

The Exercise: an Exercise generator tool for the SOURCE project

Fryni Kakoyianni-Doa¹, Eleni Tziafa², and Athanasios Naskos³

Abstract. The Exercise, an Exercise generator in the SOURCE project, is a tool that complements the properties and functionalities of the SOURCE project, which includes the search engine for the Searchable Online French-Greek parallel corpus for the University of Cyprus (SOURCE) (Kakoyianni-Doa & Tziafa, 2013), the PENCIL (an alignment tool) (Kakoyianni-Doa, Antaris, & Tziafa, 2013), and the Synonyms and the Library tools. These are designed as freely available resources for language processing, easy to use by teachers and learners. The Exercise tool enables teachers to create either online activities or print out paper-based worksheets, including a variety of texts and activities on topics, in a variety of exercise types (e.g. multiple choice, word, phrase or sentence matching, filling the gaps with missing words/phrases, text reconstruction, listening, etc.).

Keywords: exercise generator, computer-based language testing, teachers' tool, language acquisition.

1. Introduction

The SOURCE project aims at providing a whole platform of tools to assist the French language teachers in the classroom. In this paper we present the latest tool called Exercise, an Exercise generator in the SOURCE project. This is the fifth tool of the SOURCE project, which consists of:

- The Source⁴ Corpus tool, which is a search engine for the searchable online French-Greek parallel corpus for the university of Cyprus (Kakoyianni-Doa & Tziafa, 2013).

1. University of Cyprus, Limassol, Cyprus; frynidoa@ucy.ac.cy

2. University of Cyprus, Limassol, Cyprus; tziafa.eleni@ucy.ac.cy

3. Aristotle University of Thessaloniki, Thessaloniki, Greece; anaskos@csd.auth.gr

4 The Source Corpus is referred to as Source, while the project as a whole is referred to as SOURCE.

How to cite this article: Kakoyianni-Doa, F., Tziafa, E., & Naskos, A. (2016). The Exercise: an Exercise generator tool for the SOURCE project. In S. Papadima-Sophocleous, L. Bradley & S. Thoušný (Eds), *CALL communities and culture – short papers from EUROCALL 2016* (pp. 208-214). Research-publishing.net. <https://doi.org/10.14705/rpnet.2016.eurocall2016.563>

- The PENCIL tool, an alignment tool for translators, teachers and learners, which enables the creation and customisation of web-based corpora, by uploading and aligning French and Greek texts (Kakoyianni-Doa et al., 2013).
- The Synonyms tool, which provides a search engine for synonyms, based on machine translation or dictionary searches.
- The Library tool, which showcases a part of the SOURCE project corpus.
- The Exercise tool (presented in the following sections), which enables teachers to create online activities and/or print out paper-based worksheets.

Following Kakoyianni-Doa and Tziafa's (2013) methodology, "all these tools and functionalities are designed as freely available resources for language processing, along with the data to be processed, in usable formats for both teachers and learners" (p. 2) (Figure 1). As regards the parallel corpora, "the translations included are based on human understanding of textual relations, which is not the case for machine translation (yet), despite the fact that students tend to rely more and more on it" (Kakoyianni-Doa & Tziafa, 2013, p. 3). The time period covered by the corpus spans over six centuries, from the 15th to 21st century. The texts under study are instances of different domain-specific registers, so that students may compare the results and the use of each word or phrase in different contexts. The corpus consists of a fiction and a non-fiction part of 720,282 aligned sentences.

Figure 1. The SOURCE project's home page



In the following sections we focus on the construction and composition of the Exercise tool, based on a parallel corpus, “the content, annotation, encoding and availability of which are meant to serve the needs of teachers and students of French or Greek as a foreign language and also to facilitate future linguistic research” (Kakoyianni-Doa & Tziafa, 2013, p. 1). Moreover, we outline its future perspectives and applications.

2. Method

2.1. The Source Corpus as a basis

Following Kakoyianni-Doa and Tziafa (2013) methodology, the core of the project is a collection of parallel corpora, either already existing or created by the participating researchers, and aligned (at sentence level) original and translated texts, in French and Greek. The corpus development is an ongoing process, with new texts constantly added. According to previous research, “the use of corpora in the classroom can have remarkable results as regards foreign language learning” (Kakoyianni-Doa & Tziafa, 2013, p. 2; see also Hadley, 2002). Moreover, as Kilgarriff (2009) suggests, parallel corpora are easier to be disguised as dictionaries and be brought for use in the classroom. The corpus consists of different registers (Biber, 1993), in order to facilitate comparison of the results and to study the use of each word or phrase in a different context (e.g. literature, scientific, official, technical and journalistic language). Commonly used parallel corpora, mainly EUROPARL (Koehn, 2005) and the JRC Acquis corpus (Steinberger et al., 2006) from the Opus open parallel corpus (Tiedemann, 2012) were also used, as were literary works available from Project Gutenberg, or scanned, mainly from the local historical library Severeios.

2.2. The Exercise tool construction

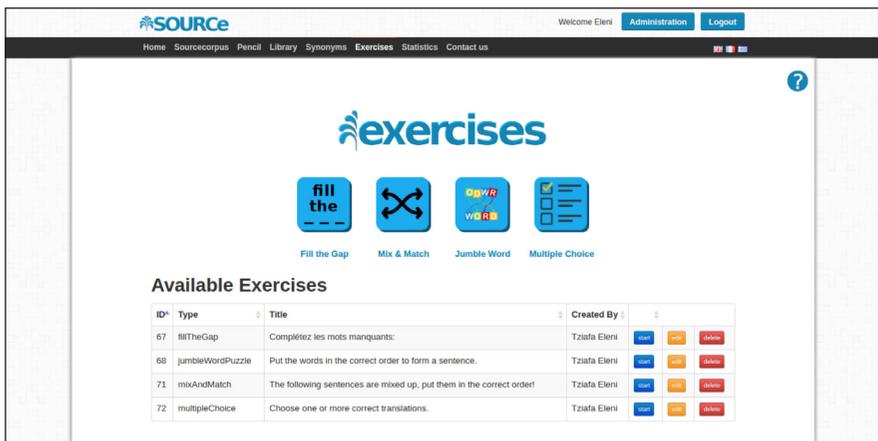
In order to support the use of corpus linguistic tools by teachers and learners with no previous expertise, we designed a simple interface through which the user may search existing corpora, upload texts, and work online with interactive exercises. The Exercise tool enables teachers to create either online activities or print out paper-based worksheets, including a variety of texts and activities on topics, in a variety of exercise types (e.g. multiple choice, word, phrase or sentence matching, filling the gaps with missing words/phrases, text reconstruction, listening, etc.). Throughout the SOURCe project, and especially in the Exercise tool, we used Java

(Enterprise Edition) for the backend and HTML5, CSS3 and Javascript for the frontend and Apache Solr for indexing, searching and sentence fetching based on a given similarity distance (used in the multiple choice exercise for the automated text selection feature, presented in the next section).

3. The Exercise tool

On the main page of the tool (Figure 2) there is a table listing the existing exercises and four buttons to create a new exercise (if the user is a teacher or administrator) or to filter the listed exercises in the table according to the exercise type (if the user is a student).

Figure 2. The exercises page (from the teacher homepage)

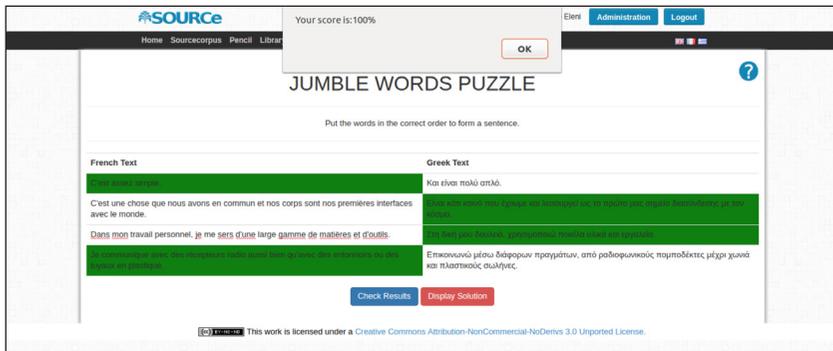


All the exercises offer (1) a common construction interface to provide the title of the exercise (in Greek and French), (2) a description, which is available only to the creator of the exercise, to take notes about the specific exercise, and (3) an automated text selection feature, which automatically fetches text from the parallel corpora collection, based on the user selection. The difference in the common automated text selection interface is that in the multiple choice exercise the teacher can select the number of suggested returned translations of every sentence based on a distance metric provided by the Apache Solr tool. The user can edit the automatically fetched text of the exercise or provide the text manually with a simple interface. All the exercises can be combined with an audio file, to produce a listening exercise. The user selects an MP3, OGG or WAV audio file and the number

of the allowed replays. The exercises can be solved either online or be given as a printout (even in cases that an audio file is attached, in order to accompany the audio file played in classroom). The combinations are innumerable. We can edit the French part, the Greek part or both, thus providing scalable difficulty levels.

When students are solving the exercise, they are able to check the results or have the solution displayed (Figure 3).

Figure 3. Solution display when the student solves a jumble words puzzle



4. Conclusions

In this paper we have presented the latest tool of the SOURCE project, the Exercise tool, which offers to French language teachers a friendly interface for online or printout exercise creation, and to students an online platform for practice. There are plenty of online exercise generators available, but they apply mostly in high-resourced languages, like English, in monolingual texts, and most of them are not free. There are very few proposals which apply on other languages (e.g. Swedish: Volodina & Borin, 2012).

Unlike other tools, the Exercise tool supports the low-resourced French and Greek languages in aligned texts. Moreover, it offers an online practice platform for students solving exercises online and checking their results in an automated manner. Finally, it is offered freely (registration required) and its source code and its assets are available under the Creative Commons license.

Our future plans include the addition of more types of texts and also exercises, along with experimental evaluations and questionnaires, in order to testify the attitude of

users toward the provided tools and also encourage teachers and students to use these resources in the classroom or beyond.

Our objective is to develop a whole platform of tools that will assist teachers to find out about, adapt and apply new tools in the classroom. To overcome the well-known problem of the existing natural language processing tools and resources not actually being included in the language learning procedure, despite their potential as learning and research tools, the main goal of the proposed project is to provide language instructors and learners with ready-made corpora and corpus-based exercises, available for use in a new learning environment. The platform provides innovative corpus-based learning activities and interactive exercises. This study could also serve as a pilot study for the creation of multilingual resources in the form of parallel corpora. This project is thus intended not only to fill a gap in the literature on corpora used in the classroom, but also to make available valuable resources, especially for a low resourced language such as Greek.

5. Acknowledgment

This project is led by Fryni Kakoyianni-Doa and is fully funded by the University of Cyprus.

References

- Biber, D. (1993). Representativeness in corpus design. *Literary and Linguistic Computing*, 8(4), 243-257. <https://doi.org/10.1093/lc/8.4.243>
- Hadley, G. (2002). Sensing the winds of change: an introduction to data-driven learning. *RELC Journal*, 33(2), 99-124. <https://doi.org/10.1177/003368820203300205>
- Kakoyianni-Doa, F., Antaris, S., & Tziafa E. (2013). A free online parallel corpus construction tool for language teachers and learners. *Procedia - Social and Behavioral Sciences*, 95(25), 535-541.
- Kakoyianni-Doa, F., & Tziafa, E. (2013). Source: building a searchable online French Greek parallel corpus for the University of Cyprus. *Revista Nebrija de Lingüística Aplicada 13* (número especial). http://www.nebrija.com/revista-linguistica/files/articulosPDF/articulo_5326df173f8bc.pdf
- Kilgarriff, A. (2009). Corpora in the classroom without scaring the students. *Proceedings of the 18th International Symposium on English Teaching and Learning in the Republic of China*. Taipei: National Taiwan Normal University. <http://www.kilgarriff.co.uk/Publications/2009-K-ETA-Taiwan-scaring.doc>

- Koehn, P. (2005). Europarl: a parallel corpus for statistical machine translation. *Proceedings of the 10th Machine Translation Summit, Phuket, Thailand* (pp. 79-86).
- Steinberger, R., Pouliquen, B., Widiger, A., Ignat, C., Erjavec, T., Tufis, D., & Varga, D. (2006). The JRC-Acquis: A multilingual aligned parallel corpus with 20+ languages. *5th Language Resources and Evaluation Conference (LREC), Genoa, Italy* (pp. 2142-2147).
- Tiedemann, J. (2012). Parallel data, tools and interfaces in OPUS. *8th International Conference on Language Resources and Evaluation (LREC)* (pp. 2214-2218). http://lrec.elra.info/proceedings/lrec2012/pdf/463_Paper.pdf
- Volodina, E., & Borin, L. (2012). Developing an Open-Source Web-Based Exercise Generator for Swedish. In L. Bradley & S. Thouësny (Eds), *CALL: Using, Learning, Knowing, EUROCALL Conference, Gothenburg, Sweden, 22-25 August 2012, Proceedings* (pp. 307-313). Dublin Ireland: Research-publishing.net. <https://doi.org/10.14705/rpnet.2012.000072>

Published by Research-publishing.net, not-for-profit association
Dublin, Ireland; Voillans, France, info@research-publishing.net

© 2016 by Editors (collective work)
© 2016 by Authors (individual work)

CALL communities and culture – short papers from EUROCALL 2016
Edited by Salomi Papadima-Sophocleous, Linda Bradley, and Sylvie Thouéšny

Rights: All articles in this collection are published under the Attribution-NonCommercial -NoDerivatives 4.0 International (CC BY-NC-ND 4.0) licence. Under this licence, the contents are freely available online as PDF files (<https://doi.org/10.14705/rpnet.2016.EUROCALL2016.9781908416445>) for anybody to read, download, copy, and redistribute provided that the author(s), editorial team, and publisher are properly cited. Commercial use and derivative works are, however, not permitted.



Disclaimer: Research-publishing.net does not take any responsibility for the content of the pages written by the authors of this book. The authors have recognised that the work described was not published before, or that it is not under consideration for publication elsewhere. While the information in this book are believed to be true and accurate on the date of its going to press, neither the editorial team, nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, expressed or implied, with respect to the material contained herein. While Research-publishing.net is committed to publishing works of integrity, the words are the authors' alone.

Trademark notice: product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Copyrighted material: every effort has been made by the editorial team to trace copyright holders and to obtain their permission for the use of copyrighted material in this book. In the event of errors or omissions, please notify the publisher of any corrections that will need to be incorporated in future editions of this book.

Typeset by Research-publishing.net

Cover design by © Easy Conferences, info@easyconferences.eu, www.easyconferences.eu

Cover layout by © Raphaël Savina (raphael@savina.net)

Photo "bridge" on cover by © Andriy Markov/Shutterstock

Photo "frog" on cover by © Fany Savina (fany.savina@gmail.com)

Fonts used are licensed under a SIL Open Font License

ISBN13: 978-1-908416-43-8 (Paperback - Print on demand, black and white)

Print on demand technology is a high-quality, innovative and ecological printing method; with which the book is never 'out of stock' or 'out of print'.

ISBN13: 978-1-908416-44-5 (Ebook, PDF, colour)

ISBN13: 978-1-908416-45-2 (Ebook, EPUB, colour)

Legal deposit, Ireland: The National Library of Ireland, The Library of Trinity College, The Library of the University of Limerick, The Library of Dublin City University, The Library of NUI Cork, The Library of NUI Maynooth, The Library of University College Dublin, The Library of NUI Galway.

Legal deposit, United Kingdom: The British Library.

British Library Cataloguing-in-Publication Data.

A cataloguing record for this book is available from the British Library.

Legal deposit, France: Bibliothèque Nationale de France - Dépôt légal: décembre 2016.