

SimpleApprenant: a platform to improve French L2 learners' knowledge of multiword expressions

Amalia Todirascu¹ and Marion Cargill²

Abstract. We present SimpleApprenant, a platform aiming to improve French L2 learners' knowledge of Multi Word Expressions (MWEs). SimpleApprenant integrates an MWE database annotated with the Common European Framework of Reference for languages (CEFR) level and several Natural Language Processing (NLP) tools: a spelling checker, a parser, and a set of transformation rules. NLP tools and resources are used to build training and writing exercises to improve MWE knowledge and writing skills of French L2 learners. We present the user scenarios, the platform's architecture, as well as the preliminary evaluation of its NLP tools.

Keywords: NLP for L2 learners, MWE, interactive Web platform, CEFR.

1. Introduction

MWE knowledge improves proficiency in writing (Paquot, 2018). Given that language learners have difficulties using MWEs, characterized by strong lexical preferences, syntactic constraints and non-compositional sense (Baldwin & Kim, 2010), learners should use them in the right context and should apply the correct morphosyntactic constraints. For instance, *jeter l'éponge* 'to abandon', is the figurative sense and the determiner is singular, while *jeter les éponges* 'throw the sponges', is used with the real sense. Collocations have strong lexical preferences (*poser une question* 'put a question', but not **demander une question* 'ask a question'). Word-for-word MWE translation generally fails (*passer l'arme à gauche* 'to kick the bucket').

1. University of Strasbourg, Strasbourg, France; todiras@unistra.fr

2. University of Strasbourg, Strasbourg, France; cargillmarion@live.fr

How to cite this article: Todirascu, A., & Cargill, M. (2019). SimpleApprenant: a platform to improve French L2 learners' knowledge of multiword expressions. In F. Meunier, J. Van de Vyver, L. Bradley & S. Thouéšny (Eds), *CALL and complexity – short papers from EUROCALL 2019* (pp. 356-361). Research-publishing.net. <https://doi.org/10.14705/rpnet.2019.38.1036>

Existing online platforms for L2 learners (Language Muse, WritingMentor) provide few lessons dealing with English MWEs. For French, projects such as Base Lexicale du Français (Verlinde, Binon, & Bertels, 2008) or DIRE Autrement (Hamel, 2010) represent MWEs' morphosyntactic and semantic features or collocation usage (Schneider & Graën, 2018). However, these resources do not propose graded CEFR exercises, except for a few websites (e.g. Bonjour de France, Le Point du FLE). To fill this gap, SimpleApprenant proposes graded exercises, annotated with CEFR levels, for MWE learning. The learner's level helps to select adequate content from the SimpleApprenant's database. Moreover, the platform provides immediate feedback by automatic error correction.

2. The SimpleApprenant platform and its scenarios

SimpleApprenant³ is an open-source Web platform, aiming to improve French L2 learners' knowledge of MWEs and writing skills. For this purpose, the platform provides CEFR level-graded exercises for learning MWEs' definitions and usage. Additionally, the platform corrects and transforms learners' productions. SimpleApprenant integrates several NLP tools and resources: an MWE database, the spelling checker LanguageTool (Naber, 2003), and the parser Mind the Gap (Coavoux & Crabbé, 2017). We also developed a set of transformation rules, requiring parsed text as input, previously checked by LanguageTool.

SimpleApprenant proposes three scenarios for learners, who freely register on the platform by indicating their CEFR level. In the first scenario, the learner should match MWEs (compatible with learners' CEFR level) with the appropriate definition or gap-filling phrase. Thus, the learner learns MWEs' definition and usage, by repeating these exercises and with positive and negative feedback (Figure 1 below).

In the second scenario, the learner writes an essay, using at least one expression from a list of MWEs, labeled with the learner's CEFR level. The teacher evaluates the essays and gives a manual feedback about the MWE usage in context. These exercises might be repeated by the learner, with new MWEs.

3. <https://simpleapprenant.huma-num.fr/SimplifyYourFrench/accueil>

Figure 1. The learner matches MWEs and their definitions (first scenario). A green message is printed if the correct answer is selected, otherwise the right answer is printed in red

SimpleApprenant

Profil

Exercices

Sondages

Déconnexion

Exercice 2

Choisissez, parmi la liste proposée, l'expression correspondant à chaque définition.

Si le contenu de l'exercice ne s'affiche pas, cela veut dire que nous ne pouvons pour le moment pas vous proposer cet exercice adapté à votre niveau.

- Avoir la liberté du mouvement des bras, des coudes.
Réponse incorrecte! La bonne réponse est : **avoir les coudées franches**
faire la claque
- Être payé pour applaudir en premier et entrainer les autres spectateurs.
Réponse incorrecte! La bonne réponse est : **faire la claque**
avoir les coudées franches
- Avoir une bonne condition physique et mentale.
Réponse correcte!
avoir la forme
- Donner suite, accepter, agréer, considérer avec bienveillance.
Réponse correcte!
faire droit
- Être attentif (à quelque chose), remarquer (quelque chose).
Réponse correcte!
faire attention

The last scenario aims to improve learners' writing skills. The learner feeds the texts to the platform, which are then processed by LanguageTool, the spelling checker integrated into SimpleApprenant. If necessary, the learner corrects the spelling errors identified by LanguageTool. Then, the corrected texts are parsed by Mind the Gap. If required, the learners apply one of the transformation rules on their parsed texts and receive the transformed text. This feedback should help the learners to avoid some grammar errors.

SimpleApprenant is currently used by French language learners and their teachers from Opole University (Poland) (A1-C1 levels) and the University of Cyprus (A1-A2 levels). We have several CEFR levels, comparable target publics (native speakers of Polish or of Greek), and the possibility to follow the same students for several years. The teachers and students use the platform during classes as an additional resource, but also at home, mainly for MWE learning or for collecting written essays. The platform is used gradually from the first (A1-A2) to the third scenario (B2-C1), according to learners' CEFR levels.

3. SimpleApprenant's architecture

For the three scenarios, SimpleApprenant uses a database of MWEs (partially annotated with the CEFR level), the spelling checker LanguageTool (Naber, 2003), Mind the Gap parser (Coavoux & Crabbé, 2017), and 21 manually defined transformation rules, requiring parsed texts as input.

We built the MWE database from Lexique-Grammaire (Gross, 1994) and from French vocabularies (Beacco, Bouquet, & Porquier, 2004). An MWE entry contains lemma, category (idiom, collocation), definition, gap-filling phrases (extracted from French Wiktionnaire), syntactic patterns, and CEFR level (Table 1). The CEFR level is automatically identified from a graded textbook corpus (Todirascu, Cargill, & François, 2019) or manually assigned from reference textbooks (Beacco et al., 2004; Gonzalez Rey, 2007).

Table 1. MWE examples

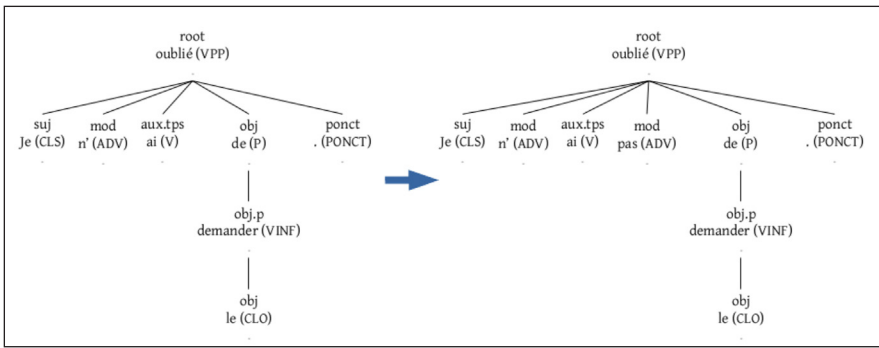
MWEs	Category	CEFR Level
Jeter l'éponge 'abandon'	idiom	B1
Faire la fête 'celebrate'	collocation	B2
Faire attention 'be careful'	collocation	A2

SimpleApprenant uses LanguageTool to detect spelling errors and Mind the Gap to create a dependency analysis of the corrected texts. The learner is asked to correct the spelling errors detected by LanguageTool. Then, the texts are parsed and the learner applies one of the transformation rules implemented in SimpleApprenant. Six deletion rules suppress adverbs and relative or participial clauses. Thirteen correction rules handle common mistakes such as verb agreement, determiner agreement, or negation errors. Complex transformation rules include passive to active voice or cleaved sentences transformed into a subject verb order structure. After the rule is applied, the learner consults the transformed text (Figure 2 below).

The transformation of learners' texts is a challenging task, due to erroneous input. Mind the Gap is a state-of-art French dependency parser: for unlabeled dependencies. For instance, the best F1 (harmonic mean of precision and recall measures) is 95.53% for reference data (Coavoux & Crabbé, 2017) but only 83.12% for learners' essays (obtained for 100 phrases of our corpus). We evaluated the transformation rules on 273 parsed sentences. Fifty-five (20.15%) were either not transformed or contained errors in the output. Out of those 55 sentences, 34 sentences (61.82%) did not show any change and 21 sentences (38.18%)

were transformed but contained errors. Deletion (82.71%) and correction rules (74.72%) are more effective than transformation rules. The rules failed because of parsing errors (due to erroneous learners' input) or syntactic patterns of the rule not matching the sentence. Even if some rules failed, agreement or negation errors are handled properly by the deletion and correction rules. As such, learners may still have feedback from these rules and see how their own text is transformed.

Figure 2. The learners apply the rule adding a second negation particle *pas* to the original dependency tree for *Je n' ai oublié de le demander* 'I do not forget to ask it', becoming *Je n' ai pas oublié de le mentionner dans mes messages*



4. Conclusion and further work

We present an online platform for French L2 acquisition, SimpleApprenant, including NLP tools supporting reformulation strategies. A large MWE database, annotated with CEFR level, is used to create exercises focusing on MWEs. The exercises are selected according to learners' CEFR levels, generated with the help of preprocessing NLP tools: a spelling checker and a parser. The evaluation of transformation rules shows that some of them should be improved before being used by teachers and learners. We are currently revising the rules to improve the system's feedback. The evaluation of the platform via Web questionnaires started with beginner and intermediate learners (A1-A2) and by teachers. The questionnaires ask the learners to classify the exercises by their difficulties and usefulness. The evaluation is still in progress and will be extended to higher learners' levels (B2-C1).

5. Acknowledgments

This research is funded by the University of Strasbourg (June 2017 - February 2019). We thank Magda Dansko (University of Opole), Fabienne Baider, and Marina Christofi (University of Cyprus) and their students for testing the platform.

References

- Baldwin, T., & Kim, S. N. (2010). Multiword expressions. In N. Indurkha & F. J. Damerau (Eds), *Handbook of natural language processing* (pp. 267-292). CRC Press.
- Beacco, J.-C., Bouquet, S., & Porquier, R. (2004). *Niveau B2 pour le français (utilisateur - apprenant indépendant) : un référentiel*. Didier.
- Coavoux, M., & Crabbé, B. (2017). Incremental discontinuous phrase structure parsing with the GAP transition. In *Proceedings of the 15th Conference of the EACL: Volume 1* (pp. 1259-1270). Association for Computational Linguistics. <https://doi.org/10.18653/v1/e17-1118>
- Gonzalez Rey, I. (2007). *La didactique du français idiomatique*. Editions Modulaires Européennes InterCommunication.
- Gross, M. (1994). Constructing lexicon-grammars. In B. T. S. Atkins & A. Zampolli (Eds), *Computational approaches to the lexicon* (pp. 213-263). Oxford University Press.
- Hamel, M.-J. (2010). Prototype d'un dictionnaire électronique pour apprenants avancés du français langue seconde. *Revue de l'APLIUT*, 29(1), 73-82. <https://doi.org/10.4000/apliut.937>
- Naber, D. (2003). *A rule-based style and grammar checker*. Ph.D. Thesis. University of Bielefeld, Germany.
- Paquot, M. (2018). Phraseological competence: a missing component in university entrance language tests? Insights from a study of EFL learners' use of statistical collocations. *Language Assessment Quarterly*, 15(1), 29-43. <https://doi.org/10.1080/15434303.2017.1405421>
- Schneider, G., & Graën, J. (2018). NLP corpus observatory – looking for constellations in parallel corpora to improve learners' collocational skills. In *Proceedings of the 7th Workshop on NLP for CALL at SLTC 2018*.
- Todirascu, A., Cargill, M., & François, T. (2019). Polylex-FLE : une base de données d'expressions polylexicales pour le FLE. *Actes de la conférence TALN 2019* (pp. 143-156).
- Verlinde, S., Binon, J., & Bertels, A. (2008). La lexicographie au service de l'apprentissage / enseignement des combinaisons de mots. *Actes du Colloque international : lexicographie et informatique* (pp. 169-175).

Published by Research-publishing.net, a not-for-profit association
Contact: info@research-publishing.net

© 2019 by Editors (collective work)
© 2019 by Authors (individual work)

CALL and complexity – short papers from EUROCALL 2019
Edited by Fanny Meunier, Julie Van de Vyver, Linda Bradley, and Sylvie Thouéšny

Publication date: 2019/12/09

Rights: the whole volume is published under the Attribution-NonCommercial-NoDerivatives International (CC BY-NC-ND) licence; **individual articles may have a different licence.** Under the CC BY-NC-ND licence, the volume is freely available online (<https://doi.org/10.14705/rpnet.2019.38.9782490057542>) 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 was not under consideration for publication elsewhere. While the information in this book is 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. 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 theme by © 2019 Frédéric Verolleman
Cover layout by © 2019 Raphaël Savina (raphael@savina.net)

Fonts used are licensed under a SIL Open Font License

ISBN13: 978-2-490057-54-2 (Ebook, PDF, colour)

ISBN13: 978-2-490057-55-9 (Ebook, EPUB, colour)

ISBN13: 978-2-490057-53-5 (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'.

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 2019.