À tradire
ISSN : 2968-3912

1 | 2022
L'apprenant de et par la traduction

Le fonctionnement de la traduction automatique neuronale

Une traduction commanditée

Sébastien Palmieri and Romain Revet

https://lodelpreprod.univ-rennes2.fr/atradire/index.php?id=193

Electronic reference
Sébastien Palmieri and Romain Revet, « Le fonctionnement de la traduction automatique neuronale », *À tradire* [Online], 1 | 2022, Online since 20 December 2022, connection on 29 September 2026. URL : https://lodelpreprod.univ-rennes2.fr/atradire/index.php?id=193

Copyright
Licence Creative Commons – Attribution 4.0 International – CC BY 4.0

Le fonctionnement de la traduction automatique neuronale

Une traduction commanditée

Sébastien Palmieri and Romain Revet

OUTLINE

Introduction
1. Les difficultés rencontrées
  1.1. Trados Studio 2021 et le format PDF : une idylle imparfaite
    Segmentation
    Mise en page
  1.2. Les difficultés d'un texte anglais non écrit par des locuteurs natifs
  1.3. Public ciblé et difficultés liées aux six chapitres précédents (non encore traduits)
  1.4. C'est la fête du qualificatif
    Anglicismes
    Traduction des exemples
    Traduction de la bibliographie
    Adaptation du style des auteurs en français
  1.5. Une terminologie non exhaustive
    Input et *output* des termes passe-partout
    Inputs
    Output
  1.6. Mise en page et éléments graphiques
    Traduction et adaptation des éléments graphiques
    Harmonisation stylistique
    Harmonisation terminologique
    Harmonisation factuelle
2. Ressources terminologiques
Sources

TEXT

Introduction

1 Dans le cadre du projet de traduction commanditée du Master TSM de l'université Grenoble-Alpes et sous la direction de Caroline Rossi, nous avons élaboré une traduction du chapitre 7 *How neural machine*

translation works (rédigé par Juan Antonio Pérez-Ortiz, Mikel L. Forcada et Felipe Sánchez-Martínez de l'Université d'Alicante, Espagne) de l'ouvrage *Translation for Multilingual Citizens*, qui est l'un des résultats du projet MultiTraiNMT (https://multitrainmt.eu) cofinancé par le programme Erasmus de l'Union européenne. Cette traduction était une commande du partenariat MultiTraiNMT, qui réunit quatre universités européennes (Universitat Autònoma de Barcelona, Université de Grenoble-Alpes, Dublin City University et Universitat d'Alacant), deux entreprises (Prompsit Language Engineering et Xcelerator Machine Translations) ainsi que plus de vingt autres membres et a pour but de « développer, d'évaluer et de diffuser des contenus en libre accès afin d'améliorer l'enseignement et l'apprentissage de la traduction automatique chez les apprenants en langues, les professeurs de langues, les traducteurs stagiaires, les professeurs de traduction et les traducteurs professionnels, dans toute l'Europe »[1].

2 Le chapitre que nous avons traduit présente à ses lecteurs les grands principes de la traduction automatique neuronale, un modèle relativement récent qu'utilisent des systèmes gratuits bien connus du grand public comme DeepL ou encore Google Translate.

1. Les difficultés rencontrées

1.1. Trados Studio 2021 et le format PDF : une idylle imparfaite

3 La première difficulté rencontrée, et pas des moindres à mon sens, est venue du fait que l'importation d'un document source au format PDF dans l'interface du logiciel de TAO Trados Studio 2021 a causé de nombreuses erreurs au sein du document. Ces erreurs se répercutant forcément dans le document cible sans intervention directe du traducteur.

Segmentation

4 Certaines parties de phrases n'ont tout simplement pas été « transférées » lors de la segmentation du texte et sont devenues manquantes dans le texte source intégré dans l'interface de Trados Studio 2021.

Entrainant mon incompréhension et m'obligeant à effectuer des recherches dans le document PDF original pour y voir plus clair et procéder aux corrections nécessaires.

Mise en page

5 La mise en page a également fait les frais de l'intégration sous Trados Studio 2021 :

- Certaines notes de bas de page n'étaient plus considérées comme telles et avaient été transférées dans le « corps du texte ». Ainsi, Après l'enregistrement au format Word, cette note de bas de page est devenue un paragraphe de corps de texte, placée seule en haut d'une page vierge.

Fig. 1



6 Il a donc fallu procéder à une correction de la mise en page afin de remettre en place de façon correcte certains éléments comme les légendes et les notes de bas de page, ou encore la bibliographie.

Fig. 2



7 Après correction de la mise en page, la note de bas de page est revenue à la place qui lui est destinée et n'est plus un paragraphe de corps de texte.

- Les styles gras ou italique n'ont pas forcément été restitués sur l'intégralité d'une phrase ou d'un titre.
  Ex. : **3 Réseaux neuronaux artificiels →** 3 Réseaux **neuronaux artificiels**
- Certaines numérotations de chapitres ont purement et simplement disparu.
  Ex. : **3.3 Couches de neurones → Couches de neurones**
- Certaines numérotations de page se sont retrouvées en plein milieu d'une page, mais pas dans le corps du texte.

- Certaines structures de schémas ont dû être retouchées, dans la mesure du possible, pour que le schéma cible soit similaire au modèle source.

8 Malheureusement, repérer et corriger toutes ces erreurs pour pouvoir, à terme, fournir un document en langue cible bénéficiant d'une mise en page à peu près identique au document source s'est avérée fastidieux et extrêmement chronophage.

1.2. Les difficultés d'un texte anglais non écrit par des locuteurs natifs

9 L'une des difficultés rencontrées également lors de cette traduction provenait du fait que le document source n'avait pas été rédigé par des locuteurs natifs anglais. Cela nous a gêné en particulier lorsque nous étions face à des structures de phrases compliquées du point de vue de l'anglais. Pour faire écho au cours de *Rédaction de contenu spécialisé en anglais* de M. André Dodeman, professeur à l'université Grenoble-Alpes, nous avons eu le sentiment qu'un locuteur natif dans la langue de Shakespeare n'aurait pas formulé les choses de cette façon. Nous avons donc été amenés à procéder à de nombreuses reformulations tout au long du texte pour essayer de rendre la lecture du document cible plus naturelle et plus agréable.

1.3. Public ciblé et difficultés liées aux six chapitres précédents (non encore traduits)

10 Bien que la majeure partie du chapitre se veuille pédagogique en introduisant pas à pas les concepts, termes, et explications correspondantes, certains passages abordaient un certain nombre d'idées ou de représentations avec un degré de technicité certain. Or, sans avoir à notre disposition les six premiers chapitres de l'ouvrage, nous avons parfois eu du mal à « cibler » le lectorat visé. Fallait-il parfois essayer de vulgariser et apporter des précisions supplémentaires à certaines explications comme nous l'a enseigné Mme Alice Carré, professeure à l'université Grenoble-Alpes, lors de son cours *Traduction technique d'anglais en français* ? Ou bien partir du principe que le lectorat visé dispose déjà de certaines connaissances sur le sujet (acquises préalablement ou lors de la lecture des chapitres précé-

dents de l'ouvrage) et maintenir le niveau technique de la phrase source ?

11 Quoi qu'il en soit, bien que nous ayons fait le choix d'essayer de restituer les propos des auteurs aussi clairement et fidèlement que possible pour le lecteur sans nous « approprier » le texte, nous ne regrettons pas de ne pas avoir préalablement interrogé le commanditaire de la traduction à ce propos. En effet, alors que nous écrivons ces lignes avec le recul nécessaire, nous avons bien conscience que notre analyse est tout autant basée sur l'expérience acquise lors de la traduction en elle-même et l'enseignement universitaire dispensé ce semestre, que sur la réflexion menée lors de la rédaction de cette partie introductive. Toutefois, l'absence de regret ne veut aucunement dire que nous referions les choses de la même façon si l'occasion se présentait. À l'avenir, nous chercherons sûrement à mieux définir « le cadre » de la traduction afin de nous assurer de mieux répondre aux attentes du commanditaire, voire essayer d'obtenir, dans le cas d'une partie d'ouvrage ou autre type de document à traduire, la totalité du ou des documents en question.

1.4. C'est la fête du qualificatif

12 Toujours au niveau des difficultés de traduction rencontrées dans ce chapitre 7, il y a un cas découlant des particularités de la langue anglaise que nous avons eu envie d'aborder ici. En effet, à la différence de la langue française, la souplesse de l'anglais permet d'enchainer les qualificatifs avec une aisance certaine et très peu de contraintes.

> *"Supercomputers were used in order to train the GPT-3 system, a process that can take several weeks or even months, but it has been estimated that learning the weights for such a model with a single powerful gaming desktop personal computer would have taken more than 350 years."*

13 Pour la traduction de cette phrase et pour une raison évoquée ci-dessous, nous avons choisi de « dissocier » *powerful* et *gaming* de la liste de qualificatifs, et tâché de les retranscrire d'une façon détournée.

> « Plusieurs superordinateurs ont été utilisés pour cette tâche, un processus susceptible de prendre plusieurs semaines, voire plusieurs mois. En effet, on estime que l'apprentissage des poids d'un tel modèle à l'aide d'un seul ordinateur personnel de bureau suffisamment puissant pour jouer[2] aurait pris plus de 350 ans. »

14 Notre décision de dissocier le mot *gaming* utilisé dans la phrase anglaise est basée sur la réflexion menée quant au choix de sa traduction. En effet, ce mot ainsi que bon nombre d'anglicismes (https://gamingcampus.fr/boite-a-outils/lexique-du-jeu-video-100-mots-du-jeu-video.html) sont régulièrement utilisés par la communauté française de joueurs de jeux vidéo, dont nous faisons partie.

15 Ainsi, bien que cela ne soit pas très correct du point de vue de la langue française, la communauté parle volontiers de *gaming*, de *gameplay*, de *game design*, de *gamers*, de FPS (*First Person Shooter*), etc. Pour ce dernier terme, il est intéressant de noter que même si l'usage de sa traduction par *jeu de tir à la 1re personne* s'est répandu, c'est l'acronyme FPS du mot anglais qui reste pourtant le plus employé.

16 Pour en revenir au mot *gaming*, nous avons hésité à le conserver tel quel dans un premier temps. En effet, c'est un anglicisme extrêmement utilisé et vous trouverez pléthore d'annonces sur des sites marchands de vente en ligne (Amazon, Fnac, etc.) pour des *fauteuils gamer/gaming*, des *accessoires gaming*, des *ordinateurs gamer/gaming*, etc. De plus, bien que la plupart des traductions proposées pour le vocabulaire (https://www.oqlf.gouv.qc.ca/ressources/bibliotheque/dictionnaires/20120701_jeu_video.pdf) du jeu vidéo aient déjà eu le droit à une proposition validée par l'Office québécois de la langue française (https://www.oqlf.gouv.qc.ca/accueil.aspx), celles-ci n'ont généralement pas été adoptées par la communauté ; l'utilisation de nombre d'entre elles ne semblant pas très naturelle du point de vue de l'usage de la langue. Nous avons finalement fait le choix de proposer une traduction pour le mot *gaming* en le séparant de la liste de qualificatifs et en utilisant une référence indirecte : *suffisamment puissant pour jouer.*

Anglicismes

17 Le texte source contenait bien d'autres termes donnant souvent lieu à des anglicismes en français : nous avons fait le choix (peut-être à tort) de conserver le terme *smartphone* dans la traduction française. Certaines traductions comme *téléphone intelligent* ou bien encore *ordiphone* existent certes, mais pour être honnête, nous avons l'impression que quasiment personne ne les utilise. Nous avons hésité également à tout simplement employer le terme *téléphone portable* ou encore *mobile (multifonction)* qui semble être la recommandation officielle d'après le *Larousse* mais, à l'heure actuelle, tous les téléphones portables ne sont pas des smartphones, et le terme *mobile (multifonction)* n'est pas très parlant. Nous avons donc décidé de conserver le terme tel quel afin de ne pas perturber le lecteur de l'ouvrage.

Traduction des exemples

18 Ce chapitre comporte également plusieurs exemples de phrases subissant une opération de traduction automatique. Si certains de ces exemples peuvent être traduits avec facilité comme « The first episode will pick up right where the previous season left off » traduit par « Le premier épisode reprend directement là où la saison précédente s'est arrêtée », d'autres sont plus compliqués, car ils sont utilisés dans la description de certains mécanismes de traduction automatique. Ainsi dans le cas de l'exemple suivant « Summer is the hottest season of the whole year », il est impossible de traduire simplement par « L'été est la saison la plus chaude de toute l'année », car le nombre de mots est alors différent en français, une caractéristique importante dans la suite du texte. Par conséquent nous avons choisi à la place de traduire la phrase par « La saison la plus chaude est toujours l'été ».

Traduction de la bibliographie

19 La seconde moitié traduite comportait les sources bibliographiques utilisées aux fins de la rédaction du chapitre. Or, les articles cités n'existent qu'en anglais, ne laissant que deux choix possibles : laisser tel quel ou procéder à une traduction personnelle. Après avoir

consulté notre tutrice, nous avons été informés de la possibilité de laisser les références en anglais, seuls certains éléments devant être traduits comme le nom des lieux. Ainsi nous avons par exemple simplement changé « Goodfellow, Ian, Yoshua Bengio & Aaron Courville. 2016. *Deep learning*. MIT Press. » en « Goodfellow, Ian, Yoshua Bengio et Aaron Courville. 2016. *Deep learning*. MIT Press. », le « & » n'étant pas utilisé en français.

Adaptation du style des auteurs en français

20 Dans la version anglaise, les auteurs s'adressent directement au lecteur en utilisant les pronoms *you* et *we*. Ceci n'est pas quelque chose de commun dans les écrits universitaires français où des tournures impersonnelles seront préférées (nous l'avons vérifié dans le corpus français Scientext). Toutefois, pour préserver ce lien direct avec le lecteur, étant donné que le texte est un texte de vulgarisation, nous avons conservé dans la traduction l'utilisation de ces pronoms. Par exemple nous avons traduit la phrase « Previously, in Section 3.3 of this chapter, we discussed the benefits of successively refining neural computations » par « Nous avons abordé précédemment dans la Section 3.3 de ce chapitre les avantages de l'amélioration successive des calculs neuronaux », plutôt que « Les avantages de l'amélioration successive des calculs neuronaux ont été précédemment abordés dans la Section 3.3 de ce chapitre », qui aurait été plus impersonnel.

1.5. Une terminologie non exhaustive

21 L'usage commercial de la TAN (Traduction Automatique Neuronale) est encore récent, et cette technologie progressant toujours plus rapidement, il n'est pas surprenant que l'évolution parallèle de la terminologie s'effectue plus lentement. Par exemple, certaines traductions de terme n'ont pas encore été trouvées ou validées par les multiples bases terminologiques faisant autorité dans le secteur de la traduction.

Input et *output* des termes passe-partout

22 Cependant, si nous devions définir les termes les plus problématiques à retranscrire, notre choix se porterait sur *input* et *output*. Non seulement ces termes sont souvent accolés à d'autres mots et employés dans tout autant de contextes différents, mais leurs interprétations varient en fonction de ces critères. Leur traduction est d'autant plus problématique avec une langue telle que la langue française qui s'accommode mal des répétitions.

Inputs

" In a network, some neurons receive external stimuli which act as *inputs* to the neural net […]"	« Au sein d'un réseau, certains neurones reçoivent des stimuli extérieurs qui servent de *données d'entrée* […] »
"Training a neural network is the process of determining the weight of the connections between its neurons so that, given a training set of *input—output* examples […]"	« Entrainer un réseau neuronal consiste à définir le poids des connexions entre ses neurones de manière que, compte tenu d'une base d'apprentissage d'exemples de *phrases sources-cibles* […] »

Output

"[…] it produces an actual *output* which is as close as possible to that in the relevant example."	« […] il génère des *propositions* réelles [a]aussi proches que possible de celles des exemples correspondants. »
"[…] which describes how far actual *outputs* are from the desired outputs […]"	« […] qui détermine le degré d'écart entre les *données de sorties* réelles et celles souhaitées […] »
"Neural networks may ideally generalise in the context of machine translation by producing *similar outputs* when fed with similar inputs […]"	« Dans le contexte de la traduction automatique, les réseaux neuronaux peuvent théoriquement produire des généralisation conduisant à des *résultats analogues* lorsqu'ils sont alimentés par des données d'entrées similaires […] »
a. Note de C. Rossi : l'usage de l'adjectif « réel » a été discuté lors de l'entretien de révision	

1.6. Mise en page et éléments graphiques

Traduction et adaptation des éléments graphiques

23 Ma partie de la traduction commanditée comporte plusieurs schémas illustrant certains exemples de mécanismes de traduction automatique neuronale. Si l'utilisation du logiciel de traduction assistée Trados m'a permis de modifier le texte de ces schémas, un test de l'enregistrement au format Word m'a montré que Trados ne pouvait pas modifier les schémas pour prendre en compte les changements dans le texte. Certains mots pouvaient donc être coupés ou bien avaient simplement disparu. Heureusement, Mme Rossi a pu récupérer auprès de ses collègues l'accès au répertoire des schémas contenus dans le manuel sur la plateforme Overleaf. Nous avons ainsi réussi à récupérer ces éléments au format SVG, que nous avons alors modifié grâce au logiciel Inkscape.

Fig. 3


Fig. 3

24 Afin que les deux parties de la traduction soient parfaitement harmonisées, nous avons consacré beaucoup de temps à la relecture commune de nos traductions respectives.

Harmonisation stylistique

25 Nous nous sommes mis d'accord à deux sur l'harmonisation des éléments de style à employer pour la traduction comme pour le problème de traduction du style des auteurs en français relevé ci-dessus ou la formulation des titres. Par exemple, Romain Revet avait tout d'abord traduit le titre de sous-section « Transformer: attention-based encoder–decoder » par « Le modèle Transformer : un encodeur-décodeur basé sur l'attention », mais dans un souci d'harmonisation avec la partie de Sébastien Palmieri, il a changé par la suite pour « Modèle Transformer : un encodeur-décodeur basé sur l'attention » en enlevant l'article.

Harmonisation terminologique

26 Afin d'assurer la cohérence terminologique pour le chapitre entier, nous avons, dès le début du projet, effectué les recherches terminologiques en binôme et avons recensé les termes sur la plateforme en ligne Google Drive. Cela a rendu impossible l'utilisation de termes français différents pour la traduction du même terme anglais dans le texte. Nous avons également relu ensemble les traductions de chacun en prêtant une attention importante à la terminologie.

Harmonisation factuelle

27 Sébastien décrit vers la fin de sa traduction le contenu d'un schéma apparaissant dans ma partie, il a donc été nécessaire d'harmoniser sa description ainsi que ce qui apparaît dans le schéma, pour qu'il n'y ait aucune différence.

2. Ressources terminologiques

28 Grâce au concours de Mme Caroline Rossi, nous avons pu bénéficier des connaissances de M. Thierry Poibeau, directeur de recherche au CNRS, directeur adjoint du laboratoire LATTICE (Langues, Textes, Traitement informatique et Traduction) et auteur de différents ouvrages sur le sujet. M. Poibeau nous a notamment expliqué que « les termes anglais sont souvent gardés tels quels en français. » Fort heureusement une grande partie des termes employés dans l'ouvrage était référencée dans des bases de données terminologiques telles que Termium plus (https://www.btb.termiumplus.gc.ca/tpv2alpha/alpha-fra.html?lang=fra), exploitée et maintenue par un ministère du gouvernement fédéral canadien, ou encore IATE (https://iate.europa.eu/home) (Interactive Terminology for Europe), la base de données terminologique de l'Union européenne.

29 Pour terminer cette partie introductive à la traduction du chapitre *How neural machine translation works* de l'ouvrage *Translation for Multilingual Citizens*, voici le glossaire réalisé à quatre mains pour cette traduction.

Anglais	Français	Sources

activation	activation	1
agglutinative language	langue agglutinante	2
architecture	architecture	1
artificial neural network	réseau de neurone artificiel	1
artificial neuron	neurone artificiel	1
attention	attention	1
attention head	tête d'attention	3
attention layer	couche d'attention	4
attention vector	vecteur d'attention	8
beam search	recherche en faisceau	1
BLEU Bilingual Evaluation Understudy	algorithme BLEU	5
byte-pair encoding	codage de paire d'octets	6
contextual vector	vecteur de contexte	4
contextual word embeddings	plongement lexical contextuel	Dérivé de *word embedding*
decoder-encoder architecture	architecture encodeur-décodeur	2
deep neural network	réseau neuronal profond	1
deep representation	représentation profonde	1
deep-learning algorithm	algorithme d'apprentissage profond.	1
degree of attention	degré d'attention	8
development corpus	corpus de développement	M. Thierry Poibeau[a]
development set	base de développement	Dérivé de *training set*
embedding	plongement	Dérivé de *word embedding*
encoding	codage	2
error function	fonction d'erreur	8
generalisation	généralisation	1
gradient	gradient	1
gradient descent	descente de gradient	1
head	tête	3
hidden layer	couche cachée	1
hidden neuron	neurone caché	1
human translation	traduction humaine	1
inhibited	inhibition	7
input neuron	neurone d'entrée	2
input sentence	phrase d'entrée	1

Language generation	génération de langage	1
layered neural network	réseau de neurones à couches	1
learning algorithm	algorithme d'apprentissage	1
learning rate	taux d'apprentissage	1
loss function	fonction de perte	8
machine learning	apprentissage automatique	2
module	module	1
multilayer	multicouche	1
natural language processing	traitement du langage naturel	1
natural neural network	réseau de neurones naturel	Dérivé de *artificial neural network*
neural machine translation	traduction automatique neuronale	1
neural network	réseau neuronal	1
NMT	TAN	2
non-contextual word embedding	plongement lexical non-contextuel	Dérivé de *word embedding*
output layer	couche de sortie	1
output neuron	neurone de sortie	1
output sentence	phrase cible	8
pre-trained model	modèle pré-entrainé	9
principle of semantic compositionality	principe de compositionnalité sémantique	10
probabilistic model	modèle probabiliste	1
recurrent model	modèle récurrent	11
semantic characteristic	charactéristiques sémantiques	1
SentencePiece	SentencePiece	12
sub-word unit	sous-mot	13
tokenization	tokenization	14
training algorithm	algorithme d'entraînement	Dérivé de *learning algorithm*
training corpus	base d'apprentissage	1
training set	base d'apprentissage	1
transformer model	modèle Transformer	11
Turing Natural Language Generation	Turing-NLG	15
vector notation	notation vectorielle	1
vector of probability	vecteur de probabilité	8
weight	poids	1

word embedding	plongement lexical	2
a. M. Thierry Poibeau, directeur de recherche au CNRS, directeur adjoint du laboratoire LATTICE (Langues, Textes, Traitement informatique et Traduction) et auteur de différents ouvrages sur le sujet sur la traduction automatique neuronale.		

Sources

30 1 : TERMIUM Plus (https://www.btb.termiumplus.gc.ca/tpv2alpha/alpha-fra.html?lang=fra)

31 2 : IATE (https://iate.europa.eu/home)

32 3 : *Les transformateurs expliqués visuellement (partie 3) : Attention multi-têtes, plongée en profondeur*. (2021, 17 januari). ICHI.PRO. Geraadpleegd op 11 April 2022, van https://ichi.pro/fr/les-transformateurs-expliques-visuellement-partie-3-attention-multi-tetes-plongee-en-profondeur-30923513022547

33 4 : Yves Mercadier. *Classification automatique de textes par réseaux de neurones profonds : application au domaine de la santé*. Intelligence artificielle [cs.AI]. Université de Montpellier, 2020. Français. ffNNT : 2020MONTS068ff. fftel-03145856f https://tel.archives-ouvertes.fr/tel-03145856/document

34 5 : *Qu'est-ce qu'un score BLEU ? - Custom Translator - Azure Cognitive Services*. (2021, 13 mei). Microsoft Docs. Geraadpleegd op 11 April 2022, van https://docs.microsoft.com/fr-fr/azure/cognitive-services/translator/custom-translator/what-is-bleu-score

35 6 : *L'évolution De La Tokenisation - Encodage De Paires D'octets En NLP*. (z.d.). Zephyrnet. https://zephyrnet.com/fr/l%27%C3%A9volution-de-l%27encodage-des-paires-d%27octets-de-tokenisation-en-nlp/

36 7 : Claude Touzet. *LES RESEAUX DE NEURONES ARTIFICIELS, INTRODUCTION AU CONNEXIONNISME : COURS, EXERCICES ET TRAVAUX PRATIQUES*. EC2, 1992, Collection de l'EERIE, N. Giambiasi. ffhal-01338010f https://hal-amu.archives-ouvertes.fr/hal-01338010/file/Les_reseaux_de_neurones_artificiels.pdf

37 8 : Pas de fiche terminologique dans le domaine de l'IA. Proposition personnelle

38 9 : Fokou, K., & → V. A. P. B. K. F. (2019, 6 November). *NLP & modèles de langue | Smals Research*. Smalsresearch. Geraadpleegd op 11 april 2022, van https://www.smalsresearch.be/nlp-modeles-de-langue/

39 10 : *La sémantique propositionnelle*. (z.d.). Lattice.CNRS. Geraadpleegd op 11 April 2022, van https://www.lattice.cnrs.fr/sites/itellier/poly_info_ling/linguistique008.html

40 11 : *Apprentissage profond et apprentissage automatique - Azure Machine Learning*. (2021, 7 September). Microsoft Docs. Geraadpleegd op 11 april 2022, van https://docs.microsoft.com/fr-fr/azure/machine-learning/concept-deep-learning-vs-machine-learning

41 12 : *Langage de programmation - SentencePiece - Introduction*. (z.d.). Gladir. Geraadpleegd op 11 april 2022, van https://www.gladir.com/CODER/SENTENCEPIECE/intro.htm

42 13 : *Codage de texte pour les tâches NLP*. (2021, 24 Augustus). ICHI.PRO. Geraadpleegd op 11 april 2022, van https://ichi.pro/fr/codage-de-texte-pour-les-taches-nlp-145588315128806

43 14 : *Le mécanisme d'attention du transformateur*. (2021, 15 November). Neuro Connection. Geraadpleegd op 11 april 2022, van https://neuroconnection.eu/le-mecanisme-dattention-du-transformateur/

44 15 : *Apprentissage profond et apprentissage automatique - Azure Machine Learning*. (2021, 7 September). Microsoft Docs. Geraadpleegd op 11 april 2022, van https://docs.microsoft.com/fr-fr/azure/machine-learning/concept-deep-learning-vs-machine-learning

NOTES

1 http://www.multitrainmt.eu/index.php/fr/accueil/presentation-du-projet

2 Note de C. Rossi : ce passage a été rediscuté et modifié lors de l'entretien de révision

AUTHORS

Sébastien Palmieri
Étudiant du master TSM de l'université Grenoble-Alpes (France)

Romain Revet
Étudiant du master TSM de l'université Grenoble-Alpes (France)