

# Multilingual CALL Framework for Automatic Language Exercise Generation from Free Text

Naiara Perez and Montse Cuadros

Vicomtech-IK4 research center

Mikeletegi 57, Donostia-San Sebastián, Spain

{nperez,mcuadros}@vicomtech.org

## Abstract

This paper describes a web-based application to design and answer exercises for language learning. It is available in Basque, Spanish, English, and French. Based on open-source Natural Language Processing (NLP) technology such as word embedding models and word sense disambiguation, the system enables users to create automatically and in real time three types of exercises, namely, Fill-in-the-Gaps, Multiple Choice, and Shuffled Sentences questionnaires. These are generated from texts of the users' own choice, so they can train their language skills with content of their particular interest.

## 1 Introduction

This paper describes a web-based computer-assisted language learning (CALL) framework for automatic generation and evaluation of exercises<sup>1</sup>. The aim of this application is mainly to allow agents in the language learning sector, both teachers and learners alike, to create questionnaires from texts of their own interest with little effort. To do so, the application includes state-of-the-art open source NLP technology, namely part-of-speech tagging, word sense disambiguation, and word embedding models. Its main features are the following:

- **Multilingual.** The platform enables to train Spanish, Basque, English and French skills. Interface messages appear in the chosen language too.

<sup>1</sup>The application is at [edunlp.vicomtech.org](http://edunlp.vicomtech.org). The web page asks for credentials to log in; the username is "vicomtech", and the password "edunlp". You will also need your own account, which you can obtain by writing an e-mail to the authors of the paper.

- **Three formats.** Users can design and answer exercises of three types: *a*) Fill-in-the-Gaps (FG): learners must fill in the gaps in a text with the correct words, based on some clues given or just the context provided by the text; *b*) Multiple Choice (MC): learners must choose the correct answers from a set of words given to fill in the gaps in a text; and, *c*) Shuffled Sentences (SS): learners must order a set of given words to formulate grammatical sentences.
- **Highly configurable.** Each exercise format offers a variety of settings that users can control. The input texts from which exercises are built are always given by the users. They also choose the pedagogical target of the exercises based on language-specific part-of-speech (PoS) tags and morphological features. Other settings include the type of clues in FG mode and the amount of distractors in MC. Furthermore, users can select the exercise items themselves or let the system do it automatically.
- **Exportable.** The exercises can be downloaded in Moodle's CLOZE<sup>2</sup> syntax to import them into Moodle quizzes, an extensively exploited platform by teaching institutions all over the world.
- **Evaluation.** The questionnaires designed can be answered in the same application, which prompts the percentage of correct answers upon submission. Correct answers are shown in green and the incorrect ones in red, so the learner can try to guess again.
- **Real-time generation, easy to use.**

<sup>2</sup>[https://docs.moodle.org/23/en/Embedded\\_Answers\\_\(Cloze\)\\_question\\_type](https://docs.moodle.org/23/en/Embedded_Answers_(Cloze)_question_type)

## 2 Related Work

There exist countless tools, both web-based and desktop software, that facilitate the creation of teaching material for general purposes, some of the best known being Moodle<sup>3</sup>, Hot Potatoes<sup>4</sup>, ClassTools.net<sup>5</sup>, and ClassMarker<sup>6</sup>. However, the focus of such tools is on enabling users to adapt the pedagogical content they are interested in, whichever it is, to certain exercise formats (i.e., quizzes, open clozes, crosswords, drag-and-drops, and so on). That is, these tools do not offer support for assessing the contents on their pedagogical suitability nor other exercise-dependent tasks, such as building clues for quizzes or distractors for multiple-choice exercises.

In the domain of language learning in particular, exercise authoring very often implies a lot of word list curation, searching for texts that contain certain linguistic patterns or expressions, retrieving definitions, and similar tasks. The availability of resources for language teachers that simplify these processes, such as on-line dictionaries or teaching-oriented lexical databases (e.g., English Profile<sup>7</sup>), depends on the target language. To the best of this paper's authors' knowledge, there does not exist at the moment an exercise generation and evaluation framework specific to learn Basque, Spanish, English, and French, that not only automates formatting the content given by the user for several exercises but also incorporates natural language processing (NLP) techniques to ease the authoring process. Volodina et al. (2014) describe a similar framework named Lärka. Lärka designs multiple-choice exercises in Swedish for linguistics or Swedish learners. The questions are based on controlled corpora, that is, the users cannot choose the texts they will be working on.

## 3 Workflow description

All the exercise formats mentioned share a common building process. Users must choose a language, provide a text –in that language–, and choose a pedagogical target from the options given. Pedagogical targets are based on PoS tags and morphological features. Different possible targets have been implemented for each language,

depending on the languages' characteristics and the richness of the parsing models available. For instance, exercises in Spanish can target the subjunctive conjugation or the definite/indefinite articles. In English, one can target, for example, past and present participle tenses. Once the initial configuration has been set, the workflow continues as follows:

### 3.1 Extraction of candidate items

The text given is segmented, tokenized and tagged with IXA pipes (Agerri et al., 2014), using the latest models provided with the tools. The tokens with a PoS tag or morphological feature selected by the user as pedagogical target are chosen as candidate items. In the case of the SS format, item candidates are sentences containing tokens with the relevant PoS tags or morphological features. If the text does not contain any candidate item, the application alerts the user that it is not possible to build an exercise with the configuration given.

### 3.2 Final item selection

The system has two ways of getting the final items from the candidates identified: the user can choose whether to select them or let the application do it randomly. In the latter case, the user can set an upper bound to the amount of items generated. The system never yields two contiguous items, since it would increase substantially the difficulty in answering them.

### 3.3 Exercise generation

Once the final items have been chosen, the actual questionnaire must be designed. This depends on the format chosen by the user and the specific settings available to that format:

**Fill-in-the-Gaps (FG).** The system substitutes the items chosen with gaps that have to be filled in with the appropriate words. Users can choose to show, for all the languages, at least three types of clues to help learners do the exercise: the lemmas of the correct words, their definition, or a word bank of all the correct words (and how many times they occur). For Spanish and English, the system is also capable of prompting the morphological features of the words to be guessed, in addition to the lemma. This feature is interesting to train on singulars and plurals, grammatical gender, verbal tenses, and so on. Moreover, depending on the pedagogical target chosen, the system automatically disables certain types of clues. It would not

<sup>3</sup><https://moodle.org/>

<sup>4</sup><https://hotpot.uvic.ca/>

<sup>5</sup><https://www.classtools.net/>

<sup>6</sup><https://www.classmarker.com/>

<sup>7</sup><http://www.englishprofile.org/>

make much sense, for instance, to give the lemmas of the correct words if the pedagogical target were prepositions, given that prepositions cannot be lemmatized. The user can also choose to not give any help.

To generate clues based on lemmas and/or morphological features, the system turns to the linguistic annotation given by the IXA pipes during candidate extraction. The annotation contains all the information necessary for each token in the text.

Retrieving definitions requires additional processing, since a word’s definition depends on the context it appears in. That is, choosing the correct definition of a word in the text provided by the user translates to disambiguating the word. The application relies on Babelfy (Moro et al., 2014) and BabelNet (Navigli and Ponzetto, 2010) APIs in order to do so. It passes the whole text as the context and asks Babelfy to assign a single sense to the words chosen as targets of the exercise. Then, it retrieves from BabelNet the definitions associated to those senses. Babelfy is not always able to assign a sense to a word; when this happens, the application returns the lemma of the word as its clue.

**Multiple Choice (MC).** Again, the exercise consists of gaps in the text to be filled with the correct words. In this case, the learner is given a set of words from which to choose an answer. This set of words contains the right answer and some incorrect words called “distractors”. When this exercise format is chosen, the system automatically generates as many distractors as specified by the user. This is achieved by consulting, for each correct answer, a word embedding model and retrieving the most similar words. The models are word2vec (Mikolov et al., 2013b; Mikolov et al., 2013a) trained on Leipzig University’s corpora<sup>8</sup> with the library Gensim for Python<sup>9</sup>. Thus, distractors tend to be words that appear often in contexts similar to the right answer, but not semantically or grammatically correct. Distractors are then transformed to the same case as the correct word and finally shuffled for their visualization.

**Shuffled Sentences (SS).** This exercise consists in ordering a set of words given to formulate a grammatical sentence. In this case, the system substitutes the sentences chosen by gaps and shows the sentence shuffled as a lead.

### 3.4 Evaluation

The user can answer the questionnaire it has designed and get it assessed by the system. For all the exercise formats, the evaluation consist in comparing the correct answers with the input received from the user. The answer is right only when it is the same to the correct answer.

## 4 The demonstrator

The application has a clean interface and is easy to follow. An exercise can be designed, answered and evaluated visiting less than four pages:

**The home page.** In the home page, the user chooses a language –Spanish, Basque, English or French–, and an exercise format –FG, MC, or SS. It leads to the exercise configuration page of the exercise chosen.

**The configuration page.** All the configuration pages share a common structure. A text field occupies the top of the page. This is where the user introduces the text they want to work with. Below the text field are the configuration options, presented as radio-button lists. All the exercise formats require that users choose a pedagogical target. Then come the format-specific settings, the only section of the page that varies.

For an FG exercise, two properties must be set: the type of clue and how the clues must be visualized. They can be shown below the gapped text with a reference to the gap they belong to, or as description boxes of the gaps (i.e., “tooltips”).

In a MC exercise, users must choose the amount of distractors they want the system to create.

For the SS mode, users can set an upper limit to the sentences that will be selected as candidates.

Finally, the user can choose whether to let the system select the items of the exercise or choose them themselves among the candidates that the system generates. In the former case, the system asks the user how many items it should create at most, and takes the user directly to the “Answer and evaluate” page. In the latter case, the user is taken to the “Choose the items” page.

If the text does not contain any token that meets the pedagogical target, the system asks the user to provide a different text or change the configuration set. That is, the application allows for users to know with a single click whether the texts they choose are suitable for the pedagogical objective they have set, without them having had previously read the text thoroughly.

<sup>8</sup><http://corpora.uni-leipzig.de/>

<sup>9</sup><http://radimrehurek.com/gensim/>

**Choose the items.** In this page users choose the items it wants to create among the words that meet the pedagogical target they chose. The interface shows the whole text with all the available candidates selected. They can be toggled simply by clicking on them. There are also buttons to select all and remove all the candidates. Once users are satisfied with their selection, they are led to the “Answer and evaluate” page.

**Answer and evaluate.** This is the page where the final questionnaire is displayed and can be filled. The appearance varies a little depending on the format chosen. FG exercises consist of the text given and gaps where the correct words should be written. If the tooltip clues option has been enabled, clues appear in description boxes when hovering over the gaps; otherwise, they appear listed below the text. Word bank clues appear boxed on to the right of the text. In MC exercises, the choices are radio-button groups listed to the right of the text. As for SS exercises, page-wide gaps replace the sentences chosen, and the shuffled words are given above the gaps.

At the bottom of this page there is a link to download the exercise in Moodle CLOZE syntax. The file that is downloaded can be imported to Moodle in order to generate a quiz.

Users can fill in the exercise they designed and submit them to get the percentage of correct answers. Furthermore, the answers are colored in green or red, depending on whether they are correct or not, respectively. This way learners can try to answer again.

## 5 Conclusions

We have described a web-based framework for language learning exercise generation. It is available for Spanish, Basque, English, and French. Users can design three types of exercises to train diverse skills in these languages. The framework is highly configurable and lets the user choose whether they want to select the exercise items or have the system do it.

As future work, the system should be improved in various ways. To begin with, the actual system delegates to the user the task of making the exercises more or less difficult by choosing the items themselves. The application will be endowed with technology that allows the users to create automatically exercises which vary in difficulty starting from the same text.

Another aspect that can be improved is the fact that exercise items are created from unigrams. It would be very interesting that the application were capable of generating multi-word candidates. This would be useful to revise collocations or phrasal verbs, for instance. In this same regard, the applicability of the system would increase if it based item candidate generation not only on PoS tags or morphological features, but on other criteria as well like the semantics of the text.

There is also room for improvement in the configurability of the application. Users should be allowed to control two key features: definition clues in FG exercises and MC distractors. Currently, the application imposes the definitions and distractors it generates, instead of presenting them as options for the users to choose.

Finally, we plan to implement more formats that add to the available three.

## Acknowledgements

This work has been funded by the Spanish Government (project TIN2015-65308-C5-1-R).

## References

- Rodrigo Agerri, Josu Bermudez, and German Rigau. 2014. IXA pipeline: Efficient and Ready to Use Multilingual NLP tools. In *LREC*, volume 2014, pages 3823–3828.
- Tomas Mikolov, Greg Corrado, Kai Chen, and Jeff Dean. 2013a. Efficient Estimation of Word Representations in Vector Space. *CoRR*, abs/1301.3.
- Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeff Dean. 2013b. Distributed Representations of Words and Phrases and their Compositionality. In *NIPS*, pages 3111–3119.
- Andrea Moro, Francesco Cecconi, and Roberto Navigli. 2014. Multilingual word sense disambiguation and entity linking for everybody. In *Proceedings of the 2014 International Conference on Posters & Demonstrations Track-Volume 1272*, pages 25–28. CEUR-WS. org.
- Roberto Navigli and Simone Paolo Ponzetto. 2010. BabelNet: Building a very large multilingual semantic network. In *ACL*, pages 216–225.
- Elena Volodina, Ildikó Pilán, Lars Borin, and Therese Lindström Tiedemann. 2014. A flexible language learning platform based on language resources and web services. In *LREC*, pages 3973–3978.