

DOI: 10.31178/RCSDLLF.13.5

Traduction juridique : limites de l'IA. Vue de la salle 404

ECATERINA GEORGESCU¹, ELENA TĂNASE²
Sous la direction de ANCA-MARINA VELICU

Abstract

Legal translation has a privileged position in the language industry these days, yet machine translation, however significant the advances in technology (neural translation, deep learning, AI, large language models) and despite the massive digitization of normative texts and case law, seems to perform less well precisely in this field. The aim of our research is to illustrate this dynamic through a case study and to investigate the reasons for it.

Keywords: AI (artificial intelligence), NMT (neural machine translation), LLM (large language models), Deep learning, legal translation

De nos jours³, la traduction juridique a une position privilégiée dans l'industrie de la langue ; or la traduction automatique, pour significatifs qu'ait été les progrès de la technologie (traduction neuronale, apprentissage profond, IA, grands modèles de langage) et en dépit de la numérisation massive des textes normatifs et de jurisprudence, semble performer moins bien précisément dans ce domaine. Le but de notre recherche est d'illustrer cette dynamique et d'en investiguer les raisons par une étude de cas.

La première section de l'article présentera les principaux repères caractérisant la position des traductions juridiques en Roumanie, la seconde

¹ MEFIC 1 – Master européen d'interprétation de conférence (anglais-français).

² MEFIC1 et MTSST1 – Master en traductions spécialisées et études terminologiques (anglais-français).

³ En 2020, les traductions juridiques et financières étaient le troisième type de traductions le plus souvent sollicitées (selon l'agence américaine Nimdzi Insights) – voir <https://atelieruldetraduceri.ro/piata-traducerilor/> (dernière consultation : 29.06.2024). En 2024, 12% des 100 bureaux de traduction les mieux cotés de par le monde fournissent y compris des services de traduction et d'interprétation juridique – voir www.nimdzi.com/nimdzi-100-top-lsp/ (dernière consultation : 30.06.2024).

présentera l'expérience initiale à l'origine de cette recherche (une tâche de traduction juridique soumise à DeepL et à ChatGPT-4 sous Copilot de Microsoft), la troisième donnera un aperçu de DeepL, de ChatGPT et de Copilot de Microsoft (les logiciels interrogés), la quatrième décrira les étapes subséquentes de l'expérience contrôlée de traduction automatique assistée par l'IA et la cinquième sera vouée à de brèves conclusions⁴.

1. Position des traductions juridiques sur le marché roumain des services linguistiques

Les traductions juridiques en général sont un passage obligé, en Roumanie, pour l'accès au statut professionnel de traducteur « autorisé » (équivalent interculturel, depuis 2016, du traducteur agréé ou assermenté français⁵). La *Loi n° 76/2016 portant modification et complément de la loi n° 178/1997 relative à l'autorisation et à la rémunération des interprètes et traducteurs utilisés par le Conseil Supérieur de la Magistrature et le Ministère de la Justice*⁶ ne reconnaît comme acte prouvant la qualification (l'expertise) du traducteur que l'attestation émise par le Ministère de la Culture, suite au concours national qu'il organise (en général) deux fois par an. Auparavant, un diplôme de master, voire de licence en langues étrangères (toutes mentions confondues, langues et littératures, LMA ou Traduction-interprétation) suffisait, l'attestation par le Ministère de la Culture ne concernant que les spécialistes d'autres disciplines. Par voie de conséquence, depuis 2016, les diplômes académiques en traduction-interprétation ne valent plus grand chose sur le marché de l'emploi roumain : d'une part, l'attestation met un signe d'égalité entre les experts de disciplines non linguistiques et les diplômés en langues et littératures étrangères, en LMA ou en Traduction-interprétation ; d'autre part, comme tout bureau de traduction a aussi une composante juridique et que même pour des traductions scientifiques, techniques, médicales, etc. on préfère les traducteurs « autorisés », seuls possesseurs d'un cachet officiel, l'embauche des titulaires de masters en traduction spécialisée ou en interprétation de conférence est en pratique conditionnée à l'attestation délivrée par le Ministère de la Culture, les diplômes académiques faisant au mieux figure de « facteur favorisant » dans le CV.

⁴ Sauf spécification contraire, toutes les sections ont été le fruit d'une rédaction collaborative des deux auteures.

⁵ Rappelons que la terminologie officielle en la matière est, selon la Société française des traducteurs (STF – <https://www.sft.fr>) : *expert traducteur près la Cour d'appel X* (forme courte : *expert traducteur*), ou, selon le cas : *expert traducteur-interprète / expert-interprète (près la Cour d'appel X)*.

⁶ Notre traduction.

Or, comment se déroule ce concours ? Les candidats ne peuvent consulter pendant l'épreuve que des dictionnaires papier publiés par des maisons d'édition reconnues. De manière exceptionnelle, sous réserve d'accord ponctuel et après évaluation préalable par la commission d'examen, ils peuvent également employer leurs glossaires papier (listes de termes bilingues, sans unités phraséologiques ni textes normalisés – au sens de la norme ISO 12616-1/2021). En septembre 2023, la commission d'examen a accepté de tels glossaires pour les traductions du français en roumain et du roumain en français. C'est là un argument pragmatique en faveur de l'utilité de nos devoirs académiques consistant à compiler glossaires et bases de termes.

Vu le cadre légal de la profession et la structuration de l'offre et de la demande de traductions sur le marché national des services linguistiques, la plupart des traducteurs roumains « autorisés » auront à faire des traductions juridiques à un moment ou à un autre de leur vie professionnelle, même en dehors de toute pratique judiciaire : actes d'états civils, diplômes, contrats, etc.

Maintenant, traduire des actes normatifs (lois, arrêtés et autres textes réglementaires, voire des codes entiers) – exercice privilégié dans nos ateliers de traduction juridique et souvent, sujet de mémoires de master à l'UB – ne représente certes pas, en pratique, le type de traduction juridique le mieux partagé. Pourtant de telles traductions *sont* réalisées par des traducteurs ou traductrices « autorisé.e.s » soit dans le cadre de projets d'envergure bénéficiant du soutien des autorités publiques, soit dans leur activité au jour le jour (traducteurs d'administration, traducteurs d'agence, voire traducteurs indépendants). Quelques exemples parmi les plus saillants :

- la traduction en roumain des Codes pénaux des États membres de l'UE, sous la coordination de Tudorel Toader⁷. Le *Code pénal* français a été traduit par Mona-Manuela Răuș, traductrice « autorisée » (experte traductrice près la Cour d'Appel de Bucarest)⁸ ;
- la traduction en français de la *Constitution de Roumanie* (malheureusement accessible seulement sur support physique, à acquérir sur le site du Moniteur Officiel de Roumanie⁹ pour le prix modique de 5.45 lei (plus 18 lei de frais d'expédition par la Poste) ;

⁷ *Codex Penal*, éd. Beck 2018, 5 volumes, en ligne sur le site du Ministère roumain de la Justice (<https://codexpenal.just.ro>).

⁸ Voir www.euroavocatura.ro/traducator.php?id=29519 (dernière consultation : 30.06.2024).

⁹ Voir <https://monitoruloficial.ro/produs/constitutia-romaniei-in-limba-franceza/> (dernière consultation : 30.06.2024).

- des traductions en français d'actes normatifs publiés sur des sites gouvernementaux, tel l'*Ordre militaire concernant les mesures de prévention de la propagation du COVID-19*, mars 2020¹⁰ ;
- la localisation en français du site web du Gouvernement roumain comportant des actes normatifs (pour le moins leur intitulé, des fois également des extraits).

Contrairement à l'anglais, le français ne jouit pas, pour l'instant, d'un dispositif de traduction de toute la législation roumaine publiée sur des plateformes juridiques payantes du type de Sintact.ro : cette plateforme propose une traduction « de qualité » roumain-anglais, quel que soit le texte normatif qu'elle héberge, avec des délais de livraisons de 24h¹¹. N'empêche que l'apparition de tels services en dit long sur le rôle et l'importance de la demande en traductions juridiques (y compris pour ce qui est des actes normatifs) sur le marché roumain actuel.

Ceci étant, le fait que dans le domaine des traductions juridiques les machines aient des résultats largement en dessous de la moyenne est à la fois inquiétant et rassurant : inquiétant parce que cela suppose des post-éditions plus lourdes de la part de l'opérateur humain ; rassurant, parce que, manifestement, nous avons encore du pain sur la planche.

2. Notre première expérience contrôlée de traduction juridique ro-fr, sur DeepL et sur Copilot de Microsoft (basé sur ChatGPT)¹²

2.1. Présentation de l'expérience

Nous avons proposé le même texte source (TS) roumain à DeepL et à Copilot de Microsoft (connecté à ChatGPT-4) : le premier alinéa du premier article traitant du vol, dans le premier chapitre du Titre II¹³ du Code pénal roumain de 2009 (version à jour). Ce texte a pour caractéristique saillante la co-occurrence des concepts juridiques *possession* et *détention* (comprendre : *détention précaire*) – voir le passage en italique sous (1) ci-après :

- (1) Articolul 228. Furtul. Luarea unui bun mobil *din posesia sau detenția altuia*, fără consimțământul acestuia, în scopul de a și-l însuși pe nedrept, se pedepsește cu închisoare de la 6 luni la 3 ani sau cu amendă.

¹⁰ www.mai.gov.ro/wp-content/uploads/2020/03/Ordre-militaire-Concernant-les-mesures-de-prevention-de-la-propagation-du-COVID-19-.pdf (dernière consultation : 30.06.2024).

¹¹ Voir www.wolterskluwer.com/ro-ro/solutions/sintact/legislatie-tradusa-in-engleza (dernière consultation : 30.06.2024).

¹² Rédaction : Ecaterina Georgescu.

¹³ *Titlul II – Infracțiuni contra patrimoniului, Capitolul I – Furtul (art. 228-256)*. (Titre II – Infractions contre le patrimoine, Chapitre I – Le Vol, articles 228 à 256).

Les traductions fournies par les deux systèmes ont respectivement été :

- (2) Le vol. Le fait de soustraire un bien meuble *à la possession ou à la garde d'une autre personne* sans son consentement, dans le but de se l'approprier indûment, est puni d'une peine d'emprisonnement de six mois à trois ans ou d'une amende. (DeepL)
- (3) Le vol. Le fait de prendre un bien meuble *appartenant à autrui* sans son consentement, dans le but de s'en emparer indûment, est puni d'une peine d'emprisonnement de 6 mois à 3 ans ou d'une amende. (ChatGPT-4 sous Copilot de Microsoft)

La première traduction confondait deux concepts juridiques distincts : *garde* et *détention* d'un bien (meuble). La seconde ignorait complètement la différence entre *détention* et *possession* d'un bien (meuble), la neutralisant sous la notion d'*appartenance*.

Ces concepts juridiques sont mentionnés dans les Codes pénaux roumain et français en vigueur sans y être définis, aussi nous sommes-nous tournées vers des textes de doctrine (lexicographie spécialisée) pour leur définition :

- (4) GARDE : obligation de veiller à la conservation d'une chose (en empêchant qu'elle ne se perde ou ne se dégrade et en attendant de la restituer) que certains contrats font naître à la charge d'une partie (investie, ou non, par ailleurs, du droit de s'en servir ; comp. *dépôt*, C. civ., a. 1915, *prêt à usage*, a. 1880) et dont l'intention est appréciée avec plus ou moins de rigueur, suivant des critères d'équité (C. civ. 1927 et 1928) et les stipulations du contrat. *Comp. usage, jouissance* (Cornu 2022/1987 : 491).
- (5) DÉTENTION : pouvoir exercé sur la chose d'autrui en vertu d'un *titre* juridique (bail, dépôt, usufruit légal, nomination judiciaire) qui rend la détention précaire en ce qu'il oblige toujours le détenteur à restituer la chose à son propriétaire (Cornu 2022/1987 : 340).
- (6) POSSESSION : pouvoir de fait exercé sur une chose avec l'intention de s'en affirmer le maître, même si (le sachant ou pas) on ne l'est pas (Cornu 2022/1987 : 792).
- (7) PROPRIÉTÉ : [...] pleine propriété, type le plus achevé de *droit réel : droit d'user, de jouir et disposer d'une chose d'une manière *exclusive et absolue sous les restrictions établies par la loi (C. civ., art. 544). V. *maîtrise*, [...] (Cornu 2022/1987 : 825).

Nous avons ensuite vérifié que les concepts désignés en roumain par *posesie/ posesor, detenje precară/ detentor precar* et *proprietate/ proprietar* sont les mêmes que les concepts de *possession*, de *détention précaire* et respectivement de *propriété* en droit français, en consultant une base de données juridique en ligne construite et gérée par les étudiants de la Faculté de Droit de l'UB : ABCJuridic <https://abcjuridic.ro/detentia-precara-stapanirea-unui-bun/>.

Aussi avons-nous pu conclure au caractère vraiment problématique des erreurs de traduction recensées dans les deux traductions automatiques. La première traduction comporte un faux sens caractérisé (l'équivalent d'une traduction de *măr* 'pomme' par *tomate*¹⁴), la seconde, une modulation (abusives) par un hyperonyme¹⁵ de langue générale.

Tout de suite après l'invite¹⁶ à traduire, nous avons formulé en français, pour Copilot, une requête d'information concernant la différence entre les deux concepts de *possession* et de *détention*, en droit français et roumain.

La réponse a prouvé que la machine *avait* accès aux informations relatives à ces concepts. Il est vrai qu'une confusion a été faite entre droit roumain et droit romain¹⁷, mais les deux notions ont bien été discriminées. À première vue du moins, tout en ayant accès à l'information sur les concepts (en mode de recherche par mots-clés), l'IA ne pouvait pas corréliser cette capacité avec l'opération de traduction, contrairement au traducteur humain, qui passe sans problèmes de la documentation-recherche-analyse conceptuelle à la recherche de variantes de réexpression en langue cible (LC) et qui peut traiter, afin de produire son texte cible (TC), y compris des *textes comparables* en LC.

2.2. Brève présentation des outils interrogés

DeepL est un système de traduction neuronale basé sur l'apprentissage profond (*Deep learning*), ChatGPT est un *grand modèle de langage* (ou *modèle massif de langage* – siglaison anglaise employée aussi en français : *LLM*), censé au demeurant être un *agent conversationnel* (synonyme français : *dialogueur*, équivalent du *chatbot* anglais), mais qui peut remplir plein d'autres fonctions, dont la traduction automatique. GPT veut dire « transformateur pré-entraîné génératif » : c'est « une famille de modèles de réseaux de neurones récurrents utilisant l'architecture des *transformers* »¹⁸. Cette architecture utilise un mécanisme d'auto-attention (ou *attention auto-régressive*) lui permettant de prédire ce qui suit non seulement compte tenu du dernier item traité (séquence de caractères, typiquement un mot ou un signe de ponctuation), mais également compte tenu du contexte : la machine peut décoder correctement les relations de dépendance entre

¹⁴ Un fruit classé par la doxa comme légume.

¹⁵ Pour la définition et des illustrations de ce procédé de traduction, voir Velicu 2022 : 75 et suiv.

¹⁶ Synonyme français de l'anglicisme *prompt*.

¹⁷ Une reformulation en roumain de la question a provoqué une réponse bien documentée sur les concepts juridiques de *posesie* 'possession' et de *detenție (a unui bun mobil)* 'détention (d'un bien meuble)' en droit roumain.

¹⁸ Voir Robert 2024, dont nous adopterons la terminologie. Wikipédia opte pour l'anglicisme *transformeur*.

mots, même s'ils se trouvent à distance les uns des autres¹⁹. Ce modèle est pré-entraîné sur d'énormes quantités de données textuelles numérisées (en langage humain), accessibles sur Internet²⁰.

Enfin, Copilot de Microsoft est un assistant numérique connecté à ChatGPT (version 4, depuis février passé) basé sur l'IA, qu'il intègre aux applications Microsoft 365. C'est donc un outil encore plus complexe que ChatGPT seul²¹ (et gratuit).

Dans un article très récent voué à la comparaison des performances de l'application spécialisée DeepL et de ChatGPT en matière de traduction automatique, l'auteure (une traductrice indépendante) arrive à la conclusion suivante :

Pour le moment, ChatGPT n'est pas un substitut adéquat des systèmes de traduction automatique conventionnels tels que DeepL, Google Translate et d'autres, car ces systèmes ont été conçus spécifiquement pour la traduction et possèdent des caractéristiques spécifiques qui en améliorent la qualité²². (Kühn-Schwendimann 2024, § "Differences between ChatGPT and DeepL", alinéa 6/7)

Parmi les caractéristiques avancées des systèmes experts de traduction automatique neuronale (TAN) sont également mentionnés « la possibilité [...] de créer un glossaire, d'offrir des suggestions de traduction alternative et de s'intégrer à des systèmes de gestion de traduction (TMS) » (*ibidem*, alinéa 3/7).

De fait, le *générateur de glossaires* de DeepL est toujours en phase de bêta-tests. De toute manière, cette option n'est à ce jour accessible que pour trois combinaisons de langues : anglais-allemand, allemand-anglais et anglais-espagnol (voir les informations sur le site du fournisseur, juin 2024). Qui pis est, elle ne peut être testée, même pour ces langues-là, qu'à partir d'un abonnement haut de gamme (plans « DeepL Pro Advanced » ou « DeepL Ultimate »²³). Ceci dit, il ne s'agit pas d'une caractéristique triviale : elle permettra (pour les couples de langues qui seront pris en compte) de créer automatiquement des glossaires à partir de bi-textes traduits avec DeepL, d'enrichir ces glossaires par insertion manuelle de données ou de télécharger dans l'application des glossaires (listes bilingues de termes) déjà compilés, sous forme de fichiers CVS ('Valeurs

¹⁹ Par exemple, dans la phrase *L'animal n'a pas traversé la rue parce qu'il était épuisé*, la machine pourra mettre en relation non seulement *animal* et *traversé* ou *traversé* et *rue*, mais également *animal*, *était* et *épuisé* (exemple adapté d'après Keldenich 2021).

²⁰ Voir Robert 2024. Aborder dans le détail le fonctionnement de ces systèmes émerge nos visées dans le présent article. Nous nous sommes contentées d'en comprendre le principe.

²¹ Voir <https://support.microsoft.com/fr-fr/topic/chatgpt-et-microsoft-copilot-quelle-est-la-différence-8fdec864-72b1-46e1-afcb-8c12280d712f>.

²² Traduit de l'anglais avec www.DeepL.com/Translator (version gratuite), notre post-édition (en italique).

²³ Environ 54 € par mois.

séparées par des virgules’). Malheureusement, pour le moment du moins, cette caractéristique n’est pas accessible pour le français ni pour le roumain (quelle que soit la vectorisation). Mais les autres caractéristiques peuvent être testées sur ce couple de langues, y compris sur la version en libre accès, et effectivement elles permettent d’obtenir plus vite un TC adéquat notamment grâce :

- à la traduction à même le TS dans le fichier d’origine, sans passer par une opération de type copier-coller (CTRL+CC) ;
- à l’option « édition dans DeepL » du TC fourni (dans la fenêtre ouverte en marge du fichier d’origine), l’exportation vers l’application-mère étant elle aussi automatique ; l’édition dans l’application permet à l’opérateur humain d’exploiter les alternatives fournies par le glossaire incorporé au système TAN ou de saisir manuellement dans le TC ses propres variantes ;
- à la commande « insertion » du texte (post-)édité, dans le fichier Word en cours, en faisant l’économie d’une nouvelle opération copier-coller.

C’est dire que, s’agissant de textes plus longs²⁴, DeepL l’emporte haut la main sur ChatGPT, pour des raisons logistiques.

À l’horizon de cette recherche, nous avons toutefois voulu interroger plutôt la *qualité des solutions de transfert interlingual* des deux systèmes, afin de voir si, et dans quelle mesure, le traducteur spécialisé (en particulier le traducteur juridique) peut désormais faire l’économie de la compilation d’outils spécifiques TAO (bases de termes, bases de données terminologiques – termes et métadonnées : définitions, explications, contextes, sources – ou mémoires de traduction). Ne pouvant pour le moment pas améliorer, sur ce cas précis, la performance de DeepL (à défaut de pouvoir y intégrer de glossaire roumain-français plus spécifique), nous avons donc poursuivi l’expérience, pour mieux saisir à la fois les limites et les mérites de la traduction automatique assistée par l’IA que fournit Copilot connecté à ChatGPT-4.

3. Poursuite de l’expérience sur ChatGPT/ Copilot

L’invite adressée à Copilot (voir expérience initiale, décrite sous 2.1. *supra*) avait été très simple et claire, mais peu contextualisée : « Bonjour. Veuillez traduire en français le texte suivant ». Nous avons donc repris l’expérience, avec une invite plus contextualisée (plus informative), en précisant cette fois-ci le type de texte (texte juridique) et sa source (article 228 du Code pénal roumain) – voir capture d’écran sous Fig. 1 *infra*. Néanmoins la traduction du TS soumis a été la

²⁴ Un professeur de traduction spécialisé ou des étudiants pendant les ateliers (dont nous-mêmes) traiteront ce texte phrase par phrase ou paragraphe par paragraphe dans la version libre de l’application ; le procès de traduction sera plus rapide qu’en régime de copier-coller extensif.

même, à ceci près que l'assistant IA a proposé une traduction des trois alinéas de l'article – voir capture d'écran sous Fig. 2 *infra*.

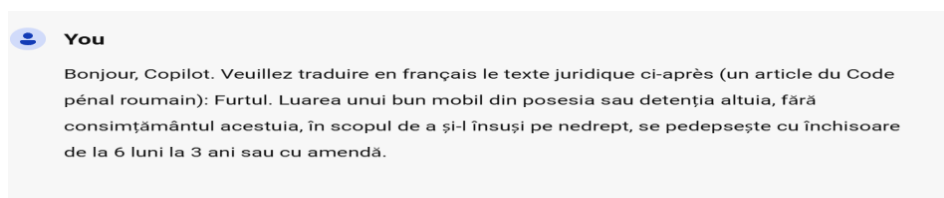


Figure 1 : invite (plus) contextualisée

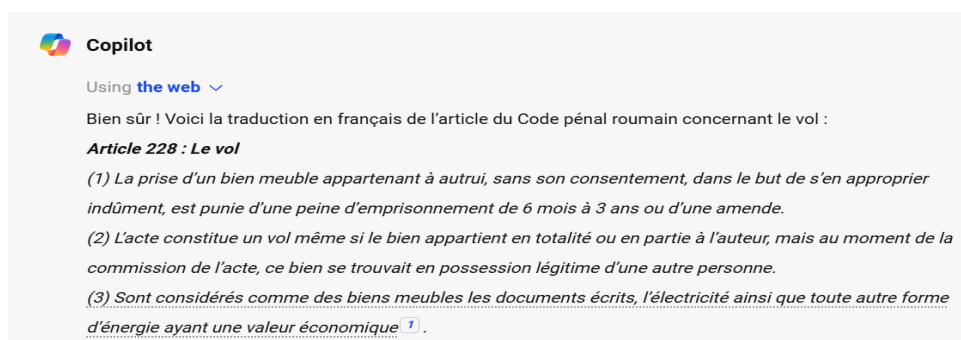


Figure 2 : Traduction du texte nommé plutôt que du TS soumis

À remarquer que la traduction de l'alinéa (2) – reproduit en roumain ci-après sous, (8) – continue de déliter la frontière entre les deux concepts de *possession* et de *détention*, omettant le concept de *détention*.

- (8) (2) Fapta constituie furt și dacă bunul aparține în întregime sau în parte făptuitorului, dar în momentul săvârșirii acel bun se găsea în posesia sau detenția legitimă a altei persoane.

La troisième étape de l'expérience est allée toujours dans le sens d'une mise en contexte plus étoffée : nous avons d'abord demandé à Copilot de discriminer les concepts juridiques de *possession* et de *détention précaire* d'un bien, pour ensuite lui faire traduire les deux premiers alinéas de l'article 228 dont nous lui avons soumis initialement le premier alinéa.

La traduction en français continue de neutraliser les concepts de *détention* et de *possession* sous l'hyponyme de langue commune *appartenir* (premier alinéa), mais rend apparemment justice à cette distinction, en traduisant (presque) littéralement la séquence roumaine *în posesia sau detenția legitimă* (aux articles définis près) par *en possession ou détention légitime* (second alinéa) :

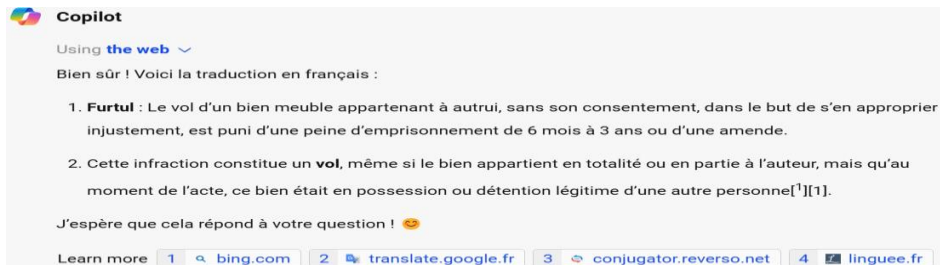


Figure 3 : Traduction des deux cohyponymes sous ambiguïté référentielle (cohyponymes ? synonymes ?)

À remarquer en outre que le *definiendum* n'est pas traduit en français, mais laissé en roumain (*Furtul = Furtul*). Cette erreur est corrélée à la modulation par l'hyponyme *Luarea* (la prise, le fait de prendre) = *Le vol* : or *le vol* était l'équivalent français attendu pour le *definiendum Furtul*, et *la prise (d'un bien meuble)* l'hyperonyme attendu dans le *definiens*. De plus, la différence spécifique *sans son consentement* est, elle, traduite, de manière tautologique, dans le *definiens*, même après cette modulation abusive (*luarea*, « la prise » = *le vol*). La prise du bien peut être faite avec ou sans consentement de son possesseur ou, selon le cas, de son détenteur, mais le *vol* est par hypothèse une prise sans consentement (c'est là le sens de la définition juridique roumaine).

Nous avons dit « apparemment », à propos du progrès que représenterait *en possession ou détention légitime* par rapport à *en possession légitime* tout court, puisque le français et le roumain (y compris juridiques) n'ont pas les mêmes contraintes en fait d'usage de la coordination en *ou* : sans reprise de l'article, *ou* peut être compris en français comme marqueur de reformulation paraphrastique (« c'est-à-dire ») et introduire un synonyme (c'est même là l'interprétation préférée). Il semble que Copilot connecté à ChatGPT-4 ait « hésité » sur la relation entre les deux termes reliés par *sau* en roumain (cohyponymes ou synonymes ?).

4. Expériences de traduction sur ChatGPT-4 et sur ChatGPT-4o (non intégrés à un autre outil)²⁵

L'invite à traduire le second alinéa de l'article 228 en fournissant le TS contextualisé (avec indication de la source) a pourtant eu un résultat différent avec ChatGPT-4 non connecté à Copilot (expérience menée par ma collègue, Ecaterina Georgescu) : le LLM a bien interprété alors la disjonction référentielle et a coordonné sous *ou* deux syntagmes nominaux définis, avec reprise de la préposition :

²⁵ Rédaction : Elena Raluca Tănase.

- (9) L'acte constitue un vol même si le bien appartient en tout ou en partie à l'auteur, mais au moment de la commission de l'acte, ce bien se trouvait *en la possession* ou *en la détention* légitime d'une autre personne.

Il faut noter que cette invite à traduire en français d'un TS roumain n'a été précédée d'aucune requête d'explication des différences entre les concepts de *possession* et de *détention* (d'un bien meuble). Mais cette invite à traduire un TS roumain a été elle-même formulée en roumain. Le décodage du TS serait-il facilité par une invite formulée dans la même langue ?

Le pas suivant a été d'inviter (à nouveau²⁶) le LLM à traduire le premier paragraphe du même article du Code pénal roumain sur le vol (voir (1), *supra*). Cette invite, formulée en roumain, a débouché elle aussi sur une réponse améliorée (par rapport aux expériences sur Copilot connecté à ChatGPT-4). Les cohyponymes *possession* et *détention* n'ont plus été neutralisés sous le verbe *appartenir*, étant nettement disjoints du point de vue référentiel (fait prouvé par les deux syntagmes nominaux définis et la reprise de la préposition *de*) :

- (10) La prise d'un bien mobilier *de la possession* ou *de la détention d'autrui*, sans le consentement de celui-ci, dans le but de se l'approprier injustement, est punie d'une peine d'emprisonnement de 6 mois à 3 ans ou d'une amende.

En ce qui concerne le second alinéa de l'article 228, le « Smart AI Translator Powered by ChatGPT » de Hix Translate (<https://translate.hix.ai>) opère également la disjonction référentielle, à la faveur de la reprise de la préposition devant le second conjoint, et ce en l'absence de toute requête verbale (sans invite à traduire), le choix de la (méta)langue (roumain plutôt que français) n'y jouant donc aucun rôle :

- (11) Le fait constitue un vol même si le bien appartient entièrement ou en partie à l'auteur, mais au moment de la perpétration, ce bien se trouvait légitimement *en possession* ou *en détention* d'une autre personne.

Le format de ce programme de TA est similaire à DeepL, avec des fenêtres TS et TC et instructions préformatées en anglais (voir Figure 4 ci-après) :

²⁶ Mais dans une nouvelle interaction, puisque la première expérience de traduction de ce texte avait été réalisée 48 heures auparavant.

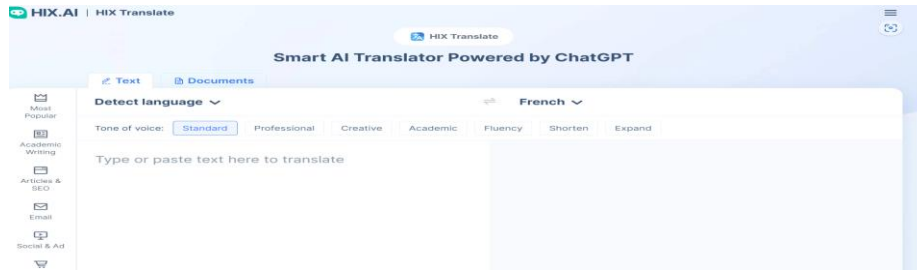


Figure 4 : Fenêtre TS/TC du système de TA Hix équipé de ChatGPT

À noter que DeepL trébuche contre la lecture synonymique des deux termes juridiques roumains *posesie* et *detenție* (voir capture d'écran sous Fig. 5) :

- (12) L'infraction constitue également un vol si le bien appartient en tout ou en partie à l'auteur, mais qu'au moment de l'infraction, le bien était *en possession légale* ou *en possession légale* d'une autre personne.

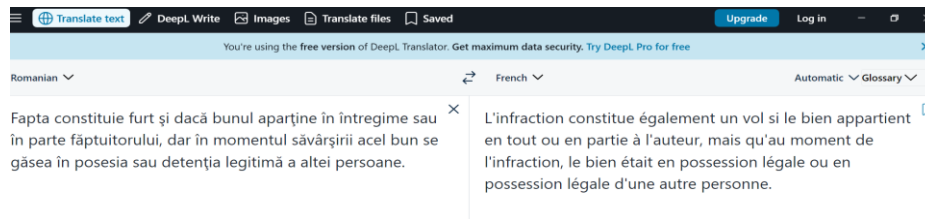


Figure 5 : DeepL, décodage erroné synonymique

Pour en avoir le cœur net au sujet du biais de la langue de l'invite à traduire, nous avons décidé de refaire l'expérience de la tâche de traduction postérieure à la tâche d'explication conceptuelle (information factuelle), à la fois sur ChatGPT-4 non connecté à Copilot, avec invite cette fois-ci en LC (en français) et sur ChatGPT-4o²⁷ (en libre accès depuis février passé) – expérience que nous avons menée nous-mêmes. Pour éviter la confusion droit roumain / droit romain (voir §2.1 et la note de bas de page 11), la question sur les relations entre les deux concepts a été formulée en roumain dans les deux cas :

- (13) E.R.T. : Bună ziua! Care este diferența dintre posesie și detenția unui bun mobil în dreptul român? Am nevoie și de sursele din care ieți informația.
 ChatGPT-4o : Bună ziua! În dreptul român, posesia și detenția unui bun mobil reprezintă două concepte distincte legate de utilizarea și controlul asupra unui bun. Iată diferențele principale dintre cele două: [...] ²⁸

²⁷ « O » vient de *omni*, 'tout' en latin : capacité à s'intégrer à tout format.

²⁸ Bonjour ! Quelle est la différence entre la possession et la détention d'un bien meuble en droit roumain ? J'ai également besoin de connaître les sources d'où vous tirez vos informations.

- (14) E.R.T. : Bună ziua! Care este diferența dintre posesie și detenția unui bun mobil în dreptul român? Am nevoie și de sursele din care ieși informația.
 ChatGPT-4 : Bună ziua! În dreptul român, distincția între posesie și detenție se referă la modul în care o persoană deține sau controlează un bun mobil. *Posesia* este situația juridică în care o persoană (numită posesor) deține un bun cu intenția de a se comporta ca și cum ar fi proprietarul acelui bun, indiferent dacă este sau nu proprietarul de drept. [...] Un aspect important al posesiei este că ea poate genera uzucapiunea, adică dobândirea dreptului de proprietate [...]
Detenția se referă la deținerea fizică a unui bun fără intenția de a se comporta ca proprietar. Detenția este de obicei recunoscută persoanelor care dețin bunul pentru un altul, cum ar fi închiriatorii, administratorii sau depozitarii. Detentorii [...] nu pot uzucapia bunul deținut [...].

En revanche, pour éviter le biais de la LS, l'invite à traduire le TS a été, elle, formulée en français :

- (15) E.R.T. : *Prière de traduire en français le texte suivant* : “Luarea unui bun mobil din posesia sau detenția altuia, fără consimțământul acestuia, în scopul de a și-l însuși pe nedrept, se pedepsește cu închisoare de la 6 luni la 3 ani sau cu amendă.”
 ChatGPT-4o : Bien sûr, voici la traduction en français du texte fourni :
 « *Le fait de prendre un bien meuble de la possession ou de la détention d'autrui, sans son consentement, dans le but de se l'approprier injustement, est puni d'une peine d'emprisonnement de 6 mois à 3 ans ou d'une amende.* »
- (16) E.R.T. : *Prière de traduire en français le texte suivant* : “Luarea unui bun mobil din posesia sau detenția altuia, fără consimțământul acestuia, în scopul de a și-l însuși pe nedrept, se pedepsește cu închisoare de la 6 luni la 3 ani sau cu amendă.”
 ChatGPT4 : Voici la traduction en français du texte : « *La prise d'un bien meuble de la possession ou de la détention d'autrui, sans son consentement, dans le but de s'en approprier injustement, est punie d'une peine d'emprisonnement de six mois à trois ans ou d'une amende.* »

Nous remarquerons la performance meilleure du modèle Omnium non seulement par rapport à Copilot connecté à ChatGPT (pas de neutralisation des cohyponymes sous le verbe *appartenir*), mais aussi par rapport à ChatGPT-4 non

ChatGPT-4o : Bonjour ! En droit roumain, la possession et la détention d'un bien meuble sont deux concepts distincts liés à l'utilisation et au contrôle d'un bien. Voici les principales différences entre les deux : [...].

*** Traduit avec www.DeepL.com/Translator (version gratuite) *** Noter le choix de vouvoyer : DeepL a bel et bien « corrigé » le choix stylistique du chercheur humain dans son interaction avec l'IA.

intégré : tour périphrastique *le fait de prendre* (vs nom d'action déverbal *la prise* + génitif, dans les tâches de traduction successives à invites en LS).

Cette expérience revoit à la baisse la pertinence du choix convergent de la langue de l'invite (invite en langue du TS), tout en confirmant le rôle de la mise en séquence des tâches : ainsi, introduire la tâche de discrimination conceptuelle *avant* la tâche de traduction optimiserait le résultat de la tâche de traduction, quelle que soit la langue de l'invite.

Nous avons constaté un résultat amélioré (préservation des deux cohyponymes) non seulement en cas de mise en séquence <tâche de traduction après tâche d'information/explication des concepts>, mais également si l'alinéa (2) était traduit avant l'alinéa (1) – expérience menée avec invites en roumain par ma collègue. Le résultat amélioré ne serait donc pas imputable à la langue de l'invite, mais au contenu du TS : si l'alinéa (2) bloque la neutralisation des cohyponymes juridiques sous l'hyperonyme non technique *appartenir*, c'est que le verbe *a aparține* y est cooccurrent, d'entrée de jeu, avec les termes *posesie* et *deținție*.

En outre, cette expérience de traduction suggère que le choix du modèle de LLM (intégré ou pas intégré à un autre outil, 4 ou 4o) affecte lui aussi la qualité des versions françaises :

(17)

Copilot basé sur ChatGPT	ChatGPT-4	ChatGPT-4o
<i>Le fait de prendre</i> (un bien meuble)	<i>La prise</i> d'(un bien meuble)	<i>Le fait de prendre</i> (un bien meuble)
<i>appartenant</i> à autrui	de la <i>possession</i> ou de la <i>détention</i> d'autrui	de la <i>possession</i> ou de la <i>détention</i> d'autrui

Le fait de +V inf est une construction typique dans les énoncés définitoires du *Code pénal* français (voir exemples (19) et (20) ci-après) ou pour décrire une infraction dépourvue de désignation nominale (exemple (18) ci-après). Cette construction y est attestée 255 fois. Par contre (*la prise de* n'y a qu'une seule occurrence (*la prise d'intérêts*) :

- (18) *Le fait de prendre* le nom d'un tiers, dans des circonstances qui ont déterminé ou auraient pu déterminer contre celui-ci des poursuites pénales, est puni de cinq ans d'emprisonnement et de 75 000 euros d'amende. (Article 434-23, *Code pénal* français)
- (19) L'escalade est *le fait de s'introduire* dans un lieu quelconque, soit par-dessus un élément de clôture, soit par toute ouverture non destinée à servir d'entrée. (Article 132-74, *Code pénal* français)
- (20) *Le fait de donner volontairement la mort* à autrui constitue un meurtre. (Article 221-1, *Code pénal* français)

L'initiation d'une nouvelle interaction avec ChatGPT-4, l'invitant directement à traduire le premier alinéa (sans titre mais surtout sans contextualisation en amont, par une question sur les concepts) confirmera le rôle pertinent de la mise en séquence des tâches : la traduction en français du syntagme multiple *din posesia sau detentia altuia* par une coordination de noms nus, sous un seul article défini et sans reprise de la préposition (*de la [possession ou détention] d'autrui*), encourageant, respectivement, la lecture paraphrastique de *ou* et synonymique des deux noms n'est plus survenue lorsque la tâche de traduction a été performée après la tâche d'information (visant à la discrimination des concepts de *possession* et de *détention* d'un bien) – voir *supra* (16) (traduction optimisée), qui fait suite à la tâche d'information sous (14). Qui pis est, nous retrouvons en (21) la même modulation abusive (à travers l'emploi de l'hyponyme dans le *definiens* : *vol... sans consentement*) que celle opérée sous Fig. 3 *supra* (traduction par Copilot) :

- (21) *Le vol d'un bien mobilier de la possession ou détention d'autrui, sans son consentement, dans le but de s'en approprier injustement, est puni d'une peine d'emprisonnement de six mois à trois ans ou d'une amende.*

Nous avons pu constater que ChatGPT neutralise les cohyponymes sous *appartenir* aussi bien lorsqu'il est intégré à Copilot de Microsoft que lorsqu'il est intégré au traducteur intelligent AI Hix :



Figure 6 : Traducteur HIX intelligent alimenté par ChatGPT

Aucune des expériences menées sur les modèles 4 et 4o non intégrés n'ayant donné ce résultat, nous nous sommes naturellement demandé si, malgré les allégations sur le site fournisseur, Copilot de Microsoft n'aura pas activé par exemple la version 3.5 de ChatGPT lors des traductions effectuées sous (3) et Fig. 2 et 3 *supra*. Une réitération de l'expérience avec ChatGPT-3.5 (non intégré²⁹) confirme cette hypothèse, puisque la traduction est exactement la même que sous (3) *supra* :

²⁹ Les trois versions 3.5, 4 et 4o sont accessibles dans le cadre d'un plan ChatGPT Plus.

- (22) Le vol. *Le fait de prendre un bien meuble appartenant à autrui sans son consentement, dans le but de s'en emparer indûment, est puni d'une peine d'emprisonnement de 6 mois à 3 ans ou d'une amende.*

Voir Fig. 7 :



Figure 7 : traduction par ChatGPT 3.5 non intégré

Un seul ennui : la modulation abusive *luarea/le vol* au niveau du *definiens* du concept de *vol* en droit pénal roumain. Cette erreur de traduction a été constatée à la fois sous Copilot (Fig. 3 *supra*) et sous la version 4 non intégrée du grand modèle de langage (exemple (21) *supra*).

C'est uniquement la neutralisation des cohyponymes sous le verbe *appartenir* (erreur de traduction) et le choix de la périphrase *le fait de prendre* (plutôt que du nom d'action déverbal *la prise de* : une solution de traduction stylistiquement consistante) qui semblent caractériser univoquement les productions de ChatGPT-3.5 (tous scénarios confondus).

Sur la foi de ces éléments, nous concluons qu'effectivement Copilot de Microsoft, à l'instar de Hix Translate assisté par l'IA, auront exploité non pas la version 4, mais la version 3.5 de ChatGPT (pour le moins, en traduisant les textes sources soumis pendant les expériences contrôlées de traduction présentées).

Nous avons aussi formulé des invites métalinguistiques, centrées sur la méthodologie de traduction et sur la mise à nu des fonctionnalités du LLM :

- (23) E.R.T. : Rebonjour ! Quand vous traduisez, exploitez-vous des textes comparables en langue cible ou seulement des mémoires de traduction ?
Merci.

Copilot (Using the web) :

Lorsque je traduis, j'utilise principalement des *mémoires de traduction*. Ces mémoires enregistrent les paires de phrases traduites (la phrase d'origine et sa traduction équivalente) dans une base de données¹. Cependant, il est important de noter que *je ne dispose pas d'accès direct à des textes*

comparables en langue cible. Si vous avez besoin d'une traduction, n'hésitez pas à me fournir le texte à traduire, et je ferai de mon mieux pour vous aider ! 😊

Pour vérifier que nous parlons bien de la même chose, nous avons sollicité à la machine la définition du syntagme *texte comparable*, et la réponse a été satisfaisante :

(24) Copilot (using the web) : Un *texte comparable* fait référence à un texte dans la langue cible (la langue vers laquelle vous souhaitez traduire) qui traite du même sujet ou contient des informations similaires à celles du texte d'origine. En utilisant des textes comparables, les traducteurs peuvent mieux comprendre le contexte et choisir des expressions appropriées pour la traduction.

5. En guise de conclusion

Nous avons pu constater, le long de cette expérimentation très ponctuelle (encore que, nous l'espérons, bien illustrative des aléas de la traduction juridique par des systèmes de TAN basés, ou non, sur ChatGPT), que certains facteurs ont influencé la qualité des traductions par ChatGPT, et d'autres, non : la contextualisation en amont (mise en séquence des tâches informatives et de traduction) et le choix de la version spécifique du LLM (3.5, 4 ou 4o), ainsi sans doute que son intégration à un autre outil de TAN ou d'IA, ont des effets certains sur les choix de mise en équivalence interlinguale opérée ; en revanche, la relation de convergence entre langues-objets (en particulier entre langue du TS) et métalangue de la communication avec la machine (langue des invites) ne semble pas jouer en soi.

Si la question sur les relations entre concepts précède la tâche de traduction, le système a davantage de chances d'extrapoler l'information à la tâche de traduction et/ou de choisir des solutions de traduction à disjonction référentielle claire entre noms de sens proche en langue commune, désormais reconnus comme termes distincts. Mais il semble incapable (à ce jour) d'alterner les modes d'opérer : les tâches de traduction sont performées par recyclage (« digestion ») des textes parallèles auxquels le LLM a (eu) accès dans sa base d'entraînement, sans recours aucun aux textes comparables en LC. Le recours à des textes comparables en LC est réservé aux tâches de génération de texte en réponse à des invites à fournir des explications, informations, etc.

Cette absence de flexibilité « cognitive » (mais peut-on vraiment parler de « connaissances », s'agissant des LLM actuels ?³⁰) donne un avantage compétitif certain à l'opérateur humain.

³⁰ Même si la linguistique pure et dure n'est pas vraiment notre point fort, les avis de Noam Chomsky et de Ian Roberts dans Chomsky 2023 nous ont quand même mis la puce à l'oreille.

Avant de clore, un mot sur l'émergence de ce projet collectif de traduction visant au final à interroger les limites de la créativité en traduction spécialisée³¹ (créativité du traducteur humain et créativité de la machine). Et une explication du sous-titre : « Vue de la salle 404 »³².

La salle 404 (rue Edgar Quinet 5-7) est le siège du CRU de l'Université de Bucarest. Elle a hébergé la plupart de nos cours en terminologie et traductions spécialisées en première année de Master d'interprétation de conférence et/ou de Traductions spécialisées et études terminologiques, ainsi que les réunions du groupe de recherche étudiante intégrée ContrasteR (Études contrastives, traduction spécialisée, terminologie).

La salle 404 n'est pas exactement une « chambre avec vue ». Elle a pourtant deux fenêtres, l'une qui ouvre sur la rue Edgar Quinet et offre une vue sur l'immeuble de la Faculté d'Architecture de l'Université d'architecture et d'urbanisme « Ion Mincu », l'autre, qui ouvre vers la cour intérieure de l'immeuble, inaccessible « à pied » à cause d'un chantier archéologique. Des pigeons viennent souvent se reposer sur un carré du toit juste sous la fenêtre vers la Faculté d'architecture et ont l'air d'écouter nos débats. Ces trois éléments ont de fait été à l'origine de notre recherche.

Le vol des oiseaux nous a fait penser au concept de vol³³. La Faculté d'Architecture, aux architectures neuronales (d'apprentissage profond). Le chantier d'archéologie, enfin, nous a fait penser à l'idée de fouille et d'enquête, mais aussi (voire surtout), à la manière dont toute pratique humaine subit l'épreuve du temps : dans leurs mémoires de master, nos collègues en Traductions spécialisées et études terminologiques s'attachent typiquement (ces dernières années) à compiler des mémoires de traduction ou des bases de données terminologiques sur les sujets les plus divers. Les mémoires de master³⁴ sont un lieu

³¹ Cette recherche a été menée dans le cadre d'un projet AUF-UB dont notre coordinatrice est l'une des co-directrices : « Les journées de la créativité, de la traduction et de la recherche » (29 mai 2024 – 29 novembre 2024), *Pour marquer le 30^{ème} anniversaire de l'AUF-ECO : Créer, transformer, innover, exceller.*

³² Qui évoque en liséré le titre d'un volume célèbre en grammaire générative où a été lancé le Programme minimaliste de recherche dans ce paradigme théorique. Nous n'avons fait que feuilleter ce volume, invoqué pendant le cours de linguistique française en L2 de Traduction-interprétation : *The View from Building 20* (Hale & Keyser 1993). Que la licence nous soit pourtant permise.

³³ « L'homonymie n'est qu'un cas de polysémie dont on ne voit pas [ou : plus ?] la motivation », écrivait Bernard Pottier (1992 : 43).

³⁴ Le mémoire de master de notre collègue, Ioana Cheroiu-Cozma (M2 en Traductions spécialisées et études terminologiques), axé sur la compilation d'une BDT et d'une mémoire de traduction en droit pénal, a d'ailleurs été l'un des déclencheurs de notre question de recherche : faut-il encore compiler des outils TAO (mémoires de traduction et bases de termes), à l'ère de DeepL et de ChatGPT ?

de réflexion (souvent assez poussée) autour de la question COMMENT : comment sélectionner les unités terminologiques, comment choisir le corpus de référence, comment structurer la base de termes, comment motiver les équivalences de traduction – à vertus surtout contextuelles, variables au gré des occurrences – enfin, comment motiver les équivalences terminologiques, virtuellement plus stables et uniformes³⁵. Mais la question POURQUOI n'y est qu'à peine adressée, et les vrais problèmes qu'elle soulève sont, le plus souvent, ignorés ou minimisés avec élégance. Notre propos ici a été, donc, *pro domo*, puisque nous avons voulu soulever y compris le défi de l'utilité (dans une perspective d'insertion professionnelle) des devoirs académiques et des sujets de recherche des mémoires de master (typiquement axés sur la production de bi-textes alignés ou de glossaires bilingues, éventuellement saisis sous forme de BDT et respectivement insérés à des mémoires de traduction). Question indiscreète s'il en est...

Références bibliographiques

- Chomsky, Noam (2023) – « Les fausses promesses de ChatGPT », *Philosophie Magazine*, Hors série « IA, le mythe du XXI^e siècle ». URL : www.philomag.com/articles/les-fausses-promesses-de-chatgpt (publié le 10 mai 2023, téléchargé le 28 juin 2024).
- Keldenich, Tom (2021) – « Le Mécanisme de l'Attention en Deep Learning – Comprendre Rapidement ». URL : <https://inside-machinelearning.com/mecanisme-attention/> (consulté le 27 juin 2024).
- Kühn-Schwendimann, Flurian (2024) – “ChatGPT or DeepL: Which is better for translation?”, *Across Language Technology*. URL : www.across.net/en/knowledge/blog/chatgpt-or-deepl-which-is-better-for-translation/ (publié le 02 février 2024, consulté le 20 mars 2024).
- Pottier, Bernard (1992) – *Sémantique générale*, Paris, PUF.
- Robert, Jérémy (2024) – « Generated Pre-Trained Transformer (GPT) : Qu'est-ce que c'est ? Comment ça marche ? », *DataScientest. Intelligence artificielle*. URL : <https://datascientest.com/generative-pretrained-transformer#:~:text=C'est%20quoi%20un%20transformateur%20génératif%20pré-entraîné%20%28GPT%29%20%3F,majeure%20dans%20le%20domaine%20de%20l'intelligence%20artificielle%20généralive> (publié le 14.06.2024, consulté le 28.06.2024)
- Velicu, Anca-Marina (2022) – *Traduire des textes fonctionnels*, București, Pro Universitaria.
- Velicu, Anca-Marina (2024) – « Les termes *prévenus* et *accusés* en droit pénal français, canadien et suisse et leurs équivalents roumains », *Çédille, revista de estudios franceses*, n° 25, p. 175-209.

³⁵ Pour la distinction entre *équivalence de traduction* et *équivalence terminologique* appliquée au langage juridique, voir Velicu 2024 : 204.