

DOI: 10.31178/RCSDLLF.12.6

TERMINOLOGUE (DCU Dublin) – un outil de gestion terminologique pas comme les autres

IOANA CHEROIU-COZMA, OANA DINU, MIRELA VASILE¹
Sous la direction de ANCA-MARINA VELICU

Abstract

The purpose of this contribution is to present the main challenges of a localization project currently being finalized – the localization, into French and Romanian, of the open-source terminology management tool TERMINOLOGUE. The main point is to assert and assess a specific algorithm for the research of interlingual equivalents of terminological units *lato sensu* (terms, collocations, idioms, standardized text)

Keywords: localization, NMT, terminology management tool, open source software, verbatim research

Notre propos dans cette contribution est de présenter les principaux défis d'un projet de localisation en cours de finalisation, réalisé dans le cadre du groupe de recherche étudiante intégrée ContrasteR du Département de langue et littérature française de la Faculté des Langues et Littératures étrangères de l'Université de Bucarest. Il s'agit de la localisation, en français et en roumain, de l'outil de gestion terminologique à code source ouvert TERMINOLOGUE, hébergé par la *Fiontar & Scoil na Gaeilge* de l'Université de la ville de Dublin DCU. Ont été impliquées dans la traduction les auteures de cet article et Anca-Marina Velicu, à titre de traductrice-pilote et de réviseuse.

La première section de l'article apportera des précisions sur les principales notions opérationnalisées (localisation, traduction automatique neuronale – TAN – et post-édition, équivalence terminologique et équivalence de traduction ; procédés de traduction), la seconde comportera un aperçu à vol d'oiseau du produit à

¹ Master de traductions spécialisées et études terminologiques, Groupe de recherche étudiante intégrée ContrasteR, Université de Bucarest.

localiser. Cette section-ci intégrera une recherche terminologique sur les concepts *open source software*, *free software*, qui nous permettront de mieux situer TERMINOLOGUE dans le paysage international des outils de gestion terminologique et d'expliquer en quoi, au juste, c'est un produit « pas comme les autres ». La tierce section ciblera la méthodologie de la traduction : y seront d'abord présentés les textes sources (TS) anglais (des séquences sous-phrastiques ou phrastiques – des « chaînes de caractères » (*strings*) correspondant aux étiquettes, commandes et autres instructions ou messages courts pré-formatés, ainsi que deux textes en discours suivi), puis l'algorithme de traduction effectivement mis en place, en insistant sur la recherche d'équivalents interlinguaux pour les unités terminologiques et phraséologiques, illustrée à l'aide d'exemples empruntés aux deux modules de localisation gérés dans le cadre du projet – la localisation de TERMINOLOGUE en français et en roumain (traduction anglais-français et anglais-roumain). Cette section comportera une très brève présentation de *DeepL*, le moteur de traduction automatique exploité. Au niveau des exemples, nous insisterons notamment sur le travail de post-édition (donc sur les modifications que nous avons proposées par rapport à la version fournie par le moteur de traduction automatique). La dernière section sera réservée à quelques conclusions relatives au projet de localisation.

1. Considérations liminaires

La localisation est « la traduction et l'adaptation globale des produits et des services à un 'locus' (latin : lieu, région, pays, continent) » (Guidère 2008 : 125). C'est une technique de traduction-adaptation efficace et fréquemment utilisée dans les domaines du marketing et de l'informatique : localisation de script, localisation de site, localisation de jeux vidéo, localisation multimédia, localisation de logiciels, etc. Cette activité exige « une *compétence* pluridisciplinaire : maîtrise des langues, de l'interculturel et des stratégies, techniques et procédés de traduction, mais aussi de l'outil informatique » (Velicu 2022 : 106).

Du fait du type de TS et du domaine (logiciel), la variation interculturelle à incidences interlinguales a été des plus modestes. Modeste, mais non pas nulle – les exemples commentés sous §3 illustreront justement le besoin de recourir à des procédés de traduction indirecte à force de prendre en ligne de compte les préférences des locuteurs-scripteurs de la langue-culture cible.

Heureusement, la localisation de la plateforme TERMINOLOGUE a été admirablement gérée par le « client » qui nous a fourni des éléments sources pré-formatés (les séquences de caractères, sous xlsx et les textes en discours suivi, sous .md). Les compétences numériques requises ont été au demeurant assez modestes (maîtrise du système Windows et du langage Markdown, habileté à naviguer sur Internet et à exploiter la traduction automatique neuronale – TAN).

Cependant, du fait de notre incompétence initiale en matière de fichiers avec l’extension .md, les plus grands défis de notre projet de traduction ont été d’ordre technique.

Markdown est un logiciel à étiquetage de fonctions pour les divers segments de texte, lisible par la machine, dédié aux sites Internet. C’est « un langage de balisage qui permet de formater du texte brut de manière simple pour un rendu en HTML ou dans un autre format »². L’objectif de Markdown est que les styles de mise en texte (hiérarchie des titres/sous-titres, styles de police – italiques, gras, italiques gras) soient rendus sur Internet avec plus de lisibilité et ce, le plus simplement possible. Par exemple, la hiérarchisation des titres requiert l’usage du symbole # à gauche, répété selon la position du titre dans la hiérarchie (## pour un titre de niveau 2, ### pour un titre de niveau 3, etc.) avec une pause entre le symbole non verbal et le premier mot du titre respectif, mais, à la différence du système HTML, l’information ne va pas *encadrer* la séquence concernée, n’étant pas répétée à droite.

Pour en savoir plus sur les langages Markdown, nous avons consulté un site d’initiation à la syntaxe de base de ce langage³ et un tutoriel avec des leçons pratiques⁴.

Un fichier Markdown contient plusieurs parties qui doivent être localisées – dont notamment les *titres*, les *paragraphes avec les styles de police gras ou italiques*, les *images*, les *éléments de liste*, les *titres de tableau*, les *entrées de tableau*, les *notes de bas de page*. D’autres parties – dont les *clôtures de code* (``), les *emoji’s*, les *expressions mathématiques* ou les *liens* (des « éléments non traduits », au sens de Gouadec 1989) – doivent au contraire être exclues de la traduction.

Nous avons fait le choix du moindre effort et avons traduit directement dans les fichiers textes ouverts à partir de .md, où la plupart des « marquages » de style (sinon tous) sont préservés en l’état :

```
# Sur *Terminologue*

## Terminologue

*Terminologue* est un outil de gestion terminologique
à code source ouvert.
```

² « Markdown : qu’est-ce que c’est ? », document consulté en ligne : <https://www.futura-sciences.com/tech/definitions/informatique-markdown-17135/> (dernière consultation : 30.06.2023).

³ “Basic Syntax. The Markdown elements outlined in the original design document”, document consulté en ligne : <https://www.markdownguide.org/basic-syntax/> (dernière consultation : 30.06.2023).

⁴ <https://www.markdwntutorial.com/lesson/1/> (dernière consultation : 30.06.2023).

C'est à peu près de cette façon-là que ce fichier sera visualisé sur Internet : police Calibri *modulo* la taille de police, qui va de 16 pour le titre à 14 pour le sous-titre et à 11 pour le paragraphe :

Sur Terminologie

Terminologie

Terminologie est un outil de gestion terminologique à code source ouvert.

C'est là le fruit d'une conversion du fichier .md, sous docx, à l'aide d'une application spécialisée en libre accès (*Aspose*⁵). Nous avons opéré ce type de conversions pour vérifier si rien d'important n'avait été perdu, car il fallait remettre les traductions sous le format exact du TS.

Dans le projet de localisation dont nous présenterons ici les principaux défis, nous avons eu recours à un moteur de traduction automatique. La traduction automatique (TA⁶) est « un logiciel qui produit automatiquement une version du TS saisi, sans interaction aucune avec l'opérateur humain (sans permettre de dialogue homme-machine), à la différence des mémoires de traduction et des systèmes de TAO [traduction assistée par ordinateur] en général » (Velicu 2022 : 254).

Il y a eu plusieurs types de systèmes de TA : la TA directe, la TA basée sur les règles, la TA basée sur l'exemple, la TA basée sur les statistiques, enfin, ces dernières années, la TA neuronale (TAN⁷). L'histoire de la TAN est relativement récente (moins de dix ans).

Robert 2010 souligne la relation intrinsèque entre traduction automatique et post-édition et distingue une post-édition *brute* (révision d'un texte cible – TC) produit par un moteur de TA) et une post-édition *évoluée* (révision d'un TC produit par un processus alliant TAO et TA). La post-édition est « la révision par un linguiste professionnel d'un texte traduit par un moteur de traduction automatique »⁸. Les langagiers eux-mêmes distinguent une *post-édition légère* et une *post-édition complète*, visant la production d'un TC aussi naturel que possible, sans ambiguïtés ni risques de malentendu (une révision par un locuteur natif de la langue cible – LC – est nécessaire).

Lors de l'analyse de notre corpus, nous allons en outre opérationnaliser les notions d'équivalence de traduction (locale, occurrence) et d'équivalence

⁵ <https://www.aspose.app/> (dernière consultation : 30.06.2023).

⁶ Angl. MT, *machine translation*.

⁷ Angl. NMT, *neural machine translation*. Pour une présentation des premiers, voir Loffler-Laurian (1996), pour une présentation mise à jour incluant les derniers, voir Velicu 2022 : 249-252 ; 261-265.

⁸ « Qu'est-ce que la post-édition ? », *Over the Word*, document consulté en ligne, <https://overtheword.com/quest-ce-que-la-post-edition/> (dernière consultation : 30.06.2023).

terminologique⁹ (généralisable et dans l'idéal invariée d'un contexte d'occurrence à l'autre) et nous nous référerons à l'inventaire des procédés de traduction de la stylistique comparée (Vinay & Darbelnet 1958).

2. Brève présentation du produit à localiser

Comme son nom l'indique, TERMINOLOGUE est un outil destiné d'abord aux... *terminologues*.

Son logiciel est développé et géré par le groupe de recherche *Gaois* de *Fiontar & Scoil na Gaeilge* de l'Université de la ville de Dublin (DCU), et son développeur principal est Michal Boleslav Měchura, linguiste et architecte de l'information.

Selon la page d'information de la plateforme TERMINOLOGUE¹⁰, il s'agit d'un outil de gestion terminologique à code source ouvert, ayant une version basée sur le nuage (<https://www.terminologie.org/>), qui permet non seulement de créer et d'éditer des bases de données terminologiques, mais offre également la possibilité de les publier en tant que microsite sur le site web de Terminologie. Par ailleurs, ces bases de données terminologiques sont programmées pour être facilement repérables par les moteurs de recherche, notamment par Google et par Bing.

TERMINOLOGUE ne nécessite aucune installation ni configuration, l'accès à la plateforme est gratuit et les utilisateurs n'ont pas besoin de compétences en matière de codage ou de programmation. Qui mieux est, elle ne collecte pas de données d'utilisation (discretion assurée).

Toutes les fonctionnalités sont expliquées, en détail, sur la page d'introduction¹¹ de la plateforme, et nous avons pris le temps de les énumérer, afin de montrer à quel point elles sont diverses et innovantes¹² :

- i. Recherche d'un terme : la plateforme dispose d'une recherche linguistique intelligente ;
- ii. Partage des termes : un terme peut être partagé entre plusieurs entrées afin de faciliter le travail des rédacteurs et d'assurer la cohérence de la base de données ;
- iii. Annotations grammaticales en ligne : des notes contenant des informations grammaticales peuvent être attachées à des mots individuels qui composent des termes complexes ;

⁹ Au sens de Velicu 2022 : 32.

¹⁰ « About *Terminologie* », <https://www.terminologie.org/docs/info/en/> (dernière consultation : 30.06.2023).

¹¹ *Idem*.

¹² Voir Breathnach 2020 pour une synthèse critique très convaincante.

- iv. Configuration des métadonnées : domaines, marques de flexion, dénomination des parties du discours, types de notes, catégories d'acceptabilité, sources, collections, commentaires, extranets ;
- v. Statut de l'entrée : (terme) vérifié ou non, publié ou non, finalisé ou non ;
- vi. Choix du format standard : multilingue, bilingue ou monolingue ;
- vii. Historique des entrées : enregistrement de chaque modification apportée à chaque entrée (avec option d'annulation des modifications) ;
- viii. Gestion des privilèges utilisateurs : il existe cinq types d'utilisateurs (lecteur, éditeur, créateur, administrateur, configurateur) ;
- ix. Gestion d'extranets : la plateforme dispose des réseaux télématiques locaux qui permettent aux utilisateurs externes de consulter un ou plusieurs termes et de laisser des commentaires sur les termes sélectionnés.

TERMINOLOGUE est actuellement disponible en irlandais, en anglais, en tchèque, en suédois, en gallois, en finnois, en russe, en espagnol, en arabe, en croate, en lituanien, en norvégien et en grec (l'ordre d'apparition des versions localisées est tel qu'indiqué sur la plateforme¹³, où – approche interculturelle et respect de l'altérité obligent – chaque langue est désignée par son nom dans l'idiome concerné).

Les domaines de spécialité dont relèvent les bases de données terminologiques hébergées sur cette plateforme sont eux aussi des plus variés : administration publique, sport, santé publique et technologies de l'information, ainsi que diverses matières scolaires. TERMINOLOGUE peut être utilisé par des terminologues qui travaillent individuellement ou en groupe et peut être exploité dans le domaine de l'éducation.

Le 23 mars dernier, nous avons eu l'occasion de participer à un séminaire sur les études terminologiques et la normalisation des termes, où nous avons été rejoints sur Zoom par Jamie Murphy, administrateur de TERMINOLOGUE, qui a fait une présentation des principes de développement et des fonctionnalités de cet outil de gestion terminologique, ce qui nous a réellement aidé.e.s à comprendre les tenants et les aboutissants de ce projet si généreux.

Pourquoi avoir dit que TERMINOLOGUE serait un outil « pas comme les autres » ? Pour l'essentiel (et mis à part ses fonctionnalités, que nous venons d'évoquer), en raison de sa vocation collaborative (et de sa licence), ainsi que de son ouverture interculturelle, dont témoignent toutes ces localisations.

¹³ “About *Terminologie*”, <https://www.terminologie.org/docs/info.en/> (dernière consultation : 30.06.2023).

TERMINOLOGUE est un logiciel à *code source ouvert* qui est en même temps un *vrai* logiciel *libre*, non pas au sens trivial (la plupart des logiciels *open source* correspondent aux critères de licence des logiciels libres), mais au sens du mouvement des logiciels libres, « un mouvement qui lutte pour la liberté et la justice » et « fait campagne pour la liberté des utilisateurs de l’informatique »¹⁴.

Si on vérifie l’extension de la licence citée dans la présentation du produit, on remarquera qu’elle accorde la permission de « traiter le logiciel sans restriction » : on a le droit de l’utiliser, de le copier, de le modifier, de le fusionner, de le publier, de le distribuer, etc., sous réserve de mentionner correctement le titulaire du droit d’auteur (*copyright*) et l’avis d’autorisation qui dégage celui-ci de toute responsabilité pour l’usage qui en sera fait.

Le logiciel libre (angl. *free software*) est en effet un logiciel qui respecte les libertés essentielles de l’utilisateur, autrement dit, l’utilisateur est libre de l’utiliser, de le modifier, de l’améliorer et d’en redistribuer des copies. En raison du terme *libre*, sa signification est souvent mal comprise, mais, de fait, *logiciel libre* devrait faire penser à la *liberté d’expression* et non à l’*entrée libre*¹⁵.

Le logiciel à code source ouvert (angl. *open source software*) est, par contre, un logiciel qui offre des avantages plutôt pratiques, dans la mesure où il peut être modifié selon les besoins de l’utilisateur et redistribué sans règlement de frais pour violation de la licence, tout en garantissant des conditions de maintenance, de compatibilité technique et de sécurité supérieures à celles de la plupart des logiciels libres. Le code source de ces logiciels est mis à disposition du public dans le cadre des licences GNU GPL qui en autorisent l’édition et la redistribution. Le concept et le terme de *logiciel libre* ont été créés par Richard Stallman en 1980 et popularisés par les campagnes de la FSF (*Free Software Foundation*). Le concept et le terme de *logiciel à code source ouvert* ont été proposés à la fin des années 1990, sous l’impulsion de Bruce Perens et Eric Raymond (dans le cadre de leur organisation *Open Source Initiative*), en réaction aux restrictions imposées par les logiciels libres¹⁶.

Il existe de nombreux sites de vulgarisation qui abordent la question des différences et des ressemblances entre ces deux types de logiciels (voir note précédente pour des exemples), qui représentent, au même titre, une alternative

¹⁴ Richard Stallman, « En quoi l’open source perd de vue l’éthique du logiciel libre », document consulté en ligne, <https://www.gnu.org/philosophy/open-source-misses-the-point.fr.html> (dernière consultation : 30.06.2023).

¹⁵ *Ibidem*.

¹⁶ Voir <https://www.geeksforgoeks.org/difference-between-free-software-and-open-source-software/> (site en anglais). Voir aussi, en français : *Logiciel « libre » et « open source » sont-ils synonymes ?* (<https://insolus.com/blog/difference-open-source-logiciel-libre>), excellente synthèse truffée de détails historiques.

aux logiciels propriétaires. Richard Stallman, principal promoteur des logiciels libres, considère, lui aussi, que « l'ennemi » n'est pas *l'open source*, mais « le logiciel privé »¹⁷.

Les deux termes sont souvent considérés synonymes, bien que les tenants du code source ouvert soulignent plutôt les avantages techniques qu'apporte le « code ouvert », et les tenants du mouvement des logiciels libres, les différences d'idéologie qui les sous-tendent. Dans son article traduit en français, Richard Stallman conclut à cet égard que « l'habitude de dire et penser *open source* fait obstacle à la compréhension de la philosophie du mouvement du logiciel libre et empêche d'y réfléchir »¹⁸.

Quant à nous, nous trouvons que TERMINOLOGUE répond parfaitement aux *desiderata* du mouvement des logiciels libres, tout en présentant les avantages techniques des logiciels à code source ouvert. Créé en anglais et en irlandais et déjà localisé (comme nous venons de le voir) pour onze autres langues, grâce à nos efforts, cet outil « pas comme les autres » sera bientôt accessible en français et en roumain également.

3. Méthodologie de la recherche d'équivalents interlinguaux pour les unités terminologiques et phraséologiques

Comme déjà évoqué dès l'introduction, notre projet de localisation a consisté dans la traduction de l'anglais vers le français et vers le roumain de deux types de textes-sources : 412 « séquences de caractères » (*strings*) correspondant aussi bien à des termes isolés (simples ou complexes) qu'à des collocations spécialisées, à des syntagmes ou autres séquences textuelles sous-phrastiques ou même à de courtes phrases (commandes, instructions ou messages courts pré-formatés) ; et deux textes en discours suivi, à hauteur respectivement de 458 et de 1959 mots (soit 2417 mots). Les séquences nous ont été livrées dans deux tableaux Excel (l'un à 95 séquences et l'autre, à 317 séquences) ; les deux textes en discours suivi nous ont été livrés dans des fichiers Markdown (extension .md). Les séquences représentent un inventaire terminologique et phraséologique exhaustif en référence à la fois aux textes en discours suivi et aux boutons ou commandes sur la plateforme. Nous avons réalisé des recherches terminologiques et phraséologiques au premier degré en vue de traduire les séquences (*strings*).

Nous avons systématiquement suivi le même algorithme de recherche d'équivalents interlinguaux terminologiques et phraséologiques (« unités terminologiques » au sens large), algorithme comprenant quatre étapes. Dans un

¹⁷ R. Stallman, « En quoi... », art. cit., <https://www.gnu.org/philosophy/open-source-misses-the-point.fr.html> (dernière consultation : 30.06.2023).

¹⁸ *Idem*.

premier temps, les phrases et les sous-phrases de TS ont été traduites en utilisant le moteur de traduction automatique *DeepL*. Pour les termes isolés (sans contexte minimal en tant que TS), nous avons recherché sur Internet des contextes d'attestation pertinents en anglais (recherche de mots clés) et nous les avons fournis comme TS au moteur de traduction automatique. Dans un deuxième temps, si la traduction proposée par la machine à l'étape précédente a été très usuelle, car employée dans des textes comparables au produit logiciel à localiser (sites ou applications), nous l'avons retenue dans le fichier d'équivalences *xlsx*. Dans tous les cas où la variante *DeepL* n'a pas été évidente, nous avons procédé à une recherche permettant la post-traduction. Nous avons interrogé les variantes machine en cliquant sur un mot de la séquence en question et nous avons validé celle ou celles que nous avons retenue(s) par des recherches guillemetées sur Google. Pour certains termes, nous avons recherché un équivalent sur IATE (*Interactive Terminology for Europe*). Nous avons respecté trois critères de choix de l'équivalent en langue cible (LC) : le nombre d'attestations, la qualité des contextes d'attestation et les critères de bonne formation des termes selon la norme ISO 704 : 2009, en particulier la préférence donnée aux structures qui privilégient la LC. Nous avons gardé à l'esprit le fait que les termes créés par dérivation, par composition ainsi que les formes abrégées en LC sont préférés aux calques structurels, qui sont à leur tour préférés aux emprunts. Finalement, nous avons inscrit et sauvé la variante choisie dans le tableau sous le document *xlsx* concerné. En résumé (et schématiquement) :

- TS anglais : phrases/ sous-phrases (combinatoire libre, collocations)/termes (simples, complexes)
- i. Version *DeepL*
 - phrases et sous-phrases : TS tel quel ;
 - termes : chercher un contexte (phrase) d'attestation pertinent en anglais et le fournir comme TS à la TA. Exemple : pour le terme anglais *password*, *A password is intended to be secret (known only by an authorized user)*, *DeepL* a proposé : *Un mot de passe est censé être secret (connu uniquement par un utilisateur autorisé)*.
 - ii. Si cette version est très commune (bien connue du traducteur, comme étant employée dans des textes comparables : sites ou applications), on la retient dans le fichier d'équivalences sous *xlsx*.
 - iii. Si par contre, cette variante n'est pas évidente, on continue la recherche (post-édition) :
 - interrogation des variantes machine (en cliquant sur un mot de la séquence en question) ;
 - et validation de la ou des variantes sélectionnées par des recherches guillemetées sur Google ;
 - recherche d'un équivalent sur IATE.
 - iv. Critères de choix du meilleur équivalent en LC :
 - le nombre d'attestations ;
 - la qualité des textes/contextes : à évaluer sur les 10 à 20 premiers résultats (sont privilégiés les contextes d'attestation comparables au texte source) ;

- critères de bonne formation des termes (ISO 704 : 2009), en particulier la préférence donnée aux structures qui privilégient la LC (« la langue en question ») : emprunt < calque (structurel) < termes créés par dérivation, composition, formes abrégées, etc. (nouvelles formes en LC).
- v. Inscription et sauvegarde de la variante de notre choix dans le tableau sous xlsx.

Selon le site web officiel¹⁹, *DeepL* est un moteur de traduction automatique neuronale qui utilise le *deep learning* ou l'*apprentissage profond*. Il est apparu en 2009 quand l'entreprise était connue sous le nom de *Linguee*. Nous sommes bien familiarisées avec cette mémoire de traduction sur laquelle la TAN continue d'être entraînée à présent. *DeepL* figure parmi les 200 sites web les plus utilisés en France, en Allemagne, en Espagne et dans d'autres pays. Parmi les fonctionnalités disponibles de la version gratuite de *DeepL* : la possibilité de choisir des options de traduction alternatives pour certains mots en cliquant simplement sur eux, le dictionnaire intégré disponible à tout usager et la possibilité de traduire des documents entiers (nombre de mots limité). Le moteur de traduction offre aussi une option payante, encore plus performante et sans limitation de taille des fichiers. La popularité de ce moteur de traduction automatique a augmenté en 2018 quand, en partenariat avec la startup française Quantmetry, *DeepL* a traduit 800 pages d'un livre écrit par trois chercheurs en intelligence artificielle (IA), décrivant des réseaux de neurones artificiels à apprentissage profond. Le moteur a fait très peu d'erreurs, ce qui a nécessité un taux de post-édition inhabituellement bas.

La TAN aura ainsi fait économiser aux humains non seulement du temps, mais aussi des ressources financières, l'entreprise s'étant concrétisée par la publication d'un manuel en français sous l'intitulé *L'apprentissage profond*, version qui a été publiée le 2 septembre 2018 par Massot Editions et Quantmetry²⁰. Nous avons utilisé pour notre projet la version publique du moteur de TAN.

Quelques exemples illustratifs feront mieux ressortir les avantages de notre algorithme de recherche d'équivalents interlinguaux. Nous nous arrêterons ici sur une collocation et sur une formule susceptible d'être employée comme étiquette de bouton de site.

Exemple 1 :

TS anglais : *Gentle introduction to (TERMINOLOGUE)*

TC français : *Introduction en douceur à (TERMINOLOGUE)*

¹⁹ <https://www.deepl.com/en/press-release> (dernière consultation : 30.06.2023).

²⁰ Julie Rousseau, « Première en France : un manuel de deep learning traduit par l'IA », document consulté en ligne, https://www.sensemaking.fr/Premiere-en-France-un-manuel-de-deep-learning-traduit-par-l-IA_a330.html (dernière consultation : 30.06.2023).

Pour choisir une variante de traduction de l'anglais *Gentle introduction (to)* en français, premièrement, nous avons eu recours au moteur de TAN *DeepL* qui a livré, pour un contexte similaire mais distinct (et très bien attesté sur Internet : *Gentle introduction to Word*) la variante *Introduction en douce*. Cette séquence, nous l'avons soumise à recherche guillemetée sur Google ; la recherche a débouché sur 2.080 attestations. Mais à y regarder de plus près, les dix premiers résultats ont systématiquement illustré le tour *introduction de...* et non *introduction à...*, la chose introduite étant soit concrète (introduction en douce de notes et de matériels de documentation à une épreuve d'examen), soit abstraite (introduction en douce de nouvelles taxes) – *en douce* signifiant dans tous ces contextes d'attestation « en cachette ». Nous en avons conclu que la version « introduction en douce » était sémantiquement inadéquate au contexte source.

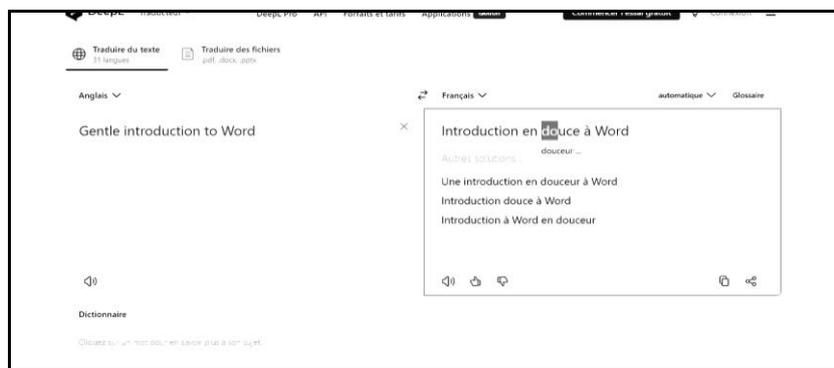


Figure 1

Puis nous avons choisi de cliquer sur le mot *douce*, pour voir d'autres synonymes proposés par *DeepL*, et nous avons retrouvé *douceur* et la variante *Introduction en douceur* (voir Figure 1 *supra*) ; c'est là une autre transposition nominale de l'anglais *gentle*, dont la recherche guillemetée a débouché sur presque cinquante fois davantage d'attestations – 93.200 résultats. Cette variante affiche deux types d'emplois, dès les dix premiers résultats : le sens concret dans le domaine technique (*introduction en douceur de qqch*), qui ne correspond pas à notre contexte source²¹, mais aussi *introduction à qqch* – avec des occurrences dans le domaine de l'informatique, par exemple *Introduction en douceur à HTML/CSS*²².

Nous avons en outre pensé à une équivalence de traduction (au sens de la stylistique comparée – traduction par une formule en contexte spécifique comparable) du type : *introduction pour les nuls*, mais cette variante n'est pas très

²¹ <https://glutz.com/ch/fr/product/2056> (dernière consultation : 30.06.2023).

²² <http://www.lespetitescases.net/introduction-en-douceur-a-html-css> (dernière consultation : 30.06.2023).

bien attestée en soi, on a d’habitude *X pour les nuls*, où *X* est un sujet ou un produit spécifique – ce qui donnerait, dans notre cas, *TERMINOLOGUE pour les nuls*. Mais, bien que *pour les nuls* soit attesté en bon français, y compris à propos de produits logiciels tel Windows, etc., et comme les auteurs de l’introduction en question auraient pu choisir en anglais aussi l’équivalent interculturel *for dummies*, et qu’ils ne l’aient pas fait, nous avons préféré *Introduction en douceur* à *TERMINOLOGUE*, afin d’éviter toute connotation fâcheuse pour les usagers du produit terminologique²³.

Voyons maintenant le traitement du même TS anglais, lors de la traduction en roumain.

TS anglais (séquence 2/317) : *Gentle introduction to TERMINOLOGUE*

TS soumis à *DeepL* : *Gentle introduction*

Variante *DeepL* n roumain : *Introducere blândă/delicată/ușoară* (voir capture d’écran sous Figure 2 ci-contre) :

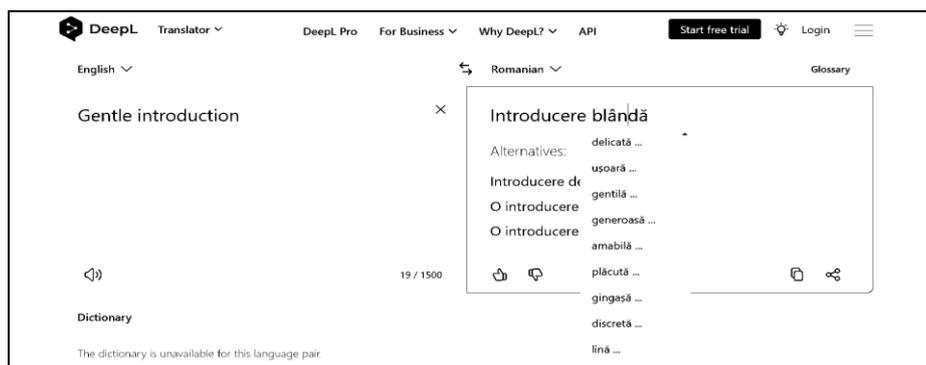


Figure 2

Nous avons constaté que les variantes de *DeepL* ne sont pas acceptables dans le contexte des nouvelles technologies de l’information. Donc nous nous sommes appuyées sur des procédés de traduction indirecte, en l’occurrence, sur la modulation. Nous avons opté pour l’équivalence *introducere pas cu pas* (‘introduction pas à pas’) qui a plus de 10 000 résultats sur Google (recherche

²³ Une discussion entre internautes à propos de la remise en forme d’un article très technique sur Wikipédia (un sujet de maths et de programmation) fait justement ressortir ce type de problèmes : les intervenants se proposent d’éviter de transformer l’article de l’encyclopédie collaborative en une simple « introduction pour les nuls » pour privilégier l’idée d’une « introduction gentille ». Voir « Discussion : optimisation linéaire », https://fr.wikipedia.org/wiki/Discussion:Optimisation_lin%C3%A9aire (dernière consultation : 30.06.2023).

guillemetée) – une formule très bien attestée dans des textes d’informatique, comme le montre l’exemple suivant :

- (1) *Khan Academy oferă un curs complet de programare web, care include module de HTML și CSS. Acesta se adresează începătorilor și vă va oferi o introducere pas cu pas în aceste limbaje*²⁴.

En français, la recherche guillemetée de *introduction pas à pas* ne débouche que sur un peu plus de 2000 résultats, mais les contextes informatiques restent bien représentés, nous sommes donc effectivement en présence d’une modulation par le synonyme :

- (2) *Introduction pas à pas de Python*²⁵

Exemple 2 :

TS anglais : *Forgot your password?*

TC français : *Vous avez oublié votre mot de passe ?*

Pour choisir une variante de traduction de *Forgot your password?*, nous avons d’abord cherché une variante de traduction sur *DeepL*, qui nous a proposé deux possibles solutions (v. Figure 3 ci-après) :

- (3) *Vous avez oublié votre mot de passe ?*
 (4) *Mot de passe oublié ?*



Figure 3

²⁴ Marian Ștefănescu, “Tutorial HTML pentru începători. Istoria limbajului și exemple practice”, document consulté en ligne : <https://code-it.ro/tutorial-htmlpartea-1-ce-este-html/> (dernière consultation : 30.06.2023).

²⁵ « Introduction pas à pas de Python – Documentation PythonScript 0.0.1. », <https://pythonscript.readthedocs.io/intropython.html> (dernière consultation : 30.06.2023).

Nous avons ensuite effectué une recherche guillemetée afin de vérifier le nombre d’attestations de chaque variante avec ses contextes d’utilisation et nous avons observé que, même si la deuxième variante (phrase nominale passive) a plus d’attestations (56.500.000) que la première (4.580.000), la variante *Mot de passe oublié* est souvent utilisée sans point d’interrogation, tandis que la première séquence (phrase active) est toujours utilisée avec un point d’interrogation (en tant que question déclarative). Voir à cet égard l’exemple suivant, retrouvé sur le site de *Microsoft*²⁶ – capture d’écran présentée sous Figure 4 :

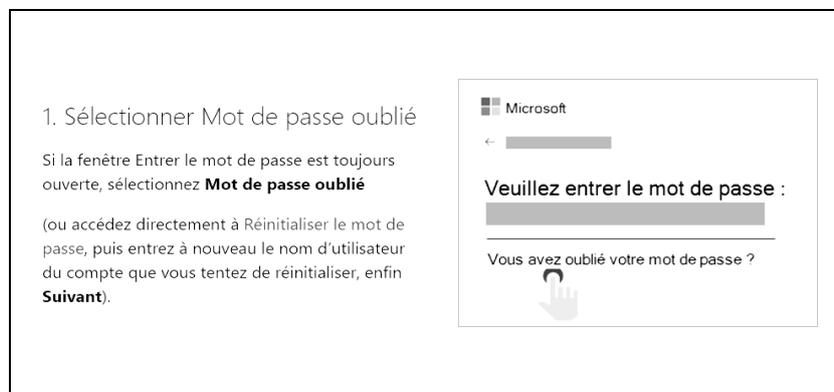


Figure 4

Nous avons donc choisi la première variante, qui est systématiquement l’équivalent contextuel de la phrase-source, utilisée pour étiqueter le même bouton sur les sites Internet concernés.

Notons d’autre part que la phrase source anglaise comportait à première vue une ellipse. Si la source de l’ellipse était la question directe canonique *Did you forget your password?*, cela aurait donné *Forget your password?*, or notre séquence source comporte, elle, une forme de passé, *Forgot*, sans sujet explicite.

En tout cas, cette formulation familière, typiquement orale, exploitée par le style « logiciel » en anglais devrait encourager le choix de la question française la plus courte (*Mot de passe oublié ?*). La modulation de voix l’emporterait alors sur la dilution (segment plus long en langue cible). C’est dire à quel point le choix du traducteur-localisateur est peu aisé.

Lors du traitement de la même séquence vers le roumain, *DeepL* proposera (dans l’ordre) trois variantes, dont les deux premières relèvent de la contraction en

²⁶ « Réinitialiser un mot de passe de compte Microsoft », <https://support.microsoft.com/fr-fr/account-billing/r%C3%A9initialiser-un-mot-de-passe-de-compte-microsoft-eff4f067-5042-c1a3-fe72-b04d60556c37> (dernière consultation : 30.06.2023).

utilisant des temps composés et la dernière, de l'implication (*your* n'y est plus explicite) :

- (5) *V-ați uitat parola?*
- (6) *Ți-ai uitat parola?*
- (7) *Ați uitat parola?*

Seules (5) et (7), situées dans le registre formel avec l'utilisation de la deuxième personne du pluriel, sont attestées en tant qu'étiquettes de boutons de site. À ce titre, ce sont des énoncés formulaires, et on peut envisager ces TC comme résultats d'équivalences – au sens de la stylistique comparée (procédé de traduction spécifique).

Des recherches guillemetées sur Google nous ont permis de vérifier que le choix de l'une ou de l'autre formule est, en grande partie, une question de lecture d'entreprise : Apple est plus inclusif et opte pour le datif possessif ; Microsoft, plus neutre, choisit la variante sans datif possessif :

- (8) *Dacă ai un dispozitiv Apple nou sau un dispozitiv Apple pe care nu te-ai autentificat cu ID-ul Apple, poți selecta „V-ați uitat parola sau nu aveți un ID Apple?” oriunde apare această opțiune. (Apple)²⁷*

Si tu as un nouvel appareil Apple ou un appareil Apple auquel tu ne t'es pas connecté avec l'identifiant [« ton identifiant »] Apple, tu peux sélectionner « Vous avez oublié votre mot de passe ou vous n'avez pas d'identifiant Apple ? » où que cette option apparaisse.

- (9) *Dacă fereastra **Introduceți parola** este încă deschisă, selectați **Ați uitat parola?** (Microsoft)²⁸*

Si la fenêtre **Introduisez la parole** est encore ouverte, sélectionnez **Vous avez oublié la parole ?**

Le taux d'attestation des deux séquences (presque 5,7 millions d'attestations pour (5) et 1,7 millions pour (7) – d'où l'on peut déduire²⁹ que la formule polie sans datif possessif recueille 2 millions d'attestations de plus) n'est pas en soi révélateur, puisque l'immense majorité des occurrences sont en combinatoire libre et en discours suivi – les emplois conditionnels-hypothétiques

²⁷ “Dacă ai uitat parola asociată ID-ului tău Apple – Apple Support (RO)”, <https://support.apple.com/ro-ro/HT201487> (dernière consultation : 30.06.2023).

²⁸ “Resetarea unei parole uitate pentru contul Microsoft – Asistență Microsoft”, <https://support.microsoft.com/ro-ro/account-billing/resetarea-unei-parole-uitate-pentru-contul-microsoft-eff4f067-5042-c1a3-fe72-b04d60556c37> (dernière consultation : 30.06.2023).

²⁹ Vu que le moteur de recherche n'est pas sélectif et ignore les majuscules, les sélections de la séquence (5) incluent celles de (7).

dominant le corpus : *în caz că...* (« au cas où »), *dacă...* (« si »), tout comme la juxtaposition de la question directe à la phrase décrivant l'action à entreprendre (si cette éventualité en vient à être réalisée). (6) est, paraît-il, réservé au discours suivi.

Quoi qu'il en soit, la hiérarchie des options *DeepL* s'avère cette fois-ci juste en termes qualitatifs, le tour à datif possessif étant nettement plus colloquial en roumain contemporain ; l'appariement stylistique du TS anglais et du TC roumain est donc parfait, au choix de la personne près, car l'ambiguïté de *your* en anglais peut prêter le flanc à discussion : vouvoiement ? tutoiement ?

4. Conclusions

La localisation de TERMINOLOGUE en français et en roumain a mis en évidence certaines tendances.

La propension du roumain vers l'emprunt (perméabilité plus grande aux anglicismes) fait pendant à la vocation (normative) du français à l'éviter : *open source software* (anglais) est rendu en français par (*le*) *logiciel à code-source ouvert* et en roumain par *soft(ul) open source*, ou carrément par *software open source* ; le bouton *set to today* est étiqueté *AUJOURD'HUI* en français, mais toujours *TODAY* en roumain (site de Microsoft)³⁰.

La proportion relative d'unités terminologiques rendues en roumain par des procédés de traduction directe (emprunt, calque, paraphrase littérale) est en général plus importante que celle des unités rendues pareillement en français, mais la traduction indirecte reste dominante dans les deux LC.

Les frontières entre équivalences terminologiques et équivalences de traduction s'avèrent assez poreuses quand il est question de séquences à usage formulaire (étiquettes de boutons). Même lorsqu'une telle séquence fait massivement l'objet d'emplois non formulaires en combinatoire libre et discours suivi, elle se comportera comme une unité terminologique (phraséologique) en tant qu'étiquette de bouton.

Références bibliographiques

- Bhreathnach, Úna (2020) – “A platform for open-source terminology management: Terminologue”, in *International VAKKI Symposium 2020 - Workplace Communication III*, February 6-7, Vaasa, Finland. https://www.uwasa.fi/en/events/vakki_symposiumi_2020
- Loffler Laurian, Anne-Marie (1996) – *La Traduction automatique*, Villeneuve d'Ascq, Presses universitaires du Septentrion.
- Robert, Anne-Marie (2010) – « La post-édition : l'avenir incontournable du traducteur ? », *Traduire* [En ligne], n° 222. URL : <http://journals.openedition.org/traduire/460> ; DOI : 10.4000/traduire.460

³⁰ “TODAY (funcția TODAY) – Asistență Microsoft”, document consulté en ligne : <https://support.microsoft.com/ro-ro/office/today-funcția-today-5eb3078d-a82c-4736-8930-2f51a028fdd9> (dernière consultation : 30.06.2023).

- Stalman, Richard (s.d.) (2022) – « En quoi l’open source perd de vue l’éthique du logiciel libre », traduction Mathieu Stumpf, document consulté en ligne : <https://www.gnu.org/philosophy/open-source-misses-the-point.fr.html#:~:text=La%20philosophie%20de%20>.
- Velicu, Anca-Marina (2022) – *Traduire des textes fonctionnels*, București, Pro Universitaria.
- Vinay, Jean-Paul, Darbelnet, Jean – 1972/1958, *Stylistique comparée du français et de l’anglais. Méthode de traduction*, Paris, Didier (Nouvelle édition revue et corrigée), document consulté en ligne : <https://archive.org/details/VinayDarbelnetStylistiqueCompareDuFranaisEtDeLanglais>