A Case-based Reasoning Approach to Validate Grammatical Gender and Number Agreement in Spanish language

Jorge Bacca, Silvia Baldiris, Ramon Fabregat, Cecilia Avila

University of Girona, Spain

Abstract — **Across Latin America 420 indigenous languages are spoken. Spanish is considered a second language in indigenous communities and is progressively introduced in education. However, most of the tools to support teaching processes of a second language have been developed for the most common languages such as English, French, German, Italian, etc. As a result, only a small amount of learning objects and authoring tools have been developed for indigenous people considering the specific needs of their population. This paper introduces Multilingual–Tiny as a web authoring tool to support the virtual experience of indigenous students and teachers when they are creating learning objects in indigenous languages or in Spanish language, in particular, when they have to deal with the grammatical structures of Spanish. Multilingual–Tiny has a module based on the Case-based Reasoning technique to provide recommendations in real time when teachers and students write texts in Spanish. An experiment was performed in order to compare some local similarity functions to retrieve cases from the case library taking into account the grammatical structures. As a result we found the similarity function with the best performance.**

Keywords — **Authoring tool, Second language acquisition, indigenous people, Case-based reasoning, local similarity functions.**

I. INTRODUCTION

IN bilingual virtual training programs for teachers that have an indigenous language as mother tongue [1] [2], there are some difficulties when teachers design and create learning objects to teach Spanish as a second language for indigenous population. Some of those difficulties were reported in [3] and are mainly related to the process of writing texts, in particular the use of grammatical gender and number in the Spanish language. The main cause of this situation is that some indigenous do not have masculine or feminine distinction, or there are particular ways to express grammatical number that differs significantly from Spanish language.

In consequence, teachers have to be aware of some rules in order to properly apply the grammatical rules of Spanish and take care of teaching them correctly to their students. Nevertheless, in some cases, indigenous teachers of Spanish language use some didactic strategies, as reading from

textbooks and the language class [4], designed to teach indigenous languages but they apply them to teach Spanish language. This situation can create some problems in students, because they do not reach a good Spanish level, so it will affect the learning process of other subjects in the future. As a result of these issues, some learning objects that are written by indigenous teachers in Spanish may contain some grammatical errors in the texts.

As a solution, in this paper we introduce Multilingual-Tiny, a web authoring tool based on the TinyMCE [5] web content editor which consist of a complete set of plug-ins and online services for teachers to support them in the learning objects design and development. Multilingual-Tiny also has a module that applies Case-based reasoning (CBR), in order to provide recommendations (based on the grammatical structure of sentences) and taking into account the previous experience of skilled teachers from writing Spanish texts and well-formed texts obtained from the Internet. All of this process support teachers of Spanish language when they are creating their learning objects, mainly when they are writing texts in Spanish language.

This document is organized as follows: In section 2, some concerns about teaching Spanish as a second language are presented. In section 3 the architecture design of Multilingual-Tiny is described, including the applied CBR cycle. Section 4 describes an illustrative scenario which present the complete process performed by Multilingual-Tiny and also how the CBR technique was applied. Section 5 describes the followed validations process as well as the obtained results. Finally conclusions are presented in section 6.

II. TEACHING SPANISH AS A SECOND LANGUAGE

Teaching Spanish as a second language to indigenous communities is not a trivial task. It supposes a challenge to governments and universities in which is important to promote effective Bilingual Intercultural Programs (BIE) and at the same time, training teachers effectively in order to introduce Spanish in a coordinated bilingualism method [4], in which both, mother tongue or L1 and second language or L2 are developed at the same time. In this context, the mother tongue (which is an indigenous language), is acquired by a natural

process [6]. The second language – L2, in this case, the Spanish language, is taught for facilitating indigenous people communication with Spanish speakers and also to receive instruction in some knowledge areas which are taught in Spanish.

Despite the efforts and advances obtained by applying the Bilingual Intercultural Programs in some countries such as Mexico and Peru, teachers of Spanish language may face some difficulties when they have to teach indigenous people how to read and write in Spanish and in the indigenous language [7] at the same time. Some of those difficulties are due to the fact that teachers of Spanish have an indigenous language as mother tongue and they learnt Spanish in a non-systematic way. The consequence is that those teachers use the same strategies for teaching both languages, so it could be counterproductive in student's learning process [8].

When teaching Spanish, teachers usually can follow two complementary strategies: reading from textbooks and the language class [5]. The former is a strategy in which teachers introduce and explains the topic in the indigenous language and then students read the book in Spanish language so that students identify vocabulary and pronunciation. Finally, the teacher explains vocabulary or concepts that students may have lost in the reading. The latter strategy is the language class, in which teachers of Spanish compare the indigenous language with the Spanish language in terms of grammar, vocabulary and structure in order to promote reflection and develop the meta-linguistic awareness [5].

In this context, in teacher's training, when universities are preparing indigenous students that will be future teachers of Spanish language for teaching in their indigenous communities, students have to develop competencies and skills in order to effectively apply the teaching strategies mentioned above and other didactic and pedagogic methods. Multilingual-Tiny, the web authoring tool developed, takes a relevant role in this task; giving recommendations to teachers to avoid grammatical errors. As a result, teachers can create quality educational content to teach Spanish and create learning objects in their mother tongue.

III. MULTILINGUAL-TINY APPROACH

A. Overview Architecture

Multilingual-Tiny is a web authoring tool developed in order to support indigenous students which will be future teachers of Spanish language in indigenous communities and teachers of this population, when they are creating learning objects, in particular, when they have to deal with some grammatical structures of sentences in Spanish. Multilingual-Tiny consist of plug-ins and online services to provide a virtual environment to design and develop learning objects in Spanish and indigenous languages and has a module based on the case-based reasoning technique, to provide recommendations in order to avoid grammatical errors and develop quality educational content.

The architecture of Multilingual-Tiny is depicted in Fig. 1. The architecture has 4 layers, from top to the bottom: The users layer, represent indigenous teachers and students that interact with Multilingual-tiny. The interface layer includes the authoring tool and shows the recommendations that come from the CBR module. The services layer provides a group of services for text processing and includes the CBR based module to provide the recommendations. Finally the data access layer includes services for data storing, such as the case library.
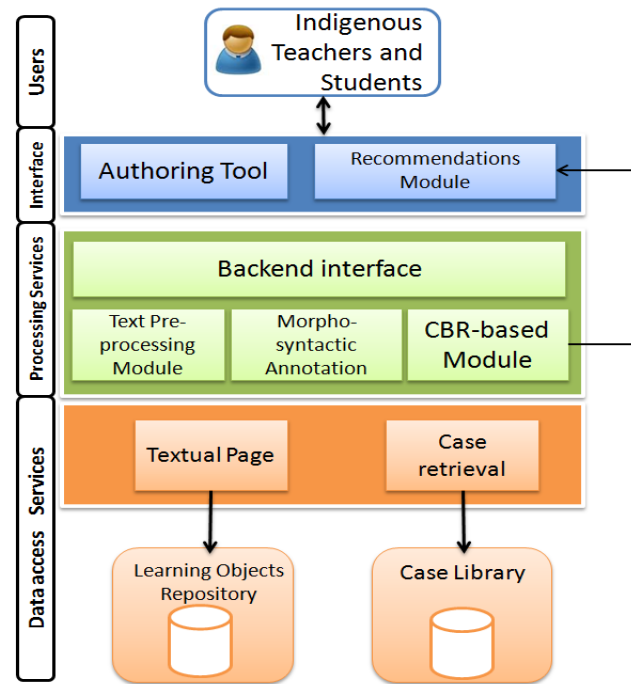

Fig. 1. Multilingual-Tiny Architecture.

B. Layer Description

The following paragraphs provide a detailed description about each of the layers mentioned before.

1) Users Layer

This layer represents the users that interact with Multilingual-Tiny, for instance, indigenous students that will be future teachers of Spanish language in indigenous communities and indigenous teachers. These users interact with the interface layer to use the service in order to create the learning objects.

2) Interface Layer and Authoring tool

Interface layer includes the authoring tool and the recommendations. The authoring tool is based on the TinyMCE [5] web content editor, which is an open source JavaScript based web editor that provides a group of services in order to create web pages without worrying about HTML code, because HTML is generated by it. The authoring tool can be integrated in the ATutor [9] e-learning platform or in other platforms.
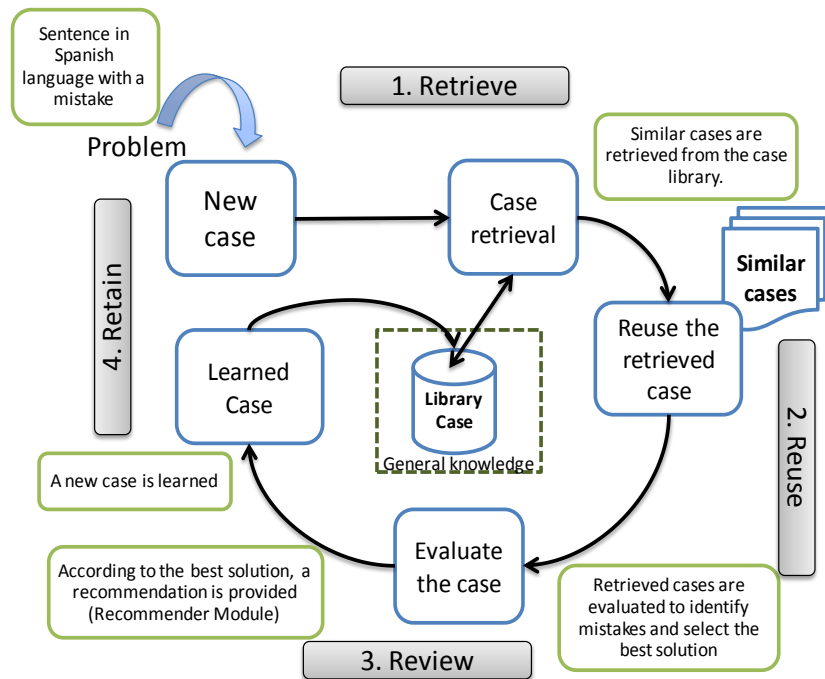
Fig. 2. CBR Cycle applied to generate grammatical recommendations for indigenous population.

As a result teachers can easily create web pages which will be part of a course in the ATutor e-learning platform as learning objects.

The authoring tool establishes communication with the Processing services layer when a learning object is being created. The text written in Spanish by indigenous teachers or students in the authoring tool is then sent to the Processing services layer to be analyzed.

The recommender module in the interface layer shows the recommendations that come from the CBR based module. These recommendations include suggestions on how to correct grammatical errors. The recommendation process is described in detail in next sections.

3) Processing Services Layer

This layer includes the services for text processing, the morpho-syntactic annotation module and the CBR based module. Those services are combined in order to provide recommendations to teachers when they are creating the learning objects to teach Spanish. The input of this layer is the text of the learning object that is being created in the authoring tool. The components of this layer are:

1) **Text Pre-processing Module:** The text pre-processing module is based on the open source FreeLing [10] library for Natural Language Processing. The input of this module is a text which has been written by the teacher as part of a learning object. This text is automatically split into sentences and the resultant sentences are split into words. This process is based on dictionaries and rules of the FreeLing library. The result of this process will be the input of the morpho-syntactic annotation module.

2) **Morpho-syntactic Annotation Module:** This module provides the morpho-syntactic annotation, which is a process of assigning tags for every word in the text, depending on the grammatical category. This process is based on the PoS (Part of Speech) tagging of FreeLing library. The input of this module is the output of the pre-processing module (which is a group of words). The PoS tagging is based on the EAGLES [11] recommendations. EAGLES define a group of standard tags for every grammatical category. As a result, each word of the text is assigned a tag depending on the context and grammatical structure of each sentence. The outputs of this module are groups of part-of-speech tags which represent a sentence. These tags will be an important component of the case representation in the case based reasoning module.

3) **Cased-based Reasoning Module:** This module is based on the Case-based reasoning technique. It takes the output of the morpho-syntactic annotation module, and executes the CBR cycle. As a result it provides the recommendations to indigenous teachers and students in order to correct grammatical errors when they write texts in Spanish language during learning objects creation. The module was built with jCOLIBRI framework [12], each case from the case library consists of a group of tags (part of speech tags) which represent a well-formed sentence. The CBR cycle, which includes 4 steps (Retrieve, Reuse, Review and Retain), is applied to grammatical sentence analysis in Spanish language and the process is depicted in fig. 2.

The main steps of this process are:

1) The Retrieve step: In this step a new case that comes from the morpho-syntactic annotation module, which is a new sentence, is compared with the cases stored in the case library by means of the similarity algorithm. As a result the most similar cases are retrieved. Both components are used:
   o Case library: Composed by a group of cases which are well-formed sentences in Spanish language obtained from a wide variety of texts from Internet. The case library is updated and new cases are stored when teachers add a new sentence structure. The case library is part of the Data Access layer which establishes communication with the services layer in order to store and retrieve cases.
   o Similarity Algorithm and Retrieve Component: Based on the JCollibri framework, the nearest neighborhood algorithm K-NN [13] is applied in order to retrieve the most similar cases when a new sentence is being analyzed. This process uses a global similarity function and a local similarity function for each attribute from the case.

2) The Reuse Step: In this step the K most similar cases obtained, by computing similarity, as described above are selected and the CBR Module organizes the cases according with the weights defined by the Morpho-syntactic Annotation Module.

3) In the next step, Review, the cases are evaluated in order to identify if the sentence is correct or if the sentence has a grammatical error. Besides, the case could be adapted or transformed to provide a recommendation about how to properly write the sentence. Further details about the overall process are depicted in section 4.

4) In the next step, which is called Retain, a new case obtained from the adaptation of the retrieved case is converted into a new case. Which is part of the recommendations provided by the recommender module and on the other hand it is stored in the case library as a new case. As a result from the process, grammatical errors in terms of using gender and number could be identified and a recommendation on how to correct it is provided to students.

4) Recommender Module

This module takes the cases retrieved from the case library as an input for providing recommendations to teachers or students on how to correct the sentence if a grammatical error in gender and number is identified. These recommendations take into account the indigenous language of teachers and students in order to explain why the sentence was incorrect from the indigenous language grammatical perspective.

IV. AN ILLUSTRATIVE SCENARIO

As well known, the CBR cycle includes 4 steps (Retrieve, Reuse, Review and Retain) as shown in Fig. 2. In this section a step-by-step illustrative case based on CBR cycle is applied in order to show how the grammatical sentence analysis in Spanish language is performed in Multilingual-Tiny to provide recommendations to students and teachers.

Step 1 - Writing the text:

Indigenous students which are preparing to be future teachers of Spanish language write a text in the web content editor when they are creating learning objects. In this step it is probably that students make mistakes in terms of grammatical issues when they write a text in Spanish but they are frequently thinking in their mother tongue which is an indigenous language. For instance:

─ *Me gustan el gatos blancos* (sentence with a mistake in Spanish).

─ *I like white cats* (English translation only for illustrative purposes).

The above sentence in Spanish has a mistake in the definite article ("el") because it is in singular form but it must be in plural form ("los").

Step 2 – Text Pre-Processing (Morpho-syntactic annotation of the initial text):

In this step the system takes the initial text and applies the morpho-syntactic part-of-speech annotation of the text according to EAGLES recommendations [11]. Taking the example mentioned above the morpho-syntactic annotation is depicted in table 1. It is important to remark that in table 1 for English language the sentence seems to be grammatically correct, but in Spanish language there is a mistake when using the definite article ("el") (which in English is "the") in singular form with a noun "gatos" (in English "cats") in plural form.

Step 3 – Case retrieval

Based on the morpho-syntactic annotation from step 2, in which each word has a specific tag (as depicted in table 1), a new case is created; this case is composed by the group of EAGLES tags. The new case could be: Case[PP1CS000, VMIP1P0, DA0MS0, NCMP000, AQAMP0]. This case is equivalent to the sentence: "*Me gustan el gatos blancos*" (in English: I like white cats). The new case is compared by means of the nearest-neighbor algorithm [13] with cases previously stored in the case library. The most similar cases are retrieved. For instance, if the following case is retrieved: Case=[PP1CS000,VMIP1P0,DA0MP0,NCMP000,AQAMP0], with a computed similarity of 96% from the global similarity function. It is important to remark that cases stored in the case library have been obtained from texts without grammatical errors.

TABLE I
MORPHO-SYNTACTIC ANNOTATION OF THE EXAMPLE.

Word – Token		Part-of-Speech tagging (EAGLES)	Meaning of the tag assigned to each word
Words in Spanish	Translation to English		
Me	I	PP1CS000	Personal pronoun, first person, common gender in singular form.
Gustan	Like	VMIP1P0	Main verb, indicative, present form, first person and plural.
El	the	DA0MS0	Define article, masculine, in singular form.
Gatos	cats	NCMP000	Common noun, masculine, in plural form.
Blancos	white	AQAMP0	Qualified adjective, masculine in plural.

Step 4 – Comparison of cases and recommendations

In this step a comparison between the new case and the most similar case retrieved is performed in order to find differences in terms of the sentence grammatical structure. By means of this comparison and the analysis performed is possible to identify for example if there are mistakes of grammatical gender or grammatical number which are common when indigenous people is learning Spanish. For instance the comparison of the example proposed ("Me gustan el gatos blancos" in English "I like white cats") with the case retrieved from the cases library is depicted in Fig. 3.

As a result of the comparison in this example, the system identifies a difference in the third element of the new case (DA0MS0) and the corresponding element of the re-trieved case (DA0MP0). Those tags are described as follows:

• DA0MS0 = Definite article (DA), Neutral (0), Masculine (M), in singular form (S), is not a possessive article (0).

• DA0MP0 = Definite article (DA), Neutral (0), Masculine (M), in plural form (P), is not a possessive article (0).

The difference was identified around the use of the grammatical number: In the new case the article is in singular form, but in the retrieved case (which has been extracted from a text correctly spelled) the article is in plural form. When the mistake has been identified, a recommendation is provided in order to correct the sentence; this recommendation takes information from the case retrieved in the CBR cycle in order to suggest the correct form that the sentence should have. As a result indigenous students and teachers can also learn by interacting with the authoring tool. Fig. 4 shows the graphical

user interface of the CBR module. In this case the interface shows the sentence that will be analyzed to identify possible mistakes in grammatical number and gender agreement.
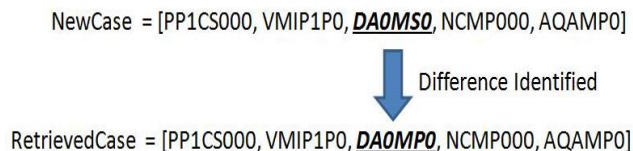

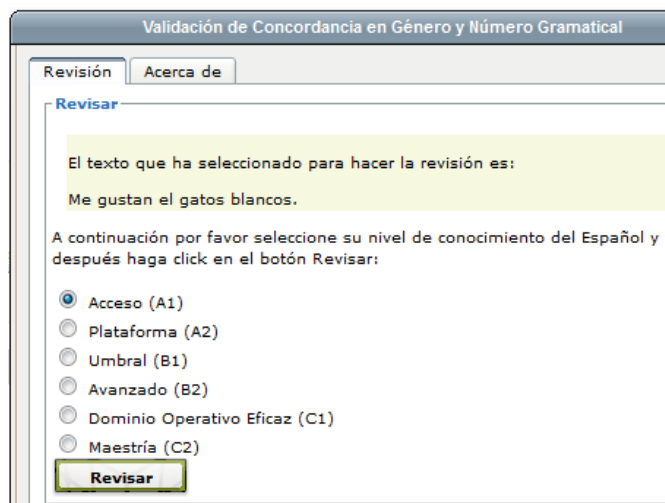Fig. 3. Comparing the example proposed with a retrieved case from case library.

Fig. 4. CBR Module graphical interface in TinyMCE.

V. EVALUATION

A. Description

The purpose of the evaluation process in to validate the main of our approach which is to support indigenous teachers and students when they write a text in a web content editor for creating learning objects. As mentioned before, the support we offer to indigenous teachers and students refers to automatically generate recommendation in order to avoid grammatical errors and develop quality educational content.

In particular, we validate the case retrieval process, because this is the process that ensures that the offered recommendation is the best one that the user could receive.

We applied the K-NN algorithm [13] in order to retrieve the most similar cases from the case library to check the grammatical number and gender agreement.

The K-NN algorithm in the jCOLIBRI framework uses a local similarity function and a global similarity function. The former is used to compute the similarity in every attribute of the cases; the latter is used to compute de global similarity considering the results of the local similarities from all the attributes of the case. We design an experiment to compare and choose the best local similarity functions that allows retrieving the most similar sentence to check the grammatical number and gender agreement. In this section we describe the methodology and the main results of the comparison.

TABLE II
TEST PERFORMED IN THE EXPERIMENT WITH VALIDATION METHOD AND VOTING METHOD APPLIED

Similarity Function	Validation method					
	Leave One Out			N-Fold Random Crossvalidation		
	Weighted Voting Method	Majority Voting Method	Unanimous Voting Method	Weighted Voting Method	Majority Voting Method	Unanimous Voting Method
Levensh.	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6
Overlap	Test 7	Test 8	Test 9	Test 10	Test 11	Test 12
Smith-Waterm.	Test 13	Test 14	Test 15	Test 16	Test 17	Test 18
Jaccard	Test 19	Test 20	Test 21	Test 22	Test 23	Test 24
Dice	Test 25	Test 26	Test 27	Test 28	Test 29	Test 30
TokensC	Test 31	Test 32	Test 33	Test 34	Test 35	Test 36

B. Methodology

In each case stored in the case library, the attribute with the highest weight is the morpho-syntactic annotation, which is basically a group of tags where each tag has been assigned to each word in the sentence according to the context and the grammatical structure. Since this group of tags is represented by means of a string data type, the local similarity function applied to this attribute should be able to compute the similarity between strings. There are many similarity functions for strings in literature some of them are described in [14], [15]. There are similarity functions based on fuzzy sets [16], and set-based string similarity [17] and [18].

For this experiment we chose four similarity functions commonly used in textual case-based reasoning. In addition we improved two similarity functions to consider the word order of the sentences during the analysis and deal with disambiguation by means of the FreeLing library. These are some important drawbacks described in [19] to be tackled in information retrieval and textual CBR. As a result 6 similarity functions were applied in the experiment. These are listed in table 3.

The validation methods used in this experiment were:
- Leave One Out
- N-Fold Random Cross-validation with 10 folds.

The voting methods selected for the K-NN algorithm were:
- Weighted Voting Method
- Majority Voting Method
- Unanimous Voting Method

The experiment was performed by means of 36 tests that combine the validation methods and the voting methods. Table 2 summarizes the group of tests in the experiment. For each test we obtain the following data:

- Precision

- Recall
- True Negatives
- False Positives
- False Negatives
- True Positives

TABLE III
SIMILARITY FUNCTIONS APPLIED IN THE EXPERIMENT

Simmilarity Function	Description	Improvement
Levenshtein	Also known as the edit distance	-
OverlapOrdered	Overlap Coefficient	Improved to consider the order of the tokens in each sentence analyzed.
Smith-Waterman	Based on the algorithm of dynamic programming for local alignment.	-
Jaccard	Jaccard coefficient for strings of characters	-
Dice	Dice coefficient for strings of characters	-
Tokens Contained Weighted First	Based on the TokensContained similarity function of jCOLIBRI framework.	Improved to consider the order of the tokens in each sentence analyzed.

C. Results

The following paragraphs summarize the main results we obtained in the experiment.

The results of tests 1,7,13,19,25 and 31 are show in fig. 5. In this case the validation method was Leave One Out and the voting method was Weighted Voting Method. On the other hand Fig. 6 shows the results of tests 4, 10, 16, 22, 28 and 34

where we used the same voting method but using the N-Fold random cross-validation method.

The F-Measure graphic in fig. 5 shows that the OverlapOrdered and TokensContained functions outperform the other functions compared, and the F-measure graphic in fig. 6 shows that Smith-Waterman, OverlapOrdered and TokensContained are functions with a better performance than the others.
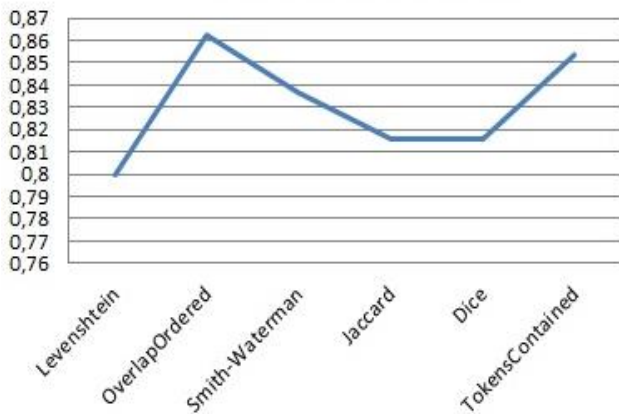

Fig. 5. Results of tests 1, 7, 13, 19, 25 and 31.
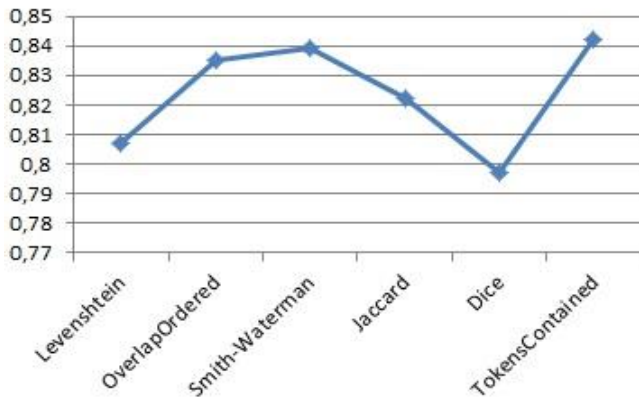
Fig. 6. Results of tests 4, 10, 16, 22, 28 and 34.

The results of tests 2, 8, 14, 20, 26 and 32 are shown in fig. 7 using the Leave One Out validation method as mentioned before, but in this case using the Majority Voting Method for the K-NN classifier.

In contrast fig. 8 shows the results of tests 5, 11, 17, 23, 29 and 35 using the same voting method but using the N-Fold random cross-validation method. In this case the results also shown that OverlapOrdered offers better performance than the other functions evaluated and TokensContained has a good performance when the voting method is the Majority Voting Method.
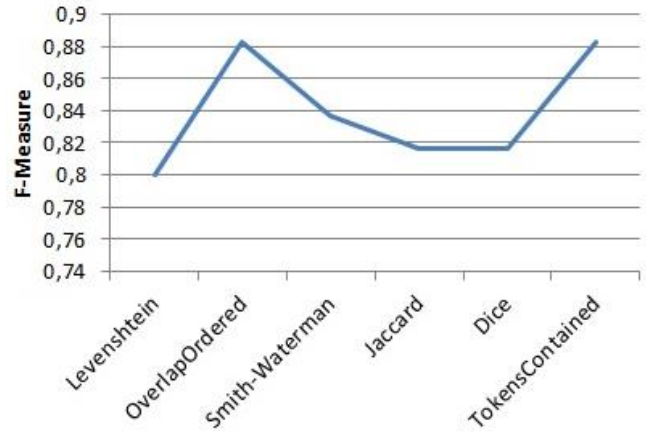

Fig. 7. Results of tests 2, 8, 14, 20, 26 and 32.
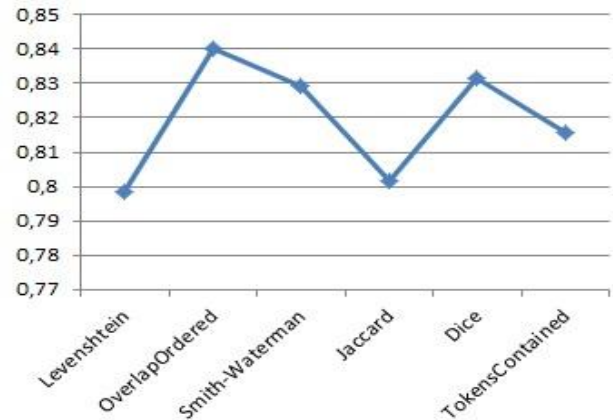

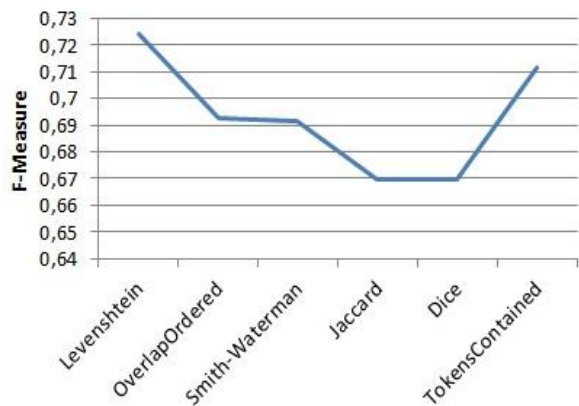Fig. 8. Results of tests 5, 11, 17, 23, 29 and 35.
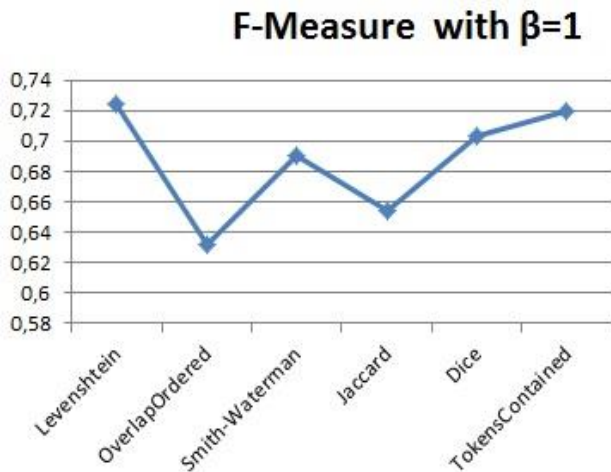

Fig. 9. Results of tests 3, 9, 15, 21, 27 and 33.

Fig. 10. Results of tests 6, 12, 18, 24, 30 and 36.

Finally the results of tests 3, 9, 15, 21, 27 and 33 are shown in fig. 9, in this case we use the same validation method (Leave one out), but we use the Unanimous Voting Method in the K-NN classifier.

In contrast, fig. 10 shows the results of tests 6, 12, 18, 24, 30 and 36 using the same voting method but the N-Fold random cross-validation method. In this case the Levenshtein distance has a better performance than OverlapOrdered in both methods of validation but Tokens Contained has almost the same performance than Levenshtein.

VI. CONCLUSION

Multilingual-Tiny as an authoring tool to support indigenous students that will be future teachers of Spanish language when writing texts in Spanish, takes a relevant role in order to help students to improve their writing skills at grammatical level so that they will be proficient teachers of Spanish. Multilingual-Tiny also provides a group of services that allow creating learning objects and design activities in the context of learning Spanish as a second language. This tool can be considered an advance in information and communication technologies to support the training process of indigenous students in the context of bilingual intercultural programs.

The case-based reasoning technique applied to the process of sentence analysis in order to identify grammatical errors mainly in terms of grammatical number and gender, is an efficient technique due to the use of the past user experience. Besides, the similarity algorithm, including the local similarity functions, the global similarity function and the retrieval process based on the K-NN algorithm in JColibri applied in the retrieval step works as it was expected in order to retrieve the most similar cases compared with a new case provided.

Evaluation developed shows that with respect to the algorithms used to retrieval cases that OverlapOrdered and TokensContained are functions with better performance to retrieve cases from the case library, so we can confirm that they are useful when we are dealing with grammatical structures of sentences in Spanish in form of part-of-speech

tags. As a result of this experiment we decided to combine both functions when performing the retrieval phase of the case-based reasoning cycle.

This strategy allows improving the system's performance in order to identify possible grammatical errors in gender and number agreement in Spanish language.

REFERENCES

[1] UNICEF & PROEIB Andes, *Atlas Sociolingüistico de pueblos indígenas en América Latina - Tomo I*. FUNPROEIB Andes, Bolivia, 2009.
[2] UNICEF – *Oficina Regional para America Latina y el Caribe*, [Online], Available: http://www.unicef.org/lac/lenguas_indigenas.pdf
[3] Trillos, M, "Bilingüismo Desigual en las escuelas de la sierra nevada de Santa Marta". THESAURUS. Tomo LI. Núm. 3, pp. 401-486. 1996.
[4] Hamel, R. E., Francis, N, "The Teaching of Spanish as a Second Language in an Indige-nous Bilingual Intercultural Curriculum". Language, Culture and Curriculum. Vol 19:2, pp. 171-188. 2006
[5] Moxiecode Systems AB - TinyMCE. [Online]. Available at: http://www.tinymce.com/
[6] Lier, L. "The Ecology and Semiotics of Language Learning – A Sociocultural Perspective". Springer Netherlands – Kluwer Academic Publishers, Boston. 2004.
[7] Valdiviezo, L. A., "Indigenous worldviews in intercultural education: teachers' construction of interculturalism in a bilingual Quechua–Spanish program. Intercultural Education.". Vol. 21,1, pp. 27-39. 2010.
[8] Arévalo, I., Pardo, K., Vigil, N, "Enseñanza de castellano como segunda lengua en las escuelas EBI del Peru". Dirección Nacional de Educación Bilingüe Intercultural.
[9] University of Toronto, ATutor Learning Management System, [Online], Available: http://atutor.ca/
[10] Padró, L., "Analizadores Multilingües en FreeLing", *Linguamatica*. 3 (2), pp. 13-20. 2011.
[11] Expert Advisory Group on Language Engineering Standards – EAGLES. "EAGLES Guidelines" [Online]. Available: http://www.ilc.cnr.it/EAGLES96/home.html
[12] J.A. Recio-García, B. Díaz-Agudo, M.A Gómez-Martín, and N. Wiratunga, "Extending jCOLIBRI for Textual CBR". In: Muñoz-Ávila, H., Ricci, F. (eds.) Case-Based Reasoning Research and Development. LNCS, vol. 3620, pp. 421-435. Springer, Heidelberg 2011.
[13] K. Yu, L. Ji, and X. Zhang, "Kernel Nearest-Neighbor algorithm", Neural Processing Letters 15, pp. 147-156, 2002.
[14] H. Núñez, M. Sànchez-Marrè, U. Cortés, J. Comas, M. Martínez, I. Rodríguez-Roda and M. Poch," A comparative study on the use of similarity measures in case-based reasoning to improve the classification of environmental system situations, Environmental Modelling & Software", vol. 19, 9, pp. 809-819, Sep. 2004.
[15] N. Elita, M. Gavrila, and C. Vertan, "Experiments with string similarity measures in the ebmt framework". In Proc. RANLP 2007 Conf., Bulgaria, Sep. 2007.
[16] J.J. Astrain, J.R. González de Mendívil and J.R. Garitagoitia, "Fuzzy automata with -moves compute fuzzy measures between strings, Fuzzy Sets and Systems", vol. 157, 11, pp. 1550-1559, Jun. 2006.

[17] M. Hadjieleftheriou and D. Srivastava, "Weighted set-based string similarity", IEEE Data Eng. Bull, 33(1), pp. 25-36, 2010.

[18] P. Achananuparp, X. Hu, and X. Shen, "The Evaluation of Sentence Similarity Measures", In: Song, I., Eder, J., Nguyen, T. (eds.) Data Warehousing and Knowledge Discovery 2008. LNCS, 5182, pp. 305-316. Springer, Heidelberg, 2008.

[19] D. Patterson, N. Rooney, M. Galushka, V. Dobrynin and E. Smirnova, "SOPHIA-TCBR: A knowledge discovery framework for textual case-based reasoning", Knowledge-Based Systems, vol. 21, 5, pp. 404-414, Jul. 2008

Jorge Luis Bacca Acosta is Systems Technologist and Telematics Engineer from the Universidad Distrital Francisco José de Caldas in Colombia, MSc. on Industrial Informatics from University of Girona in Spain, as well as Ph.D Student at University of Girona. He belongs to the BCDS (Broadband Communications and Distributed systems) group from the University of Girona. He is also a developer at the ALTER-NATIVA European Project. His research interest are technology enhanced learning, case-based reasoning, intelligent agents, semantic web and augmented reality.

Silvia Margarita Baldiris Navarro Systems and Industrial Engineer from Industrial University of Santander (UIS), Colombia Master in Industrial Informatics and Automatic from University of Girona and PhD in Technologies from University of Girona. The topics of User Modelling and Adaptations applied to learning management systems, Knowledge management and standardizations in elearning are her research interest. She has participated in several projects: Adaptation based on machine learning, user modelling and planning for complex user-oriented tasks (ADAPTAPlan), EIE-Surveyor: Reference Point for Electrical and Information Engineering in Europe, A2UN@: Accessibility and Adaptation for ALL in Higher Education, and in the European project ALTER-NATIVA.

Ramon Fabregat Gesa Ramón Fabregat is professor at the Computer Architecture and Technology Department at the University of Girona, Spain. He received his degree in computer engineering from the Universitat Autónoma de Barcelona and his doctorate degree in technology from the University of Girona. His research interests are adaptive hypermedia systems, user modeling and e-learning. He has participated to the technical program committees of several conferences. He is director of the master of computer science and automation and leads a research group on BCDS.

Cecilia Avila Garzón is a researcher at BCDS Group from Universitat de Girona (Spain). She graduated in Telematics Engineering from the Universidad Distrital Francisco José de Caldas (Colombia) and has a MSc in Industrial Informatics from the Universitat de Girona. She is a PhD student in the area of technology. His work addresses several topics as web accessibility, collaborative learning, learning objects, and work in communities. She has been researching integrated in the ALTER-NATIVA European Project as a software developer for the last 2 years.