



**HAL**  
open science

# Quedar : un verbe polysémique par excellence face à la traduction automatique

José Cornejo Cárcamo

► **To cite this version:**

José Cornejo Cárcamo. Quedar : un verbe polysémique par excellence face à la traduction automatique. Sciences de l'Homme et Société. 2023. dumas-04262522

**HAL Id: dumas-04262522**

**<https://dumas.ccsd.cnrs.fr/dumas-04262522>**

Submitted on 27 Oct 2023

**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.



# ***Quedar* : un verbe polysémique par excellence face à la traduction automatique**

**José  
CORNEJO CÁRCAMO**

**Sous la direction de**  
Caroline ROSSI  
Lucía GÓMEZ VICENTE

**Membres du jury :**  
Agnès TUTIN  
Olivier KRAIF

Laboratoire de linguistique et didactique des langues étrangères et maternelles  
(LIDILEM)

UFR LLASIC  
Département Sciences du langage  
Section Linguistique

Mémoire de master 2 mention Sciences du langage - 20 crédits

Parcours : Linguistique

Année universitaire 2022-2023





# ***Quedar* : un verbe polysémique par excellence face à la traduction automatique**

**José  
CORNEJO CÁRCAMO**

**Sous la direction de**  
Caroline ROSSI  
Lucía GÓMEZ VICENTE

**Membres du jury :**  
Agnès TUTIN  
Olivier KRAIF

Laboratoire de linguistique et didactique des langues étrangères et maternelles  
(LIDILEM)

UFR LLASIC  
Département Sciences du langage  
Section Linguistique

Mémoire de master 2 mention Sciences du langage - 20 crédits

Parcours : Linguistique

Année universitaire 2022-2023



### **DÉCLARATION ANTIPLAGIAT**

1. Ce travail est le fruit d'un travail personnel et constitue un document original.
2. Je sais que prétendre être l'auteur d'un travail écrit par une autre personne est une pratique sévèrement sanctionnée par la loi.
3. Personne d'autre que moi n'a le droit de faire valoir ce travail, en totalité ou en partie, comme le sien.
4. Les propos repris mot à mot à d'autres auteurs figurent entre guillemets (citations).
5. Les écrits sur lesquels je m'appuie dans ce mémoire sont systématiquement référencés selon un système de renvoi bibliographique clair et précis.

PRÉNOM : José

NOM : CORNEJO CÁRCAMO

DATE : 01/08/2023

# Sommaire

<b>Introduction</b> .....	<b>5</b>
<b>PARTIE 1 : ÉTAT DE L'ART</b> .....	<b>7</b>
<b>Chapitre 1 – <i>Quedar</i> : un verbe polysémique par excellence</b> .....	<b>8</b>
1. Qu'est-ce que la polysémie ? .....	8
La polysémie verbale .....	8
2. <i>Quedar</i> : entre permanence et changement .....	9
3. <i>Quedar</i> : un verbe d'état à caractère résultatif à la suite d'une dynamique des forces .....	12
<b>Chapitre 2. La traduction automatique neuronale : un outil robuste, mais pas infallible</b> .....	<b>15</b>
1. La traduction automatique (TA) de nos jours : le paradigme neuronal .....	15
2. La TAN au quotidien .....	18
3. Comment aider les utilisateurs non spécialistes à utiliser les outils de TA de manière raisonnée ? .....	22
<b>PARTIE 2 - METHODOLOGIE</b> .....	<b>24</b>
<b>Chapitre 3. Le projet LOLYTA</b> .....	<b>25</b>
1. Description du projet LOLYTA .....	25
2. Objectif du stage et tâches effectuées .....	26
3. Présentation du corpus et difficultés rencontrées .....	27
3.2.1. Première difficulté : origine des sous-titres .....	29
3.2.2. Deuxième difficulté : droits d'auteurs et notre manière de récupérer les sous-titres .....	29
3.2.3. Troisième difficulté : alignement et nettoyage du corpus .....	31
<b>Conclusion</b> .....	<b>32</b>
<b>PARTIE 3 - RESULTATS ET DISCUSSION</b> .....	<b>33</b>
<b>Chapitre 4. Présentation des résultats du projet LOLYTA</b> .....	<b>34</b>
1. L'exploitation du corpus de sous-titres .....	34
2. Les occurrences de <i>quedar</i> dans notre corpus .....	35
3. Les équivalents de <i>quedar</i> : traductions humaines vs sorties de TA .....	40
3.1.1. Les équivalents de <i>quedar</i> dans les traductions humaines du sous-corpus ES-es-FR .....	41
3.1.2. Les équivalents de <i>quedar</i> dans les traductions humaines du sous-corpus ES-ha-FR .....	42
3.2.1. Les équivalents de <i>quedar</i> dans les sorties de TA pour le sous-corpus ES-es-FR .....	43
3.2.2. Les équivalents de <i>quedar</i> dans les sorties de TA pour le sous-corpus ES-ha-FR .....	44
<b>Conclusion</b> .....	<b>45</b>
<b>Chapitre 5. Analyse des données et réflexion personnelle sur les limites de l'étude</b> .....	<b>47</b>
1. Les différents domaines de <i>quedar</i> présents dans notre corpus .....	47
2. <i>Quedar</i> et DeepL : des traductions quasi parfaites ? .....	64
3. Les limites de notre étude : des pistes de réflexion ? .....	72
<b>Conclusion</b> .....	<b>74</b>
<b>Conclusion</b> .....	<b>75</b>
<b>Bibliographie</b> .....	<b>77</b>
<b>Annexes</b> .....	<b>83</b>
<b>Annexe 1 : Occurrences de <i>quedar</i> (ES-ha-FR)</b> .....	<b>83</b>
<b>Annexe 2 : Occurrences de <i>quedar</i> (ES-es-FR)</b> .....	<b>87</b>
<b>Annexe 3 : Les avis des étudiants de L3 concernant notre intervention sur le verbe <i>quedar</i></b> .....	<b>90</b>
<b>Sigles et abréviations utilisés</b> .....	<b>91</b>
<b>Table des illustrations</b> .....	<b>92</b>
<b>Table des matières</b> .....	<b>93</b>

## Introduction

De nos jours, les étudiants et apprenants de langues étrangères font appel aux moteurs de traduction automatique (TA) au quotidien. Ils y effectuent diverses tâches sur des moteurs tels que DeepL ou Google Translate. En effet, ces moteurs ne sont plus de simples outils pour traduire, mais ils deviennent de plus en plus des outils d'aide à la rédaction en langue étrangère. Cette évolution concernant l'utilisation des systèmes de TA pourrait s'expliquer par la facilité d'usage des interfaces, mais aussi par le fait qu'ils sont librement accessibles en ligne et gratuits.

D'après l'étude effectuée par Nassau, G., Molle, N. et Kalyaniwala, C. en 2022, les étudiants traduisent des mots ou des phrases hors contexte, ou bien des paragraphes, voire des textes, en entier. Par ailleurs, les participants dans l'étude déclarent différents objectifs qui les amènent à utiliser la traduction automatique : le contournement d'une incapacité, le pragmatisme (gain du temps lors des tâches de traduction ou de rédaction), le besoin de vérification et l'enrichissement du répertoire langagier (l'apprentissage d'une langue étrangère).

Au vu de cette utilisation massive par le grand public, ici des apprenants qui ne maîtrisent pas la langue étrangère qui les intéresse, nous constatons un problème majeur. En effet, les systèmes de traduction automatique sont utiles et de plus en plus performants. Toutefois, ces moteurs de TA peuvent induire en erreur ces utilisateurs, en produisant des textes trompeurs aux niveaux grammatical et sémantique.

La traduction automatique n'est pas sans faille et le sens n'est pas toujours transmis correctement, particulièrement si un mot (ou groupe de mots) peut être traduit de différentes manières : c'est le cas de verbes polysémiques espagnols comme *quedar*. Il s'agit d'un verbe très fréquent en espagnol (Almela Pérez et al., 2005), mais qui pose des problèmes pour les apprenants dus à ses traits polysémiques (voir Gómez Vicente, 2018). Afin de résoudre ce problème, les systèmes doivent être capables d'analyser le contexte intra- ou inter-phrastique. Autrement dit, les systèmes de TA doivent identifier le contexte dans lequel le verbe se trouve pour choisir l'acception correcte, même s'il s'agit d'une tâche difficile.

Ce mémoire a deux objectifs : d'une part, nous voulons étudier la polysémie du verbe *quedar* à partir d'une étude sur corpus, en relevant les structures grammaticales qui pourraient permettre de choisir la bonne acception du verbe ; d'autre part, nous aimerions vérifier si les moteurs de TA utilisés par le grand public, comme DeepL, proposent des



sorties où le verbe est traduit correctement selon le contexte. Pour ce faire nous effectuerons une étude de cas en traduisant les phrases hors contexte (phrases isolées) et en contexte (en remplaçant la phrase dans son environnement), tout en comparant les résultats avec des traductions de référence (traductions humaines).

Les langues choisies (l'espagnol et le français) sont deux langues romanes, typologiquement similaires, mais non isomorphes pour les tournures en question, posant ainsi des difficultés pour les apprenants, mais aussi pour les machines. Ces dernières peuvent être susceptibles de s'appuyer sur une traduction « par défaut » (Halverson, 2019) au moins dans un premier temps. Autrement dit, les moteurs pourraient produire des traductions littérales, en ne sélectionnant pas le sens adéquat du verbe *quedar* par rapport au contexte en question. De plus, les langues étudiées sont des langues qui, si elles sont très bien documentées, posent encore des problèmes de traduction importants et ne donnent donc pas lieu à des résultats satisfaisants en traduction automatique neuronale (voir par ex. Ledesma & Lencinas, 2018)

Notre mémoire s'organise en cinq chapitres. Les deux premiers concernent les apports théoriques : nous réfléchissons sur la polysémie du verbe *quedar* ; ensuite, nous présenterons la traduction automatique neuronale et les difficultés que nous pouvons rencontrer. Le chapitre suivant (chapitre 3) concernera la méthodologie employée lors du stage de recherche que nous avons effectué pour mener à bien ce projet : nous décrirons le corpus du projet, la manière dont nous l'avons construit et exploité. Ensuite, nous procéderons à la présentation des résultats concernant les différents sens de *quedar* rencontrés dans notre corpus, ainsi qu'aux sorties de TA proposées par DeepL (chapitre 4). Par la suite, nous analyserons les données afin d'identifier les différents sens de *quedar* dans notre corpus et essayerons de comparer les traductions de référence avec les sorties de TA (chapitre 5).

## **Partie 1 : État de l'art**

## Chapitre 1 – *Quedar* : un verbe polysémique par excellence

Dans ce premier chapitre, notre objectif est d'introduire la notion de polysémie, et plus particulièrement celle de polysémie verbale, ainsi que faire une revue de la littérature concernant le verbe *quedar*.

### 1. *Qu'est-ce que la polysémie ?*

La polysémie est un phénomène linguistique qui a été introduit par Bréal au 19<sup>e</sup> siècle (d'après le travail de Victorri et Fuchs, 1996). En effet, Bréal présente ce phénomène en expliquant qu'une même forme (un terme) peut avoir différents sens : sens propre, sens métaphorique, sens restreint, sens étendu, sens abstrait, sens concret, notamment. Ces sens coexistent les uns avec les autres et peuvent se développer au fur et à mesure.

Plusieurs linguistes, comme Victorri et Fuchs (1996), précisent qu'il ne faut pas confondre la polysémie et l'homonymie. Ce dernier caractérise un phénomène linguistique où des mots différents partagent une même forme graphique, mais les sens sont radicalement différents. Pour illustrer cela, Redmond (2017) donne l'exemple du mot *avocat* : il fait référence à un fruit et à une profession.

Ce qui caractérise les mots polysémiques, c'est qu'ils se trouvent dans un même espace sémantique (Victorri et Fuchs, 1996) et le sens qui sera activé dépendra des interactions que ce mot aura avec d'autres mots ou groupes de mots présents dans un énoncé. Cette interaction détermine donc le sens qu'un mot polysémique prend au sein d'une phrase.

D'après les travaux de Victorri et Fuchs (1996), Kleiber (2005) et Redmond (2017), plusieurs catégories grammaticales peuvent être concernées par la polysémie. Dans ce travail, la polysémie verbale nous intéresse en particulier.

#### *La polysémie verbale*

D'après Jacquet (2003), les constructions verbales sont essentielles pour comprendre l'acception d'un verbe. Autrement dit, l'interaction du verbe avec les autres éléments sera déterminante pour connaître le sens que ce verbe a dans cet énoncé. Le verbe ne peut donc pas être isolé des autres éléments : ces derniers sont essentiels pour déterminer le sens. Autrement dit, il ne faut pas isoler les éléments d'un énoncé, mais identifier les constructions grammaticales pertinentes permettant d'analyser les ensembles pour accéder au sens véhiculé d'après Victorri et Fuchs (1996) et Vaguer (2004).

Dans ce travail, nous avons décidé d'étudier la polysémie du verbe *quedar*. Celui-ci est traduit, par des dictionnaires bilingues comme Linguee ou des concordanciers bilingues comme Tradooit, par le verbe *rester* en français. Cela dit, nous verrons tout au long de notre étude, ainsi qu'à travers de travaux de recherche comme ceux menés par Gómez Vicente (2011, 2013, 2018), Santos (2015) et González Martínez et Maldonado Soto (2017) que le verbe *quedar* est hautement polysémique. En effet, les constructions accompagnant ce verbe sont importantes pour identifier l'acception qui est employée au sein d'un énoncé.

## **2. *Quedar* : entre permanence et changement**

Les linguistes s'intéressent au verbe *quedar* en raison de sa richesse sémantique. En effet, selon le contexte d'usage et la construction syntaxique accompagnant le verbe, il peut véhiculer deux notions principales : celle de permanence et celle de changement. Cependant, il n'est pas facile de déterminer précisément le sens du verbe en raison des « domaines » (Gómez Vicente, 2013) dont il relève : à savoir l'existence, la localisation, la possession ou l'état, mais aussi le rendez-vous et le point de vue (Gómez Vicente, 2018). Ces domaines-ci sont le résultat de l'évolution de *quedar* sur le plan sémantique, notamment grâce à des extensions sémantiques qui ont eu lieu à différentes époques.

### **2.1. *Quedar* : un verbe qui évolue à travers les époques**

*Quedar* est un verbe ancien qui a été identifié pour la première fois vers le III<sup>e</sup> siècle (Santos, 2015 : 10). D'après Gómez Vicente (2013) et Santos (2015), il viendrait du latin classique *quietare*, qui signifiait se reposer, ne pas bouger ou rester en silence. Dans les siècles qui suivent, ce verbe connaît une évolution lente, mais qui lui permet d'élargir les contextes dans lesquels il est employé (Santos, 2015). Autrement dit, d'autres usages viennent s'ajouter aux usages de base : au V<sup>e</sup> siècle, on l'utilise pour exprimer *arrêter de [faire quelque chose]* ; au VI<sup>e</sup> siècle, le verbe exprime *rester* ou *rester sans bouger* ; au IX<sup>e</sup> siècle, on ajoute les notions *être d'accord en silence* ou *se laisser reposer*.

Cependant, le sens prototypique du verbe voit le jour entre le XII<sup>e</sup> et le XVI<sup>e</sup> siècles (Gómez Vicente, 2013 : 3 ; Santos, 2015 : 11) : il s'agit du sens exprimant la permanence. En effet, d'après Gómez Vicente (2013 : 3), *quedar* est le seul survivant « d'un ensemble de termes exprimant la permanence en latin ». Ce verbe aurait donc bénéficié de l'héritage sémantique des autres termes qui ont disparu.

Par ailleurs, Santos (2015 : 16) précise que cet héritage sémantique s'est manifesté à travers différentes expansions sémantiques. En effet, *quedar* a dû traverser un long processus qui aurait permis l'éloignement des usages de base et aurait provoqué donc la perte de l'association immédiate avec la fonction référentielle évoquée par le sens plus prototypique. La linguiste met donc en avant le fait que le verbe *quedar* peut être vu comme « un réseau de constructions dans lequel nous pouvons trouver un continuum sens allant du plus prototypique à celui qui semblerait n'avoir aucun lien (sur le plan synchronique) avec le sens le plus répandu, c'est-à-dire la permanence. Nous traiterons ce sens ci-après.

## ***2.2. La permanence : l'acception prototypique la plus répandue***

Comme nous l'avons vu précédemment, l'acception la plus répandue est celle de permanence. Ce sens peut être étudié à travers différents « domaines » qui ont été identifiés par Gómez Vicente (2013) : l'existence, la localisation et la possession, par exemple.

Le sens lié à l'existence serait apparu au XII<sup>e</sup> siècle (Gómez Vicente, 2013 : 3) : *quedar* voulait donc dire « rester une partie de quelque chose ». Par exemple, dans *Quedan tres galletas* (Il reste trois gâteaux), la linguiste a souligné la permanence du sujet, dans le sens d'existence, après un changement provoqué par un élément extralinguistique (peut-être qu'avant il y avait d'autres gâteaux). Cependant, ce qui est désigné et focalisé est la permanence des objets ou « entités » (Gómez Vicente, 2013 : 4) qui demeurent après le changement.

En ce qui concerne le sens lié à la localisation, apparu au XIII<sup>e</sup> siècle (Gómez Vicente, 2013 : 5), *quedar* signifiait « être ou s'arrêter (de force ou volontairement) quelque part, avec l'intention d'y rester ou de repartir » (*ibid*). Dans *Juana se quedó en casa* (Juana est restée à la maison), Gómez identifie la présence d'un complément spatial après le verbe qui situe le sujet dans une localisation précise. L'accent est donc mis sur la localisation où l'entité est restée.

Enfin, le dernier domaine qui relève de la permanence selon Gómez Vicente est la possession. D'après elle, ce sens est apparu aux alentours du XIV<sup>e</sup> siècle et faisait référence à avoir un objet dans sa possession (Gómez Vicente, 2013 : 7). Elle a pu identifier que dans cet usage, il y a obligatoirement un complément après le verbe qui « désigne l'entité que le sujet possède » (*ibid*). Par exemple : *Martín se quedó con el anillo* (Martin a gardé la bague).

Dans une étude menée plus tard, Gómez Vicente (2018) identifie le domaine du rendez-vous. Celui-ci véhicule le sens prototypique de permanence. En effet, cet usage

permet de fixer un rendez-vous dans un cadre spatio-temporel défini. Le verbe est donc accompagné par des compléments indiquant le lieu où le temps (heure, date). Elle propose l'exemple suivant pour illustrer cela : *quedamos mañana a las 9* (On se voit demain à 9 h). Le rendez-vous est donc fixé (permanence) à un moment précis.

Compte tenu de ces éléments, il faut préciser qu'il y a une deuxième notion véhiculée par *quedar*, mais qui est moins fréquente d'après les articles que nous avons consultés.

### ***2.3. Le changement : une dimension moins fréquente, mais essentielle***

Cette dimension permet de classer *quedar* comme étant un verbe de changement d'état (Gómez Vicente, 2013 et 2018 ; Santos, 2015 ; González Martínez & Maldonado Soto, 2017). Santos (2015 : 20) souligne, par ailleurs, que l'état peut être transitoire, dû à un changement ou le résultat dans une suite d'événements.

Selon González Martínez et Maldonado Soto (2017), la notion de changement d'état véhiculée par *quedar* doit être comprise comme un événement où le sujet subit un changement soit physique, soit abstrait. Par exemple, *Juana se ha quedado dormida* (Juana s'est endormie). Ici, le sujet change d'état physique : il passe d'un état éveillé à un état de sommeil. D'après Gómez Vicente, la notion de changement met en relief le nouvel état qui commence (2013 : 9). Cet usage s'est démocratisé à travers l'extension sémantique du sens prototypique. Il s'agit, en effet, de « la permanence du sujet dans un état après l'intervention d'une dynamique des forces » (*ibid*). Autrement dit, *quedar* continue d'exprimer la notion de permanence, mais après avoir subi un changement (provoqué par des facteurs extralinguistiques).

Par ailleurs, *quedar* peut aussi exprimer un changement de localisation. Cette idée est présentée par González Martínez et Maldonado Soto (2017). En effet, cette dimension a lieu quand le verbe est combiné avec d'autres verbes ou compléments de lieu. Par exemple, *Matías tiró la argollá y quedó dentro de la botella* (Matias a lancé l'anneau et il est resté à l'intérieur de la bouteille). Dans ce cas, l'anneau subit un changement de localisation : il était dehors et après il s'est retrouvé dans la bouteille. D'après les deux chercheurs, il s'agit d'un schéma résultatif où on se focalise sur le changement qui a déjà eu lieu.

Le sens, ou plutôt les sens du verbe *quedar* sont difficiles à définir. En effet, il s'agit d'un verbe qui a dû passer par différents processus d'extension sémantique. Au fur et à mesure des étapes, il y a eu d'autres usages qui sont apparus, laissant ainsi la place à de nouvelles acceptions liées aux constructions avec *quedar*. Ces acceptions et usages se sont

ajoutés à la notion de permanence de base (acception prototypique) et ont provoqué le développement des sens en lien avec la notion de changement.

### **3. *Quedar* : un verbe d'état à caractère résultatif à la suite d'une dynamique des forces**

Gómez Vicente introduit la notion de dynamique des forces dans ses articles de 2013 et 2018 pour analyser le caractère résultatif véhiculé par *quedar* dans tous ses domaines d'usage.

#### **3.1. Le concept de dynamique des forces (Talmy, 1988) appliqué à *quedar***

La dynamique des forces (*force dynamics*) est une notion venant des études en psychologie cognitive et appliquée aux études linguistiques par Talmy (1988).

En effet, d'après son article de 1988, il s'agit d'une catégorie sémantique, comme le mode ou l'aspect, qui permet d'expliquer le rapport que des entités entretiennent par rapport à une force. Cette force peut être exprimée de manière explicite dans l'énoncé ou, ce qui arrive la plupart du temps, il s'agit d'une force qui n'est pas mentionnée dans l'énoncé.

Par exemple, dans l'énoncé *The ball kept rolling against the stiff grass* (La balle continuait de rouler sur l'herbe raide) nous observons clairement la force qui s'exprime à travers le roulement de la balle. Cependant, il y a aussi une contre-force représentée par la pelouse. Talmy nomme l'entité qui exerce la force, l'agoniste, et l'entité qui s'oppose, l'antagoniste.

Cependant, comme nous l'avons mentionné, la force n'est pas forcément explicite, elle peut être due à une situation extérieure qui n'est pas mentionnée, mais qui exerce une influence sur une entité. Gómez Vicente (2018) propose l'exemple suivant : *Juan forced himself to answer the question*. Dans cet exemple, il y aurait une contre-force qui oblige Juan à répondre à la question. Cette contre-force est décrite par la chercheuse comme ce qui est attendu qu'il fasse (des attentes extérieures ou dues aux normes sociales).

Tous les domaines concernant le verbe *quedar* et ses constructions sont en lien avec des forces qui interagissent entre elles et qui provoquent un résultat exprimé par *quedar*. Autrement dit, *quedar* se situe à la suite d'une interaction des forces et véhicule le résultat provoqué par celles-ci. D'après le travail de Gómez Vicente (2013), il y aurait deux phases qui mènent à l'usage d'une construction avec *quedar* :

- La première phase se divise en deux sous-phases :

- La première sous-phase présente une situation avant qu'une dynamique des forces intervienne.
- La deuxième sous-phase se déroule lors qu'une interaction de forces a lieu : elles exercent une influence sur une entité.
- La deuxième phase se situe après la dynamique des forces et présente donc le résultat de cette interaction.

Pour illustrer ceci, prenons l'exemple du domaine de l'existence et de la phrase *Quedan tres galletas* (Il reste trois gâteaux). Cet exemple constitue la seconde phase, c'est-à-dire le résultat exprimé à l'aide de *quedar*. Étant donné que nous n'avons pas un contexte défini, nous pourrions en déduire les deux sous-phases précédent ce résultat comme suit :

- Première sous-phase : il existe un nombre  $x$  de gâteaux ;
- Seconde sous-phase : une dynamique de forces a lieu qui exerce une influence sur le nombre de gâteaux (par exemple, il y en avait dix gâteaux au départ, mais quelqu'un en a mangé sept).

Ici, l'agoniste serait le nombre de gâteaux qui existait au départ et l'antagoniste serait la personne qui mange les gâteaux. Le nombre de gâteaux restants (le résultat) est exprimé par *quedar*. Ainsi, le caractère résultatif devient évident.

### ***3.2. L'usage de quedar par rapport à d'autres verbes***

*Quedar* véhicule des sens qui ne lui sont pas exclusifs. Ceci devient évident en regardant un dictionnaire comme le *Larousse*<sup>1</sup> en ligne où nous trouvons les sens ou les usages des verbes entre crochets précédant les définitions. Dans le dictionnaire, les définitions de *quedar* sont liées aux sens qu'ils véhiculent, qui sont souvent présentés à l'aide d'autres verbes. Parmi ces verbes, nous avons repéré *estar* (être), *haber* (avoir) et *tener* (avoir ou posséder).

Il serait donc possible de remplacer *quedar* par l'un de ces verbes au sein d'un énoncé véhiculant l'un de ces sens.

Si nous comparons *quedar* et *haber* en ce qui concerne le domaine de l'existence à l'aide de la phrase analysée précédemment, nous avons les énoncés suivants :

---

<sup>1</sup> <https://www.larousse.fr/>



- **Quedan** tres galletas. (Il reste trois gâteaux).
- **Hay** tres galletas. (Il y a trois gâteaux).

Si nous n'avions pas les traductions et nous nous basions sur l'usage dans le dictionnaire, ces deux verbes seraient des synonymes sur le domaine de l'existence. Cependant, *quedar* exprime l'existence à la suite d'une interaction des forces, alors que *haber* exprime tout simplement l'existence de quelque chose.

Ces verbes peuvent véhiculer des sens similaires, mais leur usage apporte des nuances importantes sur le plan sémantique. Ces nuances peuvent causer des difficultés dans des contextes bilingues ou de traduction.

Le dernier contexte semblerait être pertinent puisque le grand public semble utiliser les moteurs de traduction automatique de plus en plus au quotidien. Ces utilisateurs n'ont pas forcément le bagage linguistique pour repérer les erreurs dans les sorties de TA et ils font confiance à ces derniers. Il est donc raisonnable de se demander comment les moteurs de TA contemporains fonctionnent.

## **Chapitre 2. La traduction automatique neuronale : un outil robuste, mais pas infallible**

Dans ce chapitre, notre objectif est de présenter, dans un premier temps, de manière succincte des éléments clés pour comprendre le fonctionnement de la traduction automatique neuronale (TAN). Ensuite, nous expliquerons rapidement en quoi les moteurs de TA peuvent paraître des outils avantageux pour le grand public (du point de vue de l'utilisateur), mais qui proposent des sorties de TA pouvant contenir différents types d'erreurs, dont des erreurs liées au sens.

### ***1. La traduction automatique (TA) de nos jours : le paradigme neuronal***

La traduction automatique (TA) a gagné en popularité et a été très utilisée dès le début des années 1990 (Barbin, 2020). Ainsi, la recherche en TA a évolué et les modèles qui intéressaient les chercheurs ne sont plus les mêmes. En 2017, Castilho et al. affirment que le modèle de TA qui attire l'attention des chercheurs est le modèle neuronal.

Les systèmes de TA neuronale fonctionnent à partir de réseaux de neurones qui jouent des rôles différents, mais qui constituent une seule architecture. Autrement dit, chaque réseau de neurones a une responsabilité : encoder la phrase source, effectuer une représentation intermédiaire, faire attention à des règles particulières qui font partie de l'algorithme, décoder les représentations pour produire une phrase cible, notamment (Koehn, 2017 ; Volkart, 2018 ; Pérez-Ortiz, Forcada & Sánchez-Martínez, 2022).

#### ***1.1. Les réseaux de neurones***

D'après Pérez-Ortiz, Forcada & Sánchez-Martínez (2022), les réseaux de neurones sont composés de calculs mathématiques. Chaque neurone possède un algorithme et ils peuvent être liés entre eux, afin de mener à bien une tâche en particulier. Ainsi, chaque réseau de neurones est organisé par couche (*layer*). Toutes les couches doivent effectuer une tâche (sauf pour les tâches intermédiaires) : la tâche effectuée par une couche est récupérée par la suivante qui s'occupe d'améliorer le résultat selon l'algorithme qui lui a été attribué. Cette manière de retravailler ce qui a été fait par la couche précédente se répète selon le nombre de couches qui composent le modèle.

Pour qu'un modèle soit efficace, il doit être entraîné à l'aide d'exemples (Koehn, 2017 ; Poibeau, 2017 ; Pérez-Ortiz, Forcada & Sánchez-Martínez, 2022). La plupart des modèles sont entraînés à l'aide de corpus parallèles, c'est-à-dire des groupes de textes en

langue source et ses traductions. Le modèle commence donc à traduire et son algorithme d'apprentissage effectue des corrections en modifiant certaines caractéristiques de chaque couche de neurones. Ceci lui permet de réduire le nombre d'erreurs par rapport au corpus d'entraînement.

Quand l'entraînement est terminé, alors le modèle peut être employé pour traduire de nouvelles phrases, similaires aux énoncés que l'on peut trouver dans le corpus d'entraînement. En effet, un modèle qui a été entraîné peut effectuer des généralisations, lui permettant de traiter de nouvelles phrases. La généralisation a lieu quand le modèle est capable de réaliser une représentation intermédiaire d'un énoncé ne figurant pas dans les exemples d'entraînement, mais qui présente des similitudes avec ces derniers (Pérez-Ortiz, Forcada & Sánchez-Martínez, 2022).

### ***1.2. Le plongement lexical***

Afin de trouver des similitudes entre les mots, ceux-ci sont placés par le système dans un espace où ils peuvent être localisés à l'aide de coordonnées spécifiques. Cet espace est un espace vectoriel où les mots sont placés par l'algorithme en fonction de leur proximité sémantique. Par exemple, si dans le corpus d'entraînement il y avait le mot *chien*, alors le système placera tout mot ayant de similitudes sémantiques proches de *chien*. Donc, lors de l'encodage d'une nouvelle phrase source où le mot *chat* est employé, le système devra placer ce mot proche du mot *chien*, étant donné que les deux sont des animaux domestiques.

Le modèle neuronal va donc placer les mots des phrases sources dans cet espace vectoriel en assignant des coordonnées à chaque mot. En fonction de leur place sur cet espace, il effectue une représentation intermédiaire qu'on appelle plongement lexical. D'après Volkart (2018) : « Un plongement lexical est une représentation du sens des mots qui se base sur l'idée que des mots qui apparaissent dans le même contexte sont similaires ».

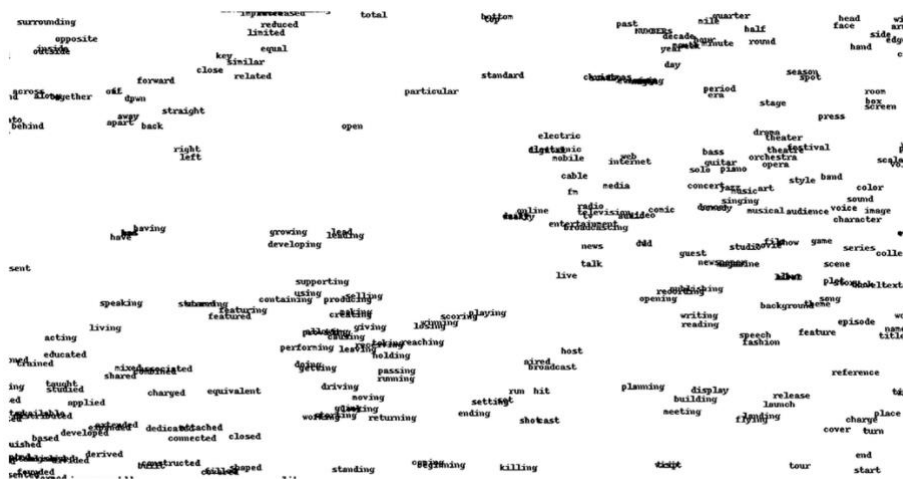


Figure 1 : Plongement lexical (Koehn, 2017 : 36)

Dans la Figure 1, nous remarquons que les mots possédant des similitudes sémantiques se trouvent proches les uns des autres. Ceci permet donc de regrouper les mots entre eux en fonction de leur sens, et d'établir des généralisations par la suite (Pérez-Ortiz, Forcada & Sánchez-Martínez, 2022). Ce mode de fonctionnement permet aux systèmes de TAN de traiter de nouveaux énoncés (Koehn, 2017). À partir de cette représentation intermédiaire, le modèle est capable de proposer des traductions (phrases cibles).

Les modèles les plus récents emploient un mécanisme d'attention. Il s'agit, en effet, d'une couche permettant de prendre en compte les plongements lexicaux effectués auparavant. Ainsi, ce mécanisme peut mener à bien un calcul pour chaque plongement lexical par mot et prédire par la suite la représentation la plus probable dans la sortie de TA (Pérez-Ortiz, Forcada & Sánchez-Martínez, 2022). Certains modèles peuvent employer différentes couches dédiées à l'attention, ce qui permet d'améliorer les résultats.

### 1.3. Le modèle le plus utilisé : Transformer

Un modèle très populaire actuellement est celui connu sous le nom de *Transformer* (Vaswani et al., 2017). Il s'agit d'une architecture du type encodeur-décodeur, employant une approche neuronale pure (Koehn, 2017).

Ce modèle possède un encodeur qui est responsable d'effectuer une représentation de chaque mot, de les placer dans l'espace vectoriel lors du plongement lexical et d'attribuer un pourcentage à chaque mot. Ceci est fait pour que le module d'attention puisse comparer ces pourcentages avec ceux des mots qui composent le corpus d'entraînement. Ainsi, ces

modules peuvent calculer successivement quels sont les mots les plus probables qu'il faut employer dans la sortie de TA.

Nous trouvons ensuite le décodeur. Celui-ci est programmé pour faire attention aux représentations vectorielles de chaque mot dans la phrase source et de les comparer avec celles des mots candidats (les mots qui seront proposés dans la sortie de TA) (Vial, 2020 ; Pérez-Ortiz, Forcada & Sánchez-Martínez, 2022).

Cependant, il s'agit bien d'une opération mathématique qui est basée sur ce qui est sauvegardé dans la mémoire du modèle, qui est à son tour construit à l'aide d'un corpus ayant un nombre limité de mots. Il ne peut donc pas avoir toutes les occurrences possibles de tous les mots. Ainsi, même des moteurs de TAN peuvent proposer des sorties de TA contenant des erreurs.

## ***2. La TAN au quotidien***

Les moteurs de traduction automatique suscitent un grand intérêt. Ceux-ci sont utilisés par un grand nombre de professionnels du domaine, mais aussi par des utilisateurs non-spécialistes.

### ***2.1. Un outil qui est de plus en plus utilisé au quotidien par un public non spécialiste<sup>2</sup>***

Loffler-Laurian (1994) explique qu'un public « non averti » a pu s'emparer d'un « instrument nouveau qui ne lui était pas destiné de prime abord » (p. 88) à partir de 1986. C'est à partir de cette époque que l'usage des moteurs de TA se démocratise parmi un public composé de personnes qui « s'amuse à traduire comme ils s'amuse avec les messageries roses ou les jeux » et de personnes « qui ont besoin de la traduction » (Loffler-Laurian, 1994 : 92).

En effet, ces utilisateurs ne maîtrisent pas forcément une langue étrangère. Ils profitent donc des moteurs de traduction automatique pour traduire des documents du quotidien tels que des modes d'emploi, écrire des courriers ou des rapports, traduire des textes pour l'école ou l'université, effectuer des recherches ou étudier (Loffler-Laurian,

---

<sup>2</sup> Nous employons le terme « non-spécialiste » comme dans l'étude de Nassau, Molle & Kalyaniwala (2022). Ce terme désigne des personnes qui ne suivent pas une formation spécialisée en langues étrangères.

1994 ; Nurminen & Papula, 2018 ; Kasperè et al., 2021 ; Nassau, Molle & Kalyaniwala, 2022).

Même si la plupart des utilisateurs n'ont pas les connaissances nécessaires pour vérifier la fiabilité des sorties de TA (Larsonneur, 2021), Nurminen & Papula (2018) démontrent qu'une majorité des utilisateurs non spécialistes possèdent des connaissances en langue étrangère et ne font appel à la TA que pour vérifier la compréhension globale du texte qui les intéresse. Ils nomment ce type de personnes comme étant des « gist MT users ».

Par ailleurs, Nassau, Molle & Kalyaniwala (2022) ont trouvé des résultats similaires à travers une étude adressée à des étudiants non spécialistes des langues. Dans cette étude, ils ont recueilli différentes raisons pour lesquelles les étudiants faisaient appel aux outils de TA :

- **Apport de vocabulaire** (synonymes, nouveaux mots) ;
- **Modèle de prononciation** (fonction de lecture du traducteur automatique afin de s'entraîner à prononcer des mots nouveaux) ;
- **Besoin de vérification** (rassure et permet de se mettre en confiance) ;
- **Occasion de manipuler la langue** (contact de plus avec la langue, une occasion supplémentaire de l'observer et d'y réfléchir).

Hellmich (2021) a trouvé des résultats similaires à travers une étude plus ciblée. Elle a décidé de s'intéresser à 26 apprenants du français aux États-Unis et qui avaient un niveau débutant (A1-A2) selon le Cadre européen commun de référence pour les langues (CECR). À travers ce travail de recherche, elle relève le besoin de chercher du vocabulaire grâce à la TA, mais aussi le besoin de vérifier leur travail et de travailler plus efficacement (gain de temps).

Par ailleurs, Yamada (2019) explore la possibilité d'utiliser les systèmes de TA pour aider les apprenants d'une langue étrangère. En effet, dans son étude, il observe que les systèmes de TA peuvent représenter une aide précieuse pour ce public d'apprenants. Bourdais & Guichon (2020) retrouvent des résultats similaires en travaillant avec des lycéens. Ces derniers utilisent des moteurs de TA au quotidien, mais possèdent des compétences linguistiques inégales, ce qui peut rendre l'usage des moteurs de TA plus difficile.

De manière générale, les moteurs de TA représentent un grand avantage pour les personnes qui cherchent à traduire un document de manière immédiate. Kasperé et al. (2021) expliquent, en effet, que ces outils facilitent l'accès à l'information, à la communication et, dans l'idéal, ils devraient proposer une version cible d'un texte source qui soit complète et qui convienne immédiatement à l'usage prévu par la personne intéressée. Cependant, les moteurs de TA disponibles pour le grand public comportent leur part de risques.

## ***2.2. Les erreurs les plus fréquentes dans les sorties de TA***

Dans les sorties de TA nous pouvons trouver d'erreurs diverses et variées. Ces erreurs-là disparaissent petit à petit des sorties de TA en raison de l'amélioration des réseaux neuronaux, des architectures mises au point et de la qualité des données d'apprentissage.

Cependant, les utilisateurs grand public ou non spécialistes ne sont pas toujours capables de reconnaître des erreurs subtiles, qui peuvent avoir des conséquences en ce qui concerne le sens (modification, modulation, manque de nuance, omissions, notamment), et au niveau morpho-syntaxique (néologismes, collocations incorrectes, fautes d'accord, notamment). Yvon (2019) pose la question suivante :

« Who will guarantee internet users that what they are reading in one seemingly pristine language accurately reproduces the source language message—of which they are necessarily unaware? »

Les chercheurs du domaine essaient d'y répondre en s'intéressant aux erreurs. En effet, l'analyse des erreurs est une tâche qui peut aider à améliorer les systèmes de TA. À travers l'annotation de corpus de sorties de TA, ils ont été capables de mettre en place des typologies d'erreurs<sup>3</sup> différentes afin d'évaluer la qualité de ces systèmes (Esperança-Rodier et al., 2019).

Il y a différentes typologies qui ont été créées au fur et à mesure, par exemple celle de Vilar et al. (2006), celle de Farrús et al. (2010) ou celle de Tezcan et al. (2018). Dans notre projet de recherche de l'année dernière, nous avons pu travailler avec la typologie du projet ExTra, qui avait pour objectif de former d'utilisateurs responsables en leur expliquant le fonctionnement de la TA sans rentrer dans les détails techniques. Cette typologie est particulièrement intéressante puisqu'elle a été conçue à partir des erreurs les plus fréquentes

---

<sup>3</sup> Une typologie d'erreurs de traduction est une liste d'erreurs qui ont été organisées par catégorie, voire par sous-catégorie, selon le type de modification qu'il faut effectuer afin de corriger le problème (Esperança-Rodier et al., 2019).

dans les sorties de TA. Elle est construite autour de quatre catégories principales d'erreurs et chaque catégorie prend en compte différentes sous-catégories.

Voici la grille de cette typologie d'erreurs :

<b>Catégorie principale</b>	<b>Sous-catégorie</b>
<b>Morphosyntaxe</b>	<ul style="list-style-type: none"> <li>• Anaphore</li> <li>• Déterminant</li> <li>• Hedging<sup>4</sup></li> <li>• Négation</li> <li>• Organisation des constituants</li> <li>• Accord</li> <li>• Composé</li> <li>• Modalité</li> <li>• Temps / Aspect / Modalité (TAM)</li> <li>• Impersonnel</li> <li>• Autre</li> </ul>
<b>Sens / Sémantique</b>	<ul style="list-style-type: none"> <li>• Perte de sens : contresens</li> <li>• Perte de sens : faux sens</li> <li>• Métaphore : verbe</li> <li>• Métaphore : groupe nominal</li> <li>• Métonymie</li> </ul>
<b>Vocabulaire</b>	<ul style="list-style-type: none"> <li>• Ajout</li> <li>• Collocation et/ou faux sens</li> </ul>

---

<sup>4</sup> Il s'agit, en effet, d'une erreur lorsque, dans la sortie de TA, la proposition de traduction est formulée de manière ambiguë par rapport à la phrase source. Il y a donc un changement en ce qui concerne le degré de certitude dans la phrase cible.



	<ul style="list-style-type: none"> <li>• Répétition</li> <li>• Mot de liaison</li> </ul>
<b>Omission</b>	<ul style="list-style-type: none"> <li>• Groupe de mots</li> <li>• Mot</li> </ul>

Figure 2 : Typologie d'erreurs du projet ExTra (tableau préparé par nos soins)

Cette typologie permet d'annoter les erreurs dans les sorties de TA. Ces annotations peuvent être utiles pour les informaticiens qui travaillent sur les architectures des moteurs de TA, notamment. En effet, ils peuvent les prendre en compte et les intégrer dans les modules chargés du mécanisme d'attention. Celui-ci pourrait donc réaliser les plongements lexicaux en tenant compte des informations annotées. Le système adapterait donc les sorties proposées en fonction des annotations faites par des humains.

Par ailleurs, il est également intéressant de mettre en garde le grand public contre ces erreurs de traduction et de lui faire comprendre comment utiliser au mieux un système de traduction automatique.

### ***3. Comment aider les utilisateurs non spécialistes à utiliser les outils de TA de manière raisonnée ?***

Hellmich (2021) propose certaines pistes qui pourraient être implémentées afin de former les étudiants non spécialistes en langue :

- Proposer des activités afin d'apprendre aux étudiants de quelle manière la quantité de texte saisie affecte la sortie de TA : trop de texte ou très peu de texte peuvent affecter la qualité de la sortie ;
- Enseigner quelques techniques pour vérifier la sortie de TA en proposant des moments d'entraînement : les erreurs récurrentes, par exemple ;
- Expliquer de manière non technique et brièvement la manière dont les algorithmes fonctionnent (Pérez-Ortiz, Forcada & Sánchez-Martínez, 2022).

Ces pistes peuvent être intéressantes afin de sensibiliser le public au fonctionnement global de la TA, aux risques liés à un usage non informé ou peu informé et à la manière dont il faudrait employer ces outils pour qu'ils soient bénéfiques pour les personnes qui l'utilisent au quotidien. Par ailleurs, ces recommandations sont en accord avec ce que Bowker & Ciro (2019) nomment « machine translation literacy » (MT literacy). Il s'agit, en effet, d'une

notion qui regroupe différentes compétences que les non spécialistes et, même les spécialistes, devraient développer pour utiliser la TA de manière raisonnée.

**Partie 2**  
-  
**Méthodologie**

## Chapitre 3. Le projet LOLYTA

Dans ce chapitre, nous allons présenter, dans un premier temps, le cadre du projet de recherche nous ayant permis de travailler sur le verbe *quedar*. Ensuite, nous allons nous pencher sur les tâches que nous avons menées à bien tout au long de notre stage de recherche. Enfin, nous présenterons le corpus que nous avons construit pour ce projet, ainsi que les difficultés rencontrées.

### 1. Description du projet LOLYTA

Le projet *LOLYTA : Lire la pOLYsémie verbale avec la Traduction Automatique*, a été créé à l'aide de mes directrices de mémoire afin de proposer un cadre d'étude de la polysémie verbale de *quedar*, en s'intéressant particulièrement au traitement du verbe en question par les moteurs de traduction automatique.

De nos jours, l'usage des systèmes de TA en ligne s'est démocratisé. Le grand public y fait appel au quotidien puisqu'il s'agit d'une solution rapide et gratuite. Alors, les usages se multiplient et les utilisateurs pensent avoir accès à des outils proposant des traductions de qualité. Cependant, ces systèmes produisent, en général, des textes ayant un niveau grammatical et orthographique correct en apparence. Les utilisateurs sont donc incités à faire confiance au contenu sémantique de ces textes. Or, la traduction automatique n'est pas sans failles et le sens des mots, des paragraphes ou du texte entier n'est pas transmis correctement.

Il nous a donc semblé important de vérifier à quel point les moteurs de TA sont fiables, notamment lorsqu'ils doivent traiter des parties du discours pouvant avoir différents sens selon le contexte d'usage. Autrement dit, l'objectif du projet est d'évaluer si les systèmes de TA en ligne sont capables de traiter des parties du discours complexes, c'est-à-dire qui ne peuvent pas être interprétées de manière littérale, mais dont le sens change en fonction de son utilisation et des autres mots ou groupes de mots dans son environnement linguistique.

Par ailleurs, comme nous l'avons vu dans les chapitres précédents, les moteurs de TAN sont capables de proposer de meilleures sorties de TA. Ceci est possible puisque ces systèmes prennent en compte l'ensemble de la phrase, voire l'ensemble du paragraphe, au moment de l'encodage et du décodage. Ainsi, ils prennent en compte le contexte général de la phrase ou du paragraphe.

Cependant, comme l'ont démontré Hutchins (2001), Kasperé et al. (2021) ou Nassau, G., Molle, N. et Kalyaniwala, C. (2022), les utilisateurs grand public ne traduisent pas toujours des phrases ou des paragraphes en entier, mais des mots ou de groupes de mots isolés. Il nous a donc paru raisonnable, pour ce projet de recherche, de vérifier s'il existe une différence entre des traductions isolées (ou hors contexte) du verbe *quedar* et des traductions en contexte (qui prennent en compte l'intégralité de la phrase où le verbe en question a été employé, ainsi que des phrases précédentes et suivantes).

## **2. Objectif du stage et tâches effectuées**

Notre stage de recherche s'inscrit au sein du projet LOLYTA, mené à bien au sein du LIDILEM (laboratoire de linguistique et didactique des langues étrangères et maternelles).

La première mission qui nous a été confiée portait sur la constitution du corpus qui allait être utilisé pour étudier la polysémie du verbe *quedar* selon différents contextes d'usage. De plus, ce corpus allait nous permettre de mettre en place une démarche empirique, c'est-à-dire qu'après la revue de littérature, nous allions nous pencher sur le corpus afin d'observer les différents cas de polysémie de *quedar*. Ensuite, nous devions exploiter le corpus afin de repérer les occurrences du verbe en question et de les classifier selon les critères sémantiques établis par Gómez Vicente (2018).

Nous devons préciser, par ailleurs, qu'il s'agit d'un corpus parallèle, c'est-à-dire un corpus composé du texte source (espagnol) et de sa traduction humaine (français). Nous avons donc dû effectuer une étape de nettoyage et de normalisation de corpus pour pouvoir, par la suite, aligner le texte source et la traduction de référence. Ce dernier élément est important puisqu'il s'agit de notre modèle de référence, nous permettant d'effectuer la comparaison entre une traduction de qualité humaine et les sorties de TA proposées par la machine.

Cette comparaison allait nous permettre de noter les erreurs de traduction dans les sorties de TA concernant le verbe *quedar* et de repérer les différentes solutions proposées par les traducteurs humains (sur le plan syntaxique et sémantique) afin de traduire correctement le sens du verbe selon le contexte d'usage.

### 3. Présentation du corpus et difficultés rencontrées

#### 3.1. Présentation du corpus parallèle

En ce qui concerne notre recherche, nous avons décidé d'étudier la polysémie verbale à travers des sous-titres de films ayant l'espagnol comme langue source. Pour ce faire, nous avons décidé de construire notre propre corpus à partir des sous-titres de films disponibles sur Netflix. Les films que nous avons choisis avaient tous des sous-titres en espagnol et en français, et ils étaient tous assez récents (entre 2020 et 2022). Afin de prendre en compte différentes variétés d'espagnol, nous avons décidé d'inclure cinq films en espagnol d'Espagne (ES-es) et cinq films en espagnol hispano-américain (ES-ha)<sup>5</sup>. Le Tableau 1 ci-après liste les films que nous avons pris en compte dans notre travail :

Es-es	ES-ha
El practicante (2019)	El cuaderno de Tomy (2020)
La familia perfecta (2021)	El comediante (2021)
A mil kilómetros de la navidad (2021)	Hasta que nos volvamos a encontrar (2022)
Amor de madre (2022)	Qué culpa tiene el karma (2022)
Hollyblood (2022)	Matrimillas (2022)

Tableau 1 : Films faisant partie du corpus étudié

Notre corpus prend en compte des variétés d'espagnol différentes afin d'avoir une représentation globale de l'usage du verbe *quedar* dans différentes parties du monde hispanophone (ici, nous avons inclus les variétés espagnole, mexicaine, argentine et péruvienne). De plus, nous avons décidé de travailler avec un échantillon d'espagnol contemporain de manière à recueillir les usages de *quedar* que nous pouvons rencontrer de nos jours.

Nous avons divisé notre corpus parallèle en deux sous-corpus tout en prenant en compte les variétés d'espagnol mentionnées. Étant donné que le nombre de films espagnols

---

<sup>5</sup> Nous avons décidé d'employer ce terme plutôt que latinoaméricain puisque l'Amérique latine (et par conséquent latinoaméricain) prend en compte d'autres pays qui n'ont pas l'espagnol comme langue officielle. Par ailleurs, ici nous tenons pas compte de toutes les variétés d'espagnol qui peuvent présenter certaines différences en ce qui concerne, notamment, le lexique. Il ne s'agit donc pas d'une variété homogène, mais d'une variété que les chercheurs regroupent afin de faire la distinction entre l'espagnol du continent américain et l'espagnol du continent européen.

était supérieur au nombre de films hispano-américains, nous avons créé un sous-corpus pour l'espagnol d'Espagne et un deuxième pour les autres variétés. Cette division facilitait notre étude et nous a permis d'évaluer, sur un petit échantillon, les usages les plus récurrents de *quedar* en fonction des régions qui nous intéressent dans ce travail. Même si cela n'est pas en lien direct avec notre hypothèse de départ, il nous a semblé intéressant de comparer les usages du verbe *quedar* entre les régions évoquées.

D'une part, le sous-corpus dédié à l'espagnol d'Espagne comportait les données suivantes (extraites de SketchEngine) :

	<b>ES-es</b>	<b>FR</b>
<b>Nombre de mots</b>	37 937	36 780
<b>Nombre de phrases</b>	7 715	7 363
<b>Nombre de tokens</b>	51 844	47 163

Tableau 2 : Données statistiques du sous-corpus Es-es–FR

D'autre part, le sous-corpus dédié à l'espagnol hispano-américain comportait les données suivantes (extraites de SketchEngine) :

	<b>ES-ha</b>	<b>FR</b>
<b>Nombre de mots</b>	40 343	38 661
<b>Nombre de phrases</b>	8 251	7 898
<b>Nombre de tokens</b>	55 470	49 490

Tableau 3 : Données statistiques du sous-corpus Es-ha–FR

En outre, nous avons décidé de créer un corpus parallèle afin de disposer de traductions humaines de référence. Celles-ci allaient être utiles au cours de l'analyse, car nous souhaitons :

- d'être en mesure de comparer les traductions humaines avec les sorties de traduction automatique ;
- d'observer les structures et les équivalents employés par les traducteurs humains par rapport à ceux employés par la machine.

Par ailleurs, le moteur de traduction automatique que nous avons décidé d'utiliser pour ce projet de recherche est DeepL<sup>6</sup>. Il s'agit d'un système qui est disponible en ligne et propose une version gratuite très sollicitée par les utilisateurs grand public. D'après leur site internet, ce moteur de TA a créé des réseaux de neurones « capables de saisir les moindres nuances et de les traduire comme aucun autre logiciel de traduction ». Par conséquent, il nous a paru l'occasion idéale de vérifier si un système proposant des traductions de 'bonne qualité' pouvait réellement identifier les différents sens du verbe *quedar*.

### ***3.2. Les difficultés rencontrées lors de la constitution du corpus***

En ce qui concerne notre corpus, nous avons rencontré différents types de difficultés tout au long de ce travail de recherche.

#### ***3.2.1. Première difficulté : origine des sous-titres***

La première difficulté concerne l'accès aux sous-titres officiels pour constituer le corpus en lui-même. En effet, nous avons décidé de travailler avec des films récents, mais les sous-titres n'étaient pas toujours disponibles sur des sites comme OpenSubtitles<sup>7</sup>. Étant donné que nous ne voulions pas faire confiance à des sites internet peu fiables, nous avons décidé de les télécharger directement sur le site de Netflix. En effet, si nous récupérions les sous-titres sur un site quelconque, nous n'étions pas absolument certains de l'origine des sous-titres. En d'autres termes, nous ne savions pas s'il s'agissait des sous-titres officiels de Netflix ou s'ils avaient été conçus par des amateurs ou des *fans*.

#### ***3.2.2. Deuxième difficulté : droits d'auteurs et notre manière de récupérer les sous-titres***

De plus, une autre difficulté, à laquelle nous n'avons pas pu trouver une solution satisfaisante, est celle des droits d'auteur. En effet, nous voulions demander l'accès aux sous-titres qui nous concernaient, mais nous n'avons pas eu de réponse de leur part, même en ayant contacté Netflix à plusieurs reprises.

Bien que nous n'ayons pas eu de retour, nous avons réussi à trouver un moyen de télécharger les sous-titres officiels.

---

<sup>6</sup> <https://www.deepl.com/translator>

<sup>7</sup> <https://www.opensubtitles.org>



D'abord, il faut se connecter sur Netflix en utilisant Firefox<sup>8</sup>. Puis, vous choisissez la série ou le film qui vous intéresse. Après avoir lancé le film, on pause la lecture de la vidéo et il faut faire un clic droit et choisir « Inspect Element » (voir Image 1). Ceci ouvre la fenêtre « Developer ». Sur cette fenêtre, il faut sélectionner l'onglet « Network » et, par la suite, chercher l'onglet XHR. Cela étant, il suffit de chercher le fichier octet-stream et le télécharger à l'aide d'un clic droit (Image 2). Le fichier téléchargé est un document au format .dms. Il faut effectuer cette étape pour les sous-titres en espagnol et en français séparément.

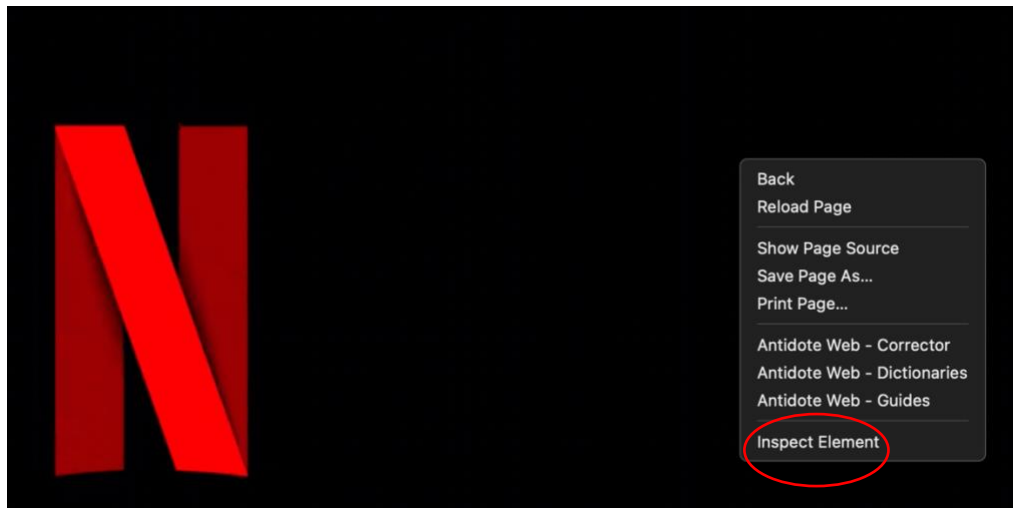


Image 1 : Sélectionner Inspect Element

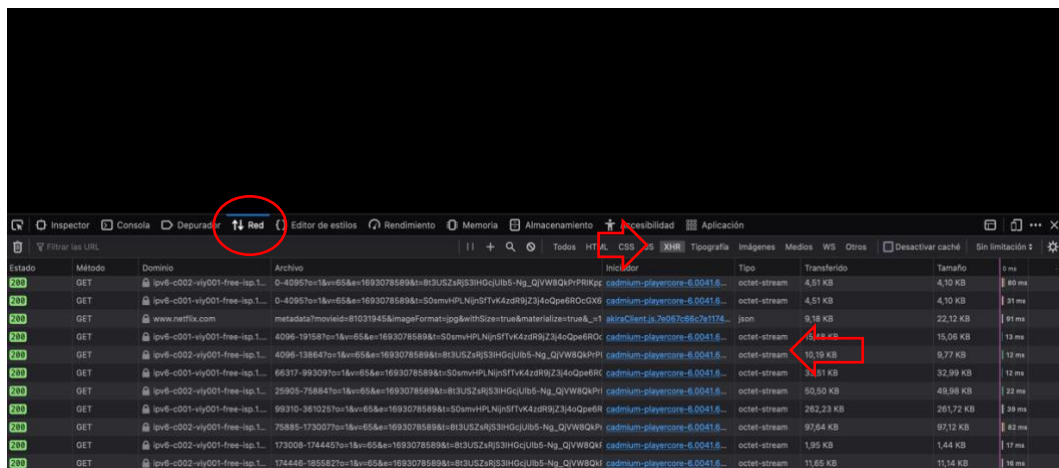


Image 2 : Cliquer sur Network et ensuite sélectionner XHR pour trouver le fichier octet-stream

<sup>8</sup> Nous avons tenté de télécharger les sous-titres en utilisant d'autres navigateurs, mais cela n'a pas fonctionné.

Une fois le fichier .dms téléchargé, il faut le convertir au format .srt. Nous avons mené à bien cette étape en utilisant le site GoTranscript<sup>9</sup>. Il suffit de téléverser le fichier .dms, puis de sélectionner le format .srt et cliquer sur « Convert » (Image 3). Cette étape nous permet de récupérer un fichier qui peut être converti, à son tour, dans un fichier au format .txt.

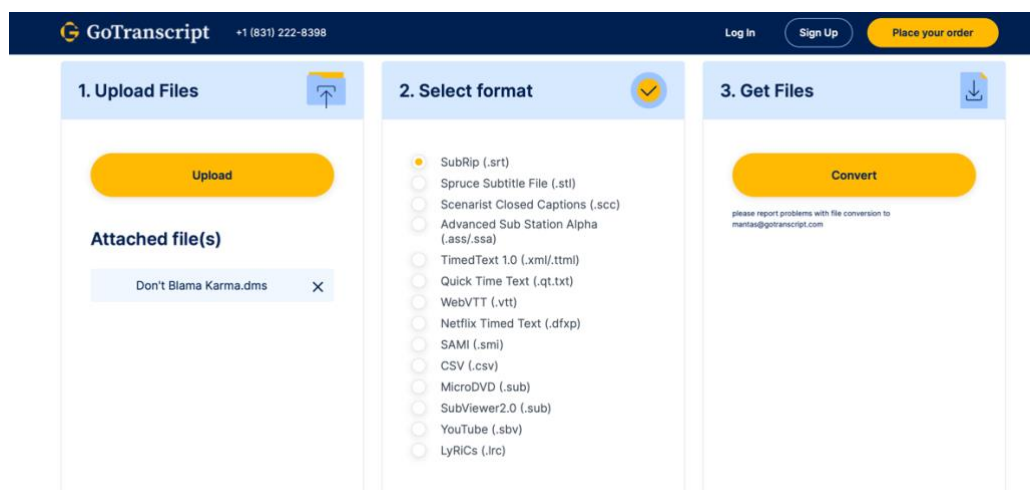


Image 3 : Capture d'écran du site GoTranscript

Le format .txt était le format qui nous intéressait. En effet, nous avons besoin de récupérer les fichiers au format .txt pour qu'ils puissent être exploitables par la suite, en les téléversant sur des sites comme Sketch Engine ou le Lexicoscope. Il s'agit de deux outils d'exploration de corpus, qui permettent de déposer ses propres corpus originaux et effectuer des analyses textuelles à l'aide de requêtes complexes. De plus, ces interfaces peuvent prendre en charge des fichiers au format .txt, c'est-à-dire le format avec lequel nous avons décidé de travailler.

### 3.2.3. Troisième difficulté : alignement et nettoyage du corpus

Étant donné que nous voulions créer un corpus parallèle, nous devions aussi trouver un moyen d'aligner les sous-titres en espagnol et ses traductions en français.

Cependant, nous nous sommes rendu compte, assez rapidement, qu'un alignement automatique n'était pas possible avec les fichiers que nous avons récupérés. En effet, ces derniers n'étaient pas divisés par segments équivalents et chaque fichier contenait des

---

<sup>9</sup> <https://gotranscript.com/subtitle-converter>

informations supplémentaires différentes. Ces dernières variaient selon les langues et le texte se trouvait, généralement, entre crochets ou en italique :

- en ce qui concerne les sous-titres en espagnol, nous avons observé des paroles de chansons, des symboles accompagnant les paroles, des bruits relevant de l'environnement, notamment ;
- les informations figurant dans les sous-titres en espagnol n'ont pas été reprises dans les sous-titres en français, créant ainsi de grands écarts entre les dialogues des personnages, rendant la tâche d'alignement plus compliquée.

Nous avons donc dû réaliser le nettoyage du corpus en ayant comme objectif final l'alignement automatique. Pour ce faire, nous avons dû travailler, simultanément, sur un texte en espagnol et sa traduction. Ceci nous permettait, dans un premier temps, de supprimer toutes les informations supplémentaires qui n'avaient pas de correspondance dans le texte traduit et vice-versa et par la suite, créer des segments équivalents entre le texte source et la traduction. Ainsi, les outils d'exploration de corpus précédemment mentionnés pouvaient prendre en charge les documents et effectuer l'alignement phrase par phrase ou segment par segment de manière automatique.

Le nettoyage de corpus a été une tâche extrêmement chronophage et répétitive, mais nécessaire pour pouvoir extraire les occurrences de *quedar* et les analyser par la suite.

## **Conclusion**

Dans ce chapitre, nous avons pu présenter l'objectif et l'intérêt du projet de stage qui nous a permis de mener à bien notre recherche. De plus, nous avons exposé la manière dont nous avons constitué notre corpus, ainsi que les difficultés rencontrées lors de ce processus : les droits d'auteur, le nettoyage et la normalisation du corpus, et l'alignement des textes source avec les traductions officielles.

**Partie 3**  
-  
**Résultats et discussion**

## Chapitre 4. Présentation des résultats du projet LOLYTA

Dans ce chapitre, nous allons décrire dans un premier temps la manière dont nous avons exploré et exploité le corpus présenté précédemment. Ensuite, nous aborderons les résultats que nous avons observés concernant le verbe *quedar*, ainsi que les structures grammaticales les plus fréquentes. Enfin, nous consacrerons une partie à la description des équivalents observables dans les traductions humaines et les sorties de TA.

### 1. L'exploitation du corpus de sous-titres

Nous avons décidé d'exploiter notre corpus à l'aide de l'outil Sketch Engine<sup>10</sup>, principalement. Même s'il s'agit d'un outil payant, nous avons pu créer un compte d'essai ayant une durée de trente jours. En effet, nous avons des connaissances sur cet outil puisque nous l'avons déjà utilisé auparavant. M. Denis Jamet, professeur de linguistique anglaise à l'université Jean Moulin Lyon 3, décrit Sketch Engine comme :

« [...] un gestionnaire de corpus et un logiciel d'analyse textuelle développé par Lexical Computing Limited depuis 2003. Son objectif est de permettre aux personnes qui étudient les langues (lexicographes, chercheurs en linguistique de corpus, traducteurs ou apprenants en langues) ou quiconque qui souhaite trouver des exemples authentiques de rechercher de grandes collections de textes selon des requêtes complexes. »<sup>11</sup>

De plus, Kilgarriff et al. (2014) précisent dans leur article que Sketch Engine est un outil dédié à l'étude des corpus offrant des fonctionnalités telles que la gestion de corpus publics et privés, la recherche des collocations et des concordances, la création d'un thésaurus à partir d'un corpus, notamment. Pour ce projet, nous avons téléversé les fichiers au format .txt pour que l'outil puisse les aligner et effectuer l'analyse textuelle, en faisant ressortir les occurrences du verbe *quedar*, ainsi que les structures grammaticales l'accompagnant.

Cependant, nous avons aussi sollicité M. Olivier Kraif, enseignant-chercheur à l'Université Grenoble Alpes, afin de déposer notre corpus sur le Lexicoscope<sup>12</sup>. D'après les articles de Kraif & Diwersy (2012) et Kraif (2016), il s'agit d'un outil conçu pour explorer et interroger des corpus multilingues à travers des requêtes donnant accès à des cooccurrents

---

<sup>10</sup> <https://www.sketchengine.eu>

<sup>11</sup> Citation prise du site de l'université Jean Moulin Lyon 3 : <https://bu.univ-lyon3.fr/sketch-engine-un-outil-pour-ceux-qui-etudient-les-langues>

<sup>12</sup> [http://phraseotext.univ-grenoble-alpes.fr/lexicoscope\\_2.0/](http://phraseotext.univ-grenoble-alpes.fr/lexicoscope_2.0/)

syntaxiques, mais aussi permettant d'effectuer des tâches de concordance, de recherche d'expressions complexes et des tableaux de cooccurrents.

Notre corpus est donc disponible dans notre collection privée où nous avons pu l'étudier. Il peut, cependant, être publié dans les collections publiques sur le Lexicoscope pour que d'autres étudiants ou chercheurs puissent l'exploiter.

## 2. *Les occurrences de quedar dans notre corpus*

Dans cette partie, nous allons décrire les occurrences du verbe *quedar*, ainsi que des structures grammaticales que nous avons pu observer dans notre corpus.

### 2.1. *Le sous-corpus ES-es-FR*

Au sein du sous-corpus dédié à l'espagnol d'Espagne, nous avons repéré 65 occurrences du verbe *quedar*. Nous avons observé ces occurrences en effectuant une requête de la forme canonique du verbe, c'est-à-dire la forme du verbe ne contenant aucune flexion.

Comme nous l'avons mentionné dans le chapitre précédent, ce sous-corpus est composé de 37 937 mots (pour la partie en espagnol). Ceci veut dire que *quedar* ne représente qu'environ 0,17 %<sup>13</sup> du pourcentage total des mots qui constituent cette partie du sous-corpus (voir Tableau 4 : Statistiques concernant les occurrences de *quedar* dans le sous-corpus ES-es-FR).

<b>Nombre de mots (ES-es)</b>	37 937
<b>Occurrences de <i>quedar</i></b>	64
<b>Pourcentage de <i>quedar</i> dans le corpus</b>	0,17

Tableau 4 : Statistiques concernant les occurrences de *quedar* dans le sous-corpus ES-es-FR

En ce qui concerne les structures grammaticales, nous avons observé des patrons syntaxiques similaires à ceux que Gómez Vicente avait déjà repérés en 2018 (et que nous avons mentionnés dans la première partie de ce travail). Voici les patrons les plus fréquents (Tableau 5) :

---

<sup>13</sup> Pourcentage calculé par Sketch Engine.

Structure grammaticale	Nombre d'occurrences
<b>Quedar + adjectif</b>	17
<b>Quedar + syntagme nominal (SN)</b>	16
<b>Quedar + syntagme prépositionnel (SP)</b>	11
<b>Quedar + adverbe</b>	7
<b>Omission du verbe</b>	13
<b>Total</b>	64

Tableau 5 : Les structures grammaticales les plus fréquentes (sous-corpus ES-es-FR)

Pour les constructions du type quedar + adjectif, nous avons, par exemple :

- <sup>14</sup>Nos **quedamos atrapados** en la nieve.  
*\*Nous sommes restés attrapés sous la neige. (traduction mot à mot)*  
*La voiture a été piégée dans la neige. (traduction de référence)*
- Pues yo ya me **quedo pagada**.  
*\*Bon moi je reste payée.*  
*J'en suis ravie !;*
- Todo para que nuestros novios **queden contentos**.  
*\*Tout pour que nos mariés restent contents.*  
*Tout est fait pour votre plaisir.*

Pour les constructions du type quedar+ SN, nous avons, par exemple :

- No **queda sitio** ya, señor.  
*\*Il ne reste pas de place, monsieur.*  
*Il n'y a plus de place, mon bon monsieur.*
- **Me quedo el anillo**.  
*\*Je reste la bague.*

<sup>14</sup> Les exemples donnés par la suite sont présentés de la même manière, c'est-à-dire : phrase source, traduction mot à mot et traduction de référence. Les traductions mot à mot sont précédées d'un \*.

*Je garde la bague.*

- **Ángel, te queda la última.**

*\*Ángel, il reste la dernière.*

*Ángel, continue.*

Pour les constructions du type quedar +SP, nous avons, par exemple :

- **Me quedo con su cuarto yo.**

*\*Je reste avec sa chambre moi.*

*Je prends sa chambre.*

- **Quedamos en la puerta del insti esta noche.**

*\*Nous sommes restés à la porte de l'école ce soir.*

*Rendez-vous ce soir devant le lycée.*

- **La Ronería queda al otro lado del barranco.**

*\*La rumerie reste de l'autre côté du ravin.*

*La rumerie se situe de l'autre côté, près d'un champ de cannes à sucre.*

Pour les constructions du type quedar + adverbe, nous avons, par exemple :

- **Yo me quedo aquí.**

*\*Je reste ici.*

*Je reste ici.*

- **Ya nos queda poquito.**

*\*Il nous reste peu.*

*C'est pour bientôt.*

- **Se te ha quedado ahí un pegote.**

*\*Il te reste un pâté là.*

*Tu avais un pâté sur le visage.*



## 2.2. Le sous-corpus ES-ha-FR

Au sein du sous-corpus dédié à l'espagnol hispano-américain, nous avons repéré 77 occurrences du verbe *quedar*. Nous avons observé ces occurrences en effectuant une requête de la forme canonique du verbe, c'est-à-dire la forme du verbe ne contenant aucune flexion.

Comme nous l'avons mentionné dans le chapitre précédent, ce sous-corpus est composé de 40 343 mots (pour la partie en espagnol). Ceci veut dire que *quedar* ne représente qu'environ 0,17 %<sup>15</sup> du pourcentage total des mots qui constituent cette partie du sous-corpus (voir Tableau 6).

<b>Nombre de mots (ES-ha)</b>	40 343
<b>Occurrences de <i>quedar</i></b>	77
<b>Pourcentage de <i>quedar</i> dans le corpus</b>	0,17

Tableau 6 : Statistiques concernant les occurrences de *quedar* dans le sous-corpus ES-ha-FR

En ce qui concerne les structures grammaticales, nous avons observé des patrons syntaxiques similaires à ceux que Gómez Vicente avait déjà repérés en 2018. Voici les patrons les plus fréquents :

<b>Structure grammaticale</b>	<b>Nombre d'occurrences</b>
<b>Quedar + syntagme prépositionnel (SP)</b>	25
<b>Quedar + adjectif</b>	18
<b>Quedar + adverbe</b>	14
<b>Quedar + syntagme nominal (SN)</b>	12
<b>Autres construction ou omission</b>	8
<b>Total</b>	77

Tableau 7 : Les structures grammaticales les plus fréquentes (sous-corpus ES-ha-FR)

Pour les constructions du type *quedar* + SP nous avons, par exemple :

<sup>15</sup> Pourcentage calculé par Sketch Engine.

- **Me quedé con** tu deseo.  
*\*Je suis resté avec ton vœu.*  
*J'ai pris ton vœu.*
- Que yo sepa, tu consultorio **queda en la otra punta de la ciudad.**  
*\*Que je sache, ton cabinet reste à l'autre bout de la ville.*  
*Ton cabinet dentaire est à l'autre bout de la ville.*
- La lavandería **queda al lado de la ferretería.**  
*\*La laverie reste à côté de la quincaillerie.*  
*La laverie est à côté de la quincaillerie.*

Pour les constructions du type quedar+ adjectif, nous avons, par exemple :

- No te puedes **quedar callada.**  
*\*Tu ne peux pas rester en silence.*  
*Tu sais pas te taire, des fois.*
- Pero ¿**quedó pajasa?**  
*\*Mais, est-il resté paille ?*  
*C'était génial ?*
- **Quedamos seleccionados** para el torneo de cocineros amateurs.  
*\*Nous sommes restés sélectionnés pour le tournoi des cuisiniers amateurs.*  
*On est pris pour le concours de cuisine.*

Pour les constructions du type quedar +adverbe, nous avons, par exemple :

- Sí, es que me **queda muy lejos**, mamá.  
*\*Oui, il me reste très loin, maman.*  
*C'est loin d'ici, maman.*
- Che, Fede, no me quiero **quedar acá.**  
*\*Écoute, Fede, je ne veux pas rester ici.*  
*Fede, pas question de rester ici.*

- Tú estuviste viajando 20 años antes de **quedarte aquí**.

*\*Tu as voyagé 20 ans avant de rester ici.*

*Tu as voyagé pendant 20 ans avant de t'installer ici.*

Pour les constructions du type quedar + SN, nous avons, par exemple :

- **Queda Puno.**

*\*Il reste Puno.*

*Alors, Puno.*

- **Queda otro paciente** todavía.

*\*Il reste un autre patient encore.*

*Il te reste un patient.*

- Porque, no sé, tal vez no limpié bien y **quedó alguna suciedad**.

*\*Parce que, je ne sais pas, peut-être je n'ai pas bien nettoyé et il est resté quelque saleté.*

*Je ne suis pas sûr d'avoir bien fait la vaisselle.*

### ***3. Les équivalents de quedar : traductions humaines vs sorties de TA***

Dans cette section, nous allons décrire, dans un premier temps, les équivalents fonctionnels du verbe *quedar* qui ont été employés par les traducteurs humains. Dans un deuxième temps, nous allons présenter les équivalents proposés par les moteurs de traduction automatique afin de pouvoir les comparer dans le Chapitre 5. Analyse des données et réflexion personnelle sur les limites de l'étude.

#### ***3.1. Les équivalents fonctionnels dans les traductions humaines***

Nous parlerons ici des équivalents fonctionnels que nous avons pu repérer dans notre corpus. Mihaescu (2012) définit l'équivalent fonctionnel par opposition à l'équivalent linguistique. En effet, ce dernier est, en réalité, la traduction exacte du mot que l'on peut trouver dans les dictionnaires. Cependant, un équivalent fonctionnel désigne, plutôt, la traduction du mot source qui remplit la même fonction dans la langue cible. Nous donnerons des exemples par la suite.

### 3.1.1. Les équivalents de *quedar* dans les traductions humaines du sous-corpus ES-es-FR

Le Tableau 8 ci-dessous regroupe les équivalents fonctionnels de *quedar*. Nous avons noté la forme lemmatisée de ceux-ci afin de simplifier le tableau.

Équivalents lemmatisés	Nombre d'occurrences
Rester	25
Être	16
Avoir	7
Se retrouver	2
Rencontrer	2
Rater	1
Comprendre	1
Se situer	1
Finir	1
Garder	1
Tomber enceinte	1
Perdre connaissance	1
Continuer	1
Se voir	1
Prendre	1
Vivre pour (faire quelque chose)	1
Aller + adjectif	1
<b>Total</b>	<b>64</b>

Tableau 8 : Les équivalents de *quedar* dans les traductions humaines (ES-es-FR)

Les trois équivalents les plus fréquents sont (par ordre croissant d'occurrence) *avoir*, *être* et *rester*. Ces trois verbes représentent les premières options de traduction que l'on peut

retrouver en effectuant une recherche rapide sur des dictionnaires bilingues comme Linguee<sup>16</sup> ou des concordanciers bilingues comme Tradooit<sup>17</sup>.

Comme nous l'avons vu précédemment, nous nous attendions à observer l'équivalent *rester* comme étant l'usage le plus répandu puisqu'il s'agit de l'équivalent littéral de *quedar* en français.

### 3.1.2. Les équivalents de *quedar* dans les traductions humaines du sous-corpus ES-ha-FR

En ce qui concerne les équivalents fonctionnels dans les traductions humaines du sous-corpus ES-ha-FR, nous avons pu repérer des résultats similaires. Le tableau ci-dessous regroupe les équivalents observés.

Équivalents lemmatisés	Nombre d'occurrences
Rester	21
Être	18
Prendre	10
Garder	4
Avoir	3
Faire	3
Attendre	2
Faire passer	1
Se retrouver	1
Se rasseoir	1
Comprendre	1
Se rassurer	1
Séjourner	1
Hériter	1

<sup>16</sup> <https://www.linguee.es>

<sup>17</sup> <http://www.tradooit.com/index.php>

Perdre	1
Se taire	1
Dire	1
S'installer	1
Cuisiner	1
Aller	1
Durer	1
Rappeler	1
Devenir	1
S'endormir	1
<b>Total</b>	<b>77</b>

Tableau 9 : Les équivalents de *quedar* dans les traductions humaines (ES-ha-FR)

Nous avons remarqué que *rester* et *être* représentent, à nouveau, les équivalents de *quedar* les plus fréquents dans ce sous-corpus. Cela dit, nous avons pu repérer d'autres équivalences fonctionnelles comme *prendre* ou *garder* qui n'avaient qu'une seule occurrence dans le sous-corpus dédié à l'espagnol d'Espagne.

### 3.2. Les équivalents fonctionnels dans les sorties de TA

#### 3.2.1. Les équivalents de *quedar* dans les sorties de TA pour le sous-corpus ES-es-FR

Le Tableau 10 ci-dessous regroupe les équivalents fonctionnels de *quedar* dans les sorties de TA. Nous avons noté la forme lemmatisée de ceux-ci afin de simplifier le tableau.

Équivalents lemmatisés	Nombre d'occurrences
Rester	36
Être	9
Avoir	7
Rencontrer	3
Se retrouver	2

Manquer	1
Tomber enceinte	1
Perdre connaissance	1
Prendre	1
S'annoncer	1
Garder	1
Aller + adjectif	1
<b>Total</b>	<b>64</b>

Tableau 10 : Les équivalents de *quedar* dans les sorties de TA (sous-corpus ES-es-FR)

Dans ce sous-corpus, nous avons repéré douze équivalents de *quedar* en français. Par rapport aux équivalents dans les traductions humaines, il y en a cinq équivalents en moins. De plus, même si *rester*, *être* et *avoir* continuent d'être les équivalents les plus fréquents, il y a un changement en ce qui concerne *rester* : ce dernier est utilisé 36 fois dans les sorties de TA. Il est donc employé dans 11 phrases de plus par rapport aux traductions humaines, c'est-à-dire une augmentation de la fréquence d'usage d'environ 144 %.

### 3.2.2. Les équivalents de *quedar* dans les sorties de TA pour le sous-corpus ES-ha-FR

Pour la partie concernant le sous-corpus ES-ha-FR, nous avons regroupé les équivalents fonctionnels dans le Tableau 11. Nous avons remarqué, comme pour le sous-corpus précédent, une réduction du nombre de lemmes employés (on passe de 24 à 21 lemmes).

Équivalents lemmatisés	Nombre d'occurrences
Rester	38
Être	15
Garder	4
Devenir	2
Se trouver	2
Avoir	1

Prendre	1
Faire passer	1
Se faire avoir	1
Se retrouver	1
Perdre	1
Obtenir	1
Se taire	1
Aller	1
Cesser	1
S'endormir	1
S'inquiéter	1
Convenir	1
Tenir	1
Passer	1
Choisir	1
<b>Total</b>	<b>77</b>

Tableau 11 : Les équivalents de *quedar* dans les sorties de TA (sous-corpus ES-ha-FR)

Nous avons repéré 21 équivalents de *quedar* en français. Par rapport aux équivalents dans les traductions humaines, *rester* et *être* continuent d'être les équivalents les plus fréquents. Cependant, il y a un changement conséquent en ce qui concerne *rester* : ce dernier est utilisé 38 fois dans les sorties de TA. Il est donc employé dans 16 phrases de plus par rapport aux traductions humaines, c'est-à-dire une augmentation de la fréquence d'usage d'environ 180 %.

## Conclusion

Dans ce chapitre, nous avons pu présenter les résultats de notre recherche. D'une part, nous avons pu expliquer la manière dont nous avons exploité le corpus, grâce à l'aide de Sketch Engine et du Lexicoscope, notamment. D'autre part, nous avons exposé les occurrences de *quedar* dans nos deux sous-corpus, en faisant attention aux structures



grammaticales les plus fréquentes avec le verbe en question. Enfin, nous avons pu présenter les équivalents fonctionnels de *quedar* dans les traductions humaines et les sorties de TA.

Tous ces éléments seront mis à profit dans le chapitre suivant, dans lequel nous analyserons nos données. Nous vérifierons ainsi si le système de traduction automatique que nous avons sollicité a été en mesure de traiter correctement les différents usages et significations du verbe *quedar*.

## Chapitre 5. Analyse des données et réflexion personnelle sur les limites de l'étude

Dans ce chapitre, nous essaierons d'analyser les données que nous avons recueillies. Premièrement, nous nous focaliserons sur les différents domaines (dont relève le verbe *quedar*) présents dans notre corpus. Nous essaierons de vérifier s'il y a un lien entre ces domaines et les structures grammaticales, qui pourraient donner des pistes pour les traduire correctement. Deuxièmement, nous nous pencherons sur les erreurs de traduction que nous avons pu repérer lors de l'analyse des sorties de TA en contexte et hors contexte. De plus, nous consacrerons une partie à la notion de *machine translationese*, c'est-à-dire le langage utilisé dans les sorties de TA, qui provoque une diminution de la diversité des équivalents de *quedar* en français.

### 1. Les différents domaines de *quedar* présents dans notre corpus

Comme nous l'avons vu dans la revue de littérature, Gómez Vicente (2013, 2018) a identifié certains domaines desquels relève le verbe *quedar*. Autrement dit, même si le verbe ne change pas, l'interprétation qu'il faut faire pour comprendre le sens du verbe dépendra du contexte d'usage. Ainsi, nous avons pu analyser nos données sous ce point de vue afin d'étudier les structures grammaticales en lien avec chaque domaine. Ceci pourrait, éventuellement, représenter des pistes pour faciliter l'apprentissage de l'espagnol ou même pour donner des idées aux informaticiens qui améliorent les systèmes de TA.

#### 1.1. Les domaines dans le sous-corpus ES-es-FR

Dans cette partie du corpus, nous avons repéré six domaines : localisation, état, existence, rendez-vous, possession et point de vue. Le domaine le plus fréquent dans nos données est la localisation (24 occurrences), suivi de l'état (18 occurrences), l'existence (12 occurrences), le rendez-vous (6 occurrences), la possession et le point de vue (2 occurrences chacun). La Figure 3 illustre cette répartition de manière graphique.

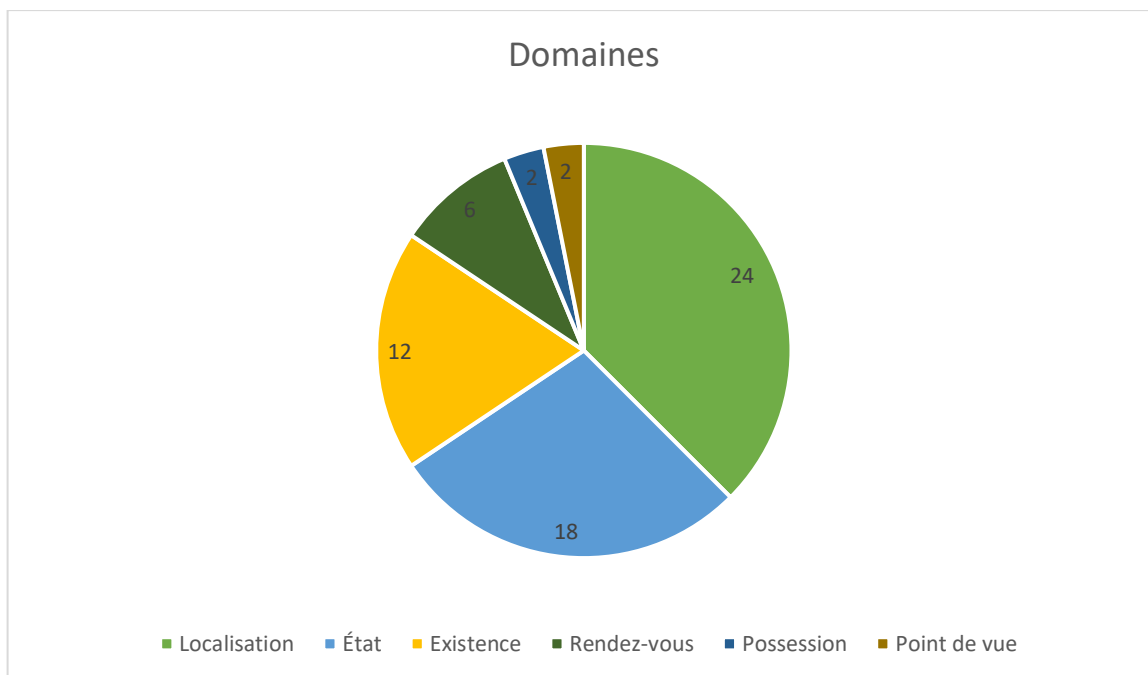


Figure 3 : Les domaines de *quedar* dans le sous-corpus ES-es-FR

Parmi ces domaines, il y a des structures grammaticales qui reviennent assez régulièrement. Nous discuterons chacun des domaines ci-dessous.

#### Le domaine de la localisation

Pour le domaine de la localisation, nous avons identifié trois constructions possibles avec *quedar* :

<b>Construction</b>	Quedar + Syntagme prépositionnel (SP)
<b>Fonction</b>	Complément de lieu
<b>Exemples</b>	<p>(27) La Ronería queda <b>al otro lado del barranco</b>.</p> <p><i>La rhumerie se situe <b>de l'autre côté</b>.</i></p> <p>(32) ¡Si lo sé, me quedo <b>en prisión!</b></p> <p><i>Je préfère rester <b>en prison!</b></i></p> <p>(62) No puedo quedarme <b>en casa</b> quitando pañales y haciéndole la cena a tu hijo.</p> <p><i>Je ne peux pas vivre pour changer des couches et faire à manger à ton fils !</i></p>

Tableau 12 : Construction avec *quedar* + syntagme prépositionnel

Dans le cas de cette construction (Tableau 12 : Construction avec *quedar* + syntagme prépositionnel), ce qui se situe à droite du verbe est une préposition du type *a* ou *en*. Il s'agit de deux prépositions qui aident à situer quelque chose ou quelqu'un dans un espace plus ou moins précis.

Si nous appliquons la notion de dynamique des forces employée par Gómez Vicente (en faisant référence au travail de recherche de Talmy (2000), il faut comprendre que ces phrases représentent le résultat suite à un échange de forces.

Par exemple, la phrase (62) est le résultat d'un événement dans le passé qui oblige le locuteur à rester chez lui. D'après la phrase, cette personne était enceinte et après la naissance de son enfant on attend à ce qu'elle reste chez elle à changer des couches et faire à manger. Autrement dit, cette naissance entraîne une situation où elle doit changer ses anciennes habitudes (à l'extérieur de la maison, probablement) et le résultat est qu'on attend qu'elle reste à un seul endroit.

Par ailleurs, étant donné que le sens sollicité par cette construction est la localisation, le rôle du syntagme prépositionnel ayant la fonction de complément de lieu est essentielle pour que la phrase soit grammaticale et que le sens global soit transmis correctement. Si on supprime ce complément, la phrase serait incomplète d'un point de vue syntaxique et sémantique.

<b>Construction</b>	Quedar + Adverbe de lieu
<b>Fonction</b>	Complément de lieu
<b>Exemples</b>	(26) Se te ha quedado <b>ahí</b> un pegote.  Tu avais un pâté <b>sur</b> le visage.  (30) Yo me quedo <b>aquí</b> .  Je reste <b>ici</b> .

Tableau 13 : Construction avec *quedar* + adverbe

Comme pour le cas précédent, ce qui accompagne le verbe (ici un adverbe du type *ahí* ou *aquí*) joue la fonction de complément de lieu. À nouveau, la présence de ce type d'adverbe est obligatoire pour que la phrase ne perde pas une information essentielle. Si on supprime l'adverbe, il y a une partie du sens qui serait perdue, notamment la permanence de quelque chose ou quelqu'un dans un espace plus ou moins précis.

De plus, ces exemples utilisant le verbe *quedar* doivent être compris comme le résultat d'une situation qui provoque la permanence dans le lieu en question. Par exemple, dans la phrase (30), le locuteur annonce qu'il va rester sur place. Ceci est dû à une situation qui a eu lieu dans le passé et qui l'incite à rester à cet endroit précis.

<b>Construction</b>	Quedarse (verbe à la forme pronominale)
<b>Fonction</b>	Noyau du syntagme verbal
<b>Exemples</b>	(8) ¿ <b>Te vas a quedar?</b>  <i>Tu restes ?</i>  (15) <b>Me quedo.</b>  <i>Je reste.</i>  (12) Vale, <b>me quedo.</b>  <i>Très bien. Je reste.</i>

Tableau 14 : Construction avec quedarse

La troisième construction relevant du domaine de la localisation est celle où le verbe *quedar* est employé à sa forme pronominale *quedarse* (voir Tableau 14). D'après Gómez Vicente (2018 : 8), le verbe est accompagné du pronom réfléchi ou pronom personnel conjoint parce que :

« (...) the pronoun allows to express a higher participation of the subject in the event. »

Autrement dit, l'emploi de ce pronom aide les locuteurs à insister sur le fait que l'événement dont ils parlent les affecte directement, et par conséquent, qu'ils jouent un rôle central dans le déroulement de l'action.

Nos exemples sont composés d'un syntagme verbal apparent et d'un syntagme nominal omis. En effet, le syntagme nominal est, souvent, un pronom personnel. Ce type de pronoms peuvent être omis étant donné que l'espagnol est une langue pro-drop. Voici la définition proposée par Jouitteau (2009-2023) sur le site ARBRES<sup>18</sup> :

<sup>18</sup> Il s'agit d'un wikigrammaire des dialectes du breton et centre de ressources pour son étude linguistique. Voici le site : [https://arbres.iker.cnrs.fr/index.php?title=Arbres:Le\\_site\\_de\\_grammaire\\_du\\_breton](https://arbres.iker.cnrs.fr/index.php?title=Arbres:Le_site_de_grammaire_du_breton)

En grammaire générative, une langue dite "à sujet nul", aussi dite "langue pro-drop" est précisément une langue qui autorise des déclaratives et des interrogatives où le sujet n'est pas réalisé.

Donc les exemples (8), (15) et (21) illustrent bien cette définition puisqu'on trouve des phrases déclaratives et interrogatives sans un sujet explicite.

Nous avons remarqué, par ailleurs, que ces trois constructions sont traduites par les verbes *rester*, *être*, *se situer* et *avoir* dans les traductions humaines. Cependant, dans les sorties de TA, ces occurrences sont majoritairement traduites par *rester* ou *être*.

#### Le domaine de l'état

Pour le domaine de l'état, nous avons identifié une construction possible avec *quedar* :

<b>Construction</b>	Quedar/quedarse + adjectif
<b>Fonction</b>	Complément du verbe
<b>Exemples</b>	(24) Todo para que nuestros novios <b>queden contentos</b> . <i>Tout est fait pour votre plaisir.</i> (25) Pues yo ya <b>me quedo pagada</b> . <i>J'en suis ravie !</i> (39) <b>Se ha quedado inconsciente</b> . <i>Il a perdu connaissance.</i>

Tableau 15 : Construction avec quedar/quedarse + adjectif

D'après Gómez Vicente, ce domaine véhicule les notions de permanence et de changement d'état. Nous les avons déjà présentés lors de la revue de littérature.

Ici, nos exemples démontrent bien ces deux notions. La notion de permanence pourrait être illustrée à l'aide de la phrase (24). En effet, le locuteur reste dans un état défini suite à un événement ou une situation. Dans ce cas, une dynamique des forces a eu lieu, ayant pour objectif que les personnes dont le locuteur parle puissent garder un état d'âme positif.

Nous trouvons aussi la notion de changement à l'aide des exemples (25) et (39). Si nous regardons la phrase (25), la locutrice est ravie à la suite d'un événement qui lui a permis de changer d'état d'esprit. Dans le cas de l'exemple (39), un événement a eu lieu, provoquant

la perte de connaissance. Avant cela, cette personne était consciente, mais après l'événement, elle ne l'était plus.

Par ailleurs, si nous essayons de supprimer ces adjectifs, ayant une fonction de complément du verbe, les phrases deviennent agrammaticales. Ces compléments sont donc nécessaires d'un point de vue syntaxique et sémantique pour qu'il n'y ait pas de perte de sens.

Il y a plusieurs équivalents fonctionnels pour ce domaine en particulier. Cependant, nous avons remarqué que les traductions humaines, tout comme les sorties de TA, proposent des traductions suivant un patron similaire : un verbe suivi d'un adjectif.

### Le domaine de l'existence

En ce qui concerne le domaine de l'existence, nous avons identifié une construction avec *quedar* :

<b>Construction</b>	Quedar/quedarse + syntagme nominal
<b>Fonction</b>	Complément d'objet direct
<b>Exemples</b>	(20) No <b>queda sitio</b> ya, señor. <i>Il n'y a plus de place, mon bon monsieur.</i> (22) No <b>queda ni una sola habitación</b> libre en toda la isla. <i>Il n'y a aucune autre chambre disponible sur toute l'île.</i> (40) Te <b>queda la última vuelta.</b> <i>Encore un tour.</i>

Tableau 16 : *Quedar/quedarse* + syntagme nominal

Comme pour les autres domaines, ici le verbe *quedar* possède une valeur résultative, c'est-à-dire qu'on ne parle que du résultat final à la suite d'un événement qui a modifié le nombre d'objets qui existait au départ. Par exemple, si nous observons l'exemple (22), nous pouvons l'interpréter de la manière suivante : au départ, il y avait un nombre  $x$  de chambres, mais il y a eu un événement qui a provoqué que ce nombre de chambres soit modifié, alors à l'aide de *quedar*, on a la possibilité de parler du nombre restant de chambres suite à cet événement. Dans cet exemple, il n'y a plus de chambres après l'événement qui a provoqué le changement d'existence.

Il est intéressant de souligner que le syntagme nominal accompagnant *quedar* a une fonction de complément d'objet direct. Celui-ci ne peut pas être supprimé de la phrase puisqu'elle deviendrait agrammaticale et on ne saurait plus de quoi on parle.

En ce qui concerne les équivalents fonctionnels dans les traductions humaines, nous avons pu repérer des verbes comme *avoir*, *commencer* ou *rester*. Similairement, si nous regardons du côté des sorties de TA, les deux verbes les plus utilisés sont *avoir* et *rester*.

Le domaine du rendez-vous

Pour ce domaine-ci, nous avons repéré une construction assez récurrente pour programmer un rendez-vous avec quelqu'un ou pour en parler du fait qu'on a un rendez-vous. Cette construction utilise le verbe *quedar* à la forme intransitive obligatoirement, c'est-à-dire qu'il s'agit d'une construction qui exclut l'usage de la forme pronominale du verbe. De plus, elle accepte deux types de syntagmes : le syntagme prépositionnel ou le syntagme adverbial (voir Tableau 17).

<b>Construction</b>	Quedar + syntagme prépositionnel Quedar + syntagme adverbial
<b>Fonction</b>	Complément de lieu ou de temps
<b>Exemples</b>	(47) Con el que habías <b>quedado en HollyBlood</b> . <i>Celui que tu voulais rencontrer au cinéma.</i> (53) Justo hemos quedado aquí para irnos de compras. <i>On va se retrouver ici pour faire du shopping.</i>

Tableau 17 : Constructions possibles pour le domaine du rendez-vous

Si nous suivons la logique des dynamiques des forces, le rendez-vous est fixé après une interaction et donc, c'est bien le résultat de cette interaction qui est représenté avec *quedar*. En effet, les exemples (47) et (53) illustrent bien cela.

Par ailleurs, ces compléments pourraient être supprimés d'un point de vue syntaxique : les phrases resteraient grammaticalement correctes. Cependant, au niveau du sens, il y aurait un élément essentiel (la date ou le lieu) qui ne permettrait pas de saisir complètement le sens de la phrase. Donc, il nous semble impossible de supprimer ces compléments d'un point de vue sémantique et pragmatique.



Parmi les équivalents fonctionnels dans les traductions humaines, nous avons identifié *se voir*, *rencontrer* ou *se retrouver*. En ce qui concerne les sorties de TA, nous avons repéré *rencontrer* et *se retrouver*, notamment.

Le domaine de la possession

Pour ce domaine-ci, nous n'avons relevé que deux occurrences. Cela dit, elles suivent la construction décrite par Gómez Vicente (2018). En effet, le verbe *quedar* est suivi d'un syntagme nominal ou d'un syntagme prépositionnel (voir Tableau 18).

<b>Construction</b>	Quedar + syntagme nominal  Quedar + syntagme prépositionnel
<b>Fonction</b>	Complément d'objet direct
<b>Exemples</b>	(37) Doscientos cincuenta, <b>me quedo el anillo.</b>  <i>250 €, mais je garde la bague.</i>  (52) <b>Me quedo con</b> su cuarto yo.  <i>Je prends sa chambre.</i>

Tableau 18 : Construction pour le domaine de la possession

Les exemples (37) et (52) illustrent un changement d'état. En effet, les locuteurs obtiennent quelque chose (une bague ou une chambre) à la suite d'un événement. D'après Gómez Vicente (2018), la préposition est facultative et donc il serait possible d'avoir : *Me quedo su cuarto yo*. De plus, à la différence du domaine lié à la prise de rendez-vous, ce domaine-ci exige l'usage du pronom réfléchi. Par conséquent, l'usage du verbe à la forme intransitive est exclu.

Dans les deux occurrences que nous avons pu repérer, le verbe *quedar* était traduit par *garder* ou *prendre* tant dans les traductions humaines, comme dans les sorties de TA.

Le domaine du point de vue

Enfin, pour le domaine concernant le point de vue, nous avons identifié deux occurrences de *quedar*, qui partagent la même construction (voir Tableau 19). En effet, il s'agit d'une construction avec un adverbe qui précède ou qui suit le verbe *quedar*. Ce changement de position s'explique simplement par le type de phrase : dans la phrase

interrogative, l’adverbe se trouve en tête de phrase et dans la phrase affirmative l’adverbe se trouve après le verbe.

<b>Construction</b>	Quedar + adverbe de manière ( <i>de modo</i> en espagnol) Adverbe interrogatif + quedar
<b>Fonction</b>	Complément du verbe
<b>Exemples</b>	(59) Pablo, ¿cómo me queda? <i>Pablo, de quoi j'ai l'air ?</i> (64) Te queda bien, pero a mí me gusta más largo. <i>Ça te va bien, mais je te préfère les cheveux longs.</i>

Tableau 19 : Constructions pour le domaine du point de vue

Le domaine concernant le point de vue est utilisé pour connaître le point de vue de son interlocuteur. La dynamique des forces n’affecte plus le sujet en lui-même, mais provoque un changement dans l’opinion ou le point de vue que l’interlocuteur a sur le locuteur.

Dans l’exemple (59), le locuteur sollicite directement l’interlocuteur pour connaître le changement de point de vue qu’il y a eu après un événement. Ici, le locuteur peut être en train d’essayer un vêtement et demande l’opinion de l’autre personne après l’avoir mis. Dans l’exemple (64), un locuteur partage directement son point de vue après un événement. On connaît donc son avis par rapport à l’interaction des forces qu’il y a eu juste avant.

En observant les traductions humaines, nous nous sommes rendu compte que *quedar* avait été traduit soit par *aller*, soit par l’expression *avoir l’air*. Le système de TA a aussi proposé ces mêmes équivalents.

### **1.2. Les domaines dans le sous-corpus ES-ha-FR**

Dans cette partie du corpus, nous avons repéré les six domaines que nous avons déjà présentés : localisation, état, existence, rendez-vous, possession et point de vue. Le domaine le plus fréquent dans nos données est l’état, (30 occurrences), suivi de la localisation (26 occurrences), la possession (11 occurrences), l’existence (5 occurrences), le point de vue (4 occurrences) et le rendez-vous (1 occurrence). La Figure 4 illustre cette répartition de manière graphique.

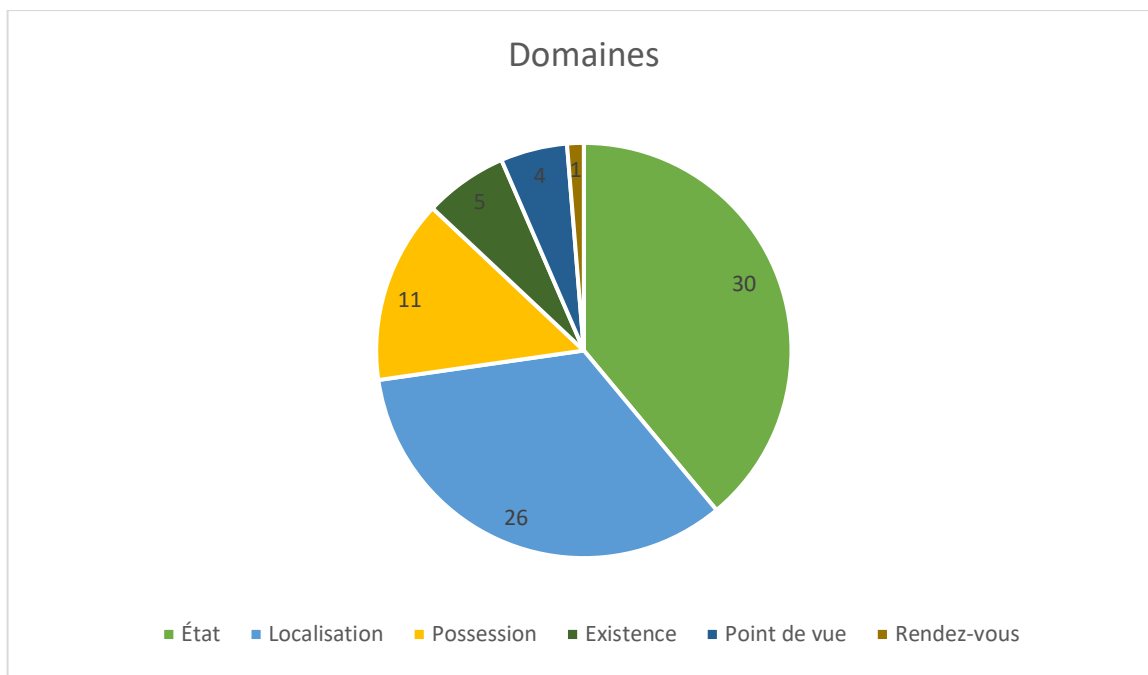


Figure 4 : Les domaines de quedar dans le sous-corpus ES-ha-FR

Parmi ces domaines, il y a des structures grammaticales qui reviennent assez régulièrement. Nous en discuterons chacun des domaines ci-dessous.

#### Le domaine de l'état

Pour le domaine de l'état, nous avons identifié deux constructions possibles avec *quedar* : *quedar* suivi d'un syntagme prépositionnel et *quedar* suivi d'un adjectif.

<b>Construction</b>	Quedar + syntagme prépositionnel
<b>Fonction</b>	Complément
<b>Exemples</b>	<p>(5) Lo mismo que un árbol que, en tiempo de otoño, <b>se queda sin hojas</b>.</p> <p><i>Tout comme un arbre qui à l'automne se retrouve sans feuilles.</i></p> <p>(57) ¡Habíamos <b>quedado en hacer una tregua!</b></p> <p><i>On a dit qu'on faisait une trêve.</i></p> <p>(66) Con ese saludo ya <b>quedaste como un goma</b>, papá.</p> <p><i>Tu as l'air bête à faire coucou.</i></p>

Tableau 20 : Construction avec *quedar* + syntagme prépositionnel

Dans le cas de cette construction (Tableau 20 : Construction avec *quedar* + syntagme prépositionnel), ce qui se situe à droite du verbe est une préposition du type *sin, en, como*. Ces prépositions sont suivies, à leur tour, soit par un verbe à l’infinitif, soit par un syntagme nominal D’après ce que nous avons pu observer, ces structures jouent le rôle de complément circonstanciel.

Si nous appliquons la notion de dynamique des forces employée par Gómez Vicente (2018), en faisant référence au travail de recherche de Talmy (2000), il faut comprendre que ces phrases représentent le résultat à la suite d’une dynamique des forces. Autrement dit, un événement a eu lieu avant qui a provoqué ce résultat-là.

Dans nos exemples, la notion de changement est véhiculée majoritairement. Cette notion est la plus fréquente en ce qui concerne le domaine de l’état (Gómez Vicente ; 2013, 2018). Si nous regardons la phrase (5), nous savons que l’automne exercera une influence sur les feuilles de l’arbre. Le verbe *quedar* se focalisera sur le fait que l’arbre perd ses feuilles à cette époque-là, c’est-à-dire qu’il met en relief ce résultat.

Il nous semble que le rôle du syntagme prépositionnel, ayant la fonction de complément, est essentiel pour que la phrase soit grammaticale et que le sens global soit transmis correctement. Si on le supprime, la phrase serait incomplète d’un point de vue syntaxique et sémantique.

Au niveau des équivalents en français, il y a différentes propositions de traduction dans les traductions humaines : *avoir, être, dire, rester* ou *avoir l’air*. De manière similaire, dans les sorties de TA, nous avons repéré les mêmes équivalents pour la plupart des occurrences.

<b>Construction</b>	Quedar + adjectif
<b>Fonction</b>	Complément du verbe
<b>Exemples</b>	(28) No te puedes <b>quedar callada</b> .  Putain, tu sais pas <b>te taire</b> , des fois ?  (36) Pero ¿ <b>quedó pajasa</b> ?

---

## C'était génial ?

(64) Gala te puso unas gotitas y **te quedaste dormido**.

Gala a mis quelque chose dans ton verre et **tu t'es endormi**.

---

Tableau 21 : Construction avec *quedar* + adjectif

Le Tableau 21 regroupe quelques exemples des constructions du type *quedar* + adjectif. Parmi ceux-ci, nous avons repéré les notions de permanence et de changement, comme pour le sous-corpus dédié à l'espagnol d'Espagne.

Étant donné que cette construction a déjà été présentée (voir Les domaines dans le sous-corpus ES-es–FR), nous n'allons pas nous attarder sur ce point.

En ce qui concerne les équivalents en français, il y a à nouveau plusieurs propositions de traduction. En effet, nous avons remarqué que le verbe *quedar* n'est pas forcément traduit, mais il y a un changement de catégorie grammaticale à travers une transposition. Les traducteurs ont décidé de trouver un verbe qui transmet la même idée en français de manière idiomatique :

- *Quedarse callada* veut littéralement dire « rester en silence ». Cette idée est transmise tout simplement par le verbe se taire.
- *Quedarse dormido* veut littéralement dire « rester dormi », et indique donc le passage d'un état éveillé à un état d'endormissement. Cette idée est simplement transmise à l'aide du verbe *s'endormir*.

Dans les sorties de TA, nous avons observé cette même technique de traduction.

### Le domaine de la localisation

Pour le domaine de la localisation, nous avons identifié les trois types de constructions avec *quedar* que nous avons présentés précédemment :

---

<b>Construction</b>	Quedar + syntagme prépositionnel Quedar + adverbe Quedarse
<b>Fonction</b>	Complément de lieu
<b>Exemples</b>	(22) La lavandería <b>queda al lado de la ferretería</b> .

---

---

*La laverie est à côté de la quincaillerie.*

(31) Tú estuviste viajando 20 años antes de **quedarte aquí**.

*Tu as voyagé pendant 20 ans avant de t'installer ici.*

(55) De verdad, **quedate**.

*Reste avec nous.*

---

Tableau 22 : Constructions avec *quedar* pour parler de la localisation (ES-ha–FR)

Comme nous l'avons discuté auparavant, ce domaine véhicule la notion de permanence dans un espace plus ou moins défini, à la suite d'une dynamique des forces qui encourage cette permanence.

Par exemple, si nous regardons la phrase (31), le locuteur parle à quelqu'un qui voyageait régulièrement, mais qui a terminé par s'installer dans un endroit précis. Ce résultat-ci est exprimé à l'aide de *quedar*.

Il y a plusieurs équivalents fonctionnels pour ce domaine en particulier. Cependant, nous avons remarqué que les traductions humaines, tout comme les sorties de TA, proposent des traductions suivant un patron similaire : un verbe suivi d'un complément de lieu, qui est nécessaire pour que la phrase soit grammaticalement correcte. Cependant, quand le verbe pronominal est employé tout seul, il est traduit par l'équivalent direct en français, c'est-à-dire *rester*.

#### *Le domaine de la possession*

Pour ce domaine-ci nous n'avons relevé l'une des deux constructions présentées dans Les domaines dans le sous-corpus ES-es–FR. Il s'agit de *quedar* accompagné d'un syntagme prépositionnel, ayant une fonction de complément d'objet direct (voir le Tableau 23).

---

<b>Construction</b>	Quedar/quedarse + syntagme prépositionnel
---------------------	---

<b>Fonction</b>	Complément d'objet direct
-----------------	---------------------------

<b>Exemples</b>	(19) Charlie <b>se queda con el secarropas</b> .
-----------------	--

*Charlie hérite du sèche-linge.*

(48) ¿Te gastás unas millas, yo **me quedo con los chicos**?

*Dépense tes points, je garde les enfants.*

---

---

(68) Pues yo **me voy a quedar con todos tus deseos** para siempre.

*Moi, je prendrai tous les tiens, pour toujours.*

---

Tableau 23 : Quedar/quedarse + syntagme prépositionnel

Comme pour les autres domaines, ici le verbe *quedar* possède une valeur résultative, c'est-à-dire qu'on ne parle que du résultat à la suite d'une dynamique des forces. Celle-ci entraîne une personne à gagner quelque chose ou à conserver quelque chose.

Il est intéressant de mettre en avant que dans ces exemples-ci, la construction privilégiée est celle qui emploie une préposition, notamment *con* après le verbe. Celle-ci peut être supprimée de la phrase et il y aurait à la place un syntagme nominal. Il est donc possible d'avoir l'exemple (19) sans la préposition *con* : *Charlie se queda el secarropas*. Cette phrase est grammaticalement correcte et est construite avec le même patron qu'en français.

En ce qui concerne les équivalents fonctionnels dans les traductions humaines, nous avons pu repérer des verbes comme *garder*, *prendre* ou *hériter*. Cependant, si nous regardons du côté des sorties de TA, nous retrouvons *garder* et *prendre*, mais aussi *rester*, *choisir*, *obtenir*, notamment.

Le domaine de l'existence

---

<b>Construction</b>	Quedar + syntagme nominal
<b>Fonction</b>	Complément d'objet direct
<b>Exemples</b>	(33) <b>Queda Puno.</b>  <i>Alors, Puno.</i>  (39) <b>Queda otro paciente</b> todavía.  <i>Tu ne peux pas, il te reste un patient.</i>  (41) ¿No <b>quedó algo</b> que haya hecho papá?  <b>Papa n'a pas cuisiné ?</b>

---

Tableau 24 : Constructions possibles pour le domaine de l'existence

Comme pour les autres domaines, ici le verbe *quedar* possède une valeur résultative, c'est-à-dire qu'on ne parle que d'un résultat à la suite d'un événement ayant modifié le nombre d'objets de départ. Par exemple, si nous observons l'exemple (39), nous pouvons

l'interpréter de la manière suivante : au départ, il y avait un nombre  $x$  de patients, mais il y a eu un événement qui a provoqué que ce nombre de patients soit modifié, alors à l'aide de *quedar*, on a la possibilité de parler du nombre restant de patients après cette dynamique des forces. Dans cet exemple, il y a encore un patient après l'événement qui a provoqué le changement d'existence.

Il est intéressant de souligner que le syntagme nominal accompagnant *quedar* a une fonction de complément d'objet direct. Celui-ci ne peut pas être supprimé de la phrase puisqu'elle deviendrait agrammaticale.

En ce qui concerne les équivalents fonctionnels dans les traductions humaines, nous avons pu repérer des verbes comme *avoir* ou *rester*. Cependant, si nous regardons du côté des sorties de TA, le seul verbe employé est *rester*.

#### Le domaine du point de vue

Pour ce domaine-ci nous n'avons relevé que trois occurrences qui suivent la construction décrite par Gómez Vicente (2018). En effet, le verbe *quedar* est accompagné d'un adverbe, qui peut se placer avant ou après (voir Tableau 25).

<b>Construction</b>	Quedar(se) + adverbe
<b>Fonction</b>	Complément du verbe
<b>Exemples</b>	<p>(27) <b>Queda bien</b>, ¿a que sí?</p> <p style="text-align: center;"><i>Pas mal, non ?</i></p> <p>(44) ¡Ese color, <b>cómo te queda</b>, por favor!</p> <p style="text-align: center;"><i>Cette couleur te va à ravir.</i></p> <p>(49) Yo te digo, Belu, si ustedes no empiezan a probar cosas nuevas, si no se renuevan, <b>la van a quedar ahí</b>.</p> <p style="text-align: center;"><i>Si vous pimentez pas un peu votre vie de couple, ça durera pas.</i></p>

Tableau 25 : Construction pour le domaine du point de vue

Le domaine concernant le point de vue est utilisé pour connaître le point de vue de son interlocuteur ou pour qu'un locuteur partage le sien. La dynamique des forces n'affecte



plus le sujet lui-même, mais provoque un changement dans l'opinion ou le point de vue que l'interlocuteur a sur l'autre.

Dans l'exemple (27), le locuteur est en train d'essayer un vêtement et partage son avis après l'avoir essayé. Dans l'exemple (49), le locuteur partage aussi son point de vue. Ce qui est intéressant, c'est qu'après le verbe, on trouve un adverbe de lieu. Ceci pourrait rendre la compréhension difficile puisqu'il ne fait pas référence à un lieu précis, mais il s'agit d'un usage métaphorique pour indiquer que ce dont le locuteur parle se terminera.

En observant les traductions humaines, nous nous sommes rendu compte que *quedar* avait été traduit par *aller*, *avoir l'air*, *durer* ou *pas mal*, notamment. Le système de TA a proposé le verbe *rester* majoritairement.

#### Le domaine du rendez-vous

Enfin, pour le domaine concernant le rendez-vous, nous avons identifié une occurrence de *quedar*, qui partage la même construction vue dans le sous-corpus ES-es-FR. Il s'agit d'une construction avec un syntagme prépositionnel, ayant une fonction de complément circonstanciel. Cette fonction permet de déplacer plus facilement le complément : soit avant le verbe, soit après celui-ci.

<b>Construction</b>	Quedar + syntagme prépositionnel
<b>Fonction</b>	Complément circonstanciel de temps
<b>Exemples</b>	(13) <b>A las diez habíamos quedado</b> , ¿qué hora es?  <i>On avait rendez-vous à 10 heures.</i>

Tableau 26 : Constructions pour le domaine du rendez-vous

Dans l'exemple (13), le complément circonstanciel de temps se trouve avant le verbe, mais il est possible de voir des phrases où ce même complément est placé après le verbe sans qu'il y ait de problème concernant la syntaxe. Par ailleurs, comme présenté précédemment, *quedar* se trouve à la forme intransitive obligatoirement, c'est-à-dire sans le pronom réfléchi qu'on peut voir présent dans d'autres domaines.

Pour ce cas en particulier, le traducteur humain a proposé l'équivalent *avoir rendez-vous*, mais il est tout à fait possible de voir d'autres équivalents qui transmettent le même message. La machine, par exemple, a proposé le verbe *se retrouver* comme dans les autres occurrences vues dans le sous-corpus ES-es-FR.

### ***1.3. Les domaines et les structures grammaticales : des indices pour traduire correctement quedar ?***

Pour créer un moteur de TA « performant », il faut d'abord l'entraîner à l'aide de corpus (monolingues, comparables, parallèles) (Yvon, 2019). Ceci veut dire que ces systèmes reprennent des structures employées par les humains (au niveau du décodeur) et, à l'aide d'un encodeur, les réutilisent pour proposer une sortie de TA. La qualité des textes qui conformeront le corpus d'entraînement est essentielle pour avoir des sorties de TA de qualité.

Comme en font la remarque Pérez-Ortiz, Forcada & Sánchez-Martínez (2022), les langues ayant de nombreuses ressources écrites ou transcrites ont l'avantage de pouvoir accéder à des textes et des traductions de qualité. Ceci facilite l'entraînement des moteurs de TA pour des langues bien dotées comme l'espagnol, le français ou l'anglais. En effet, le corpus d'entraînement permet de montrer aux réseaux de neurones des exemples de segments en langue source et une traduction humaine, qui peuvent servir de modèle pour traduire des segments similaires par la suite.

Cette méthode d'apprentissage nous fait penser aux théories basées sur la fréquence. En effet, l'une des manières d'entraîner les systèmes neuronaux est à travers la répétition, et donc à travers la fréquence. D'après Nardy & Dugua (2011) faisant référence à Bybee (2003) :

« Ainsi, dans les théories basées sur l'usage, les unités linguistiques sont conçues comme des routines cognitives, c'est-à-dire des patrons récurrents d'activation mentale, qui ne sont pas stockées dans un endroit particulier du système neural »

Si les réseaux de neurones ont déjà rencontré un patron syntaxique et que celui-ci se répète assez fréquemment, ceci devrait donc permettre de le stocker dans sa mémoire et de l'employer quand il repère cette structure dans d'autres textes. Autrement dit, « un contact plus fréquent avec certaines variantes favorisera leur ancrage dans le système et donc leur emploi » (Nardy & Dugua, 2011).

Dans le cadre de *quedar*, il serait donc possible d'entraîner les moteurs de TA avec différentes occurrences du verbe pour qu'il puisse se familiariser avec les différentes structures grammaticales et, par le biais de la répétition, apprendre les différents contextes dans lesquels elles sont utilisées. Ceci pourrait être envisageable en créant une couche (ou *layer* en anglais) qui ait la charge de noter les erreurs dans les sorties de TA concernant les

verbes employés dans différentes constructions. Par exemple, on pourrait imaginer une couche qui possède un algorithme avec les données de certains verbes et les structures grammaticales employées le plus fréquemment avec ces verbes. À travers l'entraînement, cette couche pourrait se familiariser avec des exemples venant du corpus d'entraînement. Ainsi, l'algorithme pourrait croiser les métadonnées fournies avec les phrases modèles afin d'évaluer et de rectifier les sorties de TA.

Les réseaux de neurones ou les algorithmes pourraient par la suite généraliser ce modèle, si l'étiquetage des mots et des structures grammaticales est fait correctement. Ceci permettrait à la machine de représenter le verbe *quedar* et ses co-occurents dans un espace vectoriel, en les plaçant à différents endroits de cet espace en fonction des sens. Par exemple, (25) *Me quedo pagada* et (24) *Todo para que nuestros novios queden contentos* ne seraient pas placés loin l'un de l'autre puisque les deux possèdent la même construction et relèvent du domaine de l'état, donc sémantiquement similaires. En revanche, l'exemple (22) *No queda ni una sola habitación libre en toda la isla* serait placé ailleurs dans l'espace vectoriel puisqu'il ne partage ni la même construction, ni un sens similaire (sémantiquement, cet exemple appartient à un autre domaine).

C'est grâce au développement des réseaux de neurones que la qualité de sorties de TA s'est améliorée (notamment, le modèle *Transformer* que nous avons présenté dans la revue de littérature). En effet, les systèmes de TA neuronale offrent l'avantage de prendre en compte un contexte très large lors de la traduction, car les réseaux de neurones disposent d'une certaine flexibilité lorsqu'ils rencontrent des énoncés inconnus (Koehn, 2017, 2018 ; Pérez-Ortiz, Forcada & Sánchez-Martínez, 2022), mais aussi parce qu'ils peuvent se concentrer sur des parties importantes des énoncés (Stalberg, 2020).

## **2. *Quedar et DeepL : des traductions quasi parfaites ?***

Nous avons utilisé DeepL pour recueillir les sorties de TA en français pour notre corpus.

### **2.1. *Les traductions hors contexte : un problème pour la machine ?***

Allerton (2001) expliquait que les humains sont capables de comprendre le sens des phrases grâce à nos compétences lexico-sémantiques et notre bon sens. En effet, nous arrivons à déchiffrer le sens des mots et des expressions même si nous ne les avons jamais

entendus ou vus. Nous sommes ainsi capables de comprendre les différents sens des verbes polysémiques et de choisir la construction la plus adéquate selon le contexte d’usage.

Grass (2022) ajoute :

« Chaque phrase, même si elle est porteuse de sens, ne saurait se détacher du reste du texte qui forme une sorte d’enchaînement de causes et de conséquences s’insérant dans un contexte plus large, avec lequel il entretient aussi des rapports de causalité. Pourtant, en TA, l’unité de base reste la phrase. Le rapport qu’elle entretient avec celles qui la précèdent et celles qui la suivent se limite à une recherche de l’entourage lexical le plus probable ; c’est déjà beaucoup, mais ce n’est pas suffisant pour parler d’intelligence. »

Il est ainsi, évident, que sortir des phrases de leur contexte d’usage, du texte, poserait des problèmes puisque les liens avec les autres phrases et paragraphes sont perdus. Même si les systèmes de TA s’améliorent, ils continuent de traduire au niveau de la phrase, en effectuant des recherches sur l’environnement dans lequel elle se trouve (quand un paragraphe est saisi dans le moteur de TA). Ceci aide à améliorer les sorties de TA, mais la plupart des utilisateurs sortent les phrases des paragraphes et les traduisent de manière isolée.

Nurminen & Papula (2018) présentent une étude qui démontre qu’une majorité des utilisateurs grand public font appel à la TA pour vérifier la compréhension globale d’un texte. Ces utilisateurs ont besoin des grandes lignes (« gist MT users ») d’un texte et ont tendance à saisir une phrase à la fois. Tout comme dans l’étude de Nassau, Molle et Kalyaniwala (2022) où ils ont observé que les étudiants peuvent être menés à traduire des mots ou des phrases hors contexte.

Nous avons donc fait de même. Notre objectif était de vérifier si le moteur de TA sollicité était capable de traduire correctement les différents sens des constructions avec *quedar* en français au sein des phrases isolées.

Sur 141 occurrences de *quedar*, nous n’avons repéré que 12 erreurs (voir Tableau 27) concernant les constructions syntaxiques employant ce verbe.<sup>19</sup>

ES	FR	DeepL
Ya quedó.	C'était bien ?	Déjà parti.
¡Quedó!	Oui !	C'est à gauche !

<sup>19</sup> Nous n’avons pas pris en compte toutes les erreurs de traduction présentes dans notre corpus. En effet, nous avons juste analysé les erreurs en lien avec le verbe qui nous intéresse.

Pero ¿quedó pajasa?	C'était génial ?	Mais reste-t-il de la paille ?
¿Queda?	Ah bon ?	Est-il à gauche ?
Además me queda de paso.	C'était sur ma route.	D'ailleurs, je suis en route.
¡Quedó!	Fini !	C'est à gauche !
Con ese saludo ya quedaste como un goma, papá.	Tu as l'air bête à faire coucou.	Avec ce salut, tu as déjà fait une gomme, papa.
Me quedé con tu deseo.	J'ai pris ton vœu.	J'ai respecté votre souhait.
Con el calor que hace aquí, no sabes cómo te va a quedar.	Avec la chaleur qu'il fait ici, ce sera parfait.	Avec la chaleur qui règne ici, on ne sait pas à quoi cela va ressembler.
Se va a quedar con todos mis deseos.	Elle me prendra tout.	Il va réaliser tous mes souhaits.
Pues yo ya me quedo pagada.	J'en suis ravie !	Je suis déjà payé.
Con el que habías quedado en HollyBlood.	Celui que tu voulais rencontrer au ciné.	Celle que vous avez rencontrée à HollyBlood.

Tableau 27 : Erreurs de traduction observées

Parmi les erreurs que nous avons relevées, nous en avons identifié deux types : les erreurs de sens (en jaune) et une erreur grammaticale (en vert).

L'erreur grammaticale est due à un mauvais accord de genre entre le sujet et le verbe. Même si le moteur de TA n'avait pas de contexte avant ou après cette phrase, il s'agit clairement d'un sujet masculin singulier en espagnol : *Con el*. Or, DeepL a proposé *Celle* provoquant ainsi l'erreur d'accord *rencontrée* au lieu d'accorder le participe passé au masculin.<sup>20</sup>

En ce qui concerne les erreurs de sens, nous trouvons différentes raisons pour expliquer ces fautes :

<sup>20</sup> Nous avons par ailleurs remarqué la tendance de traduire le pronom *tu* (espagnol) par le pronom vous. Même si cela n'est pas précisément une erreur, il nous semble curieux que le moteur de TA ne choisisse pas l'équivalent direct *tu*.

- d'une part, le moteur a rencontré des expressions (idiomatiques) en espagnol qu'il n'avait pas traitées auparavant comme *quedó pajasa*, *ya quedaste como un goma* ou *me quedo pagada*. Ces expressions relèvent du domaine de l'état et suivent les constructions *quedar(se)* + adjectif et *quedar* + SP (complément de manière). La solution que la machine adopte est de traduire au plus près du texte source en faisant un calque.
- d'autre part, il y a des phrases déclaratives ou interrogatives où le sujet n'est pas réalisé en espagnol du type *Ya quedó*, *¿Quedó!* ou *¿Queda?*. Étant donné que ces phrases accompagnent souvent une autre phrase, cela nous permet de comprendre de quoi on parle. Cependant, ces phrases sont traduites de manière isolée et n'ont aucun contexte permettant de clarifier le sens de *quedar*. Même pour les humains, interpréter ces phrases représenterait un véritable défi. Donc, nous pensons que les traductions proposées par DeepL proviennent certainement d'autres phrases du même style qui se trouvent dans la mémoire du logiciel qui, à son tour, les réemploie dans les sorties de TA.
- de plus, il y a des erreurs de sens qui, à première vue, ne présentent aucun problème comme : *Además me queda de paso que DeepL a traduit par* « D'ailleurs, je suis en route. ». La traduction est correcte d'un point de vue syntaxique, mais en regardant le contexte, on se rend compte que même si le verbe est au présent, il est utilisé pour raconter un fait ancré dans le passé. Donc, la machine n'a aucun moyen de comprendre qu'il s'agit d'un verbe au présent *historique* (*presente histórico*).

## ***2.2. Les traductions en contexte : de meilleurs résultats ?***

Compte tenu de ce que nous avons vu dans la section précédente, nous pouvons nous demander si le fait de donner plus de contexte serait une solution utile à ce problème. D'après Abdul-Rauf & Yvon (2020) (qui citent le travail de Kim, Thanh Tran & Ney (2019)) :

« Using simpler representations also enables to increase the size of the context: in this study, it was found that only a few previous sentences are actually useful for translation quality. »

Nous avons donc décidé d'ajouter une ou deux phrases avant et après afin de vérifier si cela constituait une véritable aide pour DeepL.

Voici le Tableau 28 avec les 12 occurrences qui n'avaient pas été traduites correctement sans contexte :

ES	FR	DeepL
<p>(a) A las chicas del piso haceles un regalo también.</p> <p><b>Charlie se queda con el secarropas.</b></p> <p>Está recontento. Es el mejor regalo.</p>	<p>Fais aussi un cadeau aux infirmières.</p> <p><b>Charlie hérite du sèche-linge.</b></p> <p>Il est ravi. C'est le meilleur cadeau.</p>	<p>Offrez aussi un cadeau aux filles de l'appartement.</p> <p><b>Charlie garde le sèche-linge.</b></p> <p>Il est ravi. C'est le meilleur cadeau.</p>
<p>(b) Vale, muchas gracias.</p> <p>Que vaya bien.</p> <p><b>Ya quedó.</b></p> <p><b>¡Quedó!</b></p> <p>¡Bien! ¡Ya vengan!</p>	<p>Merci quand même.</p> <p>Bonne journée.</p> <p><b>C'était bien ?</b></p> <p><b>Oui !</b></p> <p>Génial. Allez.</p>	<p>D'accord, merci beaucoup.</p> <p>Bonne journée.</p> <p><b>C'est fait.</b></p> <p><b>C'est fait !</b></p> <p>Bien ! Allez !</p>
<p>(c) ¡Ya vengan!</p> <p>Pero <b>¿ quedó pajasa?</b></p> <p>Obvio.</p>	<p><b>Allez.</b></p> <p><b>C'était génial ?</b></p> <p><b>Carrément !</b></p>	<p>Allez, on y va !</p> <p>Mais <b>reste-t-il de la paille ?</b></p> <p>Bien sûr.</p>
<p>(d) Queda otro paciente todavía.</p> <p><b>¿Queda?</b></p> <p>Vamos a tener que cancelar, me parece.</p>	<p>Il te reste un patient.</p> <p><b>Ah bon ?</b></p> <p>Il va falloir annuler.</p>	<p>Il reste un patient.</p> <p><b>Vraiment ?</b></p> <p>Nous allons devoir annuler, je pense.</p>
<p>(e) No te hubieras molestado. ¡Yo podía!</p> <p>No, mi amor, si no me cuesta nada.</p> <p><b>Además me queda de paso.</b></p>	<p>Il ne fallait pas. J'aurais pu le faire.</p> <p>Ça me fait plaisir.</p> <p><b>C'était sur ma route.</b></p>	<p>Vous n'auriez pas pris cette peine. J'aurais pu.</p> <p>Non, mon amour, ça ne me coûtait rien.</p> <p>D'ailleurs, <b>c'était sur mon chemin.</b></p>

<p>(f) Estoy subiendo una historia. Me etiquetás y te mato. <b>¡Quedó !</b></p>	<p>Je publie une story. Tu as pas intérêt à me taguer dedans. <b>Fini !</b></p>	<p>Je télécharge une histoire. Si vous me taguez, je vous tue. <b>Elle est en ligne !</b></p>
<p>(g) ¡Ahí está Toto! Con ese saludo ya <b>quedaste como un goma</b>, papá. Gracias, Mica.</p>	<p>Toto est là. Tu <b>as l'air bête</b> à faire coucou. Merci, Mica.</p>	<p>Voilà Toto ! Avec cette salutation, tu <b>as déjà l'air d'un caoutchouc</b>, papa. Merci, Mica.</p>
<p>(h) ¡No, Sara! ¡No! <b>Me quedé con tu deseo.</b> Pues yo me voy a quedar con todos tus deseos para siempre.</p>	<p>Non, Sara ! Non ! <b>J'ai pris ton vœu.</b> Moi, je prendrai tous les tiens, pour toujours.</p>	<p>Non, Sara ! Non ! <b>J'ai gardé ton souhait.</b> Eh bien, je vais garder tous tes souhaits pour toujours.</p>
<p>(i) Es una belleza, mamá. Con el calor que hace aquí, no sabes <b>cómo te va a quedar.</b> Qué bárbaro, Sarita.</p>	<p>C'est sublime, maman. Avec la chaleur qu'il fait ici, <b>ce sera parfait.</b> Quand même, Sarita.</p>	<p>C'est une beauté, maman. Avec la chaleur, on ne sait pas à <b>quoi ça va ressembler.</b> C'est super, Sarita.</p>
<p>(j) La maldición es cierta. <b>Se va a quedar con todos mis deseos.</b> Este es un momento muy importante para mí.</p>	<p>La malédiction est réelle. Elle <b>me prendra tout.</b> C'est un moment très important pour moi.</p>	<p>La malédiction est vraie. Il <b>va prendre tous mes souhaits.</b> C'est un moment très important pour moi.</p>
<p>(k) Se lo aseguro. Esto es inolvidable. <b>Pues yo ya me quedo pagada.</b> Vámonos, Jean-Pierre.</p>	<p>Croyez-moi, c'est inoubliable. J'en <b>suis ravie !</b> Jean-Pierre, allons-y.</p>	<p>Je vous l'assure. C'est inoubliable. Eh bien, je <b>suis déjà payé</b> pour ça. Allons-y, Jean-Pierre.</p>
<p>(l) Te demostraré que este tío no es Azrael, el vampiro que has estado buscando.</p>	<p>Je vais te montrer que ce n'est pas Azraël, celui que tu cherches depuis si longtemps.</p>	<p>Je te prouverai que ce type n'est pas Azrael, le vampire que tu cherches.</p>



Con el que habías quedado en HollyBlood.	Celui que tu <b>voulais rencontrer au ciné.</b>	Celui que tu <b>as rencontré à HollyBlood.</b>
¿Cómo sabes todo eso?	Comment tu sais tout ça ?	Comment sais-tu tout ça ?

Tableau 28 : Sorties de TA en contexte

De manière générale, les erreurs que nous avons identifiées et regroupées au Tableau 27 ont été corrigées dans les sorties de TA en contexte.

Pour les exemples (b), (d) et (f) qui contenaient des phrases avec un sujet non réalisé, le fait de remettre ces phrases courtes dans leur contexte a résolu le problème de traduction. Même si nous comparons les sorties de TA avec les traductions humaines, celles-ci ne sont pas équivalentes. En effet, il nous semble que le sens est transmis correctement dans les deux cas. Il s'agirait donc d'un débat concernant le style de la traduction, ce qui ne fait pas partie du cadre de notre étude.

De plus, l'erreur d'accord entre le participe passé et le sujet, repérée dans l'exemple (l), a été corrigée aussi. Or, une autre erreur de ce type a été introduite dans l'exemple (k) où *payé* devrait être accordé au féminin.

Par ailleurs, il y a encore des erreurs évidentes : les exemples (c), (g), (i) et (k) en sont la preuve. Il nous semble que ces erreurs concernant le domaine de l'état sont dues, comme expliqué précédemment, au fait que le système n'a jamais été confronté à ces phrases. Il encode donc une proposition basée sur les données dont il dispose, ce qui aboutit à une traduction littérale, connue sous le terme de calque.

Ces exemples sont peu nombreux et ne peuvent pas être généralisés, mais ils sont en adéquation avec ce qui a été démontré dans les études de Tiedemann, & Scherrer (2017) et Voita, E., Sennrich, R., & Titov, I. (2019) : l'ajout du contexte n'entraîne pas forcément une amélioration dans les sorties de TA. De plus, d'autres études consacrées aux erreurs dans les sorties de TA démontrent que, même si celles-ci s'améliorent, il y a encore un long chemin à parcourir avant d'avoir des sorties de TA parfaites (voir, par exemple Tezcan, Hoste & Macken, 2018 ; ou Vaupot, 2020).

### 2.3. « *Machine translationese* » : vers une simplification des usages de *quedar* ?

Comme nous l'avons vu, les systèmes de TA fonctionnent, essentiellement, en deux étapes : l'encodage du texte source et le décodage permettant de prédire les mots du texte cible (Pérez-Ortiz, Forcada & Sánchez-Martínez, 2022). La plupart des systèmes ont été

créés à partir d'un corpus parallèle d'entraînement. Ceci leur permet par la suite d'apprendre à traduire de nouvelles phrases en fonction du corpus ayant servi comme modèle. Les systèmes continuent d'apprendre avec les nouvelles phrases traduites au fur et à mesure.

Cependant, étant donné que les systèmes traduisent à partir des données dont ils disposent, les linguistes et les informaticiens ont remarqué que :

Our analysis shows that MT paradigms indeed increase/decrease the frequencies of more/less frequent words to such extent that a very large amount of words are completely 'lost in translation'. We believe, this demonstrates indeed that current systems overgeneralize and thus, we deem it appropriate to speak of a form of algorithmic bias. –Vanmassenhove, Shterionov & Way (2019)

En effet, le lexique dont les algorithmes disposent est limité pour la plupart, et ne peuvent donc pas ajouter des mots et retenir leur usage qu'après les avoir rencontrés et gardés dans la mémoire, gérée par les réseaux de neurones. Ceux-ci sont programmés pour proposer une sortie de TA et peuvent remplacer des mots qu'ils ne connaissent pas : ils effectuent un calcul pour prédire quel mot pourrait être le plus probable à l'aide des représentations intermédiaires (plongement lexical) et des classifications des mots (étiquettes).

Or, les systèmes de TA ne restituent pas exactement le sens du texte source. Ils se contentent de transmettre les idées avec un langage « standard » (Grass, 2022), c'est-à-dire les unités lexicales les plus fréquentes qui peuvent remplacer les parties du texte qu'elles n'arrivent pas à analyser correctement. Ainsi, Yvon (2019) explique que :

« Many studies have shown that this language differs from natural target-language formulations because of the presence of source-language interference through a lack of lexical diversity, a tendency toward standardization, and finally the addition of explicitations needed to ensure the source message is received correctly. »

Machine translation ou bien le langage de la machine est visible dans les sorties de TA, et nous pensons que dans notre étude, ceci devient évident en comparant les équivalents fonctionnels employés dans les traductions humaines et les sorties de TA.

En effet, si nous regardons le Tableau 8, le Tableau 9, le Tableau 10 et le Tableau 11, il y a une diversité plus importante dans les traductions humaines (TH) par rapport aux sorties de TA :

- Pour le sous-corpus ES-es–FR, nous observons 17 équivalents fonctionnels employés pour traduire *quedar* du côté des TH contre 12 équivalents employés dans les sorties de TA.

- Pour les sous-corpus ES-ha-FR, nous observons 24 équivalents fonctionnels dans les TH contre 21 du côté des sorties de TA pour traduire les constructions avec *quedar*.

Nous pensons que cette baisse de diversité s'explique par l'emploi plus fréquent du verbe *rester* dans les sorties de TA. Il s'agit de l'équivalent direct que nous pouvons trouver dans les dictionnaires ou concordanciers bilingues. Cet usage privilégié du verbe *rester* peut être dû au fait que dans le corpus d'entraînement de DeepL, cette forme est beaucoup plus présente que d'autres. Donc, le réseau de neurones du système essaie de construire des phrases grammaticalement et sémantiquement correctes avec cette équivalence.

Nous pourrions ainsi parler d'un langage de la machine puisque DeepL privilégie certaines formes ou plutôt son algorithme choisit la forme la plus fréquente pour construire des phrases. Vanmassenhove, Shterionov & Gwilliam (2021) ont trouvé que :

« Since MT systems do not necessarily distinguish between different synonymous translations (lexical richness) and morphological variants (grammatical richness), algorithmic bias could lead to the loss of morphological variety (and thus interfere with the ability of our systems to generate at all times a grammatically correct option) (...) »

En raison de sa partialité, l'algorithme fait ce qui pourrait être considéré comme un choix de traduction, tout comme les traducteurs humains le font consciemment ou inconsciemment lorsqu'ils privilégient certaines structures grammaticales, expressions, vocabulaire, notamment. Cela dit, comme l'explique Grass (2022) dans son analyse, l'intelligence artificielle ne fait pas ce choix consciemment, elle ne fait que suivre l'algorithme prévu à cet effet.

### **3. *Les limites de notre étude : des pistes de réflexion ?***

Dans le cadre de notre étude, nous avons été en mesure d'identifier à quel point le verbe *quedar* est complexe sur le plan syntaxique et sémantique. Il s'agit d'un verbe qui pose des problèmes tant aux humains comme aux machines lors des tâches de traduction automatique.

Par ailleurs, notre travail a pu s'inscrire entre l'étude de la polysémie verbale, en s'appuyant sur le travail déjà effectué par Gómez Vicente, et des travaux en traductologie. Cependant, nous avons identifié certaines limites dans notre recherche.

### ***Limite n° 1 : genre textuel restreint***

La première limite que nous avons identifiée est celle du genre textuel. En effet, notre étude est basée sur des sous-titres, donc un type de discours qui est plus ou moins naturel, mais qui n'est pas spontané. Il serait donc intéressant de pouvoir enregistrer des natifs afin d'analyser l'usage de *quedar* et les constructions les plus employées.

De plus, il serait aussi intéressant d'étudier *quedar* à travers d'autres genres textuels : ainsi il serait possible de déterminer si le verbe est utilisé dans un domaine en particulier ou si on peut le trouver un peu partout.

Du point de vue de la traduction automatique, ceci donnerait l'opportunité de traduire *quedar* dans des textes très spécialisés et des textes plus généralistes, permettant d'identifier les domaines où la TA est plus performante et dans lesquels elle fait plus d'erreurs.

### ***Limite n° 2 : consulter d'autres moteurs de TA***

Il nous semble impératif que d'autres études soient faites afin de comparer les sorties de TA de différents systèmes. En effet, chaque système possède une architecture particulière et a été entraîné avec un corpus spécifique. Il est donc évident que chaque moteur de TA pourrait proposer une sortie différente. Par conséquent, des erreurs différentes pourraient apparaître à différents moments du texte. Par exemple, un moteur peut être plus performant lorsqu'il s'agit de traduire *quedar* quand il relève du domaine de la localisation, alors qu'un autre pourrait l'être quand il s'agit de traduire les occurrences relevant du domaine du rendez-vous.

Par ailleurs, des études contrastives de ce type pourraient aider à mieux identifier les biais de chaque algorithme et d'en proposer des solutions conjointement avec les développeurs.

### ***Limite n° 3 : solliciter des professionnels pour l'évaluation des sorties de TA***

Une autre limite que nous avons identifié est le fait que nous ayons analysé les données sans avoir recours à des professionnels de la traduction.

Si nous avions eu les avis d'autres traducteurs humains en ce qui concerne les sorties de TA, nous aurions pu effectuer des analyses concernant l'accord entre annotateurs. Il est toujours intéressant de mesurer cela puisqu'il permet de déterminer si une sortie de TA est vraiment incorrecte ou s'il s'agit simplement d'un problème concernant le style de la sortie de TA ou d'une préférence personnelle du traducteur humain.

Par ailleurs, multiplier les annotateurs d'un corpus pourrait constituer un avantage puisque, même en faisant très attention, il est possible de passer à côté d'une erreur.

## **Conclusion**

Dans ce chapitre, nous avons pu présenter et analyser les données de notre corpus. Plus précisément, nous avons pu observer les différents domaines de *quedar* présents dans nos deux sous-corpus et identifier les constructions grammaticales qui peuvent être propres à chaque domaine. Ainsi, le caractère polysémique du verbe a été clairement mis en évidence.

De plus, nous avons pu relever les équivalents fonctionnels de *quedar* en français. Nous avons également pu comparer l'utilisation de ces équivalents dans les traductions humaines par rapport à ceux qui ont été obtenus par le biais de la TA. Ceci a mis en évidence quelques erreurs de traduction présentes, ainsi qu'une baisse de la diversité lexicale dans les sorties de TA.

Enfin, nous avons présenté quelques limites liées à notre travail de recherche. Ces limites concernaient les droits d'auteurs, l'intérêt de mener à bien d'autres recherches sur *quedar* en analysant d'autres genres textuels et en sollicitant d'autres systèmes de TA, et l'utilité d'inclure des traducteurs professionnels afin d'établir un accord inter-annotateur dans les futurs travaux de recherche.

## Conclusion

Pour conclure, nous pouvons dire que les systèmes de traduction automatique neuronale, comme DeepL, représentent un outil plutôt fiable au niveau des sorties de TA proposées. En effet, le système consulté a été capable de gérer, de manière générale, la polysémie du verbe *quedar*, à l'exception de quelques propositions de traduction.

Cependant, comme nous avons pu le voir, si les utilisateurs ne font pas attention à la manière dont ils utilisent ces outils, ils pourraient être amenés à faire les mauvais choix de traduction ou à choisir le sens erroné du verbe en question. En effet, les utilisateurs grand public ne sont pas formés aux possibles dangers liés aux moteurs de TA. Ils n'ont donc pas développé les bonnes pratiques pour limiter les erreurs de traduction ou de compréhension. Ainsi, les utilisateurs qui font appel aux moteurs de TA pour traduire des phrases isolées, ou hors contexte, risquent de faire confiance à des textes qui semblent être corrects, mais qui en réalité en proposent un sens incorrect du verbe polysémique en question.

Afin de limiter cela, les enseignants pourraient mener à bien deux points :

- d'une part, ils pourraient développer les connaissances des apprenants en lien à l'usage des moteurs de TA. Autrement dit, si les apprenants acquièrent et améliorent leur « machine translation literacy » (Bowker ; 2020), il se pourrait qu'ils soient en mesure de faire des choix informés et ne pas prendre de mauvaises habitudes comme décrites par Bourdais et Guichon (2020).
- d'autre part, il faudrait revoir les programmes, ou les adapter, pour que les enseignants aient le temps de présenter des verbes comme *quedar*. D'après une intervention que nous avons effectuée auprès d'une classe de L3 concernant le verbe *quedar* et ses différents usages, les étudiants l'ont trouvée intéressante et utile. En effet, nous avons identifié des avis<sup>21</sup> comme :

(11) J'ai retenu que le mot *quedar* ne se traduisait pas toujours de la même façon selon le contexte et la construction de la phrase »

(5) « Je considère qu'elle peut être très utile pour des francophones et même des hispanophones puisqu'elle peut nous faire voir la diversité de sens d'un seul mot de façon qu'on ne se trompe pas au moment de le traduire. ».

---

<sup>21</sup> Tous les avis sont disponibles à l'Annexe 3 : Les avis des étudiants de L3 concernant notre intervention sur le verbe *quedar*.

De plus, certains considèrent que ce type d'intervention devrait avoir lieu dès la première année de licence puisque les questions liées à la polysémie ne sont pas traitées explicitement ni au collège, ni au lycée, ni à l'université :

(10) L'intervention était utile et bien orchestrée cependant elle aurait probablement été appréciée dès la L1.

(4) Je pense que cette intervention serait plus utile si elle était faite en L1.

(20) (...) malheureusement ce travail on nous l'apprend généralement pas trop en classe que ce soit au collège, lycée etc.... C'est plus un travail que l'on doit apprendre de nous-même avec les séries, livres voyages etc..., en tout cas moi ça a été mon cas.

Il serait intéressant de mener à bien une étude sur l'apprentissage de *quedar* en employant une méthode basée sur des interventions explicites sur la polysémie et des ateliers où les étudiants pourraient effectuer des tâches de traduction à l'aide des machines. Ceci pourrait permettre de favoriser l'apprentissage des langues à l'aide de la TA (Yamada, 2019), tout en identifiant les différents sens de *quedar* (et les structures grammaticales les plus fréquentes), ainsi qu'en développant de bonnes pratiques sur l'usage des systèmes de TA en ligne.

## Bibliographie

Abdul-Rauf, S., Yvon, F. (2020). *Document level contexts for neural machine translation*. [Research Report] 003, LIMSI-CNRS. 72 p. (hal-03687190v2)

Allerton D. J. (2001). Ambiguity and polysemy in thin verbs. Dans : *Recherches anglaises et nord-américaines*, N°34, 2001. RANAM N°34: Linguistic ambiguity. pp. 19-27.

Almela Pérez, R., Cantos, P., Sánchez, A., Sarmiento, R. & Almela, M. (2005). *Frecuencias del español: Diccionarios y estudios léxicos y morfológicos*. Madrid: Universitas.

Barbin, F. (2020). *La traduction automatique neuronale, un nouveau tournant ?* Palimpseste. Sciences, humanités, sociétés. Université Rennes 2, pp.51-53. halshs-03603588

Bourdais, A. & Guichon, N. (2020). *Représentations et usages du traducteur en ligne par les lycéens*. ALSIC - Apprentissage des Langues et Systèmes d'Information et de Communication, Association pour le Développement de l'Apprentissage des Langues par les Systèmes d'Information et de Communication - OpenEdition, 10.4000/alsic.4533. halshs-02967171

Bowker, L., & Ciro, J. B. (2019). *Machine Translation and Global Research*. Emerald Publishing Limited.

Bowker, L. (2020). *Machine translation literacy instruction for international business students and business English instructors*. Journal of Business & Finance Librarianship, DOI: 10.1080/08963568.2020.1794739

Castilho, S., Moorkens, J., Gaspari, F., Calixto, I., Tinsley, J., & Way, A. (2017). *Is neural machine translation the new state of the art?* The Prague Bulletin of Mathematical Linguistics, 108(1), 109.

Desclés, J., & Guentchéva, Z. (2018). *La polysémie verbale appréhendée par une sémantique cognitive et formelle*.

Esperança-Rodier, E., Brunet-Manquat, F. & Eady, S. (2019). *ACCOLÉ: A Collaborative Platform of Error Annotation for Aligned Corpora*. Translating and the computer 41



Farrus, M., Costa-Jussà, M. R., Marino, J. B., & Fonollosa, J. A. R. (2010). Linguistic-based evaluation criteria to identify statistical machine translation errors. Dans V. Hansen & F. Yvon (Eds.), *Proceedings of the 14th annual conference of the European Association for Machine Translation*. European Association for Machine Translation

Gómez Vicente, L. (2011). Quedar: entre cambio y permanencia. *Estudios Hispánicos*.

Gomez Vicente, L. (2013). L'étrange polysémie du verbe quedar, ou l'expression de la permanence et du changement : étude diachronique et synchronique. *CogniTextes*, Association française de linguistique cognitive, 10.

Gómez Vicente, L. (2018). Description, acquisition and teaching of polysemous verbs: The case of quedar. *International Review of Applied Linguistics in Language Teaching*. 57. 10.1515/iral-2018-2005.

González Martínez, A. G. & Maldonado Soto, R. (2017). Quedar: cambio de estado y locación en el español de México. *Forma y función*, 30(1), 9–30. <https://doi.org/10.15446/fyf.v30n1.62405>

Grass, T. (2022). L'erreur n'est pas humaine, *Traduire* [En ligne], 246. URL : <http://journals.openedition.org/traduire/2763> ; DOI : <https://doi.org/10.4000/traduire.2763>

Halverson, S. (2019). 'Default' translation: A construct for cognitive translation and interpreting studies. *Translation, Cognition & Behavior*. 2. 187-210. 10.1075/tcb.00023.hal.

Hellmich, A. (2021). Machine Translation in Foreign Language Writing: Student Use to Guide Pedagogical Practice, *Alsic*, Vol. 24, n° 1 | URL : <http://journals.openedition.org/alsic/5705> ; DOI : <https://doi.org/10.4000/alsic.5705>

Jacquet, G. (2003). *Polysémie verbale et construction syntaxique : étude sur le verbe jouer*. 11-14.

Jouitteau, Mélanie. (éd.). 2009-2023. ARBRES, wikigrammaire des dialectes du breton et centre de ressources pour son étude linguistique formelle, IKER, CNRS, <http://arbres.iker.cnrs.fr>. Licence Creative Commons BY-NC-SA.

Kasperė, R., Horbačauskienė, J., Motiejūnienė, J., Liubiniėnė, V., Patašienė, I., & Patašius, M. (2021). Towards Sustainable Use of Machine Translation: Usability and Perceived Quality from the End-User Perspective. *Sustainability*, 13(23), 13430. <https://doi.org/10.3390/su132313430>

Kilgarrieff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., Suchomel, V.. (2014). The Sketch Engine: ten years on. *Lexicography*, 1: 7-36.

Kim, Y., Thanh Tran, D., & Ney, H. (2019). When and Why is Document-level Context Useful in Neural Machine Translation? Dans *Proceedings of the Fourth Workshop on Discourse in Machine Translation (DiscoMT 2019)*, pages 24–34, Hong Kong, China. Association for Computational Linguistics.

Kleiber, G. (2005). Quand y a-t-il sens multiple ? Le critère référentiel en question. Dans O. Soutet (Dir.), *La Polysémie* (pp. 51-74). Paris : Presses de l'Université Paris-Sorbonne.

Koehn, P. (2017). *Neural Machine Translation*. arXiv preprint arXiv:1709.07809. URL : <https://arxiv.org/abs/1709.07809>

Koehn, P. (2018). The State of Neural Machine Translation (NMT). *Omniscien*. URL : <https://omniscien.com/state-neural-machine-translation-nmt/>

Kraif, O. & Diwersy, S. (2012). Le Lexicoscope : un outil pour l'étude de profils combinatoires et l'extraction de constructions lexico-syntaxiques. Dans *Actes de la conférence TALN 2012*, Grenoble, pp. 399-406.

Kraif, O. (2016). Le lexicoscope : un outil d'extraction des séquences phraséologiques basé sur des corpus arborés, *Cahiers de lexicologie*, n° 108, Phraséologie et linguistique appliquée, p. 91-106  
DOI : 10.15122/isbn.978-2-406-06281-3.p.0091

Larsonneur, C. (2021). Neural Machine Translation: From Commodity to Commons? Dans R. Desjardins, C. Larsonneur et P. Lacour (eds). *When Translation Goes Digital. Case Studies and Critical Reflections* (p. 257-280). Editions Palgrave Macmillan.

Ledesma, M. M., & Lencinas, R. (2018). Valores semánticos de cualquiera y sus equivalentes franceses: problemas de traducción. *Nueva ReCIT: Revista Del área De traductología*, (1). Consulté sur <https://revistas.unc.edu.ar/index.php/ReCIT/article/view/20147>

Loffler-Lauria, A.-M. (1994). La traduction automatique : son utilisation par le « grand public ». *Langages*, 116, p. 87-94. <http://www.jstor.org/stable/23906709>

Meenal, S. & Govindarajan, P. (2023). The Challenges of Using Machine Translation While Translating Polysemous Words. *Studies in Media and Communication*. 11. 63. 10.11114/smc.v11i2.5944.

Mihaescu, M., (2012). Modélisation des équivalences fonctionnelles. *Synergies Roumanie* n° 7 - 2012 pp. 217-227

Nardy, A. & Dugua, C. (2011). Le rôle de l'usage sur le développement des constructions nominales chez les enfants pré-lecteurs. *Travaux de linguistique*, 62, 129-148. <https://doi.org/10.3917/tl.062.0129>

Nassau, G., Molle, N. & Kalyaniwala, C. (2022). Usages et perceptions des outils de traduction automatique : une enquête auprès d'apprenants Lansad, *Alsic* | URL : <http://journals.openedition.org/alsic/6239>

Nurminen, M. & Papula, N. (2018). Gist MT Users: A Snapshot of the Use and Users of One Online MT Tool. Dans *The 21st Annual Conference of the European Association for Machine Translation*.

Pérez-Ortiz, J. A., Forcada, M. & Sánchez-Martínez, F. (2022). How neural machine translation works. Dans Dorothy Kenny (ed.), *Machine translation for everyone: Empowering users in the age of artificial intelligence*, 141–164. Berlin: Language Science Press. DOI: 10.5281/zenodo.6760020

Poibeau, T. (2017). *Machine Translation*. Cambridge : MIT Press.

Redmond, L. (2017). *Acquisition de la polysémie du verbe prendre par des apprenants du français L2*. [Thèse doctorale]

Santos, C. (2015). Acepções do verbo Quedar: diversidade de expressão dos processos de estado. *Revista Letras Raras*. 4. 8. 10.35572/rlr.v4i3.469.

Stahlberg, F. (2020). Neural Machine Translation: A Review. *Journal of Artificial Intelligence Research* 69: 343-418.

Talmy, L. (1988), Force Dynamics in Language and Cognition. *Cognitive Science*, 12: 49-100. [https://doi.org/10.1207/s15516709cog1201\\_2](https://doi.org/10.1207/s15516709cog1201_2)

Tezcan, A., Hoste, V. & Macken, L. (2018). SCATE Taxonomy and Corpus of Machine Translation Errors. Dans G. Corpas Pastor & I. Durán-Muñoz (eds.). *Trends in E-*

*Tools and Resources for Translators and Interpreters* (p. 219-248). Brill. doi: <https://doi.org/10.1163/9789004351790>

Tiedemann, J. & Scherrer, Y. (2017). Neural machine translation with extended context. Dans *Proceedings of the Third Workshop on Discourse in Machine Translation*, pages 82–92. Association for Computational Linguistics.

Vaguer, C. (2004). *Les constructions verbales “V dans GN” Approches syntaxique, lexicale et sémantique*. [Thèse doctorale]

Vanmassenhove, E., Shterionov, D., & Way, A. (2019). Lost in Translation: Loss and Decay of Linguistic Richness in Machine Translation. Dans *Proceedings of Machine Translation Summit XVII: Research Track*, pages 222–232, Dublin, Ireland. European Association for Machine Translation.

Vanmassenhove, E., Shterionov, D., & Gwilliam, M. (2021). Machine translationese: Effects of algorithmic bias on linguistic complexity in machine translation. Dans *EACL 2021 - 16th Conference of the European Chapter of the Association for Computational Linguistics, Proceedings of the Conference* (pp. 2203-2213). Association for Computational Linguistics (ACL). <https://doi.org/10.18653/v1/2021.eacl-main.188>

Vaupot, S. (2020). Analyse des erreurs de traduction automatique pour la combinaison de langues slovène-français et perspectives pour une formation en post-édition. *Matices en Lenguas Extranjeras*, 14(2), 83–110. <https://doi.org/10.15446/male.v14n2.91299>

Vial, Loïc. (2020). *Modèles neuronaux joints de désambiguïsation lexicale et de traduction automatique*.

Victorri B., Fuchs C. (1996), *La polysémie, construction dynamique du sens*. Paris, Hermès.

Vilar, D., Xu, J., D’Haro, L.F. & Ney, H. (2006). Error Analysis of Statistical Machine Translation Output. Dans *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 06)*, 697-702.

Voita, E., Sennrich, R., & Titov, I. (2019). When a good translation is wrong in context: Context-aware machine translation improves on deixis, ellipsis, and lexical cohesion. Dans *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1198–1212.

Volkart, L. (2018). *Traduction automatique statistique vs. neuronale : Comparaison de MTH et DeepL à La Poste Suisse*. (Mémoire de Master 2, Université de Genève).

Yamada, M. (2019). Language learners and non-professional translators as users. Dans *The Routledge Handbook of Translation and Technology* (p.183-199). Routledge.

Yvon, F. (2019). Les deux voies de la traduction automatique. *Hermès, La Revue*, 85, 62-68. <https://doi-org.sidnomade-1.grenet.fr/10.3917/herm.085.0062>

## Annexes

### Annexe 1 : Occurrences de quedar (ES-ha–FR)

	ES-ha	FR	Traduction DeepL
1	Sí, es que me queda muy lejos, mamá.	<b>C'est loin</b> d'ici, maman.	Oui, c'est juste que <b>c'est</b> très loin, maman.
2	Me has hecho quedar como una psicópata.	Tu m' <b>as fait passer</b> pour une psychopathe.	Vous m' <b>avez fait passer</b> pour un psychopathe.
3	Te hubieras quedado hasta el final.	Tu aurais dû <b>rester</b> .	Vous seriez <b>restés</b> jusqu'à la fin.
4	Ni sabes dónde queda , ni sabes coger el metro.	Tu ne sais pas où <b>c'est</b> , ni comment faire.	Vous ne savez pas où <b>c'est</b> , et vous ne savez pas comment prendre le métro.
5	Lo mismo que un árbol que, en tiempo de otoño, se queda sin hojas.	Tout comme un arbre qui à l'automne <b>se retrouve</b> sans feuilles.	Tout comme un arbre qui, en automne, <b>perd</b> ses feuilles.
6	Quédate sentada, ándale.	<b>Rassieds-toi</b> , allez.	<b>Restez</b> assis, allez.
7	Me quedó todo ya.	J'ai tout <b>compris</b> .	Tout cela est <b>resté</b> gravé dans ma mémoire.
8	Porque nos la vamos a quedar .	On va la <b>garder</b> .	<b>Parce que nous allons le garder</b> .
9	Che, Fede, no me quiero quedar acá.	Fede, pas question de <b>rester</b> ici.	Fede, je ne veux pas <b>rester</b> ici.
10	¿Y me tengo que quedar acá?	Et je dois <b>rester</b> ici ?	Et je dois <b>rester</b> ici ?
11	¿Te vas a quedar a dormir?	Tu <b>restes</b> dormir ?	Vous <b>restez</b> ici ?
12	Como los indicadores no son buenos, lo más adecuado es que María se quede internada.	Ça ne s'annonce pas bien, mieux vaut que María <b>reste</b> hospitalisée.	Les indicateurs n'étant pas bons, il est préférable que Maria <b>reste</b> à l'hôpital.
13	A las diez habíamos quedado , ¿qué hora es?	On <b>avait</b> rendez-vous à 10 heures.	Nous devons <b>nous retrouver</b> à dix heures, quelle heure est-il ?
14	Quedate tranquila que yo no puedo hacer nada para acelerarlo.	<b>Rassurez-vous</b> , je ne peux pas accélérer les choses.	<b>Soyez</b> assuré que je ne peux rien faire pour l'accélérer.
15	Estamos viendo, por ejemplo, si se queda acá, si nos vamos a casa, pero acá te limpian todo, es como una luna de miel para nosotros.	On discute pour savoir si elle <b>reste</b> ici, si on rentre, mais ici, on est bichonnés, c'est comme une lune de miel. C'est ça.	Nous voyons, par exemple, s'il <b>reste</b> ici, si nous rentrons chez nous, mais ici ils nettoient tout pour vous, c'est comme une lune de miel pour nous.
16	Yo hace como siete meses tuve una operación bastante radical, una histerectomía y... en la cual me sacaron todo, me vaciaron completamente, quedé totalmente hueca.	Il y a sept mois, j'ai eu une opération assez radicale, une hystérectomie. On m'a tout retiré, on m'a vidée de fond en comble, je <b>suis</b> complètement <b>creuse</b> .	Il y a environ sept mois, j'ai subi une opération assez radicale, une hystérectomie et... on m'a tout enlevé, on m'a complètement vidée, <b>j'étais</b> totalement <b>vide</b> .
17	A la noche, con las velas encendidas, va a quedar bueno.	De nuit, avec les bougies, ce <b>sera</b> joli.	Le soir, avec les bougies allumées, ce <b>sera</b> bien.
18	Le doy mi tarjeta, vaya a hacer lo que tiene que hacer, yo me voy a quedar un rato.	Voici ma carte. Faites le nécessaire, je vous <b>attends</b> un peu.	Je vous donne ma carte, faites ce que vous avez à faire, je vais <b>rester</b> un peu.
19	Charlie se queda con el secarropas.	Charlie <b>héríte</b> du sèche-linge.	Charlie <b>obtient</b> le séchoir.

20	Tomy se queda sin padre y sin madre a la vez porque podés ir preso.	Si tu es arrêté, Tomy <b>n'aura</b> plus de parents.	Tomy <b>devient</b> orphelin de père et de mère à la fois parce qu'il peut aller en prison.
21	Tú no te estás quedando aquí, ¿no?	Vous ne <b>séjournes</b> pas ici.	Vous ne <b>restez</b> pas ici, n'est-ce pas ?
22	La lavandería queda al lado de la ferretería.	La laverie <b>est</b> à côté de la quincaillerie.	La blanchisserie <b>se trouve</b> à côté de la quincaillerie.
23	Oye, ¿y cuánto tiempo te quedabas ?	<b>Tu restes</b> jusqu'à quand ?	Et combien de temps êtes-vous <b>restés</b> ?
24	Espero que quede claro que lo que pasó anoche queda en una noche.	Ce qui s'est passé hier soir, <b>c'était</b> juste une fois.	J'espère qu'il est clair que ce qui s'est passé la nuit dernière <b>reste</b> une nuit.
25	Me quedé en el cemento, no entendí el cemento pulido.	Tu <b>m'as perdue</b> à "bloc de béton".	Je suis <b>resté</b> dans le ciment, je n'ai pas compris le ciment poli.
26	Y este es el de tu hostel, que se quedó sin tocar, ¿verdad?	Là, c'est le terrain de l'auberge qui n'a pas <b>été</b> touché ?	Et voici celui de votre auberge, qui est <b>resté</b> intact, n'est-ce pas ?
27	Queda bien, ¿a que sí?	Pas mal, non ?	<b>Ça a l'air</b> bien, n'est-ce pas ?
28	No te puedes quedar callada.	Putain, tu sais pas <b>te taire</b> , des fois ?	Vous ne pouvez pas <b>vous taire</b> .
29	Quedamos en que no te ibas a enamorar de mí.	On a <b>dit</b> que tu ne tomberais pas amoureuse de moi.	Nous <b>étions</b> d'accord pour dire que tu ne tomberais pas amoureuse de moi.
30	Entonces, me quedo .	<b>Alors je reste.</b>	Je reste donc.
31	Tú estuviste viajando 20 años antes de quedarte aquí.	Tu as voyagé pendant 20 ans avant de <b>t'installer</b> ici.	Vous avez voyagé pendant 20 ans avant de <b>rester</b> ici.
32	Si esto no se soluciona ya, perderé mi inversión y tú quedarás como rey de las remodelaciones.	Réglons ça ou je perdrai mon investissement et tu <b>resteras</b> le roi de la rénovation.	Si cela n'est pas réglé maintenant, je perdrai mon investissement et vous <b>serez</b> le roi des rénovations.
33	Queda Puno.	Alors, Puno.	Puno <b>reste</b> .
34	Ya quedó .	<b>C'était</b> bien ?	Déjà parti .
35	¡ Quedó !	Oui !	C'est à gauche !
36	Pero ¿ quedó pajasa?	<b>C'était</b> genial ?	Mais <b>reste-t-il</b> de la paille ?
37	Si la encuentro, me la quedo .	Si je la trouve, je la <b>garde</b> .	<b>Si je le trouve, je le garde.</b>
38	Y me voy a quedar hasta que llegue la grúa.	<b>J'attends</b> la dépanneuse avec toi, même si ça prend dix heures.	Et je vais <b>rester</b> jusqu'à l'arrivée de la dépanneuse.
39	Queda otro paciente todavía.	Tu ne peux pas, il te <b>reste</b> un patient.	Il <b>reste</b> encore un patient.
40	¿ Queda ?	Ah bon ?	Est-il à gauche ?
41	¿No quedó algo que haya hecho papá?	Papa n'a pas <b>cuisiné</b> ?	Il ne <b>restait</b> rien de ce que papa avait fait ?
42	Quedamos seleccionados para el torneo de cocineros amateurs.	On <b>est</b> pris pour le concours de cuisine.	Nous avons <b>été</b> sélectionnés pour le tournoi des chefs amateurs.
43	Mañana, le pedimos a Nelly que se quede y salimos.	Nelly n'aura qu'à <b>garder</b> les enfants.	Demain, nous demandons à Nelly de <b>rester</b> et nous sortons.
44	¡Ese color, cómo te queda , por favor!	Cette couleur <b>te va</b> à ravir.	Cette couleur, comment elle <b>vous va</b> , s'il vous plaît !
45	Porque, no sé, tal vez no limpié bien y quedó alguna suciedad.	Je ne suis pas sûr d'avoir bien <b>fait</b> la vaisselle.	Parce que, je ne sais pas, peut-être que je n'ai pas bien nettoyé et que des saletés sont <b>restées</b> .
46	Además me queda de paso.	<b>C'était</b> sur ma route.	D'ailleurs, je <b>suis</b> en route.

47	Que yo sepa, tu consultorio queda en la otra punta de la ciudad.	Ton cabinet dentaire <b>est</b> à l'autre bout de la ville.	Pour autant que je sache, votre bureau <b>se trouve</b> de l'autre côté de la ville.
48	¿Te gastás unas millas, yo me quedo con los chicos?	Dépense tes points, je <b>garde</b> les enfants.	Tu fais quelques kilomètres, je <b>reste</b> avec les enfants ?
49	Yo te digo, Belu, si ustedes no empiezan a probar cosas nuevas, si no se renuevan, la van a quedar ahí.	Si vous pimentez pas un peu votre vie de couple, ça <b>durera</b> pas.	Je vous le dis, Belu, si vous ne commencez pas à essayer de nouvelles choses, si vous ne vous renouvelez pas, vous <b>resterez</b> là.
50	Me quedé pensando en lo que hablamos el otro día.	Quel enthousiasme pour une simple infusion... Ça m'a <b>rappelé</b> notre discussion sur le voyage en Inde.	Je n'ai <b>cessé</b> de penser à ce dont nous avons parlé l'autre jour.
51	... quedamos seleccionados en un torneo.	On <b>est sélectionnés</b> pour un concours.	... nous avons <b>été</b> sélectionnés lors d'un tournoi.
52	Te vas a quedar acá cuidando a los nenes y impiando los pisos.	Tu <b>resteras</b> ici pour garder les enfants et nettoyer le sol.	Tu vas <b>rester</b> ici à t'occuper des enfants et à nettoyer les sols.
53	Elena, quedate tranquila que yo hablo con Mica y esto se soluciona.	Elena, je vais en parler à Mica. On va régler le problème.	Elena, <b>ne t'inquiète pas</b> , je vais parler à Mica et nous allons régler ça.
54	Quedate acá un rato.	<b>Reste</b> avec nous.	<b>Restez</b> ici un moment.
55	De verdad, quedate .	<b>Reste</b> avec nous.	Vraiment, <b>restez</b> .
56	Bueno, me quedo .	OK.	Eh bien, je <b>reste</b> .
57	¡Habíamos quedado en hacer una tregua!	On a dit qu'on <b>faisait</b> une trêve.	Nous avons <b>convenu</b> d'une trêve !
58	¡Esto no va a quedar así!	On va pas en <b>rester</b> là.	Ça ne va pas <b>rester</b> comme ça !
59	¡No va a quedar así, Belén!	On va pas en <b>rester</b> là !	Ça ne va pas <b>rester</b> comme ça, Belén !
60	¡Me voy a quedar ciego!	Je vais <b>devenir</b> aveugle.	Je vais <b>devenir</b> aveugle !
61	Pero no hay que hacerlo, eso quiero que quede claro.	Mais ne faites jamais ça, que ce <b>soit</b> bien clair.	Mais il n'est pas nécessaire de le faire, je <b>tiens</b> à le préciser.
62	Disfruten de la noche, sean felices y quédense por acá, porque en un ratito empieza.	Profitez, soyez heureux, mais ne <b>vous éloignez</b> pas trop. Ça va bientôt commencer.	Profitez de la soirée, soyez heureux et <b>restez</b> dans les parages, car tout commence dans peu de temps.
63	Quedé hecho mierda con el spray.	La bombe lacrymo m'a bien <b>fait</b> douiller.	Je me suis <b>fait avoir</b> avec le spray.
64	Gala te puso unas gotitas y te quedaste dormido.	Gala a mis quelque chose dans ton verre et tu t'es <b>endormi</b> .	Gala t'a donné quelques gouttes et tu t'es <b>endormi</b> .
65	¡ Quedó !	Fini !	C'est à gauche !
66	Con ese saludo ya quedaste como un goma, papá.	Tu <b>as</b> l'air bête à faire coucou.	Avec ce salut, tu as déjà <b>fait</b> une gomme, papa.
67	Me quedé con tu deseo.	J'ai <b>pris</b> ton vœu.	J'ai <b>respecté</b> votre souhait.
68	Pues yo me voy a quedar con todos tus deseos para siempre.	Moi, je <b>prendrai</b> tous les tiens, pour toujours.	Eh bien, je vais <b>garder</b> tous vos souhaits pour toujours.
69	Yo me voy a quedar con todos tus deseos para siempre.	Moi, je <b>prendrai</b> tous les tiens, pour toujours.	Je <b>garderai</b> tous vos souhaits pour toujours.
70	Con el calor que hace aquí, no sabes cómo te va a quedar .	Avec la chaleur qu'il fait ici, ce <b>sera</b> parfait.	Avec la chaleur qui règne ici, on ne sait pas à quoi cela va <b>ressembler</b> .
71	Se va a quedar con todos mis deseos.	Elle me <b>prendra</b> tout.	Il va <b>réaliser</b> tous mes souhaits.



<p>72        ¿Te quedas a desayunar? Ahora solo me quedan aquellos recuerdos y, en mi corazón, una voz que dice: "Te quiero".</p> <p>73</p> <p>74        Se quedó a dormir en un hotel.</p> <p>75        ¿Con cuál te quedarás ? Y me dijiste que te ibas a quedar con todos mis deseos a partir de ese momento, para siempre.</p> <p>76        ¿Y dónde se quedaron escondidos todos esos sueños que teníamos desde niños?</p> <p>77</p>	<p style="text-align: center;">Tu <b>restes</b> déjeuner ?</p> <p>Il ne me <b>reste</b> que mes souvenirs et dans mon cœur une voix qui dit : "Je t'aime".</p> <p style="text-align: center;">Elle a <b>dormi</b> à l'hôtel.</p>	<p style="text-align: center;">Vous <b>restez</b> pour le petit-déjeuner ?</p> <p>Aujourd'hui, il ne me <b>reste</b> que ces souvenirs et, dans mon cœur, une voix qui dit : "Je t'aime".</p> <p style="text-align: center;">Il a <b>passé</b> la nuit à l'hôtel.</p>
	<p style="text-align: center;">Lequel tu <b>prends</b> ?</p> <p>Tu m'as dit que tu <b>prendrais</b> tous mes vœux à partir de ce moment-là et pour toujours.</p>	<p style="text-align: center;">Lequel <b>choisirez</b>-vous ?</p> <p>Et tu m'as dit que tu allais <b>prendre</b> tous mes souhaits à partir de ce moment, pour toujours.</p>
	<p style="text-align: center;">Où <b>sont</b> cachés tous nos rêves de gamins ?</p>	<p style="text-align: center;">Et où <b>sont</b> cachés tous ces rêves que nous avons depuis notre enfance ?</p>

## Annexe 2 : Occurrences de quedar (ES-es–FR)

ES-es	Sorties de TA	FR
1 Te quedas hasta el 23, ¿no?	Vous <b>restez</b> jusqu'au 23, n'est-ce pas ?	Tu <b>restes</b> jusqu'au 23 ?
2 Este se queda .	Celui-ci <b>reste</b> .	Il <b>reste</b> .
3 Ya nos queda poquito.	Il ne nous <b>reste</b> qu'un petit peu.	<b>C'est</b> pour <b>bientôt</b> .
4 Se ha quedado para hacer serrín. ¿Cuánto tiempo queda , Ramón?	Il est <b>resté</b> pour faire de la sciure.	<b>C'est bon</b> pour la sciure. Combien de temps encore, Ramón ?
5 Es una pena que no te quedes más días porque te encantaría.	Combien de temps <b>reste-t-il</b> , Ramón ?	Domage que tu ne <b>restes</b> pas. Tu adorerais.
6 Al final, me quedo unos días.	C'est dommage que tu ne <b>restes</b> pas plus de jours parce que tu aimerais bien. Finalement, je vais <b>rester</b> quelques jours.	Je <b>reste</b> quelques jours.
7 ¿Te vas a quedar ?	Tu vas <b>rester</b> ?	Tu <b>restes</b> ? Enfin, c'est bien que tu <b>restes</b> .
8 Quiero decir, que qué bien que te quedes .	Je veux dire, c'est bien que tu <b>restes</b> .	
9 Nos quedamos atrapados en la nieve.	Nous sommes <b>restés</b> coincés dans la neige.	La voiture <b>a été piégée</b> dans la neige.
10 Te agradezco el esfuerzo, pero no me voy a quedar .	J'apprécie l'effort, mais je ne <b>reste</b> pas.	Merci d'essayer, mais je ne <b>reste</b> pas.
11 Vale, me quedo .	D'accord, je <b>reste</b> .	Très bien. Je <b>reste</b> .
12 No sé, yo creo que, al quedarme sola con mi madre tan pequeña, convertí a la gente del pueblo en mi familia.	Je ne sais pas, je pense qu'en <b>restant</b> seule avec ma mère si petite, j'ai fait des gens du village ma famille.	Je crois que m' <b>être retrouvée</b> seule avec ma mère si jeune, m'a fait voir le village comme une famille.
13 Por eso me quedé .	C'est pour ça que je suis <b>restée</b> .	Je <b>reste</b> pour ça.
14 Me quedo .	Je <b>reste</b> .	Je <b>reste</b> .
15 ¿No es mejor ir al hospital y quedarnos tranquilos?	N'est-il pas préférable d'aller à l'hôpital et de <b>rester</b> en paix ?	On ne <b>serait</b> pas plus <b>tranquilles</b> à l'hôpital ?
16 Si se quiere quedar , que se quede.	S'il veut rester, qu'il <b>reste</b> .	Elle veut <b>rester</b> .
17 Dense prisa o se quedarán sin ver el gran espectáculo de Navidad.	Dépêchez-vous ou vous <b>manquerez</b> le grand spectacle de Noël.	Dépêchez-vous où vous <b>raterez</b> le grand spectacle de Noël.
18 Quedan breves instantes para que comience la representación.	Il ne <b>reste</b> plus que quelques instants avant le début de la représentation.	Le spectacle <b>commencera bientôt</b> .
19 No queda sitio ya, señor.	Il n'y <b>a</b> plus de place, monsieur.	Je veux dire... Il n'y <b>a</b> plus de place, mon bon monsieur.
20 ¿Te has quedado ?	Vous <b>restez</b> ?	<b>Compris</b> ?
21 No queda ni una sola habitación libre en toda la isla.	Il n'y <b>a</b> plus une seule chambre sur toute l'île.	Il n'y <b>a</b> aucune autre chambre disponible sur toute l'île.
22 Alguna quedará .	Il y en <b>aura</b> .	Rien du tout ?
23 Todo para que nuestros novios queden contentos.	Tout ça pour que les mariés <b>soient</b> heureux.	Tout <b>est</b> fait pour <b>votre plaisir</b> .
24 Pues yo ya me quedo pagada.	Je <b>suis</b> déjà payé.	J'en <b>suis ravie</b> !
25 Se te ha quedado ahí un pegote.	Il vous <b>reste</b> un peu d'argent.	Tu <b>avais</b> un pâté sur le visage.
26 La Ronería queda al otro lado del barranco.	La Ronería <b>est</b> de l'autre côté du ravin.	La rumerie <b>se situe</b> de l'autre côté, près d'un champ de cannes à sucre.

28	Mamá, que quede claro. No pienso quedarme encerrada en esta habitación contigo.	Maman, <b>soyons</b> clairs. Je ne <b>reste</b> pas enfermée dans cette chambre avec toi.	Maman, <b>soyons</b> clairs. Je n'ai aucune envie de <b>rester enfermée</b> ici avec toi.
29	Yo me quedo aquí.	Je <b>reste</b> ici.	Je <b>reste</b> ici.
30	¿Tú te quedas o te vas?	Tu <b>restes</b> ou tu pars ?	Et toi, tu t'en vas ?
31	¡Si lo sé, me quedo en prisión!	Si je sais, je <b>reste</b> en prison !	Je préfère <b>rester</b> en prison !
32	Decías que si me caía me quedaría paralítico.	Tu as dit que si je tombais, je <b>serais</b> paralysée.	Tu disais que je <b>finirais paralysé</b> .
33	¿Y si me quedo ?	Et si je <b>reste</b> ?	Et si je <b>restais</b> ?
34	Pues prefiero quedarme aquí contigo, ¿no?	Je préfère <b>rester</b> ici avec toi, n'est-ce pas ?	Je préfère <b>rester</b> ici avec toi.
35	Está quedando fenomenal.	Ça <b>s'annonce</b> bien.	Ça va être <b>superbe</b> .
36	Doscientos_cincuenta, me quedo el anillo.	Deux cent cinquante, je <b>garde</b> la bague.	250 €, mais je <b>garde</b> la bague.
37	Igual no te quedas embarazada porque no quieres, amor.	Peut-être que tu ne <b>tombes</b> pas <b>enceinte</b> parce que tu ne le veux pas, mon amour.	Tu <b>tombes</b> pas <b>enceinte</b> , car tu le veux pas.
38	Se ha quedado inconsciente.	Vous avez <b>perdu</b> connaissance.	Il a <b>perdu</b> connaissance.
39	Te queda la última vuelta.	Il vous <b>reste</b> le dernier tour.	<b>Encore un tour</b> .
40	Ángel, te queda la última.	Angel, il te <b>reste</b> le dernier tour.	Angel, <b>continue</b> .
41	¿Por qué no te quedas un rato?	Pourquoi ne pas <b>rester</b> un peu ?	<b>Reste</b> donc un peu.
42	No queda nadie vivo que lo haya visto ni nadie que haya descubierto dónde se oculta ese ser tan fascinante.	<b>Personne</b> ne l'a vu et personne n'a découvert où se cache cet être fascinant.	<b>Aucune personne vivante</b> ne l'a vu ou ne sait où se cache cet être extraordinaire jusqu'à ce jour.
43	¿Te gustaría quedar conmigo esta tarde?	Voulez-vous <b>me rencontrer</b> cet après-midi ?	Ça te dirait qu'on <b>se voie</b> plus tard ?
44	Quedamos en la puerta del insti esta noche.	Je vous <b>retroverai</b> devant l'école ce soir.	<b>Rendez-vous</b> ce soir devant le lycée.
45	Mira, que te quede clarito.	Ecoutez, <b>soyons</b> clairs.	<b>Écoute-moi</b> bien.
46	Con el que habías quedado en HollyBlood.	<b>Celle que vous avez rencontrée</b> à HollyBlood.	Celui que tu voulais <b>rencontrer</b> au ciné.
47	Sara vino al hospital, me dijo que había quedado con el nuevo y que tal vez no la vería más.	Sara est venue à l'hôpital, m'a dit qu'elle <b>avait rencontré</b> le nouveau et que je ne la reverrais peut-être plus.	Sara m'a dit qu'elle allait <b>rencontrer</b> le nouveau et que je ne la reverrais plus.
48	Quedan encantadas y algo pachuchas, como una mononucleosis.	Ils <b>sont</b> tous ravis et un peu malades, comme la mononucléose.	Elles <b>sont contentes</b> et un peu K.-O., comme après une mono.
49	Mierda, no queda agua bendita.	Merde, il n'y <b>a</b> plus d'eau bénite.	Merde. Il n'y <b>a</b> plus d'eau bénite.
50	Pero no, quiero que quedemos, que quiero que me cuenten.	Mais non, je veux <b>te rencontrer</b> , je veux que tu me racontes.	Non, je veux <b>te voir</b> , j'ai envie de savoir !
51	Me quedo con su cuarto yo.	Je <b>prendrai</b> votre chambre.	Je <b>prends</b> sa chambre.
52	Justo hemos quedado aquí para irnos de compras.	On <b>se retrouve</b> juste ici pour faire du shopping.	On va <b>se retrouver</b> ici pour faire du shopping.
53	¿Debería quedarme ?	Je <b>reste</b> ?	Je ferais mieux de <b>rester</b> .
54	Me viene bien también quedarme solo un tiempo.	C'est aussi bon pour moi de <b>rester</b> seul un moment.	Ça me fera du bien <b>d'être</b> un peu <b>seul</b> .
55			

56	Ya sé que mil trenes se me pasan y que no me quedan manos para contar.	Je sais que mille trains me passent sous le nez et que je <b>n'ai</b> plus de mains pour compter.	Des centaines de trains passent. Je ne <b>peux plus les compter</b> .
57	Y tú y tu marido lo tenéis incluso mejor, que os quedáis solos, ya me entiendes.	Et toi et ton mari, c'est encore mieux, vous <b>êtes</b> seuls, tu vois ce que je veux dire.	Et toi, avec ton mari, tu vas <b>être bien</b> , si tu vois ce que je veux dire.
58	Ya que estoy aquí, me quedo .	Tant que je suis là, je <b>reste</b> .	Non, j'y suis, j'y <b>reste</b> .
59	Pablo, ¿cómo me queda ?	Pablo, de quoi j' <b>ai l'air</b> ?	Pablo, de quoi j' <b>ai l'air</b> ?
60	¿Le queda claro?	<b>C'est</b> clair ?	c' <b>est clair</b> ?
61	Queda una chica todavía.	Il <b>reste</b> une fille.	Il <b>reste</b> une candidate.
62	No puedo quedarme en casa quitando pañales y haciéndole la cena a tu hijo.	Je ne peux pas <b>rester</b> à la maison à enlever des couches et à faire à manger pour ton fils.	Je ne peux pas <b>vivre pour</b> changer des couches et faire à manger à ton fils !
63	Sara, todavía queda libre el puesto de profesora ¡y puede ser tuyo!	Sara, il y <b>a</b> encore un poste d'enseignant à pourvoir et il peut être le tien !	Le poste <b>est toujours vacant</b> . Il peut être à toi.
64	Te queda bien, pero a mí me gusta más largo.	Elle te <b>va bien</b> , mais je l'aime plus longue.	Ça te <b>va bien</b> , mais je te préfère les cheveux longs.

### **Annexe 3 : Les avis des étudiants de L3 concernant notre intervention sur le verbe quedar**

---

<b>Qu'avez-vous pensé de l'intervention ? Qu'avez-vous retenu ?</b>	
<b>1</b>	J'ai bien aimé la présentation et je l'ai trouvé intéressante et différente de nos cours traditionnels de traduction.
<b>2</b>	Très bonne présentation, bien expliquée et facile à comprendre. Les nuances de sens sont très pertinentes pour la traduction :)
<b>3</b>	Très bon powerpoint et très facile à comprendre avec des exemples simples et parlant
<b>4</b>	Je pense que cette intervention serait plus utile si elle était faite en L1.
<b>5</b>	Je considère qu'elle peut être très utile pour des francophones et même des hispanophones puisqu'elle peut nous faire voir la diversité de sens d'un seul mot de façon qu'on ne se trompe pas au moment de le traduire.
<b>6</b>	Les différentes traductions possibles.
<b>7</b>	m'a permis de mieux comprendre l'utilisation de Quedar.
<b>8</b>	J'ai trouvé le contenu très intéressant, et cette intervention a éclairci certains points de langues sur lesquels je n'étais pas sûre de moi.
<b>9</b>	Intéressante
<b>10</b>	L'intervention était utile et bien orchestrée cependant elle aurait probablement été appréciée dès la L1.
<b>11</b>	Cela est très intéressant de recevoir quelqu'un de l'extérieur pour nous éclairer sur des points clefs de la traduction qui peuvent paraître simples au premier abord. J'ai retenu que le mot quedar ne se traduisait pas toujours de la même façon selon le contexte et la construction de la phrase.
<b>12</b>	Plutôt bien. J'ai trouvé important car ça nous a permis de revoir les différentes matières d'employer le verbe quedar.
<b>13</b>	les différentes traductions du verbe quedar. avant je le comparais à permanecer
<b>14</b>	C'était très clair et ludique. La comparaison entre les verbes estar/haber et quedar étaient bien faites pour illustrer les différences de sens. Les exercices de traductions étaient agréables à faire. J'ai adoré cette intervention !
<b>15</b>	C'était sympathique
<b>16</b>	Très intéressant, permet de nous faire un rappel sur comment traduire "quedar"
<b>17</b>	bonne intervention, j'ai retenu beaucoup de choses
<b>18</b>	Très bien merci ! Je connaissais les sens de « quedar(se) » mais il était intéressant de voir les compléments.
<b>19</b>	Très intéressant
<b>20</b>	Déjà la traduction est difficile, car en traduisant certaines phrases on peut parfois perdre le sens principale. Je pense qu'il existe aussi d'autres verbes comme Quedar est que c'était hyper intéressant cette intervention malheureusement ce travail on nous l'apprend généralement pas trop en classe que ce soit au collège, lycée etc.... C'est plus un travail que l'on doit apprendre de nous même avec les séries, livres voyages etc.. , en tout cas moi ça a été mon cas.

---

## **Sigles et abréviations utilisés**

TA :	Traduction automatique
TAN :	Traduction automatique neuronale
TS :	Texte source
TC :	Texte cible
TH :	Traduction humaine
MT literacy	Machine translation literacy

## Table des illustrations

Tableau 1 : Films faisant partie du corpus étudié .....	27
Tableau 2 : Données statistiques du sous-corpus Es-es–FR .....	28
Tableau 3 : Données statistiques du sous-corpus Es-ha–FR.....	28
Tableau 4 : Statistiques concernant les occurrences de <i>quedar</i> dans le sous-corpus ES-es–FR .....	35
Tableau 5 : Les structures grammaticales les plus fréquentes (sous-corpus ES-es–FR) .....	36
Tableau 6 : Statistiques concernant les occurrences de <i>quedar</i> dans le sous-corpus ES-ha–FR .....	38
Tableau 7 : Les structures grammaticales les plus fréquentes (sous-corpus ES-ha–FR).....	38
Tableau 8 : Les équivalents de <i>quedar</i> dans les traductions humaines (ES-es–FR).....	41
Tableau 9 : Les équivalents de <i>quedar</i> dans les traductions humaines (ES-ha–FR) .....	43
Tableau 10 : Les équivalents de <i>quedar</i> dans les sorties de TA (sous-corpus ES-es–FR) .....	44
Tableau 11 : Les équivalents de <i>quedar</i> dans les sorties de TA (sous-corpus ES-ha–FR) .....	45
Tableau 12 : Construction avec <i>quedar</i> + syntagme prépositionnel .....	48
Tableau 13 : Construction avec <i>quedar</i> + adverbe .....	49
Tableau 14 : Construction avec <i>quedarse</i> .....	50
Tableau 15 : Construction avec <i>quedar/quedarse</i> + adjectif .....	51
Tableau 16 : <i>Quedar/quedarse</i> + syntagme nominal.....	52
Tableau 17 : Constructions possibles pour le domaine du rendez-vous .....	53
Tableau 18 : Construction pour le domaine de la possession .....	54
Tableau 19 : Constructions pour le domaine du point de vue.....	55
Tableau 20 : Construction avec <i>quedar</i> + syntagme prépositionnel .....	56
Tableau 21 : Construction avec <i>quedar</i> + adjectif.....	58
Tableau 22 : Constructions avec <i>quedar</i> pour parler de la localisation (ES-ha–FR).....	59
Tableau 23 : <i>Quedar/quedarse</i> + syntagme prépositionnel .....	60
Tableau 24 : Constructions possibles pour le domaine de l'existence .....	60
Tableau 25 : Construction pour le domaine du point de vue .....	61
Tableau 26 : Constructions pour le domaine du rendez-vous.....	62
Tableau 27 : Erreurs de traduction observées .....	66
Tableau 28 : Sorties de TA en contexte .....	70
Figure 1 : Plongement lexical (Koehn, 2017 : 36).....	17
Figure 2 : Typologie d'erreurs du projet ExTra (tableau préparé par nos soins) .....	22
Figure 3 : Les domaines de <i>quedar</i> dans le sous-corpus ES-es–FR .....	48
Figure 4 : Les domaines de <i>quedar</i> dans le sous-corpus ES-ha–FR .....	56

# Table des matières

<b>Introduction</b> .....	<b>5</b>
<b>PARTIE 1 : ÉTAT DE L'ART</b> .....	<b>7</b>
<b>Chapitre 1 – <i>Quedar</i> : un verbe polysémique par excellence</b> .....	<b>8</b>
1. Qu'est-ce que la polysémie ? .....	8
La polysémie verbale.....	8
2. <i>Quedar</i> : entre permanence et changement.....	9
3. <i>Quedar</i> : un verbe d'état à caractère résultatif à la suite d'une dynamique des forces.....	12
<b>Chapitre 2. La traduction automatique neuronale : un outil robuste, mais pas infallible</b> .....	<b>15</b>
1. La traduction automatique (TA) de nos jours : le paradigme neuronal .....	15
2. La TAN au quotidien .....	18
3. Comment aider les utilisateurs non spécialistes à utiliser les outils de TA de manière raisonnée ?	22
<b>PARTIE 2 - METHODOLOGIE</b> .....	<b>24</b>
<b>Chapitre 3. Le projet LOLYTA</b> .....	<b>25</b>
1. Description du projet LOLYTA .....	25
2. Objectif du stage et tâches effectuées .....	26
3. Présentation du corpus et difficultés rencontrées .....	27
3.2.1. Première difficulté : origine des sous-titres .....	29
3.2.2. Deuxième difficulté : droits d'auteurs et notre manière de récupérer les sous-titres.....	29
3.2.3. Troisième difficulté : alignement et nettoyage du corpus.....	31
<b>Conclusion</b> .....	<b>32</b>
<b>PARTIE 3 - RESULTATS ET DISCUSSION</b> .....	<b>33</b>
<b>Chapitre 4. Présentation des résultats du projet LOLYTA</b> .....	<b>34</b>
1. L'exploitation du corpus de sous-titres.....	34
2. Les occurrences de <i>quedar</i> dans notre corpus.....	35
3. Les équivalents de <i>quedar</i> : traductions humaines vs sorties de TA .....	40
3.1.1. Les équivalents de <i>quedar</i> dans les traductions humaines du sous-corpus ES-es-FR.....	41
3.1.2. Les équivalents de <i>quedar</i> dans les traductions humaines du sous-corpus ES-ha-FR .....	42
3.2.1. Les équivalents de <i>quedar</i> dans les sorties de TA pour le sous-corpus ES-es-FR .....	43
3.2.2. Les équivalents de <i>quedar</i> dans les sorties de TA pour le sous-corpus ES-ha-FR.....	44
<b>Conclusion</b> .....	<b>45</b>
<b>Chapitre 5. Analyse des données et réflexion personnelle sur les limites de l'étude</b> .....	<b>47</b>
1. Les différents domaines de <i>quedar</i> présents dans notre corpus .....	47
2. <i>Quedar</i> et DeepL : des traductions quasi parfaites ?.....	64
3. Les limites de notre étude : des pistes de réflexion ? .....	72
<b>Conclusion</b> .....	<b>74</b>
<b>Conclusion</b> .....	<b>75</b>
<b>Bibliographie</b> .....	<b>77</b>
<b>Annexes</b> .....	<b>83</b>
<b>Annexe 1 : Occurrences de <i>quedar</i> (ES-ha-FR)</b> .....	<b>83</b>
<b>Annexe 2 : Occurrences de <i>quedar</i> (ES-es-FR)</b> .....	<b>87</b>
<b>Annexe 3 : Les avis des étudiants de L3 concernant notre intervention sur le verbe <i>quedar</i></b> .....	<b>90</b>
<b>Sigles et abréviations utilisés</b> .....	<b>91</b>
<b>Table des illustrations</b> .....	<b>92</b>
<b>Table des matières</b> .....	<b>93</b>



**MOTS-CLÉS** : traduction automatique, polysémie verbale, *quedar*

## **RÉSUMÉ**

Les moteurs de traduction automatique sont utilisés au quotidien par des utilisateurs non spécialistes. Même si ces moteurs représentent une aide précieuse, Toutefois, ils ne sont pas infallibles et peuvent produire des traductions trompeuses aux niveaux sémantique et syntaxique, en particulier pour les mots polysémiques.

Cette étude se penche sur la polysémie du verbe *quedar*, en examinant si les moteurs de TA parviennent à traduire correctement les constructions employant ce verbe, hors contexte et en contexte. La quantité du texte à l'entrée (*input*) peut exercer une influence positive ou négative à la sortie (*output*). De plus, il faut faire particulièrement attention aux constructions verbales qui peuvent constituer une piste pour déchiffrer le sens dans l'énoncé : le verbe et la construction l'accompagnant forment une même unité polysémique qu'il est impossible de dissocier.

Les enseignants jouent un rôle clé : d'une part, ils peuvent sensibiliser les utilisateurs dès l'école aux problèmes liés à la TA, et d'autre part, ils devraient intégrer l'apprentissage de mots polysémiques dès les premières années d'études.

**KEYWORDS**: machine translation, polysemy in verbs, *quedar*

## **ABSTRACT**

Machine translation engines are used on a daily basis by non-specialist users. Although these engines are a valuable aid, they are not foolproof and can produce misleading translations at semantic and syntactic levels, particularly for polysemous words.

This study looks at the polysemy of the verb *quedar*, examining whether MT engines are able to correctly translate constructions using this verb, both out-of-context and in-context. The amount of text input can have a positive or negative influence on the output. In addition, particular attention must be paid to verbal constructions, which can provide a clue to deciphering the meaning within a statement: the verb and the accompanying construction form a single polysemous unit that cannot be dissociated.

Teachers can play a key role: on the one hand, they can make users aware of the problems associated with MT from school onwards, and on the other, they should integrate the learning of polysemous words right from the early years of schooling.