learners' texts, and according to existing studies about errors in basque language learning. The occurrence rates detected in the corpus are the following: spelling errors of adding (2.7%), removing (11.7%) or replacing (32.6%) a letter inside the root; wrong application of a morphonological rule (1.3%); two different morphemes are consecutive in not permitted order, adding, removing or replacing affixes (2.6%); use of incorrect affixes (4.6%); use of incorrect roots (25.2%); double words, that is, the word has to be written in two different words, joined or not by a hyphen (6.3%); typing errors (7.8%) which are not due to any misconception about the language; other errors (5.2%).

Since the differentiation between superficial errors and deep errors, and between detection and diagnosis is taken into account in the Error Analysis procedure (Ellis 85), we have chosen such a procedure as our basis for the treatment of errors.

The solution we propose in our ICALL system for the representation of the systematic incorrect knowledge is a module which will be updated by the human tutor and also by the computer tutor. Some predicted errors will be stored, however the OKER-XUXEN detector/diagnoser of errors made by language learners of basque, will be able to detect unpredicted errors and generalize them in some cases in order to be added to the incorrect knowledge module during a learning session. This process is based on three different elements: **the HIZTOKER dictionary** (erroneous morphemes which are directly linked to the corresponding correct ones in the main lexicon, *the domain*) **the ALTERATION RULES module** (two-level rules where typical morphonological errors will be represented) and **the WORD-FORMATION module** (where incorrect knowledge about word-formation will be included).

References

Agirre, E., Alegria, I., Artola, X., Diaz de Ilarraza, A., Maritxalar, M., Sarasola, K. & Urkia, M. (1992). XUXEN: A Spelling Checker/Corrector for Basque Based on Two-Level Morphology. In *Proceedings of ANLP-92*, 119-125, Trento, Italy.

Koskenniemi, K. (1983). Two-level morphology: A general computational model for word-form recognition and production. University of Helsinki, Department of General Linguistics.