AVI.CAT: TURNING GRAMMAR CHECKING INTO AN E-LEARNING EXPERIENCE

Antoni Oliver, Salvador Climent, and Marta Coll-Florit
Universitat Oberta de Catalunya (UOC)
Av. Tibidabo 39-43 08035 Barcelona (Spain)

ABSTRACT

Spelling and grammar checking has become a daily activity for almost all text processor users. Usually these tools offer limited information about the misspelling or the grammar error and in certain cases suggest one or more possible alternatives. Sometimes users make the same mistakes one day after the other because they don't know the real reason of the mistake. In these cases extended information about the error could be very handy: detailed grammatical information and exercises to improve the user's writing skills. In this paper we present AVI.cat, a grammar checker and virtual assessor for the improvement of writing skills in Catalan. The tool is based on LanguageTool grammar checker and fully integrates into OpenOffice/LibreOffice text processor. Along with the grammar checker with extended information collected from the web, the tool also offers an automatic evaluator and assessor, that can perform an automatic assessment from a collection of texts written by the user. The assessor can also give information about the use progress and suggests exercises for further improvement.

KEYWORDS

Grammar checking, writing skills, language learning

1. INTRODUCTION

In this paper we present the development of AVI.cat, a tool for grammar checking of Catalan documents that goes a step beyond simple grammar checking and acts also as an assistant for writing and grammar skills improvement. The tool is constructed from a well known free grammar checker, LanguageTool, and fully integrates into OpenOffice/LibreOffice text processors.

Catalan is a language with 10 million speakers, most of them in Spain. AVI.cat is conceived in response to the need to improve the writing skills of UOC students. UOC is a virtual university where 80% of its 60.000 students are taught in Catalan.

The use of spell and grammar checkers is a common activity among the users of text processors and other office and productivity tools. Depending on the user's writing skills, some of the corrections proposals are accepted or rejected without a full knowledge of the given error. This leads to the repetition of the same errors one document after the other. Spell and grammar checking is a good opportunity for learning grammar and improving writing skills. Nevertheless, the information given by the grammar checker is very scarce as usually it only gives a very short explanation and suggests some substitutions.

Increasing the amount of information given by the grammar checker could be useful for the user willing to know the reason of the error. This information can be a detailed grammatical explanation and some exercises. Fortunately we can find a lot of freely accessible links to grammar and exercises for several languages. The idea is to provide a link in the grammar checker dialog. If the user clicks in this link, the default web browser opens and shows the grammar information related to this error and in some cases along with some exercises.

The tool we present in this paper is developed for the Catalan language. There are several reasons for the special need of this kind of tool for this language. Firstly, the deep contact with Spanish, a very close related language, but with very different spelling and grammar rules. Most of the Catalan speakers living in Spanish territory are fully bilingual; moreover an increasing number of people writing in Catalan are new non-native speakers. This situation leads to frequent spelling and grammar errors in documents written by these users.

This poses a fundamental challenge to our society since communication skills have been rated as very important or rather important by 96% of the employers according to the Eurobarometer Survey in 2010 (European Commission 2010). Secondly, the lack of a good grammar checker for this language. Although in the last years some grammar checkers have been developed for Catalan, none of these have reached the desired level of development and integration to office productivity tools.

2. COMPONENTS

The development of a grammar checker from scratch is a very hard and time consuming task. A lot of aspects must be faced: programming and integration to office productivity tools, a deep grammatical study must be performed in order to define the errors the system must catch, a formalism for the rules must be defined and a set of rules must be written using this formalism. Fortunately, most of this work has already been done in others projects and our team could concentrate on two main tasks: the creation of a set of rules and the selection of the links to grammar resources and exercises for each rule. Lastly, we developed the automatic assessment module.

2.1 LanguageTool

Language Tool (Naber, 2003) (www.languagetool.org) is an open source grammar checker for several languages, including Catalan. In fact, LanguageTool provides a set of generic tools for developing grammar checkers for any language. LanguageTool is developed in Java and fully integrates with OpenOffice/LibreOffice. The rules are expressed in XML files and the formalism is rich and most of the errors can be expressed in such way. For more complex errors, additional rules can be written in Java.

As we have already said, LanguageTool provides a set of grammar tools for Catalan (a total of 216 rules), along with a tagger that assigns all possible POS tags to each word in a sentence.

2.2 New Rules for Catalan

We have developed a completely new set of rules for Catalan, based on the specifications of *El Corrector* (Badia et al., 2004). These specifications describe each detected error and also offer a detailed classification of errors. In total, we developed 236 rules grouped in 19 categories. Although the number of rules is not much higher than the default rules for Catalan, the inclusion of more complex patterns in the rules means that the new set is able to detect a higher number of grammatical errors. In Figure 1 we can observe a fragment of the new rules.

```xml
<category id="A2" name="Apostrofació d'articles, preposició de i pronoms febles">
...
<rulegroup id="2.2" name="Apostrofació de preposicions">
  <rule id="2.2.1" name="No apostrofació de de davant de mots amb vocal inicial">
    <pattern>
      <token>de</token>
      <token regexp="yes">[aeiouhàèéíòóú].*<exception regexp="yes">iode|iogurt|ioga|iot|en</exception></token>
    </pattern>
    <message full-explanation-url="http://cml.uoc.edu/avicat/2_2_1.html">
        Hi manca un apòstrof. Cal dir <suggestion>d'<match no="2"/></suggestion>.
    </message>
    <short>Error gramatical</short>
    <example type="incorrect"><marker>de història</marker></example>
    <example type="correct"><marker>d'història</marker></example>
  </rule>
  ...
</rulegroup>
...
</category>
```

Figure 1. An example of AVI.cat rule

The organization of the rules by categories and groups allows us a greater control on the development and facilitates the maintenance tasks. This organization is also crucial for the development of the automatic assessment module.

2.3 Links to Grammar and Exercises Resources

In Figure 1 we can also observe a new feature of the new set of rules: the inclusion of the *full-explanation-url* attribute. This attribute provides a link to the grammar and exercises associated with the detected error. This link points to a specific web page for each rule. This web page is hosted in one of our servers and redirects to grammar and exercises. This fact also facilitates the maintenance of the system, as we can change the final web pages as many times as we want, with no modification in the XML containing the rules. Some rules share the same final pages, but we can change the content simply editing the html files in our server.

In Figure 2 we can see AVI.cat in action. As we can observe, the dialog is almost the same than the LanguageTool dialog, but includes a *More...* link. Clicking on this link the default browser will open the page containing the detailed information.
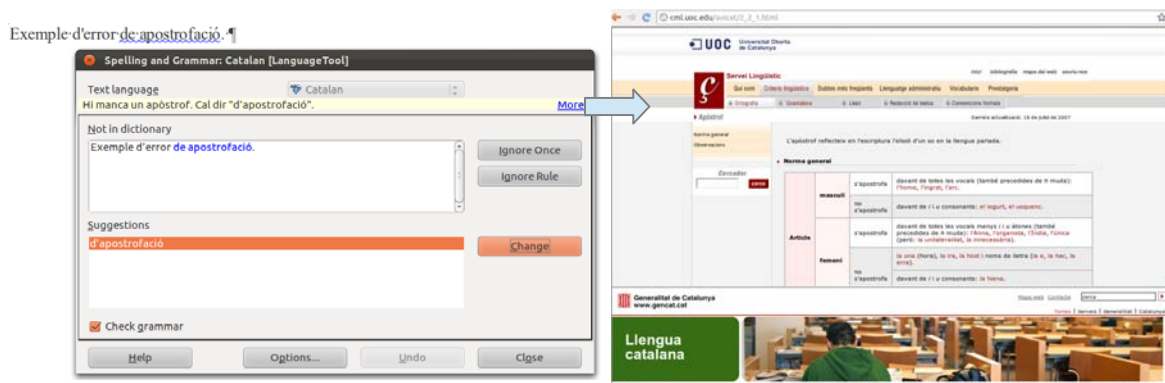

Figure 2. AVI.cat in action

2.4 Automatic Assessment Module

The automatic assessment module, under development, takes a set of documents and counts the number of errors for each category, rule group and rule id. Two new parameters will be included in each rule: an error severity score and a confidence score. These parameters will allow the calculation of a weighted score for each error category and a global value as well. The application will store the values for each user and will be able to detect the improvements, if any, of the user's writing skills. The assessment module will also suggest some grammatical readings and exercises.

3. USAGE SCENARIOS

In this section we present some of the scenarios where AVI.cat can be a very useful tool.

AVI.cat can work as a traditional grammar checker, so it can be used by any user of text processors writing in Catalan. The system will detect grammatical errors as the user types, as well as it can be executed under demand.

The system can be also used by language learners of any level. Although when developing the system we had in mind a Catalan native speaker, the range of detected errors is wide enough for the tool to be used by foreign learners of Catalan.

At UOC we have also successfully used the tool in subjects related to language technologies and computational linguistics. As the tool provides a tagger and a powerful formalism to detect language patterns, it can be used in several learning activities. One of the most successful activities is the creation of new rules for AVI.cat and the testing of the rules in real texts.

We are also working in the integration of AVI.cat to virtual classrooms at UOC as an advisor for the lecturer, as explained in the following section.

4. FUTURE DEVELOPMENTS

We plan to further develop AVI.cat, adding new functionalities as well as creating rules for other languages.

One of the extensions that has already got founds for the development is the language assessment and monitoring at the Virtual Campus of the UOC. This extension will allow to check all the documents written by the students as practical works and to send reports to the tutors in case of detecting low writing skills.

This new module will be also developed for English. This means that new rules for this language must be developed, or at least, the default rules should be revised and grouped in several categories.

One important aspect of this kind of tools is the improvement of the existing rules and the creation of rules for other errors. We plan to develop an on-line environment where users can suggest new errors to detect, submit and test new rules and also report errors.

5. CONCLUSION

In this paper we have briefly presented a tool for the improvement of writing skills in Catalan. The tool acts as a grammar checker but provides much more detailed information about the errors. Thus, AVI.cat can also work as a virtual assessor for the student or as an assessment and monitoring tool for the lecturer by analyzing a set of documents. With AVI.cat we can turn the everyday activity of revising texts into a profitable e-learning activity.

ACKNOWLEDGEMENT

REFERENCES

Badia, T., A. Gil, M. Quixal, and O. Valentín. 2004. "NLP-enhanced Error Checking for Catalan Unrestricted Text." In *Proceedings of the Fourth International Conference on Language Resources and Evaluation, LREC*, 1919–1922.

Domingo, Judith, Guillem Massó, Òscar Puente, Bernat Grau, Toni Badia, and Martí Quixal. 2008. *Especificacions D'errors d'El Corrector*. GLiCom, Grup de Lingüística Computacional. Universitat Pompeu Fabra.

European Commision. 2010. Eurobarometer survey: Employers' perception of graduate employability. http://europa.eu/rapid/pressReleasesAction.do?reference=MEMO/10/638&type=HTML . Last accessed 23.03.2012

Naber, D. 2003. A Rule-based Style and Grammar Checker. Master's thesis, University of Bielefeld.