

Conception d'un outil d'aide à la compréhension écrite pour les apprenants de français langue étrangère

Yuming Zhai, Gabriel Illouz, Anne Vilnat

► **To cite this version:**

Yuming Zhai, Gabriel Illouz, Anne Vilnat. Conception d'un outil d'aide à la compréhension écrite pour les apprenants de français langue étrangère. 9ème Conférence sur les Environnements Informatiques pour l'Apprentissage Humain, Jun 2019, Paris, France. hal-02265646

HAL Id: hal-02265646

<https://hal.archives-ouvertes.fr/hal-02265646>

Submitted on 11 Aug 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Conception d'un outil d'aide à la compréhension écrite pour les apprenants de français langue étrangère

Yuming Zhai, Gabriel Illouz et Anne Vilnat

LIMSI-CNRS, Univ. Paris-Sud, Univ. Paris-Saclay, France
{prenom.nom}@limsi.fr

Résumé. En vue d'aider les apprenants de français langue étrangère en compréhension écrite, nous concevons un outil qui propose des réécritures en contexte pour un mot ou une suite de mots. Nous nous appuyons sur les équivalences de traduction pour proposer ces réécritures, et un système de classification automatique des procédés de traduction au niveau sous-phrastique permettra de contrôler et d'interpréter le processus.

Mots-clés. TAL en EIAH, big data pour l'éducation, dispositif d'entraînement

Abstract. In order to help learners of French as a foreign language in reading comprehension, we design a tool which propose rewritings in context for a word or a sequence of words. We will rely on translation equivalences to propose these rewritings, and an automatic classification of sub-sentential translation processes will allow us to control and to interpret the searching process.

Keywords. NLP in Intelligent Learning Environment, big data for education, training device

1 Introduction

L'acquisition des langues étrangères est importante pour les étudiants, surtout quand ils veulent poursuivre des études dans un pays étranger. Au cours de l'apprentissage d'une langue étrangère, avoir recours à sa langue maternelle ou à une autre langue plus proche et déjà apprise est une pratique courante [9]. La traduction est ainsi la méthode la plus utilisée pour mieux assimiler et comparer les connaissances sur la langue. Une autre stratégie consiste en la reformulation paraphrastique [8,10,5]. Une telle compétence aide les apprenants à élargir leur vocabulaire et le répertoire d'expressions, à prendre l'habitude de réfléchir dans la langue étrangère au lieu de la langue maternelle. Elle est aussi importante pour la production et la compréhension écrite : pour simplifier ou rendre plus complexe leurs énoncés [2,3]; pour transformer le langage du texte en ses propres termes, ce qui est crucial pour vérifier leur compréhension et la lier avec leurs connaissances préalables en vue de faire des inférences [6].

Compte tenu des recherches précédentes, nous proposons la conception d'un nouvel outil pour aider les apprenants (*a priori* chinois) de Français Langue Étrangère (FLE) en compréhension écrite, via la réécriture des mots ou des suites de mots en contexte. Nous ciblons les apprenants chinois, pour tester notre hypothèse que l'apprentissage d'une troisième langue étrangère (français) peut bénéficier de l'apprentissage d'une deuxième langue étrangère similaire (anglais).¹ Étant donné ce but précis, les réécritures proposées doivent être contrôlées et de bonne qualité.

1. Une majorité des étudiants chinois apprend l'anglais depuis l'école primaire.

L'extraction de réécritures suivra l'approche d'extraction des paraphrases par des traductions communes dans des corpus parallèles bilingues [1]. Le développement de cet outil sera aussi un cadre d'application possible de nos travaux de recherche en détection automatique de procédés de traduction au niveau sous-phrastique [12,13]. Cela distingue les traductions littérales de celles non littérales [11,4], en les étiquetant comme par exemple *Généralisation*, *Particularisation*, *Modulation Sémantique*, etc. Nous intégrons la classification automatique de procédés de traduction pour avoir un meilleur contrôle sémantique pendant la recherche de réécritures, en vue de faciliter la compréhension des justifications du système et de comparer les connaissances dans des langues différentes.

2 Difficultés en compréhension écrite

Les difficultés en compréhension écrite évoluent selon le niveau de langue des apprenants. Par exemple, les débutants peuvent confondre des mots homographes, sans pouvoir distinguer leur catégorie grammaticale et leur usage en contexte. Pour les apprenants dans un niveau intermédiaire même avancé, les phénomènes qui leur posent des difficultés sont : 1) des mots inconnus des domaines généraux ; des mots déjà vus mais avec un sens inconnu ou figuré dans un contexte donné ; 2) des termes spécifiques d'un certain domaine ; 3) des expressions figées, des idiomes ; 4) des structures syntaxiques complexes, des phrases longues avec une logique complexe. Les difficultés sur la structure syntaxique dépassent la portée de notre travail. Cet outil vise à aider les apprenants à mieux comprendre le texte au niveau sous-phrastique, en proposant des réécritures en contexte.

3 Conception du prototype

Outils existants Face aux définitions parfois obscures des dictionnaires, les apprenants peuvent être découragés en tentant d'identifier le sens exact selon le contexte par eux-mêmes. *DeepL*, le concurrent de *Google Translate*, permet de varier les expressions sous-phrastiques dans la langue cible. *Linguee*, un concordancier bilingue qui souligne la traduction correspondante (mais elle pourrait être erronée), montre l'usage des segments en contexte. *Writefull* exploite les informations de fréquence à partir de larges bases de données textuelles, permettant aux utilisateurs de vérifier l'usage de segments en production écrite grâce à ses multiples fonctions. *Rewordify* simplifie automatiquement un texte anglais en enlevant les ambiguïtés, pour faciliter la lecture aux apprenants débutants ou intermédiaires.

Notre conception De notre côté, nous visons à proposer des réécritures françaises en contexte en nous basant sur la désambiguïsation sémantique, la traduction et l'alignement de mot automatique [1,7]. Nous nous appuyons sur des traductions pivots anglaises et chinoises pour les apprenants chinois. Nous utiliserons aussi l'information sur les procédés de traduction au niveau sous-phrastique pour contrôler le processus.

Notre outil sera développé sous forme d'une application web. Chaque apprenant peut s'inscrire sur le site pour sauvegarder son historique de requêtes. Nous avons besoin des informations telles que sa langue maternelle, les autres langues étrangères maî-

Conception d'un outil d'aide à la compréhension écrite pour les apprenants de FLE

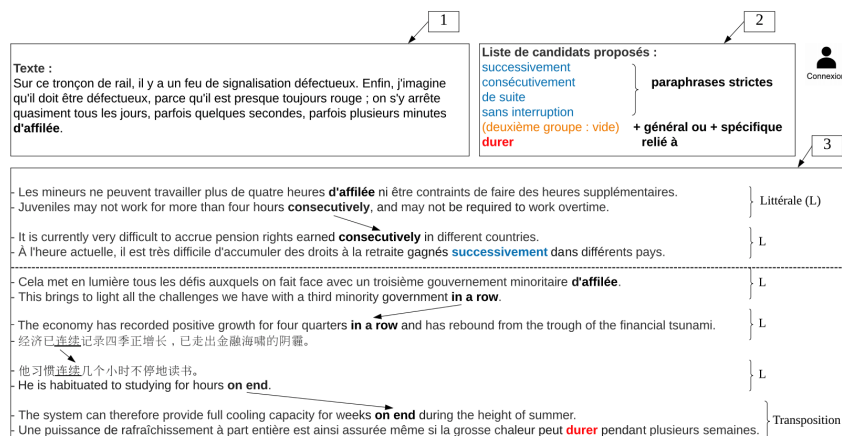


Fig. 1: Interface du prototype : réécrire en contexte pour mieux comprendre.

trisées et le niveau atteint, *etc.* pour créer son profil. L'interface la plus importante est montrée dans la figure 1. L'apprenant charge un texte dans le bloc 1 (au moins une phrase complète), et à chaque fois il sélectionne un mot ou une suite de mots qui lui pose des difficultés de compréhension. Le système récupère la phrase entière où cette entrée cherchée apparaît, afin de proposer une liste de candidats selon le contexte, affichés dans le bloc 2.

Les réécritures seront classées et contenues dans trois couleurs différentes : 1) réécritures avec une équivalence sémantique stricte, à savoir des paraphrases, et la phrase après la substitution reste grammaticale ; 2) réécritures en relation d'implication (plus générale ou plus spécifique), si elles existent ; 3) réécritures qui sont reliées avec le mot ou la suite de mots d'une certaine manière (*e.g.* ils appartiennent au même champ sémantique ; il faut adapter la structure de la phrase pour la substitution).

L'apprenant peut cliquer sur chaque candidat proposé, et le bloc 3 s'affichera pour justifier la réécriture. Ici l'entrée cherchée est *d'affiliée*, et nous montrons le processus pour trouver le candidat *successivement* et *durer*, respectivement. Pour chaque paire de traduction trouvée, nous appliquons une classification automatique pour informer si elle est littérale ou non littérale. C'est le chemin avec chaque étape typée jusqu'au candidat qui permet d'interpréter et de classer les résultats. Par exemple, la locution adverbiale *on end* est traduite par un verbe français *durer* dans la dernière paire de phrases. La classe grammaticale change avec le procédé de traduction *Transposition*, ainsi ce candidat n'est pas une paraphrase stricte en contexte. L'apprenant peut décider de n'afficher que les paires de traduction non littérale.

Nous garderons les traces d'utilisation des apprenants pour étudier la façon dont ils apprennent, voir quel procédé de traduction est plus utile pour aider la compréhension, *etc.* À long terme, nous souhaitons que cet outil puisse aussi aider les apprenants à maîtriser la traduction selon les différents procédés de traduction.

4 Conclusion

Nous proposons la conception d'un nouvel outil pour aider les apprenants de FLE en compréhension écrite, via la proposition de réécritures sous-phrastiques en contexte. L'extraction de réécritures s'appuiera sur des traductions communes dans des corpus parallèles bilingues, et nous intégrerons la classification automatique de procédés de traduction pour classer les candidats et rendre la recherche plus interprétable. Nous planifions une évaluation de notre outil à la fois quantitative avec des tests en compréhension écrite disponibles, pour observer si les apprenants obtiennent de meilleures notes en utilisant l'outil, et qualitative en classe avec la supervision d'un enseignant qui peut proposer ses propres tâches pédagogiques en utilisant ce site.

Références

1. Bannard, C., Callison-Burch, C. : Paraphrasing with bilingual parallel corpora. In : Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics. pp. 597–604. Association for Computational Linguistics (2005)
2. Chachu, S. : The intermediary role of reformulation in learning French as a foreign language : The case of students at the University of Ghana. SHS Web of Conferences **38**, 00007 (01 2017). <https://doi.org/10.1051/shsconf/20173800007>
3. Chen, M.H., Huang, S.T., Chang, J., Liou, H.C. : Developing a corpus-based paraphrase tool to improve EFL learners' writing skills. Computer Assisted Language Learning (2013). <https://doi.org/10.1080/09588221.2013.783873>
4. Chuquet, H., Paillard, M. : Approche linguistique des problèmes de traduction anglais-français. Ophrys (1989)
5. Eshkol-Taravella, I., Grabar, N. : Repérage et analyse de la reformulation paraphrastique dans les corpus oraux. In : TALN2014. pp. 304–315 (2014)
6. Kletzien, S.B. : Paraphrasing : An effective comprehension strategy. The Reading Teacher **63**(1), 73–77 (2009)
7. Kok, S., Brockett, C. : Hitting the Right Paraphrases in Good Time. In : Human Language Technologies : Conference of the North American Chapter of the Association of Computational Linguistics, Proceedings, June 2-4, 2010, Los Angeles, California, USA. pp. 145–153 (2010)
8. Martinot, C. : La reformulation : de la construction du sens à la construction des apprentissages en langue et sur la langue. Corela. Cognition, représentation, langage (HS-18) (2015)
9. Nation, P. : The role of the first language in foreign language learning. Asian EFL **5**(2), 1–8 (2003)
10. Rossari, C. : Les opérations de reformulation : analyse du processus et des marques dans une perspective contrastive français-italien. Sciences pour la communication, Peter Lang (1994)
11. Vinay, J.P., Darbelnet, J. : Stylistique comparée du français et de l'anglais : méthode de traduction. Bibliothèque de stylistique comparée, Didier (1958)
12. Zhai, Y., Max, A., Vilnat, A. : Construction of a Multilingual Corpus Annotated with Translation Relations. In : First Workshop on Linguistic Resources for Natural Language Processing. pp. 102–111 (2018)
13. Zhai, Y., Safari, P., Illouz, G., Allauzen, A., Vilnat, A. : Towards Recognizing Phrase Translation Processes : Experiments on English-French. In : (Preprint version) 20th International Conference on Computational Linguistics and Intelligent Text Processing (2019)