Automatic Exercise Generation in an Essay Scoring System

Itziar ALDABE[a], Montse MARITXALAR[a*], Olatz PEREZ DE VIÑASPRE[a], Larraitz URIA[b]
[a]*IXA NLP Group, University of the Basque Country UPV/EHU, Spain*
[b]*IKER (UMR5478), IKERBASQUE*
*montse.maritxalar@ehu.es

Abstract: In this paper, we present an automated essay scoring system for language learners to improve their writing skills. The system combines statistical and Natural Language Processing (NLP) techniques to compute the evaluation measures of the essays. It also incorporates a module to automatically generate multiple-choice questions (MCQ) to test the language level in Basque. The automatic generation of the options that constitute the MCQs is based on the linguistic information compiled in the rules written for a Basque grammar checker, which detects, among others, determiner errors.

Keywords: generation of MCQs, evaluation of essays, NLP in ICALL

Introduction

In the last two decades several research has been developed on computer-based automated essay scoring (AES) systems [4] which provide students with feedback to improve their writing. Although some research communities have criticized this type of system, recent studies [2] demonstrate that AES systems help improving the school performance. One of the advantages is that they measure all essays using the same scoring model. This way, AES systems offer "objective" data to improve on the controversial task of manual essay evaluation.

The essay scoring system presented in this paper combines statistical and Natural Language Processing (NLP) techniques to compute the evaluation measures of the essays. The application, named IAS, uses a client-server architecture [1]. On the one hand, the client interprets the results calculated by the server in order to decide the evaluation measures that the application will give to the user as the result of the evaluation. On the other hand, the server includes a request manager which triggers its language modules depending on the language requested by the client. Those language modules are responsible for the linguistic process of the essay. By means of this linguistic process the texts are analysed and students' errors are detected. Both the interface and the application itself are multilingual. So far, we have integrated NLP tools for two very different languages, such as Basque and Spanish.

In this article, we present an improvement of the system, where a new module to automatically generate MCQ questions for Basque has been incorporated. The automatic generation of the options that constitute the MCQs is based on the linguistic information compiled in the rules written for a Basque grammar checker, which is also integrated in IAS.

1. Automatic Generation of MCQs to Improve on the Use of Determiners

The experience acquired by teachers when teaching a language is a valuable source when establishing the criteria for creating tests. In contrast, a learner corpus containing a collection of errors is an alternative way of defining the heuristics required to generate learning material. In our opinion, the use of learner corpora for creating exercises offers interesting advantages: (i) error analysis studies on learner corpora show us a realistic picture of the written output of learners; (ii) they provide us with objective data about the typical error types; and (iii) the development of grammar checkers based on corpora to detect and diagnose errors provides us with information on error diagnosis that can be used for the automatic generation of learning material.

Determiner errors are relatively common among Basque language learners. The Basque learner corpora collected in the last years have provided us with enough data to perform a deep study of this phenomenon. Based on that information, we have created a grammar to automatically detect some types of determiner errors using the Constraint Grammar (CG) formalism [3]. Based on the linguistic information compiled in the rules of the grammar checker, IAS generates MCQ tests to check the use of determiners in Basque.

The next example presents an automatically generated MCQ to deal with the R_DET error type (repetition of the determiner). The source sentence, which is a text written by a learner of Basque, contains a R_DET error (option d) and the system generates the rest of the options, i.e. the correct answer (option a), and the two distractors (options b and c).

Nire bizitzaren .... orain dela 5 urte gertatu zen (... *in my life took place 5 years ago).*

a) egun zoriontsu **bat** (**one** happy day)

b) egun**a** zoriontsu**a bat** (**a one** happy the day)

c) egun**a** zoriontsu **bat** (**one** happy the day)

d) egun zoriontsu**a bat** (**the one** happy day)

In order to create the above presented MCQ example, the system first detects two determiners in the same DP in the source sentence (option d). Then, it deletes the determiner from the adjective (zoriontsua - zoriontsu) to generate a correct DP (option a). The new DP is set as the correct answer of the MCQ. In order to generate new errors, i.e., the distractors, different strategies are applied: (i) to add determiners to various words in the same DP; (ii) to delete the determiner to all the elements in the DP; (iii) to change the declension mark of the phrase; (iv) to change the order of the words inside the phrase, etc. In the example above, both distractors are created by adding the definite article -a to different words inside the DP (options b and c).

2. Experiment

This experiment deals with the correct use of determiners in Basque by means of the use of IAS. Here we present the results related to the R_DET error type, since it is one of the most common error. A linguist first supervised the automatically detected errors in a text written by a Basque language learner. Then, she manually added new error instances in order to create different types of determiner errors. After that, the system automatically generated the MCQs embedded in the text.

The linguist was responsible for creating the different types of R_DET errors which were previously observed and detected in the Basque learner corpus. She had to establish different examples of R_DET errors, obtaining the highest level of casuistry. In total, she marked 15 Determiner Phrases (DPs) out of a total of 68 within the text, that is to say,

22.05% of the DPs. Once erroneous DPs were defined, the system generated the corresponding correct answer for each of them as well as two more distractors. The test generated for the experiment was answered by thirty low-language-level learners from Argentina, who took part in this experiment on-line. Those students were not therefore the writers of the Basque learner corpus. The experiment was carried out with low language level learners since the most R_DET errors have been detected in the mentioned corpus (proportionally and with a statistically significant difference) at this level.

Based on the fact that 60% of the students (18) correctly answered more than 11 MCQs, we could say that the generated test was not difficult for most of them. There was a significant difference in percentages when comparing the number of errors made by the Argentinian students when filling the MCQs and those errors detected in the Basque language corpus. Based on the results of the students who answered incorrectly at least 4 MCQs, the percentage of R_DET errors per DP of Argentinian students was 5.88% or higher. In contrast, the percentage in the low-level sample of the Basque language corpus was 0.77%. These results can lead us to consider (i) that offering this type of test could cause the emergence of new errors that students would not produce in an open task; or (ii) that perhaps foreign language learners of Basque make different errors than those made by second language learners of Basque.

3. Conclusions

IAS is an automated essay scoring system for students of Basque to improve their writing. The system is the core of the first bilingual web application developed to handle Basque and Spanish essays. IAS incorporates a module which automatically generates MCQ to test the Basque language level. The automatic generation of the options that compose the MCQs is based on the linguistic information compiled in the rules written for a Basque grammar checker, which is also integrated in the system.

The results of the experiment leave open a new research challenge related to the type of exercise that should be proposed to improve the writing of language learners. In addition, we should study the type of learner corpus to be used as the basis for the generation of language exercises.

Acknowledgements

This work has been supported by Olatz Perez de Viñaspre PhD grant for the Basque Government and the OpenMT-2 project (TIN2009-14675-C03-01).

References

[1] Castro-Castro, D., Lannes-Losada, R., Maritxalar, M., Niebla, I., Pérez-Marqués, C., Alamo-Suarez, N.C., & Pons-Porrata, A. (2008) A Multilingual Application for Automated Essay Scoring. *Lecture Notes in Advances in Artificial Intelligence - LNAI 5290* - Springer New York 243-251.

[2] Grimes, D. y Warschauer, M., (2010). Utility in a Fallible Tool: A Multi-Site Case Study of Automated Writing Evaluation. *Journal of Technology, Learning, and Assessment*, 8 (6).

[3] Karlsson F., Voutilainen A., Heikkila J. & Anttila A. (1995) Constraint Grammar: a Language Independent System for Parsing Unrestricted Text. Mouton de Gruyter.

[4] Landauer, T. K., Laham, D., Foltz, P. W.: Automated Essay Scoring and Annotation of Essays with the Intelligent Essay Assessor. In: Mark D. Shermis and Jill C. Burstein (Eds.). Automated Essay Scoring: A Cross Disciplinary Perspective, pp. 87-112. Mahwah, NJ: Lawrence Erlbaum Associates (2003)