



HAL
open science

La réécriture monolingue ou bilingue facilite-t-elle la compréhension ?

Yuming Zhai, Gabriel Illouz, Anne Vilnat

► To cite this version:

Yuming Zhai, Gabriel Illouz, Anne Vilnat. La réécriture monolingue ou bilingue facilite-t-elle la compréhension ?. 6e conférence conjointe Journées d'Études sur la Parole (JEP, 31e édition), Traitement Automatique des Langues Naturelles (TALN, 27e édition), Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition), 2020, Nancy, France. pp.370-378. hal-02784787v1

HAL Id: hal-02784787

<https://hal.science/hal-02784787v1>

Submitted on 7 Jun 2020 (v1), last revised 23 Jun 2020 (v3)

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.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

La réécriture monolingue ou bilingue facilite-t-elle la compréhension ?

Yuming Zhai Gabriel Illouz Anne Vilnat
Université Paris-Saclay, CNRS, LIMSI, 91400, Orsay, France
{prénom.nom}@limsi.fr

RÉSUMÉ

La capacité en compréhension écrite est importante à développer pour les apprenants de langues étrangères. Cet article présente une expérience pour vérifier si les paraphrases fournies en contexte facilitent la compréhension des apprenants. Les paraphrases ont été extraites automatiquement d'un corpus parallèle bilingue. Suite à l'analyse des résultats, nous proposons des pistes d'enrichissement d'un outil conçu préalablement, pour automatiser la sélection de réécritures dans un futur travail, tout en caractérisant mieux différents types de réécritures.

ABSTRACT

Does monolingual or bilingual rewriting facilitate comprehension ?

Reading comprehension skills are important to develop for foreign language learners. This article presents an experiment to verify whether paraphrases provided in context make comprehension easier for learners. The paraphrases were automatically extracted from a bilingual parallel corpus. After analysing the results, we propose directions to enrich a previously designed tool, in order to automate the selection of rewritings in the future, while better characterizing different types of rewritings.

MOTS-CLÉS : extraction de paraphrase, compréhension écrite, TAL pour l'enseignement.

KEYWORDS: paraphrase extraction, reading comprehension, NLP for education.

1 Introduction

L'apprentissage des langues étrangères est important pour les étudiants, surtout quand ils veulent poursuivre des études dans un pays étranger. En effet, les études et l'intégration dans une société étrangère nécessitent un niveau de langue intermédiaire voire avancé. Au cours de l'apprentissage d'une langue étrangère, il est important de développer la capacité en compréhension écrite. Selon le niveau des apprenants, les difficultés en compréhension écrite évoluent. 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 ou avancé, [Yilmaz Güngör \(2015\)](#) a mené une enquête auprès de 23 étudiants de français langue étrangère (FLE) dans le cadre d'un cours de compréhension écrite. Cela montre leurs difficultés dans les catégories suivantes : lexical, grammatical, contextuel et socio-culturel. Nous listons ici quelques phénomènes concrets : des mots inconnus des domaines généraux ; des mots déjà vus mais qui apparaissent dans un contexte donné avec un sens inconnu ou figuré ; des termes spécifiques d'un certain domaine ; des nuances et des subtilités de langue ; différents registres de langue ; des expressions figées, des idiomes ; des structures syntaxiques complexes, des phrases longues avec une logique complexe, etc.

L'enquête menée par [Dagiliené \(2012\)](#) suggère que la traduction intégrée dans les activités quotidiennes des cours d'anglais s'avère utile pour le progrès des étudiants dans diverses compétences de langue. Selon [Martinot \(2012\)](#), un apprenant de langues étrangères a intérêt à prendre conscience des différentes procédures reformulateurs qu'il peut appliquer aux énoncés de la langue seconde. Compte tenu des rôles importants des traductions et des reformulations en acquisition de langues étrangères, notre problématique de recherche est d'appliquer des travaux en traitement automatique des langues (TAL) pour faciliter la compréhension écrite aux apprenants de FLE.

Dans notre publication précédente ([Zhai *et al.*, 2019b](#)), nous avons proposé la conception d'un outil dans ce but, via la proposition de réécritures des mots ou des suites de mots en contexte. Les recherches sur l'extraction de paraphrase et la reconnaissance des procédés de traduction constituent la base pour le développement de cet outil.

Dans cet article, nous décrivons une expérience utilisant des réécritures extraites automatiquement d'un corpus parallèle bilingue. Elle est menée avec la participation des apprenants chinois de FLE. Nous proposons des pistes d'enrichissement de l'outil conçu suite à l'analyse des résultats. Nous ciblons les apprenants chinois pour observer l'influence de la maîtrise d'une première langue étrangère (anglais) sur l'apprentissage d'une deuxième langue étrangère similaire (français), puisqu'une majorité des étudiants chinois apprend l'anglais depuis l'école primaire.

2 Travaux précédents

En TAL, la paraphrase est définie comme une forme alternative exprimant, dans une même langue, le même contenu sémantique, la même information ou la même idée que la forme originale ([Barzilay & McKeown, 2001](#); [Fujita, 2005](#); [Callison-Burch, 2007](#); [Bhagat, 2009](#); [Madnani & Dorr, 2010](#); [Bouamor, 2012](#)). Dans le cadre de notre travail, nous nous focalisons sur la paraphrase linguistique au niveau sous-phrasique et nous utilisons le terme « réécriture » pour désigner les reformulations textuelles qui ne gardent pas strictement la même sémantique que le segment original.

Concernant l'extraction de paraphrase, [Bannard & Callison-Burch \(2005\)](#) ont proposé d'exploiter les techniques de traduction automatique pour extraire des paraphrases à partir de corpus parallèles bilingues. Leur hypothèse est que deux segments monolingues sont des paraphrases potentielles s'ils partagent des traductions communes (appelées aussi traductions "pivots") dans une autre langue. Actuellement, la plus grande ressource de paraphrases, PPDB (*ParaPhrase DataBase*) est construite en se basant sur cette méthode ([Ganitkevitch *et al.*, 2013](#); [Pavlick *et al.*, 2015b](#)). En revanche, [Pavlick *et al.* \(2015a\)](#) ont révélé qu'il existe d'autres relations que l'équivalence stricte (paraphrase) dans PPDB.

Des traductions "pivots" non littérales dans des corpus parallèles bilingues peuvent influencer l'équivalence stricte entre les candidats de paraphrase extraits. Néanmoins, elles n'ont pas reçu assez d'attention pendant l'extraction des paraphrases dans des corpus parallèles bilingues. Afin d'étudier systématiquement les traductions non littérales, nous avons proposé une typologie de procédés de traduction adaptée à un corpus anglais-français de *TED Talks*, en nous fondant sur les théories développées en traduction ([Vinay & Darbelnet, 1958](#); [Chuquet & Paillard, 1989](#); [Molina & Hurtado Albir, 2002](#)). Nous avons aussi présenté une classification automatique des procédés de traduction en utilisant ce corpus annoté manuellement ([Zhai *et al.*, 2019a](#)).

3 Extraction automatique de paraphrases

Étant donné notre but : aider les apprenants en compréhension écrite, les paraphrases proposées en contexte doivent être contrôlées et de bonne qualité. Afin d'étudier en même temps l'utilité des traductions (réécritures bilingues), nous avons suivi l'approche d'extraction de paraphrases via la méthode par pivot dans des corpus parallèles bilingues (Bannard & Callison-Burch, 2005).

Nous avons adapté un système de traduction automatique statistique, développé par Gong *et al.* (2013). Le corpus utilisé contient celui de *TED Talks*¹ et celui de *Tatoeba*². Ce corpus anglais-français contient 397k paires de phrases parallèles (sans doublon). L'alignement automatique de mots a été effectué par l'outil *FastAlign* avec les paramètres par défaut (Dyer *et al.*, 2013). Pendant cet alignement, la langue source est le français, et la langue cible est l'anglais, afin d'utiliser l'anglais comme une langue "pivot" pour générer des paraphrases françaises.

Étant donné une entrée (un mot ou une expression, *ex.* « de par le monde »), ce système peut réaliser ces trois tâches :

- 1) Afficher toutes les phrases françaises où l'entrée apparaît, ainsi que les phrases correspondantes en anglais. Ce concordancier bilingue peut faciliter l'examen manuel du corpus parallèle.
- 2) Extraire toutes les traductions anglaises possibles pour cette entrée. (*ex.* « *over the world, around the globe* », etc.)
- 3) À partir de chaque traduction en anglais, récupérer leur rétro-traduction en français, à savoir des candidats de paraphrase. (*ex.* « à travers le monde, aux quatre coins du monde » etc.)

Suite à cette extraction automatique, nous avons effectué une sélection manuelle des paraphrases et des traductions adéquates. Quand ce système ne fournissait pas ces informations, à cause de la taille limitée du corpus ou de la difficulté d'alignement automatique de mots, nous avons eu recours à la ressource en ligne *Linguee*³. Par exemple dans cette phrase : *Quand il n'écrivait pas, il racontait des récits édifiants sur les prouesses accomplies par des personnages dans sa tête.* Le mot « édifiant » n'existe pas dans notre corpus, mais grâce à *Linguee*, nous avons pu trouver ses traductions anglaises (dans ce contexte) « *amazing* » ou « *stunning* », ce qui peut être retraduit en français comme « étonnant », « incroyable », « époustoufflant », etc. Dans certains cas, nous n'avons pu extraire que des traductions anglaises, parce qu'il n'existe pas de paraphrase (même avec l'aide de *Linguee*), par exemple les mots « pétrole », « charbon ».

Enfin, les paraphrases proposées ont été classées selon leur niveau de difficulté, avec l'aide de la ressource FLELex⁴ (Francois *et al.*, 2014), qui est le premier lexique classé pour le FLE qui indique les fréquences normalisées de mots dans chaque niveau de A1 à C2⁵.

1. Nous avons utilisé deux corpus anglais-français de *TED Talks*, qui ont été publiés pour la campagne d'évaluation IWSLT en 2013 et 2016 (<https://wit3.fbk.eu/>).

2. *Tatoeba* est un site de collection de phrases multilingues et de leurs traductions (<https://tatoeba.org/eng>).

3. <https://www.linguee.fr/>

4. <http://cental.uclouvain.be/flelex/>, la version FLELex_CRF contient 2038 expressions multi-mots.

5. <https://www.france-langue.fr/niveaux-de-francais/>

4 Expériences avec les apprenants chinois

Pour étudier notre hypothèse que des réécritures monolingues ou bilingues facilitent la compréhension, nous avons mené deux expériences sur un test de compréhension écrite avec des étudiants chinois adultes. Les tests se sont déroulés sous forme de QCM (sur la plateforme *Moodle*). Nous avons contacté trois enseignantes de FLE de trois établissements chinois.⁶ Nous avons échangé des idées sur la préparation des textes et des questions avec ces enseignantes. Après les tests, les étudiants ont dû répondre à un questionnaire sur leur profil (niveau d'étude en français et en anglais) et leur satisfaction. 15 étudiants de niveau intermédiaire (A2) et 11 étudiants de niveau avancé (B2) ont participé.⁷

Nous avons préparé un texte pour chaque niveau selon la grille pour l'auto-évaluation du CECR⁸. Le texte du niveau A2 (548 tokens) est issu d'un site de ressource pour l'éducation, et nous l'avons légèrement modifié. Celui du niveau B2 (300 tokens) a été écrit par nous à partir d'une liste de mots et de segments. Pour chaque groupe d'étudiants, le test est divisé en trois phases, avec pour chacune une version différente du texte qui a été présentée aux étudiants : 1) version originale (sans aucune information complémentaire) ; 2) les étudiants peuvent regarder les paraphrases des mots ou des expressions mises en gras (pas uniquement ceux qui sont pertinents pour répondre aux questions), qui apparaissent dans une fenêtre quand le curseur est passé dessus ; 3) en plus de ces paraphrases en français, les traductions anglaises sont ajoutées.

À chaque phase, les étudiants répondent au même ensemble de questions conçues par les auteurs. Notre hypothèse est la suivante : grâce aux informations complémentaires, les apprenants peuvent répondre correctement à de plus en plus de questions. Nous avons respecté les principes suivants pour la conception des questions et des options de réponses : 1) les questions concernent les mots ou les expressions potentiellement difficiles, mais pas la logique du texte ; 2) les paraphrases et les traductions anglaises fournies ne doivent pas contenir exactement les mêmes chaînes de caractères que celles présentes dans les réponses proposées ; 3) les options de réponse doivent être assez proches les unes des autres pour que les distracteurs soient bons. Les distracteurs ne doivent pas permettre de trouver la bonne réponse par déduction de façon évidente.

Nous avons limité le temps pour chaque test : 30 minutes pour le premier, et 15 minutes pour chacun des deux suivants. L'utilisation des dictionnaires a été interdite pendant les tests (peu importe le format). Nous mettons à disposition le matériel utilisé dans ces expériences⁹ : les textes en trois versions en format HTML, les questions de compréhension, les listes des paraphrases et traductions proposées, le questionnaire et les réponses anonymisées.

5 Analyse des résultats

Résultats du groupe A2 Les résultats par participant sont montrés dans la figure 1(a). Pour deux étudiants (N° 4, 15), des informations complémentaires semblent apporter une aide supplémentaire. La figure 1(b) montre le nombre d'étudiants ayant choisi la bonne réponse pour chaque question.

6. Université des langues étrangères de Dalian, Alliance Française de Dalian et Collège de technologie professionnelle de la ville de Ningbo.

7. 44 étudiants ont été contactés, en revanche, seuls 26 étudiants ont fait le test malgré des relances.

8. <https://www.coe.int/fr/web/portfolio/self-assessment-grid>

9. https://github.com/YumingZHAI/reecriture_comprehension

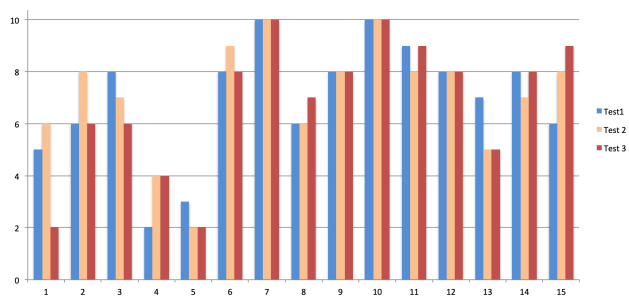
Nous voyons que la deuxième question est la plus difficile. Elle porte sur la phrase suivante et sur le mot « survivre » : [...] *il fera bientôt beaucoup trop chaud sur la Terre pour que certaines espèces puissent survivre*.

La question est : *Quand il fera bientôt beaucoup plus chaud, certains animaux vont :*

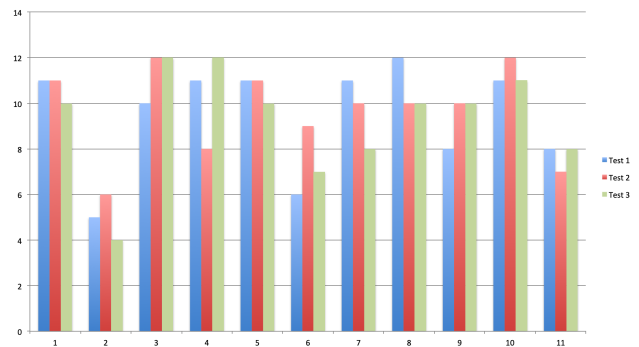
a) disparaître b) souffrir c) changer de lieu de vie

Les paraphrases données sont « continuer à vivre » et « toujours en vie », et la traduction anglaise est « survive ». La plupart des étudiants ont choisi « changer de lieu de vie » or la bonne réponse est « disparaître ». Nous supposons que la difficulté ne concerne non seulement le mot « survivre », mais aussi l'expression « trop + adj. + pour que ... puisse ... ».

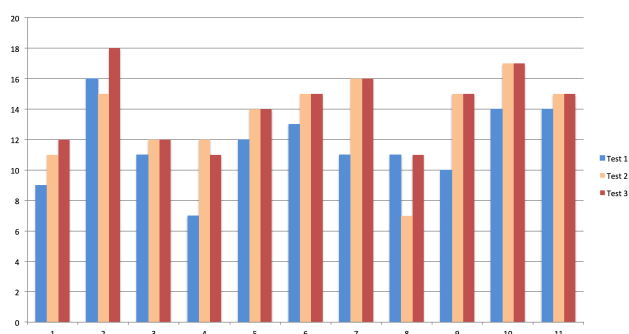
Nous avons analysé leurs réponses au questionnaire. En général, le niveau d'anglais n'est pas très élevé pour plusieurs étudiants, surtout ceux qui ont eu des moins bons résultats. Ceux qui ont le mieux répondu, ont dit qu'ils apprennent l'anglais et le français en les comparant, tandis que les autres étudiants ne veulent pas mélanger l'apprentissage des deux langues. La majorité des étudiants ont confirmé que la maîtrise de l'anglais favorise l'apprentissage du français, surtout pour le vocabulaire et la grammaire. Dix personnes (sur quinze au total) ont indiqué qu'un outil proposant des paraphrases en contexte pourrait les aider dans la lecture. Neuf personnes trouvent que la proposition des traductions anglaises est utile dans un tel outil, mais qu'il serait encore plus utile si une phrase avec un contexte était donnée en même temps.



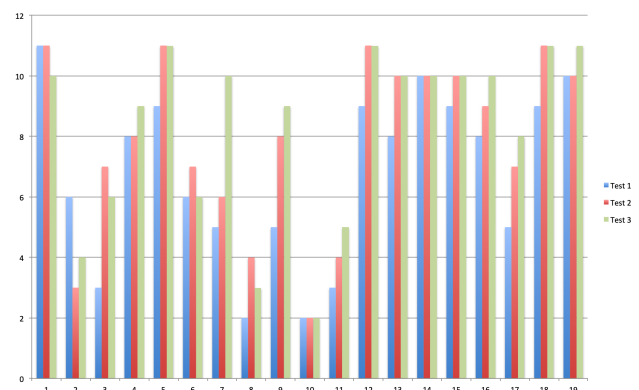
(a) Nombre de bonnes réponses par participant : 15 participants (axe X), 11 questions (axe Y)



(b) Nombre de personnes qui répondent correctement par question : 11 questions (axe X), 15 participants (axe Y)



(c) Nombre de bonnes réponses par participant : 11 participants (axe X), 19 questions (axe Y)



(d) Nombre de personnes qui répondent correctement par question : 19 questions (axe X), 11 participants (axe Y)

FIGURE 1: Résultats pour les apprenants de niveau A2 (en haut) et B2 (en bas)

Résultats du groupe B2 Les résultats par participant et par question sont montrés dans les figures 1(c) et 1(d). Pour les questions N° 1, 14, 19, la performance est déjà très élevée au test 1.

À part l'étudiant N° 8, tous les autres étudiants ont eu des meilleures notes avec des informations complémentaires par rapport au premier test. Pour trois étudiants (N° 1, 2, 8), l'ajout des traductions anglaises apporte de l'aide supplémentaire en plus des paraphrases. En général, l'évolution des performances du groupe B2 est plus homogène que celle du groupe A2.

La question la plus difficile est la question 10, peu importe la version du texte, seuls deux étudiants ont donné la bonne réponse. Elle concerne cette phrase : *On y voyait un ivrogne errer avec son calepin à la main, avec **dans son sillage** certains de ses compagnons.*

La question est : *Où voyait-on les compagnons de cet homme ?*

a) autour de lui b) derrière lui c) devant lui d) ce n'est pas dit

Cela montre que la paraphrase (« sur ses traces ») et les traductions (« *in the tracks/wake of* ») ne sont pas suffisamment claires pour aider la compréhension. Cette difficulté est aussi due aux principes de la conception de cette expérience, où les informations complémentaires ne doivent pas rendre la réponse évidente.

Dans la figure 1(d), nous constatons que pour la deuxième question, l'ajout d'informations nuit à la compréhension. Cette question porte sur la phrase suivante : *Ainsi, j'ai pu voir à la **lisière** d'une ville, dans un pâté de maisons isolé [...]*

La question est : *Où se déroule l'histoire ?*

a) au centre-ville b) à la campagne c) aucune des deux réponses

Nous avons fourni « frontière » et « bordure » comme paraphrase, et « *edge* » comme traduction. La bonne réponse est « aucune des deux », mais « à la campagne » a été choisie par la majorité des étudiants à la troisième phase du test. Seuls quatre étudiants ont saisi la nuance entre la « lisière d'une ville » et « la campagne », sachant que cette notion existe aussi en chinois.

Parmi ces 11 étudiants en niveau B2, trois sont en quatrième année de licence du département langue française en Chine. Les autres poursuivent leurs études en France, dont certains ont eu des bonnes notes en TOEIC.

Quatre étudiants (N° 2, 6, 7, 8) préfèrent ne pas mélanger l'anglais et le français pendant l'apprentissage, alors que les autres les apprennent par comparaison. Les étudiants confirment que la maîtrise de l'anglais favorise l'apprentissage du français, à part les similarités en vocabulaire et en grammaire, ils ont aussi mentionné l'étymologie et les expressions figées. Excepté deux étudiants qui préfèrent toujours utiliser un dictionnaire, les autres expriment leur envie d'avoir un outil qui propose des paraphrases en contexte pour élargir leur vocabulaire. L'ajout des traductions anglaises dans un tel outil est apprécié par huit étudiants, tandis que les trois autres préfèrent n'avoir affaire qu'à une seule langue étrangère à la fois.

6 Bilan et perspectives

Les résultats montrés dans la figure 1 valident notre hypothèse (et de façon plus évidente pour le groupe B2) : les paraphrases facilitent la compréhension des apprenants. Pourtant, l'hypothèse que les traductions anglaises fournissent encore de l'aide supplémentaire n'est pas encore complètement validée. Nous pensons qu'une expérience de plus large envergure, avec plus de textes et de participants, est nécessaire pour mieux étudier nos hypothèses. Nous testerons des procédures plus méthodologiquement robustes que celles utilisées dans notre première tentative. Par exemple, demander à différents groupes d'étudiants de faire le test avec différents niveaux d'informations et de

comparer les résultats doit être considéré.

Pour notre travail futur, des corpus parallèles plus volumineux seront exploités pour extraire des paraphrases. Dans cette expérience, l'extraction des paraphrases a été automatique mais le filtrage a été manuel. L'utilité des paraphrases pour faciliter la compréhension étant vérifiée, nous tenterons d'automatiser ce filtrage dans le futur. Pour cela, nous intégrerons la reconnaissance automatique des procédés de traduction (*ex.* classification entre traduction littérale et non littérale) (Zhai *et al.*, 2019a). Notre hypothèse est que pendant l'extraction de réécritures via la méthode par pivot, la reconnaissance des procédés de traduction nous permettra de caractériser les réécritures dans les trois cas suivants : 1) réécritures en équivalence sémantique stricte, à savoir des paraphrases, et la phrase après la substitution reste grammaticale ; 2) réécritures en relation d'implication (plus général ou plus spécifique), si elles existent ; 3) réécritures reliées avec le mot ou la suite de mots d'une certaine manière (par exemple, ils appartiennent au même champ sémantique ; il faut adapter la structure de la phrase pour la substitution). Cela nous permettra d'éviter les contresens, de classer les réécritures et d'étudier quel type de réécriture conviendrait mieux à chaque type d'apprenants.

Pour le développement de l'outil, nous suivons la chaîne de traitement présentée dans la figure 2. Nous sommes aussi conscients des aspects suivants à prendre en compte pendant l'implémentation et l'évaluation d'un tel outil : l'adaptation du système face aux demandes différentes des apprenants et l'évaluation longitudinale de l'outil selon les différents niveaux des apprenants.

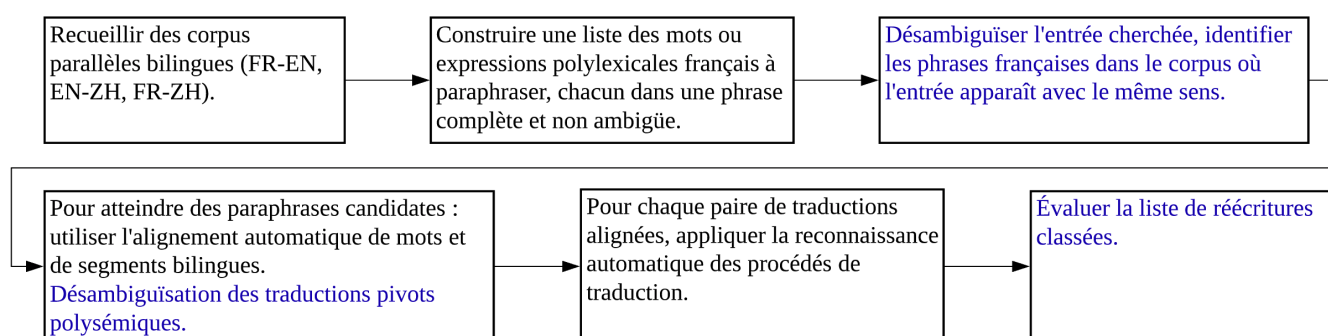


FIGURE 2: Chaîne de traitement pour le développement de l'outil (en noir : les blocs prêts ou en développement ; en bleu : les blocs à développer)

Remerciements

Nous remercions les relecteurs anonymes pour leurs remarques détaillées et constructives. Nous exprimons notre gratitude aux participants, et aux enseignantes qui ont aidé à solliciter les participants en Chine : Chunhong Yu, Jingjing Liang, Xinyan Li, Miao Wang et Xiaoya Xu. Nous remercions l'agence ANR pour son financement à travers le projet ALECTOR (ANR-16-CE28-0005).

Références

BANNARD C. & CALLISON-BURCH C. (2005). Paraphrasing with bilingual parallel corpora. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, p. 597–604 : Association for Computational Linguistics.

- BARZILAY R. & MCKEOWN K. (2001). Extracting Paraphrases from a Parallel Corpus. In *Association for Computational Linguistic, 39th Annual Meeting and 10th Conference of the European Chapter, Proceedings of the Conference, July 9-11, 2001, Toulouse, France.*, p. 50–57.
- BHAGAT R. (2009). *Learning Paraphrases from Text*. Thèse de doctorat, Los Angeles, CA, USA. AAI3368694.
- BOUAMOR H. (2012). *Etude de la paraphrase sous-phrastique en traitement automatique des langues. (A study of sub-sentential paraphrases in Natural Language Processing)*. Thèse de doctorat, University of Paris-Sud, Orsay, France.
- CALLISON-BURCH C. (2007). *Paraphrasing and Translation*. Thèse de doctorat, University of Edinburgh, Edinburgh, Scotland.
- CHUQUET H. & PAILLARD M. (1989). *Approche linguistique des problèmes de traduction anglais-français*. Ophrys.
- DAGILIENĖ I. (2012). Translation as a learning method in English language teaching. *Kalbu studijos*, **21**, 124–129. DOI : [10.5755/j01.sal.0.21.1469](https://doi.org/10.5755/j01.sal.0.21.1469).
- DYER C., CHAHUNEAU V. & SMITH N. A. (2013). A simple, fast, and effective reparameterization of IBM model 2. In L. VANDERWENDE, H. D. III & K. KIRCHHOFF, Édts., *Human Language Technologies : Conference of the North American Chapter of the Association of Computational Linguistics, Proceedings, June 9-14, 2013, Westin Peachtree Plaza Hotel, Atlanta, Georgia, USA*, p. 644–648 : The Association for Computational Linguistics.
- FRANCOIS T., GALA N., WATRIN P. & FAIRON C. (2014). FLELex : a graded Lexical Resource for French Foreign Learners. In N. C. C. CHAIR), K. CHOUKRI, T. DECLERCK, H. LOFTSSON, B. MAEGAARD, J. MARIANI, A. MORENO, J. ODIJK & S. PIPERIDIS, Édts., *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland : European Language Resources Association (ELRA).
- FUJITA A. (2005). *Automatic Generation of Syntactically Well-formed and Semantically Appropriate Paraphrases*. Thèse de doctorat, Ph. D. thesis, Nara Institute of Science and Technology.
- GANITKEVITCH J., VAN DURME B. & CALLISON-BURCH C. (2013). PPDB : The paraphrase database. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics : Human Language Technologies*, p. 758–764.
- GONG L., MAX A. & YVON F. (2013). Improving bilingual sub-sentential alignment by sampling-based transpotting. In *Proceedings of IWSLT*, Heidelberg, Germany.
- MADNANI N. & DORR B. J. (2010). Generating phrasal and sentential paraphrases : A survey of data-driven methods. *Computational Linguistics*, **36**(3), 341–387. DOI : [10.1162/coli_a_00002](https://doi.org/10.1162/coli_a_00002).
- MARTINOT C. (2012). De la reformulation en langue naturelle, vers son exploitation pédagogique en langue étrangère : pour une optimisation des stratégies d'apprentissage. *Synergies Pologne*, **9**, 63–76. <http://gerflint.fr/Base/Pologne9/martinot.pdf>.
- MOLINA L. & HURTADO ALBIR A. (2002). Translation Techniques Revisited : A Dynamic and Functionalist Approach. *Meta*, **47**(4), 498–512. DOI : [10.7202/008033ar](https://doi.org/10.7202/008033ar).
- PAVLICK E., BOS J., NISSIM M., BELLER C., VAN DURME B. & CALLISON-BURCH C. (2015a). Adding semantics to data-driven paraphrasing. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1 : Long Papers)*, volume 1, p. 1512–1522.

PAVLICK E., RASTOGI P., GANITKEVITCH J., DURME B. V. & CALLISON-BURCH C. (2015b). PPDB 2.0 : Better Paraphrase Ranking, Fine-Grained Entailment Relations, Word Embeddings, and Style Classification. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, ACL 2015, July 26-31, 2015, Beijing, China, Volume 2 : Short Papers*, p. 425–430.

VINAY J.-P. & DARBELNET J. (1958). *Stylistique comparée du français et de l'anglais : méthode de traduction*. Bibliothèque de stylistique comparée. Didier.

YILMAZ GÜNGÖR Z. (2015). La compréhension des textes en français langue étrangère : quelles difficultés ? *Journal of International Social Research*, **8**(40).

ZHAI Y., ILLOUZ G. & VILNAT A. (2019a). Classification automatique des procédés de traduction. In *26th Conférence sur le Traitement Automatique des Langues Naturelles*, Toulouse, France. HAL : [hal-02265644](https://hal.archives-ouvertes.fr/hal-02265644).

ZHAI Y., ILLOUZ G. & VILNAT A. (2019b). Conception d'un outil d'aide à la compréhension écrite pour les apprenants de français langue étrangère. In *9ème Conférence sur les Environnements Informatiques pour l'Apprentissage Humain*, p. 379–382, Paris, France. HAL : [hal-02265646](https://hal.archives-ouvertes.fr/hal-02265646).