

e-Learning for Quechua's Grammar

Jose Lozano

Universidade Federal do Rio Grande do Sul
Rua Tocantins, Agronomia
Porto Alegre, Brazil
(55)(51) 91809242
josemachino@gmail.com

Edith Rivero

Universidad Nacional San Antonio Abad del Cusco
Av. La Cultura S/N
Cusco, Peru
(51)984420442
edith.tupac@gmail.com

ABSTRACT

Nowadays, many Native American languages have been disappeared because of globalization. The ones that still alive are struggling with strong social pressure from the dominant languages (mainly English, Spanish and Portuguese). There are few institutions that try to boost the learning of these indigenous languages. Also, there are few resources for learning such as books, institutes, and technological tools. So, in the mission of preserve these languages alive, the teaching of these languages by internet could be an option. Thus, we developed a web system for the largest Native American language: Quechua. The system allows user to learn Quechua's grammar. It includes two games to practice the vocabulary, a dictionary and a syntactic analyzer which shows the structure of a sentence as a tree. In this work, we present background information of Peruvian languages, how the system's architecture is designed and a description of the learning process through the system in its three sections: expert module, instructor module and student module. Corresponding tests are showed. Finally, we present the conclusions and future work.

Categories and Subjects Descriptors

K.3.1 [Computer and Education]: Computer uses in Education.

General Terms

Algorithms, Languages

Keywords

Quechua Grammar, e-learning, NLP

1. INTRODUCTION

Native American languages have been lost popularity because of strong social pressure from the dominant languages, and many of these indigenous languages disappeared. One of the most spoken Native American Language is Quechua. However, the lack of resources about grammar and the availability of Quechua's teachers complicate the learning process of this language compared to other languages referenced in the context of globalization, which incorporate technological tools that improve its accessibility and assist in the process of learning.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

XXXXXXXXXXXXX – As informações serão preenchidas no processo de edição dos Anais.

Quechua language has 37 variations or dialects according to Torero [15]. This set of variations is due to the inexistence of standardization, causing a complication in the development of a learning system for all language dialects. However, we can try to understand the pattern for all these variations and establish the one that should be learned. Moreover, forcing in this way it will create problems between native speakers. Thus, we focus our attention in the Quechua's grammar that is taught in Cusco-Peru: The dialect Qosqo-Qollao which belongs to Quechua-IIC [12]. All grammar rules are according to the "Academia Mayor de la Lengua Quechua (AMLQ)"¹ which is the unique organization with the responsibility to propagate Quechua language. It is important to mention that as many other Native American languages, Quechua was transmitted from one generation to the following just by speaking. The reason of this affirmation is because the absence of writing resources.

There are many ways of learning grammar such as showing the list of morphemes that form a word, or organizing a word within a sentence, or making the syntactical three of a word. There are natural language processing (NLP) tools that can help a person to understand the grammar through morphological and syntactical analyzers. Despite having NLP tools for understanding the Quechua's grammar, there is no web system for Quechua that combines them within a learning context.

In section 2 will be presented the state of the art. After that, in section 4 we will describe the Quechua language, its structure and complexity. Then, a description of our web system is presented in section 5, including some screenshots of its performing. Finally, the last section exposes its implementation and testing.

2. RELATED WORK

Comparing with foreign languages, a very little work has done in the area of natural language processing and e-learning systems for Peruvian indigenous languages. Within NLP tools, Rios [12] developed a spell checker for Quechua which uses a morphological analyzer. Her morphological analyzer was developed in Foma [5]. Also, Rios developed a parallel Treebank Spanish-Quechua [13]. There exists also a program made in python called Antimorfo which uses finite state transducers. However, it only can be used by console.

In other cases, Finite State Transducers (FST) are applied because it is possible to define a relation between a set of strings [6], like a machine that reads one string and generate another. It can be used

¹ <http://www.mcultura.gob.pe/academia-mayor-de-lalengua-quechua>

as a recognizer, generator and translator. So, the importance of FST in natural language processing tools and tutoring systems is relevant. For building FST, we need to define regular expressions which are structured into a formula in a special language that is used for specifying simple classes of strings [6].

Taking in consideration web based applications, there is E-learning. It is the acquisition and use of knowledge distributed and facilitated primarily by electronic means. This form of learning currently depends on networks and computers, but likely evolves into systems consisting of a variety of channels and technologies. E-learning can take the form of courses as well as modules and smaller learning objects [3]. In this context, we have the Moodle platform which provides a wide range of types of activities like taking quizzes, managing communication through forums, video conferences and others.

On other side, Intelligent tutoring systems (ITS) [9] enables learners to practice their skills by carrying our tasks with highly interactive learning environments. ITS are composed of three types of knowledge, organized into four separate software modules. The expert module is a computer presentation of a domain expert's subject matter knowledge (declarative knowledge) and problem-solving ability (procedural knowledge). Learner model is the level where the student interacts with the tutoring system. The model evaluates each learner's performance and his/her behavior in order to determine his/her level of knowledge, perceptual abilities, and reasoning skills. Instructional Model contains knowledge for making decisions about instructional tactics. It relies on the diagnostic processes of the learner model for making decisions about what, when and how to present information to a learner.

We can see a clear example of what can we do for preserving a language using programming tools. A bolivian girl using Scratch [10] built an interactive tutor which contents Aymara's grammar, toy stories, parts of the body, riddles and songs. Also, there are web pages that show grammar structure, but they are static pages. There are videos in youtube which teach some grammar rules.

3. CONTEXTUAL SITUATION OF INDIGENOUS LANGUAGES IN PERU

Recently, the government of Peru emitted the Law 29735 in July 2011 for preservation, development, revitalization, and use of indigenous languages for future generations. The law incorporated every distinct ethno linguistic variations as one of Peru's official languages. From the set of native languages in Peru, it has lost at least 37 native languages like Uro and Paquina, but even with that huge loss, there still remain 47 native languages in the country.

Clearly, one example of the difficulty in reaching a consensus is the Quechua language. It has had an official alphabet since 1985, but there are many variations of the language throughout the mountainous areas where it is spoken. So far, two variations of Quechua have been recognized: Qollao and Chanka. The reason because we focus our work in Qollao dialect is because it belongs to Cusco-Peru while the Chanka's dialect belongs to Ayacucho-Peru.

Referring to propagation of Quechua language, The National Institute of Statistics and Informatics of Peru (INEI) made a census in 2007. It revealed that more than 4 million Peruvians belong to ethnic groups that still use their native language. Out of that population, 83 percent spoke Quechua, 11 percent spoke Aimara, and 6 percent used another native language as their

mother tongue [8]. So, the government has produced textbooks for teaching indigenous languages. Moreover, it manifests the importance of having more translations of documents to the indigenous languages and e-learning systems for assisting teaching process is fundamental for the survival of Peruvian native languages. Any kind of prevention of Peruvian languages from disappearing will help to retain Peruvian cultural integrity because language diversity is linked to the expression of individual and collective identity, as well as a different way of conceiving and describing reality.

4. QUECHUA LANGUAGE

Among the Quechua language we have two branches: Quechua I and Quechua II. This last one has three dialects A, B and C. Within these dialects we have variants and Qosqo Qollao belongs to Quechua-IIC, spoken in Cusco-Peru [12]. Quechua is an agglutinative language where morphemes are postponed to the root forming a word. The Quechua's grammar is composed of four parts: morphology, syntax, phonology and orthography. Our intention is to cover morphology and syntax. Morphology [1] is concerned with the form and lexical meaning of a word and syntax deals with the word structure in sentences and phrases. Furthermore, morphology directly interacts with syntax through inflection in terms of agreement or case marking [11]. It is necessary to keep in mind the two level rules of morphology, consisting of a lexicon and a phonemic surface representation [7]. We will use five categories: verbs, nouns, adverbs, adjectives and pronouns. Verbs are declined to form other verbs and to conjugate we add suffixes as show in table 3. Nouns are declined, showing case and number. We illustrated the case in table 2. It must be said that some suffixes must be postponed in a certain order. For example, when we say *warmikunapaq* (for women) is correct, but *warmipaqkuna* is incorrect. Here, the consideration of sequence of these two suffixes *kuna* and *paq* is relevant.

Table 1: Identification of subject and object

Subject	noqa(I)	tayta(father)	wasi(house)
Object	warmita	paywan (he)	Brasilman

Table 2: Case example

Noun	wasi(house)	qolqe(money)	killa(moon)
Case	manta	rayku	pi
Noun+Case	wasimanta	qolqerayku	killapi

4.1. Structure

It is important to know the basic structure of the sentence which will be registered in our rules for the context free grammar (CFG) [6]. The basic structure of a sentence is subject-Object-Verb. The identification of the subject can be a pronoun, name of person, thing or animal. The object is identified by morphemes like *-ta-*, *-wan* and *man* which are postponed to the subject.

To identify a verb, we need to know the morphemes that can be postponed to a verb. In the following Table 3 we present the conjugation of the verb 'munay'(love).

4.2. Complexity

Due to a Quechua is a highly agglutinative language. A noun can be categorized as a subject or object. Within subject it can be categorized as common, proper, or compound. Similarly, verb

may be declined to be an adjective. However, some suffixes are not applied to certain verbs as showed in Table 4. The complexity of the sentence is when each word has many suffixes we have to determine the function is completed in the sentence according to the suffixes.

Table 3: Conjugation of Verb munay (love)

	Past	Present	Future
Noqa	Munarani	munani	munasaq
Qan	Munaranki	munanki	Munanki
Pay	Munaran	munan	Munanqa
Noqanchis	munaranchis	munanchis	munasunchis
Noqayku	Munarayku	munayku	munasaqku
Qankuna	munarankichis	munankichis	munankichis
Paykuna	Munaranku	munanku	munanqaku

Table 4: Complexity

Verb	adjective	Adjective
munay	munaycha (pretty)	munasqa(dear)
qelqay	-	qelqasqa(written)

5. OVERVIEW OF THE SYSTEM

The web system is composed of two main parts: The student interface where the user interacts with lessons of Quechua's grammar (Student Module), and the interface where the teacher inserts new material according to the category to be taught (Instructor Module). Internally, in the server side it runs a third module (Expert Module) which contains grammatical rules. We choose these structure for simulate ITS, even though it has not the capability of advising in a smart way what students should improve in their learning.

5.1. Expert Module

In this module, we describe the process of establish the grammar rules and the generation of the executable FST. First, we choose one of the categories and register a new lexicon. After that, we register the grammatical rules, obtained from [2, 4, 14]. Then, it is generated the FST with Foma and stored in the server. Foma has functions of analyzing and generating a word, which will be used when the server requests it. Then, we made it reviewed by an expert. In this case, the expert is a teacher member of AMLQ who verifies the correctness of rules and lexicons. This verification is done before uploading to the server.

5.2. Instructor Module

This module corresponds to the part of register and updates the content of the lessons and exercises. We structured it in two parts, a basic level and intermediate level. In the basic level, we restrict the instructor to upload examples and exercises. In the intermediate level, the instructor can add new lessons to each category, definitions within the lessons, and exercises. For inserting content, we use the CKEditor² which allows defining some functions in order to let the instructor just to apply it. For example in Figure 1, we want to outstand the importance of the morpheme **paq** and **man**, so we use the icon that has the sun image. This icon will produce the intermittent effect over the word which will be shown in the student module.

² <http://ckeditor.com/>



Figure 1: CKEditor



Figure 2: Example tab

Each lesson that is registered, the instructor must select a label which is configured in a table of labels used by the expert. Also, the teacher has a section to check commentaries given by users. Thus, the teacher can improve their content about exercises and lessons to be more explicit.

5.3. Student Module

The student module contains. It is organized in lessons, practice, vocabulary, and syntactic analyzer. In the lessons, it contains three tabs which are definitions, examples and exercises. In the exercises, we define two types one is to fill space blanks and the other is a multiple choice. When the student completes the exercise, the answer is sent to the server which analyses each word of the answer and check if it is correct or not. In that process, it is used the morphological analyzer and generator for evaluating the answers. For each answer, the FST takes as input the word and the label, corresponding to lesson that is being attended, and analyzes or generates the word depending of the exercise. Then, it compares the FST result against the answer and stores the result. After that the server sends the results and in the user interface is shown the score of the exercise as illustrated in Figure 3. Also, their punctuation is stored, letting to keep track of

the performance of the user in each lesson. So, every time that the user signs in, the system shows the user status which is illustrated in Figure 6. The practice tab contents the two games which are the memory game, shown in Figure 5, and the meaning match game. For the vocabulary tab, the user searches the meaning of a word, and the system show the possible meaning. The vocabulary table contains 3000 words with their respective meaning. The last tab contains the syntactic analyzer, which shows the structure of a word as a tree as seen in Figure 4.



Figure 3: Exercise example

5.4. Architecture and system’s functionality

The systems functionality consists of five main functional entities which are register class theory, register class exercise, practice grammar, practice vocabulary, register member. The first two are oriented to be used by a teacher and the remaining, by the user. The architecture is organized around a set of server-side components. These components are learning, morphological, syntactic, admin, user controller, dbquechua and FST.

The learning controller receives the client request and select what controller is required such as admin, user or morphological. If the request is to qualify an exercise, the morphological controller is invoked. If the request is to analyze a sentence, the syntactic controller is invoked together with the morphological controller. If the request is to retrieve or record information in the database, then the learning controller is invoked. Only the user controller has access to syntactic controller. The morphological analyzer invokes the FST to process the word. Finally the web browser component, which is on the client side, sends the request to the server.

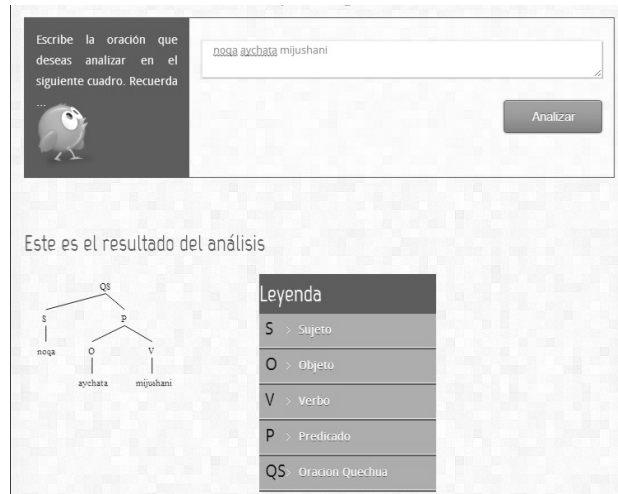


Figure 4: Syntactic Analyzer

6. IMPLEMENTATION AND ASSESSMENT

The morphological analyzer was tested according to the grammatical rules. We present the analysis in Table 5. For example the verb **mijun** is analyzed and show the structure that is shown as **miju+Vprim+Indicativo+Presente+3ra.Sg**.

Vprim means that is composed of one root. *Indicativo* means the classification of the verb. *Presente* means the tense of the verb. The term *3ra.Sg* is the subject of the verb. Words are analyzed in this way. According to the input the FST will show the structure. Sometimes a word can have two or more structures, because it can be in many classifications such as **allin**. And these results are in accordance with AMLQ.

Table 5: Results of morphological analyzer

Input Word	Output
noqa	noqa+Pronom+Personal+1ra.Sg
mijun	miju+Vprimi+Indicativo+Presente+3ra.Sg
paykuna	paykuna+Pronom+PErsonal+3ra.PI
pujllani	pujlla+Vprimi+Indicativo+Presente+1ra.Sg

We present the analysis of sentences in Table 6. The interpretation of the output is that each word appears with syntactic category labels, which is correct according to AMLQ.

Table 6: Results of syntactic analyzer

Input Sentence	Output
Noqa alqoyman pujllani	[QS[S noqa][P[O alqoyman][V pujllani]]]
Waldircha aychata mijun	[QS[S Waldircha][P[O aychata][V mijun]]]
Paykuna allin mijun	Verb and gender do not correspond.



Figure 5: Memory game



Figure 6: User Status

For example the label *S* means substantive, *P* is predicate, *V* is verb, *O* is object. Also, we can see that one sentence is not syntactically correct because **paykuna** is plural and **mijun** is a verb conjugated in singular.

In the phase of implementing, we consider the principles of designing interface because it is important to capture the attention of the user with a good combination of colors and styles. Thus, we integrate the morphological and syntactical analyzer into web environment.

The user testing took place together to fifteen children under twelve years old who already have taken some computer lessons which gave them basic knowledge of computer interaction and they can operate a web system. We choose this age because in our country children who are in the last year of elementary school start to receive lessons of Quechua. We use the system to practice and reinforce what they have learned.

The structure of testing is divided in measuring the performance and grade of satisfaction of children. So, we developed some questions structured in a survey that was given to children. The Figure 7 shows a summary of results obtained.

During the test application we have appreciate a special enthusiastic in children when they interacted with games because facial expressions shown happiness of getting points. Another thing we have noticed after the test application was some interventions of children that asked us about including some other alternatives in the teaching process like voice to help them in the pronunciation and more games.

According to the previous questionnaire and observations, we realized that it is needed to implement more communicative approaches like helpers as kind of assistance, and a section where teacher can clarify exercises.

Questions	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
Do find comfortable the navigation between pages?	0	0	0	3	12
Do you find problems in understanding the questions?	0	0	0	2	13
Do you consider the combination of colors appropriately?	0	2	3	10	0
Have you noticed any error during the interaction of the web system?	10	0	0	4	1
Whenever you push a button, do you wait a lot of time getting an answer?	11	1	1	2	0

Figure 7: User Status

Furthermore, the web system was tested by experts in linguistics who teach Quechua. Their opinions validate the dialect studied, and they suggested including all Quechua's dialect variations. For that, we can implement a normalizer as a future work. In the context of system portability, it is accessible in the most known web browsers, and its interface is able to be resizable in mobile devices.

7. CONCLUSION AND FUTURE WORK

The web system for teaching Quechua's grammar was implemented towards provide a nice interface to the user considering the principles of human computer interaction. We believe that is one more step to recover the Inca's language. The system is based on a morphological analyzer and syntactic analyzer; both tools can be used in other tasks of natural language processing. Furthermore, a learning approach applied to teach grammar was successfully employed. We take advantage of the results achieved in these areas and adapt them to the eLearning context.

The use of FST in the implementation of the morphological analyzer allows us to use it as a morphological generator as well. Needless to say, this method is effective in the construction of morphological analyzers. Also, the inclusion of a syntactical analyzer helps to show users how a Quechua sentence is structured. We expect that user of the system will be able to understand the way a sentence in Quechua is written, and make use of Quechua correctly.

As a communicative approach, social plugins can be added to share knowledge and inquiries in order to engage users in learning. Also, users could customize the interface layouts to improve the way of interaction with the system. We are considering the integration of a mentor entity, which guide students learning and provide feedback about results, would

enforce the system. This could be supported jointly to a speech tool for helping in the pronunciation of word and phrases in order to make the system interactive and useful.

8. ACKNOWLEDGEMENTS

We express our gratitude to Miguel Sanchez, member of Academia Mayor de la Lengua Quechua who verified the grammar rules and author of theory and exercises in Quechua. Also we thank the Linguistic group of Quechua (San Antonio Abad of Cusco University).

9. REFERENCES

- [1] Bauer, L. Encyclopedia of Language and Linguistics. 2006.
- [2] Choque, D.T. Manual para el aprendizaje del idioma Quechua Inicial, Primaria, Secundaria y Superior. 2012.
- [3] e-learning: A programmatic research construct for the future. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.99.6062rep=rep1type=pdf>. Accessed: 2013-09-01.
- [4] Fernandez, L.A.H. Qheswa Simi Hamut'ana Kuraq Suntur. Peru, 2005.
- [5] Foma. <https://code.google.com/p/foma/>. Accessed: 2013-06-01.
- [6] Jurafsky, D. & Martin, J. H. Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition. 2008.
- [7] Koskenniemi, K. Two-level Morphology: A General Computational Model for Wordform Recognition and Production. PhD thesis, 1983.
- [8] Peru's native languages are fighting for survival. <http://www.peruthisweek.com/blogs-perus-nativelanguages-are-fighting-for-survival-100088>. Accessed: 2013-07-01.
- [9] Phobun, P. & Vicheanpanya, J. Adaptive intelligent tutoring systems for e-learning systems. Pages 4064 - 4069, 2010.
- [10] Scratch. <http://scratch.mit.edu/>. Accessed: 2013-07-01.
- [11] Spiegel, S.R. Machine learning for the analysis of morphologically complex languages. PhD thesis, 2011.
- [12] Rios, A. Spell checking an agglutinative language: Quechua. In 5th Language and Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics, pages 51-55. Fundacja Uniwersytetu im. A. Mickiewicza, 2011.
- [13] Rios, A. & Göhring, A. & Volk, M. Parallel treebanking spanish-quechua: How and how well do they align. In The 10th International Workshop on Treebanks and Linguistic Theories (TLT10), 2012.
- [14] Taller Capacitación Docente. Municipalidad del Cusco, February, 2012.
- [15] Torero, A. Los dialectos Quechuas. Anales Científicos de la Universidad Nacional Agraria N2, 1964.