Cahiers de Traduction

Volume 31 Issue 02/November 2025 PP 481- 499

دفاتر الترجمة
Cahiers de Traduction

ISSN 1111- 4606
EISSN 2602- 6023

« Traduction Automatique Neuronale versus Traduction Humaine :DEEPL dans la traduction des collocations en statistique du français vers l'arabe »

« Neural Machine Translation vs. Human Translation: Evaluating DeepL in Rendering French–Arabic Statistical Collocations »

GUIDOUM Nadéra[1], SI BACHIR Zina

Université d'Alger 2 nadira.guidoum@univ-alger2.dz
Université d'Alger 2 zina.sibachir@univ-alger2.dz

Reçu le:28/08/2025 *Accepté le:25/09/2025* *Publié le 20/11/2025*

Résumé :

Le présent article analyse de manière comparative la traduction automatique neuronale fournie par le logiciel DEEPL et la traduction humaine. L'étude porte sur un corpus de collocations spécialisées en statistique, traduites du français vers l'arabe. Ainsi, l'évaluation repose sur des critères de fidélité sémantique, de précision terminologique. Les résultats mettent en évidence la performance de DEEPL dans la production d'équivalent globalement cohérents. Cependant, certaines limites apparaissent dans le traitement des collocations complexes, où la traduction humaine se distingue par une meilleure adéquation contextuelle. Ce constat confirme la valeur ajoutée de l'expertise humaine pour accompagner ou corriger les traductions automatiques.En conclusion, l'étude souligne la complémentarité entre approches neuronales et traduction spécialisée dans le domaine scientifique.

Mots clés : Traduction Automatique Neuronale, Collocation, Typologie des Erreurs, Etude Sémantique, DeepL.

Abstract :

This article provides a comparative analysis of neural machine translation provided by DEEPL software and human translation. The study focuses on a corpus of collocations specialised in statistics, translated from French into Arabic. The evaluation is based on criteria of semantic fidelity and terminological accuracy. The results highlight DEEPL's performance in producing generally consistent equivalents. Nevertheless, some limits appear in the processing of complex collocations, where human translation stands out for its better contextual adequacy. This study confirms the added value of human expertise in supporting or correcting machine translations.

In conclusion, the study highlights the complementarity between neural approaches and specialised translation in the scientific field.

Keywords : Neural Machine Translation, Collocation, Error Typology, Semantic Study, DeepL

[1] **L'auteur correspondant : GUIDOUM Nadéra**

Cahiers de Traduction
Volume 32 Issue 01/November 2025 pp 481-499
ISSN 1111- 4606 EISSN 2602-6023

1. Introduction :

La traduction automatique (TA) a connu des avancées majeures au cours des dernières années, passant des approches basées sur des règles à des modèles statistiques, pour finalement aboutir à l'ère des systèmes de traduction automatique neuronale TAN (NMT, pour *Neural Machine Translation*). Fondés sur des réseaux de neurones profonds, ils ont marqué une avancée majeure en produisant des traductions plus harmonieuses et précises sur le plan contextuel.

Contrairement aux méthodes traditionnelles qui décomposent le processus de traduction en étapes distinctes (telles que l'analyse syntaxique et la génération de phrases), la TAN utilise un modèle unique pour apprendre directement les correspondances entre les langues source et cible. Cette approche end-to-end permet de capturer des structures grammaticales complexes et des nuances linguistiques qui étaient difficiles à modéliser avec les techniques précédentes.

Ainsi, les Transformers, proposés par des chercheurs de Google en 2017, a été introduit précisément pour améliorer la traduction automatique. Les Transformers sont un ensemble d'architecture de réseau de neurones pour la traduction automatique neuronale, qui repose exclusivement sur des mécanismes d'attention, car ils adoptent une structure encodeur (dispositif matériel ou logiciel qui convertit une information d'un format ou d'un code vers un autre) –décodeur (reconvertit des données encodées vers un format interprétable par le système ou l'humain), ont joué un rôle clé dans cette évolution. Grâce à leur mécanisme d'attention, ces modèles peuvent se concentrer sur les parties pertinentes de la phrase source pour générer des traductions de haute qualité. Ces progrès ont non seulement amélioré la précision des traductions, mais ont également ouvert la voie à des applications dans des domaines spécialisés, tels que la traduction médicale, juridique et technique. (Vaswani, A. 2017:3).

Mattieu Guidère mettait en lumière le retour en force de la traduction automatique, après une période de résultats décevants et de critiques sur sa fiabilité. (2010 : 57).

Parmi les défis qui se présentent à la TAN, la traduction des collocations représente un enjeu majeur dans le domaine de la traduction.

Le terme « collocation » est traditionnellement utilisé pour désigner un phénomène linguistique dans lequel une unité lexicale présente une disposition à s'associer à d'autres éléments lexicaux. Il représente une relation lexicale caractérisée par la cooccurrence, qui relie des mots ayant plus ou moins une force associative. Par conséquent, on peut affirmer que le verbe « maigre » est exclusivement associé au nom « viande », alors que l'adjectif « fort » peut présenter des collocations telles que « pluie », « homme ».

La notion de « collocation » a été largement reconnue par les linguistes comme une contribution issue de la théorie firthienne, bien que sa référence initiale remonte à 1605 (Wouden 1997:7). En effet, Firth (1957) a été le premier chercheur à considérer la tendance typique d'un élément lexical à s'associer à d'autres lexèmes comme faisant partie intégrante de son interprétation sémantique. Au fil des époques de nombreux linguistes à l'instar de Halliday 1966, Sinclair 1966, ont approfondi et développé la notion de collocation. Des lexicologues et lexicographes,

Cahiers de Traduction
Volume 32 Issue 01/November 2025 pp 481-499
ISSN 1111- 4606 EISSN 2602-6023

dont Cruse, Seaton, Benson, Carter et McCarthy, entre autres, ont manifesté un intérêt marqué pour l'exploration des modèles de collocation.

Cependant, contrairement à des considérations essentiellement théoriques, les traducteurs rencontrent souvent des difficultés persistantes pour identifier les collocations appropriées (Newmark, 1988).

Dans la présente étude, nous nous proposons d'effectuer une analyse sémantique sur la traduction des collocations du français vers l'arabe visant à analyser, décrire et évaluer la manière dont les unités polylexicales cooccurrantes sont traduites du français vers l'arabe, en tenant compte des rapports de sens entre ces unités. La première partie sera consacrée à un bref rappel historique de l'évolution de la T.A jusqu'à l'apparition de la traduction neuronale au 21ème siècle. Dans la seconde partie, nous aborderons le phénomène collocatif, en nous axant sur les différents définitions et classifications. **Notre démarche** d'analyse consistera à prendre un échantillon constitué d'une série de termes extrait d'un corpus web, à savoir le Glossaire des Termes Statistiques édité par the International Statistical Institut, que nous allons traduire à l'aide d'un logiciel de TAN à savoir Deepl, et procéder à la comparaison des output du logiciel aux termes originaux proposés par l'ISI.

Suivant cette démarche, nous allons tenter de répondre à la question suivante : les logiciels de traduction neuronale sont – ils en mesure de proposer des traductions reconnues, ou bien même proposer une traduction fiable qui véhiculent les spécificités des termes en statistiques.

2. Traduction automatique : notions de base.

Le présent volet de notre étude vise à présenter la traduction automatique dans son ensemble et la traduction automatique neuronale de manière particulière. Tout d'abord, une définition de la notion de traduction automatique s'impose, avec un bref rappel historique. Nous clôturons ce volet par la TAN et une présentation du logiciel DeepL.

2.1 Définition :

Définie comme traduction effectuée par un logiciel conçu pour analyser un texte ou des termes dans une langue de départ et produire un texte dans une langue d'arrivée sans intervention humaine (DESLILE, 2013 : 687), la traduction automatique se présenterait comme étant la prise en charge totale par une machine du processus traductionnel. Ainsi, traducteur ou l'utilisateur n'aura qu'à insérer son texte en langue source dans un logiciel pour obtenir la traduction dans la langue cible. BOUILLON et CLAS, la définissent à leur tour la traduction automatique : « comme *l'application de l'informatique à la traduction des textes d'une langue naturelle de départ (ou langue source LS) dans une langue d'arrivée (ou langue cible LC). (Bouillon, Clas 1993 : 15)*

Cahiers de Traduction
Volume 32 Issue 01/November 2025 pp 481-499
ISSN 1111- 4606 EISSN 2602-6023

En addition aux précédentes définitions, nous retrouvons la schématisation proposée par Hutchins et Somers en 1992, ayant pour but l'explication du processus de la traduction automatique et l'intervention de la traduction humaine, grâce à l'illustration suivante :

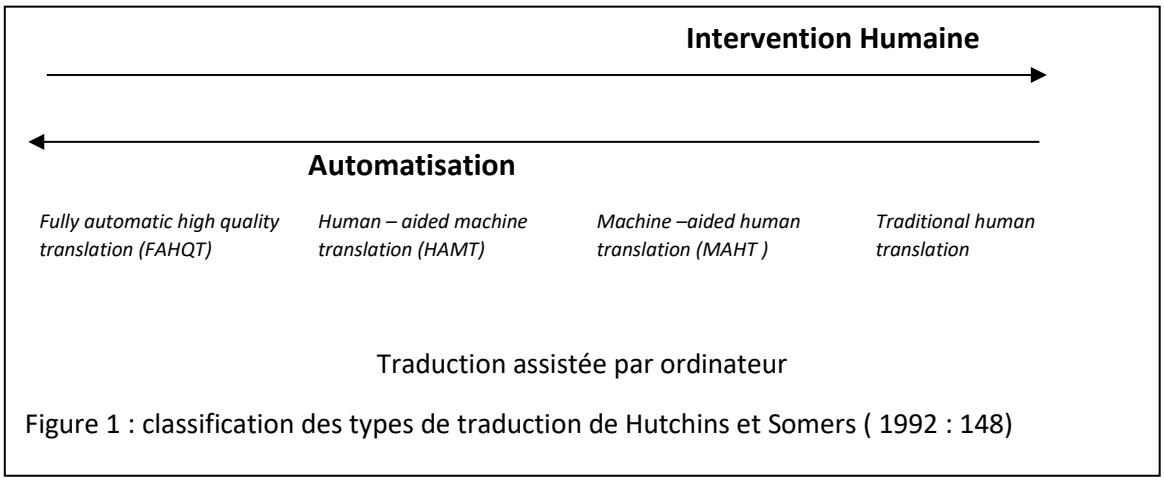


Figure 1 : classification des types de traduction de Hutchins et Somers (1992 : 148)

- *Fully automatic high quality translation (FAHQT) : Traduction entièrement automatique de haute qualité (TEAHQ).*
- *Human – aided machine translation (HAMT): Traduction automatique assistée par un humain (TAAH).*
- *Machine – aided human translation (MAHT) : Traduction humaine assistée par ordinateur (THAO).*
- *Traditional human translation = traduction humaine traditionnelle.*

Dans le schéma précédent, nous remarquons que la classification de la fonction que prennent à tour de rôle la machine et l'humain dans le processus de traduction est bien établie.

Le point de départ étant la traduction humaine traditionnelle que nous connaissons depuis l'antiquité, cette forme de traduction a évolué durant le 20ème siecle. De la traduction humaine au sens strict du terme, a Traduction humaine assistée par ordinateur (THAO) désigne une traduction effectuée par un être humain en se servant de l'outil informatique tel que les logiciels de bureautique et de traitement de texte. A ce stade, les logiciels de traduction tels que Systran ou Reverso ou Google traduction n'existait pas encore. Il ne s'agira pas d'introduire un texte de la langue source dans un logiciel et recevoir sa traduction dans la langue cible. Cependant, avec le progrès technologique qu'a connu le monde, la traduction automatique assistée par l'humain connaitra indéniablement un essor important. Ainsi, l'être humain n'effectuera pas une traduction à proprement parler, mais peut intervenir avant, et dans ce cas il s'agira de la pré- édition ; ou à la fin du processus ce qui est désigné par la post – édition. La post-édition (PE) désigne le processus de relecture, de révision et de correction des erreurs produites par la traduction automatique (TA). Elle s'inscrit dans une dynamique de collaboration entre l'humain et la machine. (BELAIDI, 2024 : 9). Le point d'arrivée du schéma précédent étant la traduction entièrement automatique de haute qualité désigne une traduction automatique réalisée dans sa totalité par un logiciel sans l'intervention de l'humain, et dont le résultat serait concluant.

Cahiers de Traduction
Volume 32 Issue 01/November 2025 pp 481-499
ISSN 1111- 4606 EISSN 2602-6023

Ce cheminement allant de la traduction humaine exclusive vers la traduction automatique exclusive, s'est opéré sur plusieurs décennies, voire un siècle. La section suivante nous permettra de faire un petit survol historique sur l'évolution de la traduction automatique.

2.2 Rappel historique de la T.A :

La traduction automatique a été l'un des plus vieux rêves du monde moderne. Au XXe siècle, ce rêve est devenu réalité, sous forme de programmes informatiques capables de traduire une grande variété de textes d'une langue naturelle à une autre. Mais, comme toujours, la réalité n'est pas parfaite. Il n'existe pas de "machines à traduire" qui, en appuyant sur quelques boutons, peuvent prendre en charge n'importe quel texte dans n'importe quelle langue et produire une traduction parfaite dans une autre langue sans intervention ou assistance humaine.

Les premières tentatives d'automatisation" du processus de traduction entre les langues naturelles remontent au milieu du XVIIe siècle, lorsqu'un moine allemand de la ville de Spire, Johannes Becher, a écrit une brochure sur son invention d'un méta-langage mathématique destiné à décrire le sens de phrases écrites dans n'importe quelle langue. Le méta-langage en question était composé de chaînes de nombres pour la signification des mots et accompagné de listes d'équations attribuant par exemple des mots allemands et latins à des expressions mathématiques de significations, de sorte que les phrases dans l'une de ces langues pouvaient être traduites dans l'autre langue de manière "mécanique" grâce à ces équations. (FREIGANG, 2001 : 1).

Cependant les premières machines à traduire à proprement dites firent leur apparition entre au cours de la première guerre mondiale au Etats-Unis et au Royaume-Unis ayant pour principal but le décryptage des communications ennemies. Rapidement, les scientifiques se sont rendus compte que ces machines pourraient servir dans le domaine de la traduction. Le mathématicien Warren Weaver a formulé cette idée pour la première fois en 1947. (Hutchins, 1986).

La publication par Weaver en 1949 d'un mémorandum sur l'idée d'utiliser les ordinateurs modernes pour traduire est considérée comme étant le point de départ de la traduction automatique, proposant d'utiliser les méthodes cryptographiques.

Le dit mémorandum parle de la possibilité d'aider des humains à communiquer entre eux grâce à des ordinateurs puissants travaillant avec calculateurs électroniques. Il permit la création de plusieurs centres de traduction automatique, particulièrement dans les universités, ayant pour vocation principale la traduction des textes scientifiques et techniques du russe vers l'anglais. (Léon, 2005, p.14).

La fin des années 50 vit l'apparition de quelques systèmes opérationnels dont la qualité des traductions obtenues laissait à désirer. En 1964, plusieurs agences qui subventionnaient les recherches dans ce domaine, en créant un comité dénommé ALPAC (Automatic Language Processing Advisory Commitee), qui avait pour mission principale d'étudier la possibilité d'arriver à une traduction automatique complète (Koehn, 2010).

Cahiers de Traduction
Volume 32 Issue 01/November 2025 pp 481-499
ISSN 1111- 4606 EISSN 2602-6023

Cependant, dans un rapport publié en 1966, l'ALPAC suggère la diminution des investissements dans le domaine de la traduction automatique en faveur de la recherche linguistique et l'amélioration de la traduction humaine (Koehn, 2010).

Ce n'est qu'à la fin des années 70, qu'un regain d'intérêt pour la traduction automatique fit son apparition, avec en autre l'utilisation par la Commission Européenne acquiert le système de TA SYSTRANT, ouvrant la voie à l'apparition de plusieurs systèmes de traduction automatique durant les années 80. (Bouillon et Clas, 1993).

Durant les années 2000, nous assistons à une accessibilité du grand public aux logiciels de TA grâce à leur gratuité (nous pouvons citer Google Translate et Reverso à titre d'exemple).

Bien que la TA ait connu un grand essor, elle ne fait toujours pas l'unanimité auprès des traducteurs et du grand public, car bien souvent les traducteurs ont recours à la post – édition pour parfaire le travail effectué par la machine.

Après l'ère dominante des systèmes statistiques à base de segments (phrase-based SMT), la traduction automatique neuronale (TAN) a introduit un modèle de bout en bout où un unique réseau encode puis décode la phrase source, supprimant la fragmentation en modules séparés. La percée majeure est venue avec le mécanisme d'attention, qui permet au décodeur de "se focaliser" dynamiquement sur différentes parties de la phrase source à chaque pas de génération, levant la contrainte du vecteur de contexte de longueur fixe et améliorant fortement la qualité sur les phrases longues. Des comparaisons systématiques ont démontré que, sur des ensembles de tests multi-domaines, les systèmes neuronaux surpassent généralement les systèmes phrase-based lorsqu'ils sont entraînés à grande échelle, tout en posant de nouveaux défis (domaine, rareté lexicale, longues phrases, collocation, expressions figées ….). (BAHDANAU, 2016).

2.3 Présentation de DeepL :

En 2017, l'entreprise allemande DeepL GmBH, précédemment Linguee GmBH, a lancé un service de traduction automatique. Au départ, DeepL ne prenait en charge que sept langues (français, anglais, allemand, espagnol, italien, néerlandais et polonais), avec 42 combinaisons. Le logiciel de traduction automatique TA se basait sur les réseaux de neurones qui fonctionnent sur un super ordinateur basé en Islande d'une puissance 50.000 fois plus que la puissance moyenne d'un ordinateur grand public. Dès son lancement Dès son lancement, les résultats des traductions proposés par DeepL surpassent ceux des autres systèmes tels que Google Translate, ou d'autre logiciel. L'évaluation humaine des traductions de DeepL montrent une nette préférence des traducteurs humains pour DeepL. A l'heure actuelle, DeepL propose 30 langues, avec des variations géolinguistiques pour l'anglais britannique, l'anglais américain, portugais, portugais brésilien et pas moins de 800 combinaisons linguistiques. Au début de son lancement, DeepL n'était accessible qu'en ligne, mais en mars 2018 la société a lancé DeepL Pro, un service qui permet d'intégrer DeepL à différentes applications et logiciels.

Cahiers de Traduction
Volume 32 Issue 01/November 2025 pp 481-499
ISSN 1111- 4606 EISSN 2602-6023

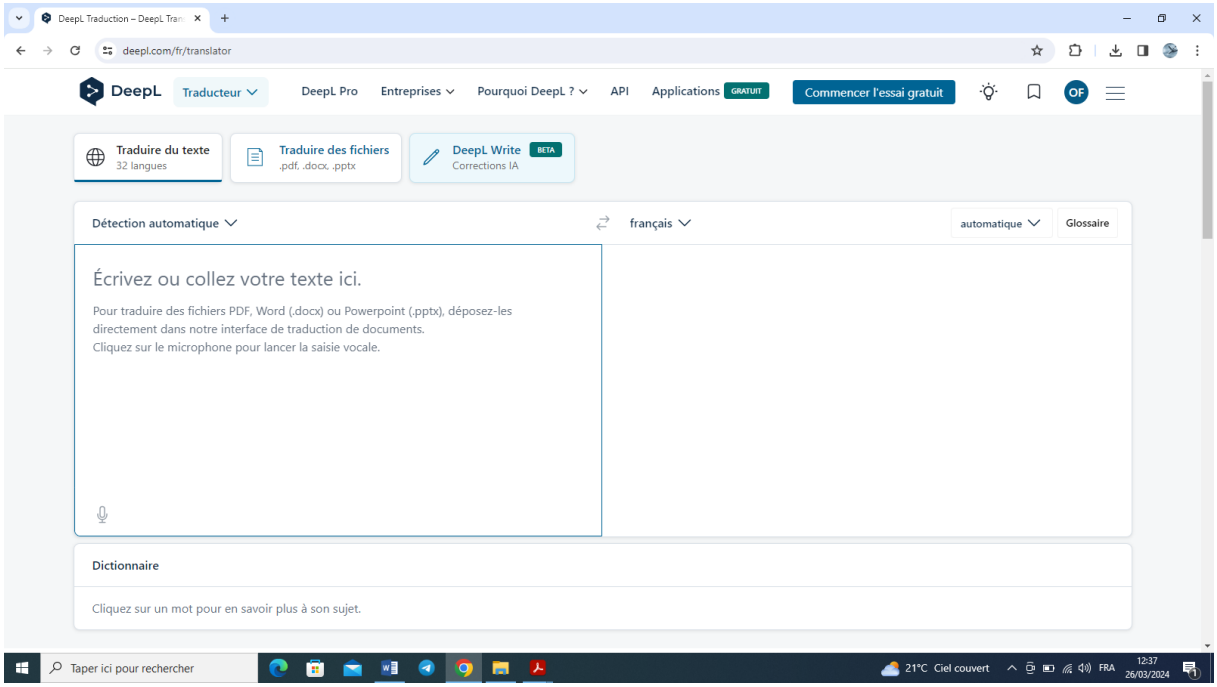


Figure 2 : Interface de DeepL dans sa version gratuite.

3 . Collocation : éléments définitoires et classification.

3.1 Eléments définitoires :

Une collocation est une expression récurrente familière (Gledhill, 2000 : 11). Pour bon nombre de linguistes, les collocations se rapportent à un éventail de syntagmes pluri-lexicaux couramment reconnus et dénommés : phrases toutes faites, clichés, expressions figées, formules, collocations libres ou liées, idiomes, syntagmes lexicaux, tournures, unités phraséologiques etc. La notion de collocation a fait l'objet de définitions diverses, lesquelles varient selon les objectifs particuliers de l'analyste. Les phrasiologues et les lexicographes – entre autres – étudient la manière dont les unités lexicales se combinent. Ainsi, les adjectifs strong et powerful présentent un sens proche mais une compatibilité différente avec certains noms : strong argument, powerful argument par opposition à strong tea / powerful tea, strong car / powerful car. (Gledhill, 2000 : 12)

Une définition globale et précise des collocations n'est pas aisée à établir. Ainsi bon nombre de linguistes ont proposé leur propre définition. Nous en citeront quelqu'un :

Les fondements historiques : Charles Bally et les « groupements usuels » :

Charles Bally pose les premières bases théoriques de l'étude des collocations dans son *Traité de stylistique française* de 1909. Il conceptualise ces phénomènes comme des « groupements usuels » ou « séries phraséologiques », qu'il définit en ces termes : « Il y a série ou groupement usuel lorsque les éléments du groupe conservent leur autonomie, tout en laissant voir une

Cahiers de Traduction
Volume 32 Issue 01/November 2025 pp 481-499
ISSN 1111- 4606 EISSN 2602-6023

affinité évidente qui les rapproche, de sorte que l'ensemble présente des contours arrêtés et donne l'impression du « déjà-vu » » (Netzlaff, 2004, p. 21).

L'approche contextualiste de John Rupert Firth :

John Rupert Firth développe une conception radicalement différente en ancrant les collocations dans une théorie contextuelle du sens. Pour Firth, les collocations constituent « la compagnie que les mots gardent » (*the company that words keep*) ou encore « des mots actuels en compagnie habituelle » (*actual words in habitual company*) (Benmahammed, 2019, p. 108).

Cette approche s'inscrit dans une vision fonctionnaliste du langage où « le sens ne peut être déterminé qu'en le rapportant à son contexte de situation » (Delhay, 1996, p. 25). Selon Firth, « avoir un sens, pour un énoncé, c'est fonctionner de manière appropriée dans un contexte situationnel » (Delhay, 1996, p. 24). Cette théorie contextuelle implique que les collocations ne peuvent être comprises indépendamment de leur environnement discursif.

Cependant, cette approche a fait l'objet de critiques notables. John Lyons (1966) reproche à la théorie contextuelle du sens de Firth de ne pas permettre « une analyse sémantique globale » et de négliger « la compatibilité sémantique et la relation du sens entre les mots » (Benmahammed, 2019, p. 108).

La systématisation de Franz Josef Hausmann :

Franz Josef Hausmann propose une définition qui est devenue référentielle dans la lexicographie didactique. Selon lui, une collocation est « une combinaison qui, dans l'usage, est préférée aux autres combinaisons possibles et devient norme ». Plus précisément, les collocations sont des « combinaisons usuelles préférées par la norme à d'autres combinaisons qui, d'après le système, seraient aussi envisageables » (Netzlaff, 2004, p. 10- 11).

La contribution majeure de Hausmann réside dans sa conceptualisation de la relation asymétrique entre les constituants. Il distingue la *base*, « à faible dépendance sémiotaxique », du *collocatif* ou *collocateur*, « à forte dépendance sémiotaxique » (Netzlaff, 2004, p. 12). Cette hiérarchisation permet de qualifier la collocation de « relation orientée » où « la base est sélectionnée uniquement pour son sémantisme, le collocatif est choisi en plus en dépendance de la base » (Netzlaff, 2004, p. 12).

Hausmann souligne également que les collocations sont « transparentes » mais « non prédictibles », ce qui les distingue à la fois des locutions (opaques) et des combinaisons libres (prédictibles) (Netzlaff, 2004, p. 12-13).

Igor Mel'čuk : les fonctions lexicales

Igor Mel'čuk, dans le cadre de la Théorie Sens-Texte, développe une approche formalisée en intégrant les collocations dans la catégorie des semi-phrasèmes. Sa définition met l'accent sur la contrainte lexicale : les collocations sont des « groupes lexicalement contraints, c'est-à-dire dont la combinatoire n'est déterminée ni par leur sémantisme ni par leurs propriétés syntaxiques » (Netzlaff, 2004, p. 19).

Cahiers de Traduction
Volume 32 Issue 01/November 2025 pp 481-499
ISSN 1111- 4606 EISSN 2602-6023

Mel'čuk définit formellement la collocation selon trois conditions simultanées :

1. Le sens global inclut le sens du constituant A.
2. A est sélectionné de façon régulière et non contrainte
3. B n'est pas sélectionné de façon régulière et non contrainte, mais en fonction de A et du sens à exprimer (Netzlaff, 2004, p. 24)

Benson, Benson et Ilson : distinction grammaticale et lexicale

Benson, Benson et Ilson (1986) proposent une typologie fonctionnelle en distinguant les collocations grammaticales des collocations lexicales. Selon eux, une collocation est « une combinaison de mots récurrente et semi-figée » (Bertrand, 1998, p. 15). Les collocations lexicales, qui nous intéressent particulièrement, sont « des combinaisons de lexèmes » qui sont « arbitraires et imprévisibles » (Bertrand, 1998, p. 16).

Anthony P. Cowie : gradation et restrictions

Anthony P. Cowie développe une approche graduelle en distinguant les « open collocations » (collocations ouvertes) des « restricted collocations » (collocations restreintes). Selon Cowie, les collocations se situent sur une échelle allant de celles « à très vaste champ collocationnel » à celles qui « ne se combinent qu'avec un seul mot » (Netzlaff, 2004, p. 25).

Cowie souligne également la difficulté d'établir des frontières nettes entre collocations et expressions idiomatiques.

3.2 Classification :

Le phénomène collocatif a donné lieu à plusieurs classifications, à l'instar des définitions et qui a suscité un large débat entre les linguistes. Aussi, nous proposons quelques éléments de classification qui nous semble des plus pertinentes pour notre travail.

Anthony Paul Cowie :

Cowie propose de diviser les collocations comme suit :

- Les collocations libres : ce sont les combinaisons qui comprennent des mots ayant une signification tellement générale qu'ils peuvent être associés à d'innombrables mots, comme le verbe « effectuer » qui est associé à « une recherche », « un prélèvement » ou « une enquête ».
- Les collocations restreintes : ce sont des combinaisons dans lesquelles le sens du mot « a » détermine le choix de l'élément « b », comme dans « rendre service ».

Franz Josef Hausmann : (Hausmann 1121, p. 1616)

Hausmann divise les collocations en six types, d'après la structure grammaticale suivante :

- Nom + adjectif : célibataire endurci
- Nom + verbe : la colère s'apaise
- Verbe + nom : retirer de l'argent
- Verbe + adverbe : il pleut à torrent
- Adverbe + adjectif : grièvement blessé
- Nom + préposition + nom : une bouffée d'oxygène

Sur la base de cette classification syntaxique, Newmark (Newmark, 1122,
p. 211, estime que les collocations verbales les plus courantes en anglais sont celles dans

Cahiers de Traduction
Volume 32 Issue 01/November 2025 pp 481-499
ISSN 1111- 4606 EISSN 2602-6023

lesquelles le deuxième élément est un nom (adjectif + nom ou nom + nom et verbe + nom). De même, Corpas (Corpas 1111, p. 11 71) adopte la classification de Hausmann (1985) et l'applique aux syntagmes verbaux en espagnol.

Morton Benson Benson et al. (Benson et al., 1121, p. 160 que les collocations verbales se divisent en deux catégories principales, selon Larivière (Larivière 1112, p. 172), qui sont les plus connues dans le domaine des collocations verbales courantes dans la langue courante, à savoir :

- Les collocations grammaticales : elles se composent d'un mot dominant (nom, adjectif ou verbe) suivi d'un élément qui est souvent une préposition, comme : s'abstenir de, attentif à. Hausmann (1121, p. 1611)
- Les collocations lexicales : elles se composent généralement de deux unités lexicales de même importance, telles que « déployer des efforts », lexicales d'importance égale, comme « déployer des efforts », « gros fumeur ».

Ce type de collocation est divisé, selon Benson et al en sept types similaires à ceux de Hausmann, à l'exception du type « verbe + nom » qui se subdivise en fonction du sens du verbe en deux types : verbe (au sens de création et d'activation) + nom, verbe (au sens d'élimination et d'annulation) + nom.

4. Typologie des erreurs dans la TAN :

Dans le but de détecter les erreurs dans un processus de TAN, il est important de disposer d'une traduction de référence afin de comparer les résultats des output de la TAN avec la traduction humaine de référence.

Nous estimons qu'un système de traduction n'est en aucun cas sans erreurs, et présente certaines ambiguïtés. Ainsi, nous proposons de reproduire le schéma de classification présenté par Valar et al, 2006 : 697 – 698. Dans un premier temps ; les erreurs sont divisées en cinq grandes catégories, à savoir :

1 / Les mots manquants,

2 / Mots incorrects,

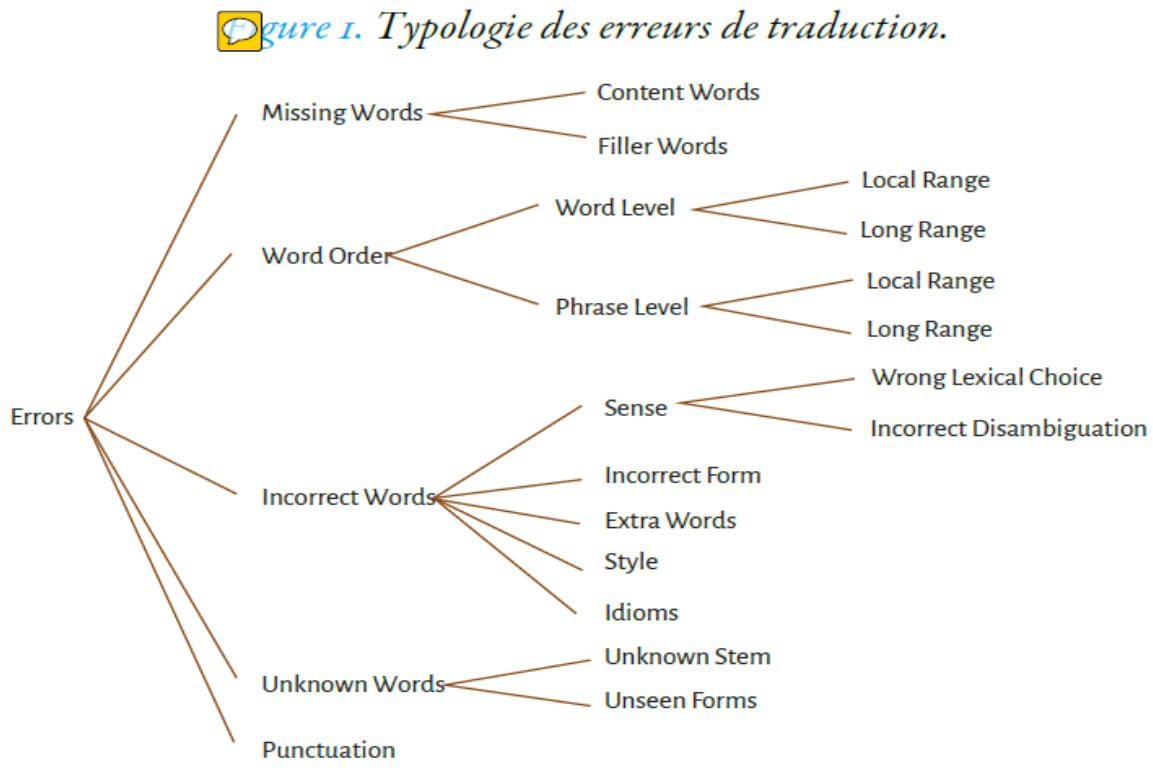
3 / Ordre des mots,

4 / Mots inconnus,

5 / Ponctuation.

Cahiers de Traduction
Volume 32 Issue 01/November 2025 pp 481-499
ISSN 1111- 4606 EISSN 2602-6023

Figure 1 : Typologie des erreurs de traduction automatique



Source : récupérée de Vilar et al. (2006, p. 699).

1 / Les mots manquants : Une erreur de « mot manquant » se produit lorsqu'un mot manque dans la phrase générée. Nous pouvons distinguer deux types d'erreurs : lorsque le mot manquant est essentiel pour exprimer le sens de la phrase, et lorsque le mot manquant est uniquement nécessaire pour former une phrase grammaticalement correcte, mais que le sens est préservé.

2 / L'ordre des mots : pour cette catégorie, nous distinguons une réorganisation basée sur des mots ou des phrases. Dans le cadre de notre travail nous nous intéresseront sur les mots car il s'agit de collocation, mais bien que la phrase reste correcte sur le plan grammatical, cependant nous remarquons une déperdition sur le plan sémantique.

3 / Les mots incorrects : nous faisons face à cette catégorie lorsque le système est dans l'incapacité de trouver un équivalent adéquat, et procède ainsi à une traduction littérale à savoir mot à mot.

4 / Les mots inconnus : Les mots inconnus constituent aussi une source d'erreurs. On peut ici distinguer les mots ou racines totalement inconnus des formes inhabituelles liées à des racines déjà connues. Une sous-catégorie de ce type d'erreur revêt une importance particulière lors de la traduction du français vers l'arabe. Dans la plupart des langues européennes, particulièrement celles partageant le même alphabet, il suffit de recopier les noms propres inconnus dans la

Cahiers de Traduction
Volume 32 Issue 01/November 2025 pp 481-499
ISSN 1111- 4606 EISSN 2602-6023

phrase traduite sans modification. En revanche, les caractères arabes ne peuvent pas être directement retranscrits en français correctement ; ils nécessitent une adaptation, parfois basée sur la prononciation. Ainsi, dans le cas de cette combinaison linguistique, il est également nécessaire de distinguer les personnes, lieux, organisations et autres noms propres inconnus.

5 / Les erreurs de ponctuation : certaines erreurs de ponctuation peuvent survenir, mais compte tenu la qualité de la traduction automatique neuronale, elles ne représentent que des perturbations mineures dans le cas de textes long. Dans le cadre de notre travail, qui est axé principalement sur le terme composé extrait d'un glossaire cette catégorie d'erreur ne sera pas prise en considération.

5. Méthodologie et analyse :

Notre choix s'est porté sur une liste de termes que nous avons extraits du glossaire des termes statistiques édité par l'International Statistical Institute, disponible sur son site officiel. Le nombre de termes présentant un phénomène collocatif dépassant les 3000 termes. Notre échantillon de termes correspond à la typologie des erreurs. Une traduction de ces termes par le logiciel Deepl, et sa comparaison avec ce qui est proposé dans le dit glossaire, va nous permettre de répondre au questionnement de savoir si la traduction exclusivement automatique des termes composés est-elle possible ?

Premier exemple : coefficient de corrélation d'échantillon – mot manquant

Cette analyse compare deux traductions :

Traduction ISI : معامل ارتباط العينة (complète)

Traduction DeepL : معامل الترابط (incomplète)

La traduction DeepL présente une omission significative en ne traduisant que partiellement le terme original. Le logiciel a produit "معامل الترابط" (coefficient de corrélation) mais a complètement omis la partie "d'échantillon" (العينة en arabe).

Le dictionnaire encyclopédique des Statistiques de Yadolah Dodge définie « le coefficient de corrélation » comme une mesure de l'intensité de la relation linéaire entre deux variables aléatoires. (YADOLAH, 2007 : 76).

Il existe plusieurs types

de coefficient de corrélation : C.C simple, C.C de rangs de KENDALL, C.C multiple, C.C d'échantillon.

DeepL a omis la partie "d'échantillon" (العينة) dans sa traduction, se limitant à "coefficient de corrélation" (معامل الترابط). Il s'agit ici d'une omission partielle laquelle un composant sémantique important est manquant.

Cette omission a un impact, car la traduction devient générique, ainsi tout type de coefficient de corrélation n'a pas de spécification alors qu'il peut se baser sur un échantillon uniquement au lieu de prendre en charge toutes les données à analyser.

Cahiers de Traduction
Volume 32 Issue 01/November 2025 pp 481-499
ISSN 1111- 4606 EISSN 2602-6023

En statistique, la distinction entre toutes les données et un échantillon est cruciale. Cette perte d'un élément du terme pourrait induire des erreurs d'interprétation dans un contexte d'analyse de données.

Second exemple : fonction de décision aléatoire – mot incorrect

Traduction ISI : دالة القرار العشوائية
Traduction DeepL : وظيفة القرار العشوائي
Le concept de « fonction » en arabe : دالة vs وظيفة
La distinction entre **دالة** et **وظيفة**constitue le cœur de cette problématique traductionnelle.
دالة (fonction)

- Étymologiquement dérivé de la racine د-ل-ل (d-l-l) signifiant « indiquer, guider, montrer le chemin »
- En mathématiques, ce terme correspond exactement au concept occidental de « fonction » comme relation univoque entre ensembles, et utilisé systématiquement dans la littérature mathématique arabe moderne pour désigner les fonctions au sens technique, ce qui implique une correspondance précise et indiscutable entre langue source et langue cible

وظيفة : (fonction)

- Étymologiquement de la racine و-ظ-ف (w-ẓ-f) signifiant assigner une tâche, ayant pour premier de travail, emploi, devoir, responsabilité.
- Connotation plus générale de « rôle » ou « fonction sociale » plutôt que mathématique

En théorie statistique de la décision, une « fonction de décision aléatoire » (randomized decision function) désigne une règle qui associe des probabilités à différentes décisions déterministes. Cette notion technique, développée par Abraham Wald, constitue un élément fondamental de l'analyse décisionnelle moderne.(MORLAT,1964 : 4)

Le terme anglais « function » dans ce contexte ne se réfère pas à un rôle social mais bien à une relation mathématique formelle entre l'espace des observations et l'espace des décisions. Cette précision conceptuelle justifie l'usage du terme **دالة** plutôt que **وظيفة**.

L'Institut international de statistique, fondé en 1885, joue un rôle central dans la standardisation de la terminologie statistique mondiale. Ses glossaires multilingues, notamment en arabe, visent à harmoniser l'usage terminologique pour faciliter la communication scientifique internationale.

La proposition ISI **دالة القرار العشوائية** s'aligne avec la terminologie mathématique déjà établie. De plus, la cohérence terminologique avec les autres concepts statistiques utilisant **دالة est établie; comme suit :**

Fonction D'autocorrélation	**دالة الارتباط الذاتي**

Cahiers de Traduction
Volume 32 Issue 01/November 2025 pp 481-499
ISSN 1111- 4606 EISSN 2602-6023

Fonction D'autocovariance	دالة التغاير الذاتي
Fonction De Configuration (Calcul Des Cumulants)	دالة النمط
Fonction De Coût	دالة الكلفة
Fonction De Décision Admissible	دالة قرار مقبولة
Fonction De Décision Aléatoire	دالة القرار العشوائية

L'emploi du terme : **وظيفة** entraîne une certaine ambiguïté conceptuelle, entre le concept mathématique précis et une notion plus générale, générant ainsi perte de la rigueur scientifique nécessaire à la compréhension de la langue de spécialité.

L'analyse de cette erreur traductionnelle relève de l'importance terminologique dans la transmission des savoirs scientifiques. Le choix entre **دالة** et **وظيفة**n'est pas simplement une question de préférence linguistique mais une précision conceptuelle.

Aussi, la recommandation de l'ISI de traduire par **دالة القرار العشوائية** s'impose comme la traduction correcte, respectant à la fois les conventions mathématiques arabes et les exigences de précision scientifique.

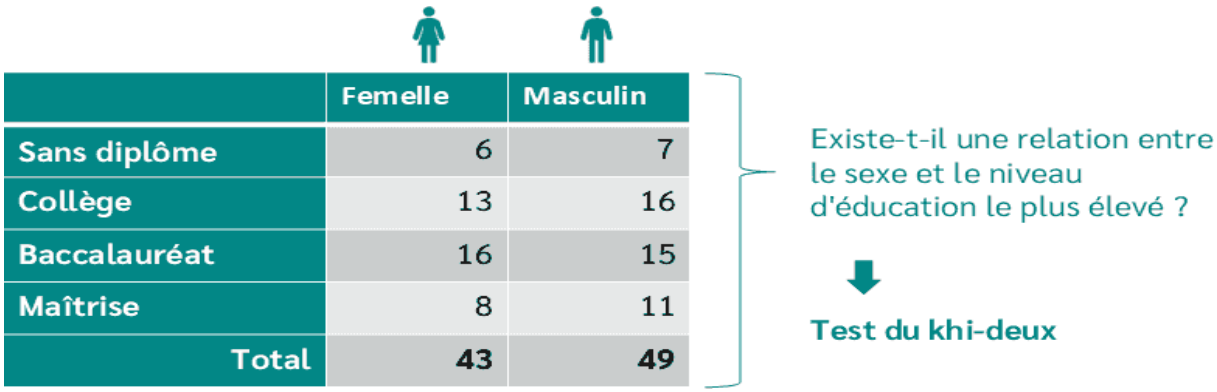
Troisième exemple : Chi deux test ou le test de khi-deux

Définition:

Traduction de Deepl : اختبار تشي اثنين

Le test du khi-deux est un test d'hypothèse utilisé pour les variables catégorielles avec une échelle de mesure nominale ou ordinale. Le test du khi-deux vérifie si les fréquences observées dans l'échantillon diffèrent significativement des fréquences attendues. Les fréquences observées sont donc comparées aux fréquences attendues et leurs écarts sont examinés.

Cahiers de Traduction
Volume 32 Issue 01/November 2025 pp 481-499
ISSN 1111- 4606 EISSN 2602-6023



Supposons que nous voulions étudier s'il existe un lien entre le sexe et le niveau d'éducation le plus élevé. Pour ce faire, nous créons un questionnaire dans lequel les participants cochent leur sexe et leur niveau d'études le plus élevé. (https://datatab.fr/tutorial/chi-square-test) (exemple à titre explicatif).

Analyse de la traduction du terme "Chi deux test" par DeepL (arabe : اختبار تشي اثنين)

Nature de l'erreur

La traduction automatique du terme statistique « Chi deux test » (ou « test du Chi² ») en arabe par DeepL donne « اختبار تشي اثنين », soit « test Chi deux ». Cette traduction est littérale et ne respecte pas l'usage terminologique reconnu dans les écrits scientifiques arabes, où l'on utilise les expressions « اختبار كاي مربع » ou « اختبار X^2 » pour désigner le test du Chi². « تشي اثنين » est une translittération partielle et l'ajout de « اثنين » (deux) ne correspond pas à la convention linguistique scientifique arabe.

D'après la typologie des erreurs en traduction automatique exposée par Valar (Vilar) et al., 2006, p.697-698, cette erreur correspond à la catégorie **de mot inconnu**

Il peut arriver que le système rencontre un mot qu'il ne connaît pas. Dans ce cas, l'instruction est de conserver tel quel tout terme absent de son dictionnaire. Ainsi, certains mots restent non traduits dans la version brute. C'est alors au post-éditeur de les traduire. Parfois, il suffit simplement de remplacer le mot source par son équivalent dans la langue cible. Mais, bien souvent, ce type d'oubli ou d'erreur provoque des ajustements plus complexes au niveau de la morphologie et de la syntaxe, car l'absence de reconnaissance d'un terme perturbe fréquemment les relations et les attributions à différents niveaux de la phrase.

Dans ce cas précis, **DeepL ne reconnaît pas "Chi deux" comme un terme technique ou spécialisé**. Au lieu de restituer le mot arabe spécialisé, l'outil propose une translittération hybride (traduction partielle) qui aboutir à une **incohérence terminologique**, rendant le terme vague voire incompréhensible pour le lecteur cible.

Le système traite "Chi" comme un élément inconnu et le laisse tel quel ou le translittère, ce qui est typique du comportement décrit dans la typologie de Valar et al.

Cahiers de Traduction
Volume 32 Issue 01/November 2025 pp 481-499
ISSN 1111- 4606 EISSN 2602-6023

Problème de dépendance morpho-syntaxique : En ajoutant « اثنين », DeepL transfère la structure du français sans adaptation, faussant l'attribution du sens et la dépendance syntaxique dans la langue cible.

Cette erreur altère la clarté du message et l'acceptabilité terminologique scientifique. Elle nécessite une post-édition où un traducteur humain reconnaîtrait et corrigerait le terme.

Le cas de « Chi deux test → اختبار تشي اثنين » illustre parfaitement le problème du vocabulaire spécialisé inconnu du système de traduction automatique, car l'absence d'un terme dans le syst-me entraine soit la non-traduction, soit une translittération inadaptée, ce qui engendre une charge supplémentaire sur le post-éditeur pour rétablir l'usage correct et conventionnel.

Quatrième exemple : niveau de qualité acceptable – ordre des mots

Traduction ISI : (للمواصفات) مستوى القبول للجودة AQL – Acceptable Quality Level

Traduction DeepL : مستوى الجودة المقبول

L'erreur commise par DeepL concerne l'ordre des mots. DeepL traduit littéralement « niveau de qualité acceptable » par « مستوى الجودة المقبول » (« niveau de la qualité acceptable »), alors que la traduction conventionnelle, notamment dans les écrits normatifs (ISO, contrôle qualité), privilégie une traduction du type « مستوى القبول للجودة » (« niveau d'acceptabilité de la qualité »). La solution standard met l'accent sur l'acceptabilité, c'est-à-dire la part de la qualité jugée tolérable ou conforme à un seuil défini.

Cette modification de l'ordre des mots change le sens : la version DeepL focalise sur la qualité acceptée, alors que la version ISI exprime un seuil ou un critère d'acceptabilité. En contexte industriel et statistique, l'exactitude terminologique est essentielle, et l'ordre des mots véhicule la nuance du concept : il ne s'agit pas de toute « qualité acceptable », mais d'un « niveau d'acceptabilité de la qualité » ou « niveau de tolérance de qualité ».

La traduction proposée par DeepL peut induire une compréhension erronée, en parlant de la qualité qui serait « acceptable » en général, alors que l'expression vise un seuil chiffré et réglementé (AQL – Acceptable Quality Level), crucial dans le contrôle qualité industriel.

Non-respect des usages professionnels et normatifs : La traduction ISI reprend une formulation normalisée et reconnue sur le plan international. Celle proposée par DeepL, bien qu'intelligible, manque de précision terminologique.

D'après la typologie de Vilar et al. (2006 : 697–698), cette erreur relève de la catégorie c) Erreur de réorganisation de la phrase (ordre des mots). Le système de traduction automatique n'a pas réorganisé les constituants de la phrase source de façon appropriée pour la langue cible :

« Les erreurs d'ordre des mots se produisent lorsque le système traduit les segments lexicalement mais ne les place pas dans l'ordre syntaxique naturel de la langue cible. »

C'est un cas typique où la traduction brute ne tient pas compte de la structure normée dans la langue cible, ce qui altère le sens technique du terme.

Cahiers de Traduction
Volume 32 Issue 01/November 2025 pp 481-499
ISSN 1111- 4606 EISSN 2602-6023

Cette erreur, de type « ordre des mots », soulignée par la typologie de Vilar et al., démontre la limite des outils de traduction automatique dans la gestion des terminologies spécialisées notamment le phénomène collocatif et des structures syntaxiques propres aux domaines techniques.

D'autres exemples :

Analyse de la traduction automatique de collocations scientifiques simples :

Terme En Français	Traduction de l'ISI	Traduction de DeepL
Corrélation Linéaire	ارتباط خطي	ارتباط خطي
Distribution Normale A Plusieurs Variables	التوزيع الطبيعي المتعدد المتغيرات	التوزيع الطبيعي المتعدد المتغيرات
Enquête Exploratoire	بحث استطلاعي ؛ عينة استطلاعية	بحث استطلاعي ؛ عينة ستطلاعية
Limites De Probabilité	حدود الاحتمال	حدود الاحتمال
Modèle Randomisé	نموذج عشوائي	نموذج عشوائي

Caractérisation linguistique et typologique

- **Nature des collocations** : Les termes considérés sont des collocations scientifiques bien formées, où chaque constituant possède un équivalent lexical et conceptuel reconnu en arabe. Leur structure syntaxique est simple (N + adj, N + complément), et la terminologie est largement normalisée dans les disciplines scientifiques.
- **Fidélité de la traduction automatique** : DeepL propose ici, pour chacun des exemples, une traduction directement équivalente à celle du glossaire ISI ou équivalente à l'usage professionnel arabe. Cela s'explique par :
  - La présence d'équivalents lexicaux directs pour tous les composants du groupe nominal.
  - L'usage international et la formalisation des constructions linguistiques (collocations) dans les langues source et cible.
  - L'absence de termes polysémiques ou ambigus pour ces concepts.

Analyse des erreurs et satisfaction terminologique :

Aucune erreur catégorisable selon Vilar et al., 2006 : Selon la typologie de Vilar et al. (2006), l'absence d'erreur dans ces cas tient à la simplicité lexicale et compositionnelle du groupe des termes :

1. **Pas de mot inconnu** : Tous les constituants sont présents dans le dictionnaire et la base de données du système.
2. **Pas d'erreur d'ordre des mots** : La structure syntaxique étant la même en arabe et en français, le transfert syntaxico-sémantique est immédiat.

Cahiers de Traduction
Volume 32 Issue 01/November 2025 pp 481-499
ISSN 1111- 4606 EISSN 2602-6023

3. **Pas d'omission, ni de faute morpho-syntaxique** : La traduction respecte à la fois la morphologie et la syntaxe cible.
4. **Pas d'erreur de collocation** : DeepL reconnaît la collocation comme entité prête à traduire.
    - **Justesse grâce à la composition sémiotique simple** : La réussite de la traduction n'est pas due à une sophistication terminologique du système, mais à la nature « transparente » de la collocation, facilement décomposable et recomposable dans la langue cible à savoir l'arabe. almaany+3

La traduction automatique de collocations scientifiques telles que « corrélation linéaire » ou « limites de probabilité » aboutit à des résultats corrects quand la collocation se compose exclusivement de termes standards, chacun ayant un équivalent dans la langue cible, et que la structure syntaxique n'exige pas de réorganisation. Ce succès n'est pas une garantie de compétence terminologique avancée du système, mais une conséquence de la simplicité lexicale et grammaticale. Cela souligne l'importance de la structure des collocations dans la qualité de la traduction automatique et l'enjeu pour l'élaboration de glossaires spécialisés d'utiliser des terminologies à haut taux de transparence sémantique et morphologique.

6. Conclusion :

Dans cet article, nous avons tenté de répondre à la question de savoir si les logiciels de traduction neuronale peuvent proposer une traduction des collocations égale à celle de l'être humain. Nous avons proposé l'analyse qualitative d'une série de termes extraits du glossaire édité par l'International Statistical Institute traduits automatiquement avec l'outil de TA neuronale DeepL. Il ressort de notre étude que la traduction automatique neuronale (TAN) permet un gain de temps lorsqu'il s'agit de corpus volumineux et capable de traiter rapidement et massivement de listes termes, tout en offrant une traduction souvent grammaticalement correcte même lorsqu'il s'agit de collocation. Cependant, la TAN demeure sujette à des erreurs spécifiques qui diffèrent de celles des anciennes approches de TA.

Même si la TAN contribue largement à améliorer la qualité des traductions, il n'en demeure pas moins qu'elle n'est pas exempte d'erreur, ni de limitations contextuelles ou terminologiques.

Dane le même ordre d'idée, Anthony Pym estime que les logiciels de traduction automatique désormais accessibles au grand public peuvent sembler utiles, car ils permettent aux lecteurs de comprendre globalement un texte rédigé dans une langue étrangère. Cependant, les textes générés par ces logiciels ne peuvent en aucun cas être considérés comme équivalant à une véritable traduction, ce qui signifie qu'ils doivent être révisés par un traducteur professionnel. (PYM, 2011 : 5).

7. Liste Bibliographiques :

- Bahdanau, D. (2016). *Neural Machine Translation by jointly Learning to align and translate*. ICLR, San Diego, USA. http:/arxiv.org/pdf/1409.0473
- Belaidi Ahmed . Ghadi Ahmed (2024), Une Traduction Automatique Meilleure Grâce Aux Textes Parallèles, Cahiers de Traduction, Alger, Volume 29, n° 01.

Cahiers de Traduction
Volume 32 Issue 01/November 2025 pp 481-499
ISSN 1111- 4606 EISSN 2602-6023

- Benmahammed, Y. (2022). Aperçu sur les collocations en linguistique moderne مجلة *البحوث و الدراسات الإنسانية*. 696-665 , https://asjp.cerist.dz/en/article/178464
- Bertrand, C. (1998). *Étude comparative des combinaisons lexicales dans deux domaines de spécialité : collocations lexicales et collocations conceptuelles* (Thèse de maîtrise, Université de Montréal). http://hdl.handle.net/1866/29624
- Bouillon , P., & Clas , A. (1993). *La Traductique*. Montréal : AUPLEF UREF .
- DEEPL , T. DEEPL., https://www.deepl.com/fr/translator (consulté en décembre 2023).
- Delhay, C. (1996). J.R. Firth : contexte et collocation. *Scolia : Sciences Cognitives, Linguistiques Et Intelligence Artificielle*, (9), 23-35. https://doi.org/https://doi.org/https://www.persee.fr/doc/scoli_1253-9708_1996_num_9_1_948
- Delisle, J. (2013). *La Traduction raisonnée* 3ème edition. Canada: Les Presses de l'Université d'Ottawa.
- Freigang, K. (2001). Automatisation of Translation: Past, Presence, and Future. *Tradumatica*. https://doi.org/https://doi.org/http//fti.uab.es/tradumatica/revista
- Gledhill, C. (2021). The Discourse function of collocation in research article introductions. *English for Specific Purposes*, (19), 115-135. https://doi.org/https://doi.org/0000-0001-7834-3511
- Guidere, M. (2010). *Introduction à la traductologie*, 2ème édition. De Boeck.
- Hutchins , J., & Somers, H. (1992). *An Introduction to Machine Translation*. London : Academic Press Limited .
- Kohen, P. (2000). *Statistical Machine Translation*. Cambridge: Cambridge University Press.
- Larivière, L. (1997). Valeur sémantique du verbe dans les collocations verbales spécialisées. *Traduction, Terminologie, Rédaction*, (11), 173-197. https://doi.org/https://doi.org/https://doi.org/10.7202/037321ar
- Netzlaff, M. (2004). *La collocation adjectif-adverbe et son traitement lexicographique : Français – allemand – espagnol* (Thèse de Doctorat, Friedrich-Alexander-Universität Erlangen). https://www.researchgate.net/publication/33739539_La_collocation_adjectif-adverbe_et_son_traitement_lexicographique_francais_-_allemand_-_espagnol
- Newmark, P. (1988). *A Textbook of translation*, 1ère édition. London and New York : Prentice Hal.
- Pym, A. (2011). What technology does to translating. *The International Journal for Translation and Interpreting*, *1*(3), 1-9. https://doi.org/https://doi.org/https://www.researchgate.net/publication/287006159_What_technology_does_to_translating
- Van Der Wouden , T. . (1997). *Negative Contexts : Collocation, Polarity and Multiple Negative* (1 ط). London and New York: Routledge.
- Vaswani, A., & Shazeer, N. (2023). *Attention Is All You Need*. 31st Conference on Neural Information Processing Systems , Long Beach, CA, USA. http://arxiv.org/abs/1706.03762
- Vilar, D., & Xu, J. (2006). *Error analysis of statistical machine translation output*. Presented at the International Conference on Language Resources and Evaluation. https://www.researchgate.net/publication/307174612
- Yadolah, D. (2007). *Statistique : Dictionnaire encyclopédique*. France : Springer.