Chapter 80
"Evaluator": A Grading Tool for Spanish Learners

Paz Ferrero
Universidad Autónoma de Madrid, Spain & Ciudad Universitaria de Cantoblanco, Spain

Rachel Whittaker
Universidad Autónoma de Madrid, Spain & Ciudad Universitaria de Cantoblanco, Spain

Javier Alda
University Complutense of Madrid, Spain

ABSTRACT

Computational linguistics can offer tools for automatic grading of written texts. "Evaluator" is such a tool. It uses FreeLing as a morpho-syntactic analyzer, providing words, lemmas, and part of speech tags for each word in a text. Multi-words can also be identified and their grammar identified. "Evaluator" also manages leveled glossaries, like the one developed by the Instituto Cervantes, as well as other electronically available dictionaries. All these glossaries enable the tool to identify most words in texts, grading them into the six levels scale of the Common European Framework of Reference for Languages. To assign a lexical level to the text under analysis, a statistical distribution of leveled qualified lemmas is used. Other ways to assign a lexical level to a text by using corpora of a preset level are also suggested. The syntactic analysis is based on a collection of grammar structures leveled by following the descriptors given by the Instituto Cervantes. These grammar structures are identified within the text using quantitative indices which level a text by comparing it with a given corpus. Finally, semantic identification is done using semantic fields as defined by the Instituto Cervantes. Latent Semantic Analysis is also used to group texts dealing with the same topic together. All these methods have been tested and applied to real texts written in Spanish by native speakers and learners.

INTRODUCTION

This contribution is based on the PhD dissertation titled *Definition and analysis of quantitative linguistic parameters for automatic assessment tools applicable to Spanish as a Foreign Language* (Ferrero 2011), in which the three authors of this chapter were involved. It describes and applies some automatic methods to quantitatively evaluate reference levels of texts written in Spanish by

DOI: 10.4018/978-1-4666-6042-7.ch080

learners of Spanish as second foreign language to obtain the "Diploma de Español como Lengua Extranjera" (DELE), according to the six main levels described in the Common European Framework of Reference (CEFR): A1-A2, B1-B2, C1-C2. Official human evaluators had previously and independently scored those texts we analyzed automatically. The reference level is given with a confidence parameter that measures the reliability of the level given. In this section, we present an overview of the process, before explaining the sources we have used and describing in detail the modules which make up "Evaluator." We have, then, designed a prototype whose morpho-syntactic analyzer, FreeLing (Padró, 2006; Padró, 2009), is the tool used as the first step to process the lemmas of the Spanish texts. After this, those lemmas are automatically assessed lexically, syntactically, and semantically.

The evaluation of lexical competence is based on the classification of lexis into the levels given by the Plan Curricular del Instituto Cervantes (PCIC), our golden standard for the study. For those words not included in the PCIC list, other criteria have been developed based on the presence of those words in a combination of selected glossaries. Moreover, multiword units are also evaluated to refine the determination of the lexical level of the text. After this, the whole text is assessed with a quantitative level and given a numerical grade by applying a method derived from Zipf's law (Zipf, 1932).

The syntactic level is given by identifying categorized syntactic structures in the text being evaluated. A wide collection of syntactic structures has been created and leveled based on the PCIC, with the level being assigned depending on the complexity of the structure. The identification of these leveled structures is compared against a reference corpus with a predetermined level.

At the semantic level, we identify the content of a text following the semantic fields given by the PCIC. Then, after using customized corpora, it is possible to correlate the semantic fields of the text being tested quantitatively with the corpora. Besides this, Latent Semantic Analysis (LSA) groups texts in semantic clusters that can be identified and used for the evaluation of semantic content.

The methods developed have been tested on texts written by learners of Spanish. The results obtained using the automatic graders correlate well with the scores given by the human graders. This system is conceived to help learners of Spanish to self-assess their texts, giving them a level. "Evaluator" is able to distinguish texts among several different text contents, and to check grammatical complexity and lexical level of those texts following descriptors from the Plan Curricular of Instituto Cervantes (PCIC).

The PCIC is based on the Common European Framework of Reference (CEFR). The CEFR document recognizes that "establishing cut-off points between levels is always a subjective procedure" (Cervantes, 2006, p. 32) when situating a student text at a specific level. However, using linguistic descriptors and an automatic tool to do the grading, assessment should respect a standard, and be objective and systematically reliable. Our aim with this tool was to achieve this reliability in the assessment of learners' written texts. Figure 1 shows the general design of the tool we have been describing.

We conceive "Evaluator" as a compact tool (see Figure 1). At present, in its experimental stage, we input a text and the tool outputs a leveled text according to its vocabulary, grammar, and content independently. In this contribution we describe the basic mechanism of "Evaluator" as a prototype. However, our aim is to create an automatic tool to evaluate texts written by learners of Spanish as a second language which can be used by the learners themselves to assess their progress.

STATE OF THE ART

From the beginning of this research in 2006 until its completion in 2011, we have been tracking the development and research in automatic evaluation