LE PROJET AFPC : CORPUS PARALLÈLE ARABE-FRANÇAIS

Géhane ESSAWY
Academy of Arts, Égypte
jessawy@edu.ksu.sa

Hind ALOTAIBI
King Saud University, Royaume d'Arabie saoudite
hialotaibi@ksu.edu.sa

Hakima GUELLA
Université de la Princesse Nourah, Royaume d'Arabie saoudite
hguella@pnu.edu.sa.

Résumé : Le recours à la TAO étant indispensable, les corpus parallèles jouent un rôle de plus en plus important dans la pratique de la traduction et la didactique de la traduction, car ils soutiennent l'approche Data-Driven Learning DDL. Cet article propose un projet basé sur une approche d'alignement de corpus parallèle français-arabe et décrit un projet en cours à la Faculté de Langues et de Traduction de l'Université Roi Saoud pour compiler un corpus parallèle arabo-français de 3 millions de mots qui servira de ressource pour la formation à la traduction. Nous allons expliquer les différentes étapes de ce projet, y compris le choix des textes, la segmentation…etc. Ensuite nous allons montrer son apport à la formation des traducteurs arabophones. Le corpus sera disponible gratuitement en texte intégral au format XML et à travers une interface web conviviale qui fournit un concordancier prenant en charge les besoins de recherche bilingues et plusieurs options de filtrage.

Mots-clés : corpus parallèle, traductologie de corpus, traitement de l'arabe, corpus bilingue, corpus français-arabe.

Abstract: As CAT is essential, parallel corpora play an increasingly important role in translation practice and translation didactics, as they support the Data-Driven Learning DDL approach. This article proposes a project based on an French-Arabic parallel corpus alignment approach and describes an ongoing project at the Faculty of Languages and Translation of King Saud University to compile an Arab-French parallel corpus of 3 million words that will serve as a resource for translation training. We will explain the different stages of this project, including the choice of texts, segmentation… etc. Then we will show its contribution to the training of Arabic-speaking translators. The corpus will be available free of charge in full text in XML format and through a user-friendly web interface that provides a concordancer supporting bilingual search needs and several filtering options.

Keywords: parallel corpus, corpus translation studies, treatment of Arabic, bilingual corpus, Arabic-French corpus.

BDD-A33052 © 2021 Editura Universităţii din Suceava
Provided by Diacronia.ro for IP 216.73.216.75 (2026-10-30 00:14:35 UTC)

1. Introduction

Avec l'arrivée de la traduction automatique neuronale en 2015, nombre d'entreprises de services linguistiques et traductionnels ont de plus en plus recours à la TAO (Traduction assistée par ordinateur). L'importance d'une « culture de l'informatique » (Arrouart, 2003) inclut les mémoires de traduction et encore, la création et l'exploitation de corpus dans l'enseignement et la pratique de la traduction et des langues.

Si les uns tentent de surévaluer la qualité de la traduction automatique au détriment de la traduction humaine, l'intégration des outils informatiques à l'activité traductionnelle est indéniablement fructueuse. Une approche raisonnée (Look, 2019) de l'utilisation de la TAO assurera l'équilibre souhaité entre la démystification de la machine (Moorkens,2018) et la maîtrise des enjeux techniques offerts par ces nouveaux outils afin d'être capable d'analyser et de manipuler ce processus complexe qu'est l'activité traduisante. Le développement des *soft-skills* implicant l'usage outils à disposition permettrait au traducteur de gérer des projets de traduction qui dépassent le simple transfert linguistique pour expérimenter des outils génériques.

Les corpus linguistiques sont utilisés depuis assez longtemps et ont été des ressources indispensables pour étudier les cooccurrences et l'évolution de la langue. A côté des mémoires de traduction, les outils issus de la linguistique de corpus deviennent un élément enrichissant de toute dernière génération offrant des possibilités d'investigation importantes pour tout traducteur.

L'acte de traduire implique non seulement une parfaite maîtrise et de la langue source et de la langue cible, mais aussiune exigence de se familiariser avec une langue -, ses caractéristiques, ses nuances, ses péripéties, ses tournures, ses idiomes; il a fallu donc trouver un moyen moderne d'exploration plus profonde des différentes langues : le corpus.Le corpus peut donc être défini comme étant une base de données formée de textes électroniques stockés selon des critères bien précis dans la recherche linguistique (Leech, 2010). Les outils informatiques d'aide à la traduction occupent ainsi une place grandissante dans l'univers du traducteur.

Au cours des vingt dernières années (Loock, 2019; Bowker, Pearson, Laviosa, 2002 ; Zanettin, Bernardini, Stewart, 2003), les travaux sur les corpus à fin didactique se sont multipliés. Profiter des corpus de textes a pour objectif de se familiariser avec cette notion et utiliser les corpus existants pour pouvoir interpréter le processus traductif, démarche inhérente à l'enseignement et à l'apprentissage de la traduction.

Il s'avère donc nécessaire de « comprendre les contraintes, les pressions et les motivations qui influencent spécifiquement l'acte traductionnel et sous-tendent cette forme unique de « communication », comme l'avait précisé Mona Baker (1998), l'une des fondatrices des travaux à partir de corpus.

Une description des travaux antérieurs sur l'importance de l'exploitation des corpus en traduction et en traductologie sera suivie par une présentation de corpus parallèles du couple des langues français-arabe. La troisième partie de

BDD-A33052 © 2021 Editura Universităţii din Suceava
Provided by Diacronia.ro for IP 216.73.216.75 (2026-10-30 00:14:35 UTC)

notre travail d'expérimentation sera consacrée à la méthodologie adoptée dans notre étude et aux différentes étapes de la conception du projet AFPC. La dernière section portera sur les conclusions, les recommandations et les nouvelles perspectives qu'offre le présent travail.

2. Travaux antérieurs

Une enquête menée en Europe (Mellange, 2006) a montré que les traducteurs professionnels souhaitent être informés sur les outils liés à l'exploitation de corpus surtout dans des domaines spécifiques. D'autres travaux ont démontré l'importance des corpus dans différents domaines tels que l'enseignement des langues (Botley,1996), la traductologie (Baker 1993 ; 1995 ; Olohan, 2004), la traduction automatique Rauf et Schwenk ; 2011 ; Tian et autres, 2014), des études terminologiques (Bowker Pearson, 2002) et également la linguistique comparée (Johansson, 1999).

L'apport que représentent les corpus à la traduction est indéniable. Les études lancées par Look (2018) et par Lynne Bowker (1998) ont prouvé que l'utilisation d'un corpus comparé (bilingue portant sur un seul domaine spécifique) a largement amélioré la pratique traductionnelle par rapport aux ressources conventionnelles comme les dictionnaires car, elle a permis une mise en contexte qui a facilité la compréhension du domaine et par la suite assuré une meilleure maîtrise du choix des termes et des idiomes.

Les corpus sont, sous une forme plus ou moinstransparente, omniprésents dans tout l'environnement du traducteur (Loock 2016) ; à part les bases de données terminologiques, les mémoires de traduction, les sites de traduction automatiques, les corpus- comparables ou parallèles- trouvent leur place dans l'univers traductologique. Les corpus peuvent être unilingues, utilisant des textes d'une seule langue, ou bilingues, utilisant deux langues différentes, voire multilingues.

3. Les corpus bilingues/ parallèles :

Les corpus bilingues sont des documents archivés ou mis en ligne que les traducteurs peuvent utiliser comme référence lors de leur activité traductionnelle. Sinclair les définit comme étant « des textes authentiques traduits en une (ou plusieurs) langue(s) étrangère(s) », (Sinclair, 1995). Les concordanciers bilingues des deux versions d'un texte sont alignés section par section; la version traduite du passage voulu apparaît immédiatement quand on utilise un concordancier.

Ce type de corpus permet d'observer les caractéristiques linguistiques de la traduction, confronter un texte écrit dans sa langue originale et sa traduction en langue cible et de découvrir les stratégies mises en œuvre par les traducteurs. Le chercheur pourrait ainsi comparer les aspects linguistiques de chacune des deux langues, leurs valeurs sémantiques, leurs références culturelles ainsi que leurs variantes syntaxiques. Ceci conduit indéniablement à une meilleure compréhension et de la langue source et de la langue cible et à une maitrise

BDD-A33052 © 2021 Editura Universităţii din Suceava
Provided by Diacronia.ro for IP 216.73.216.75 (2026-10-30 00:14:35 UTC)

parfaite des systèmes linguistiques propres à chacune des deux langues choisies et en fin de compte à répondre à des difficultés liées aux références culturelles.

Les Corpus parallèles sont utilisés en traduction tant au niveau théorique qu'au niveau pratique (Hunston, 2002) :

- Le plan théorique s'intéresse aux processus traductifs, aux moyens utilisés pour formuler une même idée dans deux langues différentes tout en établissant des comparaisons entre les aspects de chacun des textes, original et traduit.
- Le plan pratique, lui, se concentre sur les différentes utilisations des corpus dans le développement des outils à la traduction assistée par ordinateur, comme les mémoires de traduction, les systèmes de gestion de terminologie, l'enseignement de la traduction et surtout la formation des traducteurs.

Gouadec a confirmé que la formation des traducteurs inclut l'initiation à exécuter des tâches bien précises dans des contextes définis en utilisant des outils et des systèmes préétablis. (Gouadec, 2002).

Donc, les corpus parallèles sont utilisés à des fins didactiques et font partie des outils utilisés par les étudiants pour réaliser un projet en traduction qui vise à développer des corpus spécialisés (juridiques, médicaux, etc…). Les formateurs assurent une initiation à la conception de corpus, aux critères généraux qui conditionnent cette activité, aux différents outils de recherche, aux logiciels appropriés, etc…

Ainsi, ces corpus parallèles contribuent à aider les étudiants à différencier les caractéristiques du texte original et du texte traduit, à repérer les techniques du transfert du sens, à analyser l'influence de la langue maternelle sur la langue seconde, à distinguer les homonymes, comprendre les expressions figées et les expressions courantes propres à chacune des deux langues utilisées.

L'intérêt de ce genre de corpus c'est qu'ils offrent, à l'encontre des dictionnaires bilingues, des exemples concrets de pratiques traductionnelles qui mettent en relief le contexte qui a conditionné les choix du traducteur dans un cadre spatio-temporel bien déterminé.

Pearson (2003) estime que les corpus parallèles permettent aux étudiants de comparer leurs propres traductions avec celles des traducteurs professionnels et par la suite découvrir et acquérir les techniques et les stratégies traductives pratiquées par ces traducteurs. Ce genre de corpus pourrait également enrichir les connaissances des étudiants dans le domaine de la traduction spécialisée et de pouvoir facilement discerner les particularités, la terminologie et les structures propres à chaque domaine (Kübler, 2011).

Plusieurs travaux ont été consacrés aux apports des corpus au développement des compétences des traducteurs et ont encouragé leur exploitation et leur intégration dans les programmes de formation des étudiants-traducteurs (Bernardini, 2004). Frankenberg-Garcia et Santos (2003) ont utilisé le corpus parallèle « Compara » pour initier les étudiants aux variantes et aux différents aspects des langues anglaise et portugaise. L'objectif

BDD-A33052 © 2021 Editura Universităţii din Suceava
Provided by Diacronia.ro for IP 216.73.216.75 (2026-10-30 00:14:35 UTC)

était de savoir comment traiter des différences linguistiques, syntaxiques et extra-linguistiques que pose la traduction entre ces deux langues. Les deux chercheurs ont prouvé que l'usage du corpus parallèle a largement contribué à la formation des traducteurs et ont incité les responsables à introduire les corpus parallèles dans les curriculums universitaires de traduction en tant que ressource pédagogique indispensable.

Dans une étude récente, Salhi (2013) a démontré que les corpus parallèles peuvent également contribuer à interpréter le texte original ; il a effectué son étude sur le corpus des Nations Unies « EAPCOUNT », et il a à titre d'exemple analysé la polysémie du mot *destruction* dans les différents contextes dans la langue d'origine. Il a ainsi mis l'accent sur l'importance d'un corpus parallèle pour éclaircir les perspectives sémantiques du texte original, perspectives qui pourraient être bénéfiques dans le traitement de la langue cible et du texte traduit.

De sa part, Rodriguez (2014) a fait preuve de l'importance des corpus parallèles dans la formation des traducteurs de langues anglaises et espagnoles. Son étude a démontré que le corpus parallèle a profondément sensibilisé les étudiants aux structures comparatives et aux expressions figées et courantes propres à chacune des deux langues. Un questionnaire effectué auprès des sujets de l'étude a révélé qu'ils ont acquis, grâce à l'utilisation du corpus, plus de confiance en eux-mêmes lors de leur pratique traductive.

En 2014, Haiyan Li et Zhongxin Dai ont utilisé un corpus parallèle de textes anglais traduits en chinois pour la formation de quatre-vingt-dix étudiants en traduction. Ilsont divisé les étudiants en deux groupes ; le premier groupe a reçu un enseignement traditionnel, et le deuxième a bénéficié d'une méthode d'enseignement fondée sur la recherche dans les bases de données des corpus parallèles. Des tests ont été effectués avant et après la formation ; les résultats ont prouvé que le deuxième groupe a donné des meilleurs résultats du fait qu'il a été initié à l'usage des corpus.

Heylen et Verplaetse (2015) ont utilisé les corpus parallèles dans les cours de traduction médicale et analysé leur influence sur le comportement traductionnel des étudiants. Des étudiants en master, répartis en deux groupes, devaient traduire de l'anglais vers l'allemand des textes médicaux en utilisant des corpus parallèles. Le groupe ayant utilisés les corpus parallèles s'est montré plus performant que l'autre groupe surtout au niveau terminologique.

Cependant, et malgré l'intérêt accordé aux corpus bilingues dans les milieux académiques, la langue arabe est en grande pénurie de ce type de ressources linguistiques et pédagogiques indispensables à la formation des futurs traducteurs (Al-Ajmi, 2011 ; Al-Sultaiti, Atwel, 2006).L'élaboration d'un wordnet pour l'arabe (Elkateb et al., 2006) constitue une ressource importante mais demeure, à elle seule , insuffisante.

La mise en œuvre de cours de traduction introduisant les corpus étant rare dans le monde arabe, le but du présent travail est de fournir un outil informatique moderne qui permettrait une nouvelle approche de la

BDD-A33052 © 2021 Editura Universităţii din Suceava
Provided by Diacronia.ro for IP 216.73.216.75 (2026-10-30 00:14:35 UTC)

traductologie, de l'enseignement de la traduction et de l'apprentissage des langues, en intégrant cette discipline de traductologie de corpus qui permettrait des'interroger sur les pratiques et les difficultés de la pratique traductionnelle. Or, faciliter les modes de recherche et mettre à la disposition de l'utilisateur des outils numériques, n'est pas une tâche facile étant donné que plusieurs variantes contextuelles affectent l'aspect sémantique d'un terme ou d'une expression pour ce couple de langues. Tel est le cas des corpus bilingues pour le binôme « arabe-français ». L'introduction de notre corpus AFPC permettrait à l'utilisateur d'accéder gratuitement à l'intégralité du site et de découvrir les hyperliens dans les textes du corpus afin d'étendre sa connaissance terminologique contextualisée.

Notre projet, qui constitue un travail d'expérimentation, a donc une portée didactique; il prétend répondre aux besoins des étudiants en traduction quant aux homonymes, à la polysémie, aux nuances entre les termes, aux systèmes d'équivalents, d'autant plus que les systèmes linguistiques de l'arabe et du français sont largement différents.

Mais, avant d'étaler les particularités et les avantages d'un tel outil pédagogique, il s'avère indispensable de passer en revue les différentes étapes de la conception de ce corpus AFPC. La section suivante nous informera sur les étapes importantes et les détails techniques de ce projet qui préconise l'auto-apprentissage et la structuration des connaissances.

4. La conception du corpus parallèle français-arabe AFPC :

La revue de la littérature a démontré la grande nécessité d'un corpus parallèle arabo-français avec des matériaux de haute qualité source et cible et un alignement de haute qualité. Le but de ce projet est de créer une ressource utile pour l'enseignement des langues et la formation à la traduction. Par conséquent, le corpus doit tenir compte des facteurs suivants (Biber, 2003):

Taille: Il semble qu'il existe un débat sur la taille d'un corpus (voir Krishnamurthy, 2001; Leech, 1991). Cependant, plusieurs se sont convenus que la taille du corpus dépend de l'objectif auquel il est destiné et de plusieurs facteurs pratiques, tels que les droits d'auteur et la disponibilité.

Nombre de corpus précédemment réalisés sont de taille limitée, ce qui restreint leur efficacité. Par conséquent, la première phase de ce projet consiste à collecter 3 millions de phrases; ce nombre va augmenter dans les étapes ultérieures. La collecte de documents et de leurs traductions est faite par des dizaines de volontaires, essentiellement des étudiantes à l'Université Roi Saoud, et révisée par les auteures de cet article.

Représentativité: Selon Biber (1993), la représentativité désigne «la mesure dans laquelle un échantillon inclut toute la gamme de variabilité d'une population». Deux facteurs principaux influencent la représentativité d'un corpus: l'équilibre, qui se réfère à la variété des genres inclus dans un corpus; et l'échantillonnage, qui se réfère à la façon dont les morceaux de texte pour

BDD-A33052 © 2021 Editura Universităţii din Suceava
Provided by Diacronia.ro for IP 216.73.216.75 (2026-10-30 00:14:35 UTC)

chaque genre sont sélectionnés. La plupart des corpus arabe-français ont tendance à être restreints en terme de genres et de types de texte, ce qui affecte négativement leur utilité au niveau de l'apprentissage de la langue et en tant que ressources de formation à la traduction. Ils ont également tendance à être de nature statique, ce qui diminue leur représentativité. Le corpus proposé vise à couvrir un large éventail de domaines (juridique, médical, religieux, littéraire, économique…etc.) et de types de textes (rapport, compte-rendu, revue, livres, sites internet…etc.) avec des mises à jour régulières pour maintenir sa nature dynamique et ouverte.

Qualité: Comme ce corpus est destiné à être utilisé à des fins pédagogiques, la qualité doit être d'une importance capitale. Des échantillons de textes humains de haute qualité ont été compilés et alignés manuellement afin d'assurer une grande précision. Les auteures de cet article, aidées par une équipe de volontaires de l'Université Roi Saoud, ont mené un travail long et minutieux afin de vérifier que les sources retenues soient fiables, que les textes traduits soient publiés par des entités officielles et des traducteurs professionnels.

Disponibilité: La plupart des corpus parallèles arabophones décrits dans la section précédente ne sont pas disponibles gratuitement. De plus, nombre d'entre eux sont coûteux (par exemple des produits LDC) ou ont une interface non conviviale (par exemple OPUS), ce qui les rend difficiles à utiliser pour des non-experts. Par conséquent, notre projet vise à attirer un large éventail d'utilisateurs en fournissant un site Web gratuit, convivial, accessible en ligne et facile à utiliser comme outil de recherche.

Le projet s'est déroulé en deux phases: la première phase comprenait la collection des données, la compilation et l'alignement d'échantillons de texte de haute qualité traduits par l'homme provenant de divers types de sources. La deuxième phase a consisté à concevoir une interface web avec un concordant bilingue où les utilisateurs peuvent explorer le contenu de l'AFPC en français et en arabe. Les deux phases seront détaillées ci-dessous.

Étape 1

Le projet a commencé par la compilation, le nettoyage et l'alignement d'échantillons de texte de haute qualité traduits par l'homme et contenant différents types et genres de texte. Cette phase a consisté à préparer les fichiers et à segmenter, aligner et vérifier les textes.

Préparation des dossiers: les textes ont été recueillis d'une manière unidirectionnelle, c'est-à-dire en français (langue originale) avec leurs traductions officielles en arabe à partir de plusieurs sources: documents imprimés, tels que des livres, des magazines et des journaux et sites Internet.



BDD-A33052 © 2021 Editura Universităţii din Suceava
Provided by Diacronia.ro for IP 216.73.216.75 (2026-10-30 00:14:35 UTC)

Le logiciel OCR (ABBYY FineReader et Readiris 15) a été utilisé pour convertir des documents imprimés en textes lisibles par machine, car les deux programmes supportent l'arabe et le français. Les textes ont été classés selon les huit domaines suivants: social, biographique, littéraire, administratif, médical, juridique, religieux et scientifique. Les textes ont été segmentés par niveau de phrase, alignés manuellement, puis stockés sous forme de fichiers MS Excel. Chaque fichier contient les métadonnées suivantes dans les deux langues: Titre, Auteur, Éditeur, Année de publication, Pays, Sexe de l'auteur, Médium, Domaine et Sujet. Les métadonnées permettront aux utilisateurs du corpus de sélectionner les textes en fonction de leurs besoins spécifiques.

Segmentation: L'étape suivante dans la construction du corpus parallèle était la segmentation. Les textes ont été divisés en segments courts au niveau d'une phrase ou d'un paragraphe. Il existe certaines applications disponibles aujourd'hui qui peuvent effectuer ce processus automatiquement, comme un logiciel de mémoire de traduction (par exemple, WinAlign, Memsource). Ces applications sont considérées comme indépendantes du langage et peuvent gérer différents types de fichiers texte. Cependant, les résultats ne sont pas exempts d'erreurs. C'est parce que, dans la plupart du temps, il est difficile pour de telles applications de déterminer ce qu'est une phrase. Les signes de ponctuation, tels que les points, les points virgules, les points d'exclamation ou les points d'interrogation, sont couramment utilisés pour indiquer la fin d'une phrase, mais pourraient être difficiles à identifier. La période, par exemple, peut également indiquer une décimale ou une abréviation telle que Dr. ou Ms., et peut ne pas nécessairement signifier la fin d'une phrase. Par conséquent, il a été décidé d'effectuer le processus de segmentation manuellement afin d'assurer l'exactitude et la qualité des textes.

L'alignement: Quah (2006, p.100) a défini l'alignement comme «le processus de relier un segment de langue source à son segment de langue cible correspondant». Ce processus peut être exécuté à différents niveaux, c'est-à-dire au mot, à la phrase, au paragraphe, ou au niveau du texte. Cependant, la plupart des corpus parallèles alignent les textes au niveau de la phrase ou du paragraphe.

L'alignement des textes sources et des traductions était un processus difficile; les traducteurs ne traduisent pas nécessairement les textes d'une manière prévisible et linéaire. Frankenberg-Garcia et Santos (2003) ont remarqué que les traducteurs ont tendance à diviser les phrases source-texte en deux ou plusieurs phrases. Donc, plusieurs alternatives étaient possibles : joindre deux ou plusieurs phrases source-texte ensemble les rendant comme une seule phrase dans la traduction; laisser les choses tel quel; réorganiser les phrases de différentes façons, et/ou insérer des éléments qui n'étaient pas présents dans le texte source.

BDD-A33052 © 2021 Editura Universităţii din Suceava
Provided by Diacronia.ro for IP 216.73.216.75 (2026-10-30 00:14:35 UTC)

Pour ce projet, il a été décidé d'effectuer manuellement le processus d'alignement. L'alignement manuel consiste à parcourir le texte séquentiellement et à relier la première phrase du texte source avec la première phrase du texte cible, et ainsi de suite. L'alignement automatique gère les traductions terminées, divise les textes source et cible en segments et lie les segments correspondants. Il existe des outils d'alignement disponibles sur le marché qui permettent aux utilisateurs de vérifier l'alignement automatique et de modifier manuellement et de réorienter les segments qui ne correspondent pas ensemble. Malheureusement, plusieurs de ces outils ne peuvent pas prendre en charge des langues qui ne sont pas basées sur l'alphabet romain (comme l'arabe) et ne produisent donc pas de résultats précis lorsque la langue source et la langue cible ont des structures et des directions de texte différentes, comme l'arabe et le français.

Étape 2

La deuxième phase, comprend l'un des objectifs du projet et consiste à faciliter la consultation du contenu du site et permettre aux traducteurs et aux formateurs d'exploiter le contenu du corpus via une interface web conviviale. L'interface Web offre un concordancier bilingue, qui est un outil de recherche généralement utilisé pour étudier le contenu d'un corpus parallèle (Bowker, 2002). Un tel outil récupère toutes les occurrences d'un mot de recherche spécifique dans un contexte donné dans les langues source et cible. Les utilisateurs peuvent également affiner leur recherche en fonction du domaine, de l'année de publication, du pays, du sujet, du support ou du sexe de l'auteur.

Ce présent travail a donc pour objectif d'aboutir à une compréhension plus fine du texte source et d'optimiser l'aspect naturel du texte cible en fournissant un outil moderne qui incite à la réflexion, l'analyse, l'esprit critique et l'interprétation, pour une meilleure compréhension du processus traductif. Ceci car les différents sens d'un mot donnent souvent lieu à des différentes traductions dans la langue cible ; nous supposons que ce corpus-outil à l'enseignement de la traduction, pourra prendre en compte le caractère polysémique et de la langue arabe et du français.

Pour illustrer cette idée, nous allons analyser deux termes en utilisant deux entrées dans les deux langues : une recherche en arabe et une autre en français. Une recherche toute simple du terme « ***organisation*** » en français pourrait éclairer l'apprenti-traducteur sur les différentes nuances du mot puisqu'à plusieurs reprises le terme est traduit par un équivalent différent en arabe selon le contexte :

Dans la **figure 1**, le mot *organisation* est associé à l'Organisation des Nations Unies et à l'Organisation Mondiale de la Santé, c'est-à-dire, au sens premier du terme qui est une instance ou un organisme international ; ainsi, le terme équivalent en arabe est منظمة désigne un organe institutionnel. Mais, sur la **figure2**, le terme *organisation* est utilisé dans le sens de *association*, et par la suite l'équivalent arabe vient refléter cette nuance en utilisant le terme جمعية qui veut

BDD-A33052 © 2021 Editura Universităţii din Suceava
Provided by Diacronia.ro for IP 216.73.216.75 (2026-10-30 00:14:35 UTC)

dire *association*. La **figure 3** nous propose une troisième nuance du terme *organisation* en désignant par ce terme le fonctionnement de la Cour Internationale de justice, et ainsi l'équivalent arabe vient souligner cette nuance par l'usage du mot تنظيم. Quant à la **figure 4**, elle montre une quatrième nuance du terme organisation qui est utilisé dans le sens de communauté (organisation sociale), et dont l'équivalent arabe tient compte de ce sens en optant pour le terme مجتمع. Donc quatre termes différents en arabe ont été utilisés pour traduire le même mot en français ; seule la mise en contexte et la consultation des différents exemples appartenant à des domaines variés a permis de distinguer les différentes nuances du terme.

المصدر	فى الفرنسية	فى العربية
من كتاب التقرير السنوي للبنك الدولي لعام 2017 , المزيد من المعلومات	Le Mécanisme de financement mondial, qui est hébergé par le Groupe de la Banque mondiale, est le dispositif de financement et de mise en œuvre de l'initiative « Chaque femme, chaque enfant » de l'Organisation des Nations Unies	يعد صندوق التمويل العالمي، الذي يقع مقره بمجموعة البنك الدولي، الذراع التمويلية والإطار التنفيذي لحركة "كل امرأة وكل طفل" التي أطلقتها الأمم المتحدة
من كتاب التقرير السنوي للبنك الدولي لعام 2017 , المزيد من المعلومات	Depuis 2015, la Banque s'est engagée, avec l'Organisation mondiale de la Santé, à produire tous les deux ans un Rapport de suivi mondial consacré à la couverture maladie universelle	وبالاشتراك مع منظمة الصحة العالمية، التزم البنك بأن يصدر كل عامين منذ عام 2015 تقريرا للرصد العالمي يركّز على تحقيق التغطية الصحية الشاملة

Fig.1

من كتاب منظمات محلية في غواتيمالا ترحب بالفارين من عنف العصابات في أميركا الوسطى , المزيد من المعلومات	Aujourd'hui, Toribio est le coordonnateur d'un nouveau centre, ouvert le mois dernier, dans la ville d'El Ceibo, au Guatemala, par l' organisation Pastoral de Movilidad Humana avec l'aide du HCR, l'Agence des Nations Unies pour les réfugiés.	واليوم، يعمل توريبيو كمنسق لمركز جديد تديره الجمعية الرعوية الإنسانية لمساعدة الأشخاص المتنقلين في غواتيمالا والتي فتحت أبوابها الشهر الماضي في السيبو بدعم من المفوضية

Fig.2

BDD-A33052 © 2021 Editura Universităţii din Suceava
Provided by Diacronia.ro for IP 216.73.216.75 (2026-10-30 00:14:35 UTC)

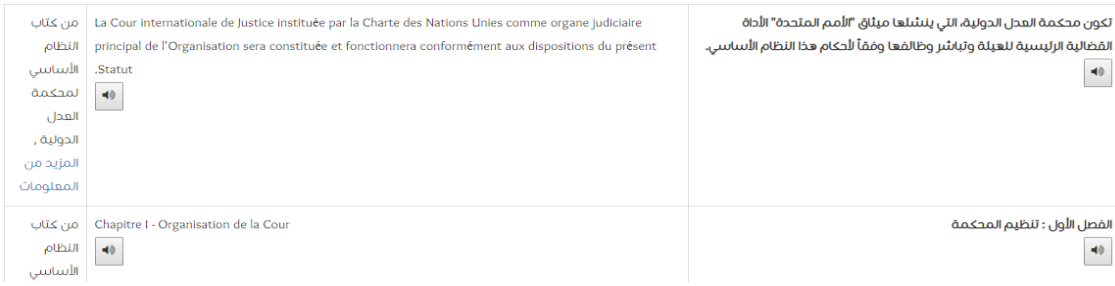


Fig.3

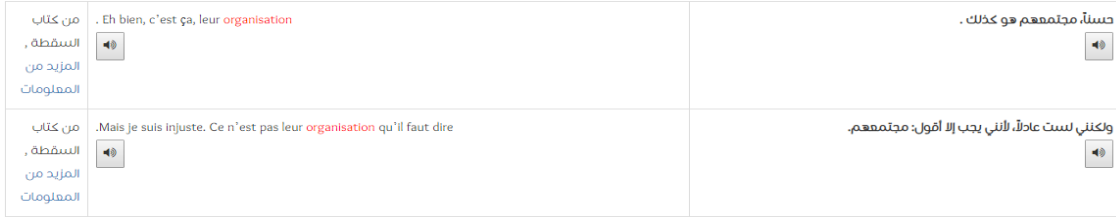


Fig.4

Le même type de recherche a été utilisé par mot-clé, en arabe cette fois; c'est le mot *kanoun* قانون. La recherche affiche 168 résultats pour ce mot-jusqu'à cette phase du projet ; l'utilisateur pourra ainsi consulter tous les segments qui incluent ce terme dans des contextes et des domaines variés. Parmi les traductions proposéespour ce terme en français, nous avons trouvé respectivement le terme *code* (fig.5 et 6, associé à "code militaire" et « code de procédure civile »), le terme *loi* (fig.7), le terme *droit* (fig.8).



Fig.5

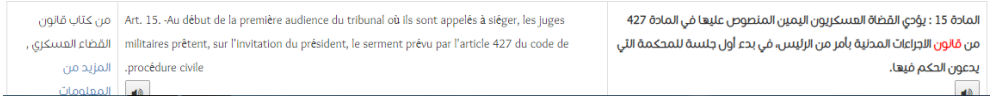


Fig.6

BDD-A33052 © 2021 Editura Universităţii din Suceava
Provided by Diacronia.ro for IP 216.73.216.75 (2026-10-30 00:14:35 UTC)



Fig.7



Fig.8

Ainsi, les relations sémantiques qui lient les différents termes à leur contexte et aux autres mots qui lui sont associés dépendent du contexte et du type de texte ainsi que des connaissances de l'utilisateur qui sont censées être modifiées au cours de sa recherche terminologique.

Cette approche fonctionnelle met en relief les contenus spécialisés, les visées discursives, la contextualisation du discours dans le système linguistique cible. Ainsi l'Apprentissage des techniques de gestion de corpus par les futurs traducteurs s'avère indispensable pour une gestion autonome de l'activité traductionnelle dans sa dimension pragmatique. Une fois le site finalisé, et afin de démontrer l'apport de cette ressource pédagogique, une étude de terrain sera mise en place pour tester l'impact de l'utilisation de ce corpus parallèle dans les classes de traduction pour mettre en relief les potentialités et les éventuelles faiblesses que pourrait représenter ce projet qui sera soumis périodiquement à des mises à jour adéquates.

Le défi majeur serait donc d'être capable d'assurer un enseignement de la traduction qui harmonise la traduction humaine, la traduction automatique et la traduction assistée par ordinateur pour aboutir à une « traduction augmentée » (Loock, 2019) qui dégage *la plus-value* de l'humain valorisant ses capacités créatives par rapport à la machine. Une meilleure performance de la pratique traductionnelle est l'objectif principal de ce travail qui favorise l'auto-apprentissage de la traduction en offrant un outil pédagogique facile à manipuler et présentant un large corpus de textes bilingues.

Ce projet AFPC est donc un corpus inédit et simple à manipuler pour un apprenti-traducteur ; il constitue une ressource appropriée pour surmonter les contraintes terminologiques, linguistiques, discursives et lexico-sémantiques du métier de la traduction.

Conclusion

Dans cet article, nous avons décrit un projet en cours à la Faculté de langues et de traduction de l'Université Roi Saoud afin de compiler un corpus parallèle français-arabe de 3 millions de mots qui servira de ressource pour l'apprentissage de la traduction. Le corpus peut être utilisé comme corpus comparable pour étudier les différences entre la langue traduite et la langue

BDD-A33052 © 2021 Editura Universităţii din Suceava
Provided by Diacronia.ro for IP 216.73.216.75 (2026-10-30 00:14:35 UTC)

source. Le corpus a été vérifié manuellement à différentes étapes, y compris les traductions, la segmentation de texte, l'alignement et la préparation des fichiers pour améliorer leurpertinence. Le corpus est disponible en format XML et à travers une interface web conviviale. L'interface Web fournit un concordancier qui prend en charge les requêtes de recherche bilingues et plusieurs options de filtrage. Ce corpus parallèle est destiné à favoriser les nouveaux modes d'ingénierie de la connaissance et faciliter l'accès à l'information afin de servir de ressource didactique dans l'enseignement de la traduction et les cours de formation des traducteurs.

Bibliographie

Abdul Hay A., O. Kraif. (2013) : *Constitution d'une ressource sémantique arabe à partir de corpus multilingue aligné. TALN-RECITAL* 17-21 juin 2013, Les Sables d'Olonne, 299-312.

Al-Ajmi, H. (2003) : « Compiling an English–Arabic parallel text corpus », M. Murata, S. Yamada, & Y. Tono (Eds.), *Proceedings of Asian Association for Lexicography*, 51–54.

Al-Ajmi, H. (2004): *A new English–Arabic parallel text corpus for lexicographic applications*. *Lexikos*, 14 (1).

Alkahtani, S., & Teahan, W. J. (2015, December): « A new parallel corpus of Arabic/English », *Proceedings of the 8th Saudi Students Conference in the UK*. World Scientific, 279.

Arrouart Catherine.(2003) : « Les mémoires de traduction et la formation universitaire : quelques pistes de réflexion », *Meta*, vol. 48, n° 3, Les Presses de l'Université de Montréal.

Aston, G. (1999) : *Corpus use and learning to translate*. *Textus*, 12 (2), 289–314.

Baker, M. (1999) : « The role of corpora in investigating the linguistic behaviour of professional translators », *International Journal of Corpus Linguistics*, 4 (2), 281–298.

Baker, M. (1998) : « Réexplorer la langue de la traduction : une approche par corpus », *Meta* 43(4), 480-485.

Baker, M. (1995): « Corpora in translation studies: An overview and some suggestions for future research », *Target*, 7 (2), 223–243.

Baker, M (1993): « Corpus linguistics and translation studies: Implications and applications », Mona Baker, Gill Francis et Elena Tognini-Bonelli, dir. *Text and Technology*. Amsterdam/Philadelphie: John Benjamins, 223-250.

Ballard, Michel, Pinera-Tresmontant Carmen, éds., (2008): *Les corpus en linguistique et en traductologie*, Arras, Artois Presse Université.

Bernardini, S. (2004): « Corpus-aided language pedagogy for translation education », K. Malmkjær (Ed.), *Translation in undergraduate degree programmes,* Amsterdam and Philadelphia: John Benjamins Publishing, 97-111.

BDD-A33052 © 2021 Editura Universităţii din Suceava
Provided by Diacronia.ro for IP 216.73.216.75 (2026-10-30 00:14:35 UTC)

Bernardini, S., & Castagnoli, S. (2008): « Corpora for translator education and translation practice », E. Yuste-Rodrigo (Ed.), *Topics in language resources for translation and localization,* Amsterdam and Philadelphia: Benjamins Publishing, 39–55.

Bernardini, S. (2016): « Discovery learning in the language-for-translation classroom: corpora as learning aids », *Cadernos de Tradução*, 36 (1), 14-35.

Boulton, A., & Tyne, H. (2013): « Corpus linguistics and data-driven learning: a critical overview », *Bulletin suisse de Linguistique appliquée*, 97, 97-118.

Botley, S., Glass, J., McEnery, A., & Wilson, A. (1996) : *Proceedings of TALC 96*. Lancaster, UK: UCREL Technical Papers, Volume 9.

Bowker, L., & Pearson, J. (2002): *Working with specialized language: A practical guide to using corpora*, London: Routledge.

Chujo, K., Kobayashi, Y., Mizumoto, A., & Oghigian, K. (2016): « Exploring the Effectiveness of Combined Web-based Corpus Tools for Beginner EFL DDL », *Linguistics and Literature Studies,* 4 (4), 262-274.

Cohn, T., & Lapata, M. (2007): « Machine translation by triangulation: Making effective use of multi-parallel corpora », *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*, 828-735.

Dehghan, A. & Darasawang, P. (2014): « Independent learning through the use of data driven learning », *Proceedings of the International Conference: DRAL 2. Place2: ILA 2014*.

Elkateb S., W.Black, H.Rodriguez, M.Alkhalifa, P. Vossen, A. Pease et C.Fellbaum (2006): « Building a WordNet for Arabic », *Proceedings of The Fifth International Conference on Language ressources and Evaluation* (LREC 2006).

Eisele, A., & Chen, Y. (2010): « MultiUN: A multilingual corpus from United Nations documents », *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC),* European Language Resources Association (ELRA), 2868-2872.

Flowerdew, L. (2015) : « Data-driven learning and language learning theories: Whither the twain shall meet », A. Leńko-Szymańska, & A. Boulton (Eds.), *Multiple affordances of language corpora for data-driven learning,* Amsterdam: John Benjamins, 15-36.

Flowerdew, L. (2015): « Corpus-based research and pedagogy », *EAP: From lexis to genre.* Language Teaching, 48(1), 99-116.

Frankenberg-Garcia, A. (2015): « Training translators to use corpora hands-on: Challenges and reactions by a group of thirteen students at a UK university », *Corpora*, 10 (3), 351–380.

Frankenberg-Garcia, A., Flowerdew, L., & Aston, G. (Eds.). (2011): *New trends in corpora and language learning*. London: Continuum.

Frérot, Cécile (2010) : « Outils d'aide à la traduction : pour une intégration des corpus et des outils d'analyse de corpus dans l'enseignement de la traduction et la formation des traducteurs », *Les Cahiers du GEPE*, Corpus

BDD-A33052 © 2021 Editura Universităţii din Suceava
Provided by Diacronia.ro for IP 216.73.216.75 (2026-10-30 00:14:35 UTC)

et mémoires de traduction, Strasbourg: Presses universitaires de Strasbourg, URL: http://www.cahiersdugepe.fr/index.php?id=1164.
Frérot, Cécile (2016): « Corpora and corpus technology for translation purposes in professional and academic environments. Major achievements and new perspectives. » , *Cadernos de Tradução*, 36 (1), 36-61.
Gouadec, D. (2002): « Training translators: Certainties, uncertainties, dilemmas. », *Training the language services provider for the new millennium,* Oporto: University of Porto, 31-41.
Granger, S., & Lefer, M. A. (2016): « From general to learners' bilingual dictionaries: Towards a more effective fulfillment of advanced learners' phraseological needs. », *International Journal of Lexicography*, 1(10), 1-17.
Guzman, F., Sajjad, H., Abdelali, A., & Vogel, S. (2013): « The AMARA corpus: Building resources for translating the web's educational content. », *Proceedings of the International Workshop on Spoken Language Translation, IWSLT 2013*. Heidelberg, Germany.
Haiyan, L., & Zhongxin, D. (2014): « Effectiveness of self-built Chinese–English corpus on assisting translation teaching », *International Journal of Humanities and Social Science*, 4 (7), 96–98.
Hassan, S., & Atwell, E. S. (2016): « Design and implementing of multilingual Hadith corpus », *International Journal of Recent Research in Social Sciences and Humanities*, 3 (2), 100–104.
Hu, K. (2016): « Corpus-Based Study of Translation Teaching », *Introducing Corpus-based Translation Studies*, Springer Berlin Heidelberg, 177-191.
Johansson, S. (2007): *Seeing through multilingual corpora: On the use of corpora in contrastive studies*. Amsterdam and Philadelphia: John Benjamins.
Johns, T. (2002): « Data-driven learning: The perpetual challenge », B. Kettermann, & G. Marko, (Eds.), *Teaching and learning by doing corpus linguistics,* Amsterdam: Rodopi, 107-117.
Karras, J. N. (2016): « The effects of data-driven learning upon vocabulary acquisition for secondary international school students in Vietnam », *ReCALL*, 28 (02), 166-186.
Kilgarriff, A. (2013): « Terminology finding, parallel corpora and bilingual word sketches in the Sketch Engine », *Proceedings of ASLIB 35th Translating and the Computer Conference*, London.
Kübler, N. (2011): *Working with different corpora in translation teaching. New Trends in Corpora and Language Learning*. London: Continuum, 62-80.
Kübler, N. (2011): *Corpora, Language, Teaching, and Ressources: from Theory to Practice*, Bern, Peter Lang.
Kübler, N. (2013) : « Traduction pragmatique, linguisique de corpus, traducteur: un ménage à trois explosif ? », communication au colloque *Tralogy II : Anticiper les technologies pour la traduction*, 17 et 18 janvier 2013, Paris, vidéo en ligne : http://webcast.in2p3.fr/videos-traduction_pragmatique.

BDD-A33052 © 2021 Editura Universităţii din Suceava
Provided by Diacronia.ro for IP 216.73.216.75 (2026-10-30 00:14:35 UTC)

Laviosa, S. (1998): « L'approche basée sur le Corpus / The Corpus-based Approach », *Meta* 43(4).

Lee, J. H., Lee, H., & Sert, C. (2015): « A corpus approach for autonomous teachers and learners: Implementing an on-line concordancer on teachers' laptops. », *Language Learning & Technology*, 19 (2), 1–15.

Leech, G. N. (2010): « Corpus linguistics », K. Malmkjaer (Ed.), *The Routledge linguistics encyclopedia* (3rd ed.), London: Routledge, 103-113.

Lin, M. H. and Lee, J.-Y. (2015): « Data-driven learning: Changing the teaching of grammar in EFL classes », *ELT Journal*, 69 (3): 264–274.

Loock, Rudy (2018) : « Traduction automatique et usage linguistique: une analyse de traductions anglais-français réunies en corpus », *Meta*, 63(3), 786-806 ; https://doi.org/10.7202/1060173ar.

Loock, Rudy (2016) : *La traductologie de corpus*. Villeneuve d'Ascq : Presses universitaires du Septentrion.

Marco, J., & van Lawick, H. (2015): « Enhancing translator trainees' awareness of source text interference through use of comparable corpora », *Multiple Affordances of Language Corpora for Data-driven Learning*, 69, 225.

Molés-Cases, T. & Oster, U. (2015): « Webquests in translator training: Introducing corpus-based tasks », A. Leńko-Szymańska & A. Boulton (Eds.), *Multiple affordances of language corpora for data-driven learning,* Amsterdam: John Benjamins, 201–224.

Olohan, M. (2004): *Introducing corpora in translation studies*. London: Routledge.

Pearson, J. (2003*):* « Using parallel texts in the translator training environment », F. Zanettin, S. Bernardini, & D. Stewart (eds), *Corpora in Translator Education*, Manchester: St. Jerome, 15–24.

Perez, M. M., Paulussen, H., Macken, L., & Desmet, P. (2014): « From input to output: The potential of parallel corpora for CALL. », *Language Resources and Evaluation*, 48 (1), 165–189.

Quan, Z. (2016): « Introducing "mobile DDL (data-driven learning)" for vocabulary learning: an experiment for academic English. » *Journal of Computers in Education*, 3 (3), 273-287.

Rodríguez-Inés, P. (2014): « Using corpora for awareness-raising purposes in translation, especially into a foreign language (Spanish–English). », *Perspectives*, 22 (2), 222–241.

Salhi, H. (2013): « Investigating the complementary polysemy and the Arabic translations of the noun Destruction in EAPCOUNT. », *Meta: Translators' Journal*, 58 (1), 227–246.

Sauron, Véronique (2007) : « Les nouvelles technologies dans l'enseignement de la traduction : l'exemple de la traduction juridique », Lavault Elisabeth. (éd.), *Traduction spécialisée : pratiques, théories, formations*, Peter Lang, Berne, 207-224.

Sellami, R., F. Sadat, L. Hadrich Belguith (2013) : « Traduction Automatique. Statistique à partir de corpus comparables: Application au couple de langues arabe-français. », *CORIA,* Avril 2013, 109-118.

BDD-A33052 © 2021 Editura Universităţii din Suceava
Provided by Diacronia.ro for IP 216.73.216.75 (2026-10-30 00:14:35 UTC)

Sinclair, J. (2005): « Collocation Studies », in Ramesh Kishnamurthy (ed.), *English Collocation Studies: The OSTI.*

Sinclair, J. (1991) : *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Sinclair, J. (1995): « Corpus typology: A framework for classification. », *Stockholm Studies in English*, 85, 17–33.

Sinclair, J. M. (Ed.) (2004): *How to use corpora in language teaching.* Amsterdam: John Benjamins Publishing.

Singer, N. (2016): « A Proposal for Language Teaching in Translator Training Programmes Using Data-Driven Learning in a Task-Based Approach. », *International Journal of English Language & Translation Studies,* 4 (2), 155-167.

Teubert, W. (2002): « The role of parallel corpora in translation and multilingual lexicography », B. Altenberg, & S. Granger (Eds.), *Lexis in contrast*, Amsterdam: John Benjamins, 189-214.

Verplaetse, H., & Heylen, K. (2015): « Parallel corpora for medical translation training: An analysis of impact on student performance. », *Proceedings of the 4th International Conference on Corpus Use and Learning to Translate (CULT)*, Alicante, Spain.

Wang, X., Gao, Y., & Hao, T. (2016): « The Construction of a Customized Medical Corpus for Assisting Chinese Clinicians in English Research Article Writing », *China National Conference on Chinese Computational Linguistics,* Berlin: Springer, 241-252.

Zanettin, Frederico et al. F. (eds). (2003): *Corpora in Translator Education*, Manchester, St.Jerome.

BDD-A33052 © 2021 Editura Universităţii din Suceava
Provided by Diacronia.ro for IP 216.73.216.75 (2026-10-30 00:14:35 UTC)