



**HAL**  
open science

## Un nouveau besoin dans l'industrie: une aide au “ rédacteur traduisant ”

Claire Lemaire

► **To cite this version:**

Claire Lemaire. Un nouveau besoin dans l'industrie: une aide au “ rédacteur traduisant ”. 24e Conférence sur le Traitement Automatique des Langues Naturelles (TALN); Orléans, France – 26-30 juin 2017, Jun 2017, Orléans, France. hal-01664363

**HAL Id: hal-01664363**

**<https://hal.science/hal-01664363>**

Submitted on 14 Dec 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Un nouveau besoin dans l'industrie : une aide au « rédacteur traduisant »

Claire Lemaire<sup>1</sup>

(1) ILCEA4, Université Grenoble Alpes, 38058 Grenoble, France  
claire.lemaire1@univ-grenoble-alpes.fr

## RESUME

---

Nous mettons en relief, grâce à une expérimentation avec questionnaires et corpus parallèle, une situation nouvelle en entreprise de rédaction multilingue, pour laquelle il n'existe pas de technologie TAL dédiée. Nous suggérons de tirer profit de cette situation inédite de rédacteur traduisant, pour utiliser l'expertise du rédacteur pendant le processus de traduction et nous préconisons de développer une TA permettant une édition en cours de processus.

## ABSTRACT

---

**A new need in companies: writing assistance for "translating writers".**

We highlight, thanks a experiment through questionnaires and parallel corpus, a new situation of multilinguistic writing in French companies, for which no dedicated NLP technology is provided. We propose to take advantage of this unprecedented situation of having a "translating writer", to use the writer expertise during the translation process and we suggest to create a new MT that allow editing during the process.

**MOTS-CLES :** TAL et multilinguisme, TAO, TA, TAFD.

**KEYWORDS:** NLP and multilinguism, CAT, MT, DBMT.

## Introduction

Les technologies de TAL actuelles sont prévues, soit pour des rédacteurs monolingues qui n'effectuent pas de traduction, soit pour des traducteurs qui ne rédigent pas le texte devant être traduit. Nous pensons qu'il est temps de s'intéresser à une production bilingue. Nous justifions notre approche par une enquête de terrain révélant une situation de plus en plus fréquente dans les entreprises où il est demandé aux cadres de produire des documents bilingues, essentiellement en français et en anglais.

Dans une première partie nous présentons deux enquêtes effectuées auprès de cadres en entreprise et auprès de traducteurs professionnels sur leurs usages en outils d'aide à la traduction. Dans une deuxième partie, nous décrivons les différents outils et leurs différents types d'utilisateurs, et la naissance d'un nouveau type d'utilisateur industriel. Dans la dernière partie, nous discutons de fonctionnalités qui pourraient répondre à ce nouveau besoin.

# 1 Questionnaires

Les questionnaires sont largement utilisés pour étudier des pratiques de traducteurs et depuis longtemps : Slocum a effectué un sondage sur la TA (Slocum, 1985), Vasconcellos un autre sur les usages de la TA (Vasconcellos, 1993), Lagoudaki a observé l'apport de la traduction automatique (Lagoudaki, 2008), Chan s'est penché sur ce qu'apporte une certification en traduction (Chan, 2010), Dam et Zethsen ont recueilli des commentaires de traducteurs sur leur statut social (Dam, Zethsen, 2010), Frérot et Karagouch ont étudié l'adéquation entre formation universitaire et besoins du monde professionnel (Frérot, Karagouch, 2016). Les questionnaires sur les pratiques linguistiques des salariés ont été utilisés par un certain nombre de chercheurs : Harzing et Pudelko ont eu recours aux questionnaires sur les pratiques linguistiques en entreprise pour pointer les barrières langagières (Harzing, Pudelko, 2012) et proposé des solutions pour pallier le grave manque à gagner engendré par des déficiences linguistiques en entreprise (Tenzer, Pudelko, Harzing, 2014).

## 1.1 Objectif visé

L'objectif visé par ces questionnaires est de modifier un outil d'aide à la traduction dédié à des traducteurs professionnels de façon à ce qu'il soit utilisable par des traducteurs non professionnels. Ces travaux ont été financés par un éditeur de logiciel, lequel a également financé plusieurs autres études : ainsi il a par exemple intégré une fonctionnalité de TA à son logiciel déjà commercialisé depuis plusieurs années de mémoire de traduction. Plus précisément, l'objectif de l'éditeur est de toucher, au sein d'une entreprise ayant déjà des traducteurs professionnels utilisant son outil, des collègues cadres traduisant, afin de leur proposer son logiciel.

## 1.2 Méthode utilisée

Nous avons envoyé des questionnaires à des cadres travaillant en entreprise, n'exerçant pas dans le domaine de la traduction et n'ayant pas non plus suivi de formation en traduction, devant traduire dans le cadre de leur profession au moins une fois par semaine et utilisant des technologies de TAL. Les cadres sélectionnés travaillent avec au moins deux langues parmi l'allemand, l'anglais et le français. Ainsi, 34 cadres ont été sélectionnés. Le questionnaire est composé de 5 sections visant principalement à récolter des informations sur les usages des outils en général et les fonctionnalités en particulier des technologies de traduction. Parallèlement, nous avons envoyé un questionnaire semblable à un échantillon de traducteurs professionnels représentant 8,5 % des traducteurs de la SFT<sup>1</sup>, soit 10 traducteurs, travaillant entre l'allemand, l'anglais ou le français.

## 1.3 Résultats obtenus

L'outil principal utilisé par le panel de traducteurs est SDL Trados pour 71 % des traducteurs ; puis en deuxième position, viennent Transit de Star et Wordfast (d'Yves Champollion) avec chacun 11 % d'utilisateurs. En comparaison, la première technologie utilisée par les experts métier en entreprise est Google Translate, avec 45 %, suivi de Word (Bing) avec 14 % et Reverso avec 13 %.

---

<sup>1</sup> Société française des traducteurs, syndicat, qui entre autres, répertorie les traducteurs en France

La majorité des traducteurs professionnels interrogés (70 %) n'utilisent pas d'outil secondaire. Cependant, 18 % des répondants utilisent SDL Trados comme second outil et 6 % utilisent Wordfast. Les experts métier, quant à eux, sont plus nombreux à utiliser un outil secondaire, à savoir un dictionnaire en ligne pour 30 %, puis trois des outils cités précédemment : Google Translate à 20 %, puis Linguee et Reverso à 13 % chacun. La traduction automatique semble ainsi omniprésente chez les experts-métier alors qu'elle n'est qu'anecdotique chez les traducteurs professionnels spécialisés.

En dehors de l'outil principal, les fonctionnalités les plus utilisées chez les cadres sont le concordancier à 74 %, puis le correcteur orthographique et grammatical à 70 %, puis les dictionnaires bilingues, également à 70 %, suivis de la traduction automatique (TA) à 52 % et enfin des dictionnaires auto-suggest à 39 %. Chez les traducteurs, les fonctionnalités les plus utilisées sont le correcteur orthographique et grammatical, puis les dictionnaires bilingues, chacun à 90 %, puis le concordancier, employé à 70 %, puis les dictionnaires auto-suggests 40%, et enfin la fonctionnalité de traduction automatique, utilisée par 10 % des traducteurs interrogés.

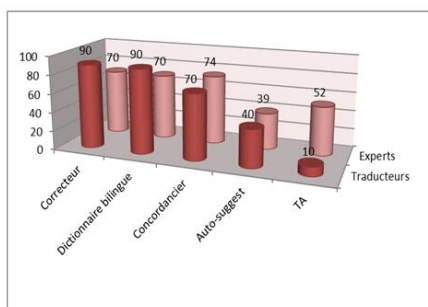


Figure 1 : Pourcentage d'utilisation des fonctionnalités par les cadres (« experts ») et les traducteurs

Le résultat sur les fonctionnalités le plus flagrant est donc, comme pour l'outil principal, la différence d'utilisation de la TA, avec 52 % chez les cadres contre 10 % chez les traducteurs professionnels. La comparaison entre les autres fonctionnalités montre peu de différence entre les pratiques des experts et celles des traducteurs interrogés. En effet, les trois premières fonctionnalités sont les mêmes, les correcteurs orthographiques et grammaticaux, les dictionnaires bilingues, et les concordanciers.

Dans un second temps, pour affiner l'analyse des différences entre les deux populations, nous avons été dans une entreprise en immersion pendant trois mois pour chercher des textes bilingues rédigés par ces cadres. Nous avons ainsi découvert un nouveau type d'utilisateur : un cadre ayant un bon niveau dans une autre langue et à qui il est demandé de produire directement des documents bilingues, sans avoir recours à un traducteur, sans soutien linguistique et sans temps alloué spécifiquement à la tâche de traduction.

## 2 Outils existants pour le rédacteur traduisant

### 2.1 Deux paradigmes

Lorsqu'il s'agit de produire des traductions de textes, les systèmes existant aujourd'hui peuvent être classés selon deux paradigmes : (1) les outils qui proposent plusieurs traductions de phrases plus

ou moins proches de celle que l'utilisateur cherche à traduire (diffusion) ou à comprendre (acquisition) selon la situation ; (2) ceux qui proposent à l'utilisateur une traduction personnalisée et le plus souvent unique du segment venant d'être saisi. Notons cependant que la plupart des systèmes ont une option permettant de proposer plusieurs possibilités de traduction, surtout lexicales, lorsque le contexte ne permet pas au système de désambiguïser avec sûreté. Ces deux types d'outils ont un paradigme commun de logique séquentielle constitué de deux étapes : une production dans une langue 1, puis une production dans une langue 2.

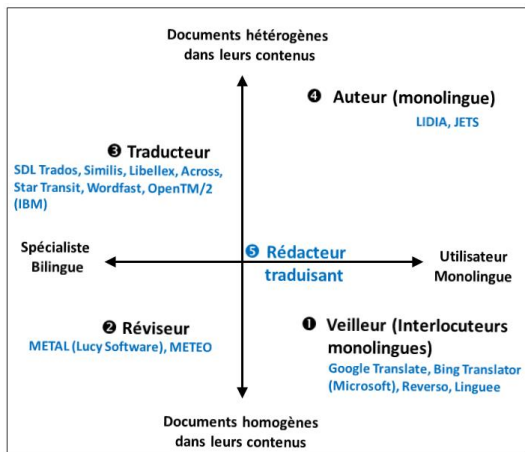


Figure 2 : Les utilisateurs déjà connus et le rédacteur traduisant

## 2.2 Quatre types d'utilisateurs

Si l'on classe maintenant les systèmes existants selon les utilisateurs, d'après Blanchon (Blanchon, 2004), nous avons : (1) La traduction du veilleur monolingue ou la traduction de dialogues oraux finalisés<sup>2</sup> (qui correspond à une traduction 100% automatique, c'est-à-dire la TA). Les situations pour lesquelles cette traduction est prévue sont la veille ou la navigation sur Internet par exemple. (2) et (4) La traduction du réviseur bilingue et la traduction de l'auteur monolingue qui produisent respectivement des traductions « brutes », destinées à être révisées, et des traductions ne nécessitant pas de révision, en faisant appel à l'auteur au moyen de désambiguïstation interactive en langue source (qui correspond à une traduction 75-90% automatique). (3) La traduction du traducteur. Elle est constituée d'outils destinés à un utilisateur humain bilingue (qui correspond à une traduction 15-50% automatique). En 2017, à notre connaissance, les outils conçus continuent de s'adresser à un utilisateur soit absolument monolingue, soit bilingue spécialisé en traduction. Nous avons situé par rapport au schéma de Blanchon, quelques outils : Google Translate, Bing Translator (Microsoft), METAL (Lucy Software), Meteo, SDL Trados 3, Similis, Libellex, Across, Star Transit, Wordfast, OpenTM/2 (IBM)<sup>4</sup>, Lidia et Jets.

<sup>2</sup> c'est-à-dire des systèmes prévus pour des interlocuteurs monolingues dans différentes langues, tels que des systèmes de réservations de billets d'avion par exemples

<sup>3</sup> ou SDL Studio ou SDLX5

<sup>4</sup> anciennement TM/2

Notre cas d'étude présente un cinquième type d'utilisateur, ni absolument monolingue, ni bilingue spécialisé en traduction, devant traiter des documents au contenu varié mais tout de même issus d'un même domaine de spécialité.

### 3 Outil à créer pour le rédacteur traduisant

Cette situation nouvelle de traduction par le rédacteur, sans budget ou temps supplémentaire, et sans que le rédacteur soit en contact avec le lecteur en langue cible, présente un atout majeur. En effet, contrairement au traducteur classique, qui n'a pas le droit de modifier le texte source, le rédacteur traduisant en est maître. Or s'il est connu que le besoin de certaines informations nécessaires au lecteur cible n'apparaît qu'au cours de la traduction, posant le problème de la désambiguïsation, il est tout à fait nouveau de pouvoir agir sur le texte source en fonction du résultat obtenu en langue cible. Ainsi le rédacteur traduisant pourrait, après une première TA, découvrir au vu du résultat, qu'il lui faut retoucher le texte source pour : (1) ajouter des informations extralinguistiques, c'est-à-dire désambiguïser ; (2) ajouter d'autres informations linguistiques apparues manquante par l'alignement (conséquent à la TA) et reprendre des fautes d'orthographe ou de grammaire : c'est-à-dire améliorer la rédaction. Puis le rédacteur pourra effectuer une deuxième TA.

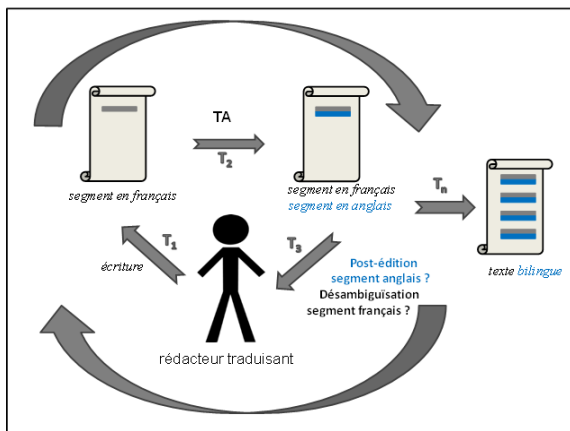


Figure 3 : Pré-édition et post-édition itérative de chaque segment

#### 3.1 Désambiguïsation interactive

Avec cette pré-édition et post-édition permettant d'agir à la fois sur le texte source et sur le texte cible, le rédacteur traduisant pourrait lever les ambiguïtés au fur et à mesure du processus de traduction, non pas préalablement au transfert de l'ensemble du texte mais au vu d'un résultat partiel (segment par segment), de façon itérative jusqu'à ce qu'il soit satisfait. Le rédacteur-traduisant améliore d'autant plus facilement le texte en langue cible et celui en langue source, qu'il s'agit de son domaine d'expertise. En effet, le traduisant possède une excellente connaissance de toutes les informations extralinguistiques que peut fournir le rédacteur, et le rédacteur prend conscience, au vu des résultats de la TA, des besoins en information que ce processus de traduction requiert. Nous sommes convaincu qu'un utilisateur même ne connaissant pas parfaitement la langue d'arrivée pourra ainsi résoudre un nombre considérable d'erreurs. Le cadre endosse ainsi tour à tour le rôle de

rédacteur et de traducteur, mais en général pas celui de « réviseur ». Nous entendons ici par réviseur, un senior très compétent dans le domaine. Sa langue maternelle est la langue cible, et il maîtrise la langue source au moins au niveau C1. Dans notre cas, le profil type du rédacteur traduisant fait qu'il ne peut pas être un réviseur.

À notre connaissance, le seul outil de TA avec désambiguïsation interactive est une maquette ancienne, LIDIA-1, développée en Ariane-G5, un système de développement et d'exploitation d'applications propriétaire, constitué de 5 langages de programmation<sup>5</sup> conçus pour le développement de technologies de la traduction (Blanchon, 1994) et (Boitet, Blanchon, 1994). Le rédacteur aidait le système par une désambiguïsation de tous les termes qui étaient susceptibles de poser un problème au cours de la transposition dans les autres langues. Il s'agissait d'une désambiguïsation par étapes, qui avait lieu avant la fin de l'étape d'analyse (passage à une structure abstraite multiniveau de Vauquois) et donc avant l'étape de transfert (quelle que soit la ou les langues cible). La désambiguïsation reposait sur un processus de clarification, fondé sur un mécanisme de reconnaissance de patrons. À chaque patron, était associée une méthode de production de dialogue en langue naturelle et demandait ainsi à l'utilisateur traduisant de fournir des informations supplémentaires permettant de lever les ambiguïtés du texte en langue source. Il fallait alors prévoir toutes les ambiguïtés possible à l'avance et les lever avant la phase de transfert.

Nous préconisons de construire un système beaucoup plus simple, avec une seule paire de langue (dans les deux sens) et non plus avec une désambiguïsation interactive de tout le texte en une fois en amont d'un unique passage en TA, mais avec une désambiguïsation d'un seul segment, avant chaque passage en TA. L'environnement technologique consisterait également en un moniteur interactif lié à un système de gestion de base de données et à un traitement de textes. Le processus de traduction automatique se décomposerait également en trois étapes successives, classiques en TA : l'analyse de la langue source, le transfert et la génération en langue cible. L'unité de traduction en revanche ne dépasserait pas le segment. L'analyse de la langue source resterait monolingue. Lors de la phase de transfert, les traitements grammaticaux seraient également séparés des traitements lexicaux, et la spécialisation des grammaires aurait lieu en fonction de la typologie, et la spécialisation des dictionnaires aurait lieu en fonction du domaine. La génération en langue cible serait également monolingue. En fonction du résultat, et de la prise de conscience de certaines ambiguïtés, il est en revanche possible en plus de la désambiguïsation interactive, de modifier le segment source. Certaines ambiguïtés ne seraient donc pas traitées informatiquement, mais seraient annulées par la reformulation d'une phrase univoque en entrée du système.

### **3.2 Aide à la rédaction**

En outre, il s'agirait d'agir sur des fonctionnalités de : (1) déclenchement systématique par défaut d'un affichage des synonymes possibles des mots source et des mots cible par info-bulle ; (2) déclenchement systématique de la détection de langue et l'affichage par défaut de la correction grammaticale et orthographique au cours de la frappe ; (3) création automatique de dictionnaires auto-suggest, en les créant systématiquement par défaut, que les utilisateurs les demandent ou non. De nombreux outils de traitement de textes possèdent déjà la plupart de ces fonctionnalités, pour le rédacteur traduisant, qui les maîtrise mal car n'étant pas spécialiste de la rédaction, comme un universitaire où un traducteur, il faudrait les activer par défaut. Par ailleurs, les travaux de Hong-Thai Nguyen (Nguyen, 2009), mais également de Phap-Cong Huynh (Huynh, 2010), et plus récemment de Ying Zhang (Zhang, 2016), montrent que ces efforts ergonomiques sont étudiés et partiellement

---

<sup>5</sup> ATEF, ROBRA, EXPANS/TRANSF, SYGMOR et TRACOMPL

développés (dans l'outil « aximag »), en particulier avec l'introduction de « mini-dictionnaires », un pour chaque segment, préparé à l'avance. À l'instar de ce que proposent des outils tels que Leslie ou Antidote, l'outil du rédacteur traduisant proposerait une cohérence lexicale, d'ajouter des virgules, de couper les phrases en deux, ou encore de mettre des parenthèses, suivant les résultats sémantiques obtenus lors de la phase de transfert du segment.

Par ailleurs nous proposons de créer deux nouvelles fonctionnalités. Ainsi, il s'agirait de créer des fonctionnalités de : (1) recherche d'information lexicologique, c'est-à-dire une recherche de vocabulaire de spécialité sur Internet directement en fonction des ajouts de l'utilisateur aux dictionnaires auto-suggest. Les outils linguistiques sont généralistes, or un cadre en entreprise utilise un vocabulaire récurrent ; (2) d'affichage, c'est-à-dire de présentation bilingue du document avec affichage ou masquage d'une langue ou de l'autre. Il s'agirait de concevoir un mode d'affichage d'une seule langue. Il serait ainsi plus aisé pour le rédacteur-traducteur de vérifier ce que le lecteur anglophone verra apparaître en anglais, bien qu'il n'y ait pas de mise en regard du texte ; puis de vérifier ce que le lecteur francophone verra apparaître.

## Conclusion

Les résultats de notre expérimentation à l'aide de questionnaires et d'un corpus de traduction nous a fait découvrir une situation tout à fait inédite de traduction par un traducteur ayant la possibilité de retoucher le texte source. Nous proposons de tirer profit de cette situation de rédacteur traduisant, pour utiliser l'expertise de l'utilisateur et de l'injecter dans le processus de traduction par deux moyens : une désambiguïsation interactive et une modification du texte source. Nous proposons également une aide à la rédaction bilingue mettant en évidence les manques de traduction et autres lacunes lexicales, orthographiques et grammaticales par de la recherche d'information, et un nouvel affichage en plus d'autres fonctionnalités traditionnelles.

Les perspectives de recherches sont la création d'un « proof of concept » sur le même domaine de spécialité, le même couple de langue et le même sens de traduction du système de désambiguïsation interactive évoqué et des fonctionnalités d'aide à la rédaction bilingues proposées. Ensuite, si les tests sont concluants, bien que nous ayons l'intuition qu'il serait impossible d'intégrer l'ensemble des fonctionnalités proposées à un traitement de texte (par exemple la désambiguïsation interactive, pour des raisons techniques, au niveau de la performance en particulier), les recherches pourraient se poursuivre par une étude permettant de choisir si l'ensemble ou partie de ces fonctionnalités seraient à intégrer à l'outil déjà commercialisé ou de créer un nouvel outil dédié pour les cadres.



## Références (style *Titre sans numéro*)

- BLANCHON H. (2004). Comment définir, mesurer et améliorer la qualité, l'utilisabilité et l'utilité des systèmes de TAO de l'écrit et de l'oral. Une bataille contre le bruit, l'ambiguïté, et le manque de contexte. HDR, Université Joseph Fourier, Grenoble.
- BLANCHON H. (1994). *LIDIA-1 : une première maquette vers la TA Interactive" pour tous"*. Thèse de doctorat, Université Joseph-Fourier-Grenoble I, Grenoble.
- BOITET C., BLANCHON H. (1994). Promesses et problèmes de la « TAO pour tous ». Après LIDIA-1, une première maquette". *Langages. Le traducteur et l'ordinateur, sous la direction de Jean-René Ladmiral, 116*, pp. 20–47.
- CABRÉ M. T. (2007). Constituer un corpus de textes de spécialité. *Cahiers du CIEL*, 37–56.
- CHAN A. L. (2010). Perceived benefits of translator certification to stakeholders in the translation profession: A survey of vendor managers. *Across Languages and Cultures, 11*(1), 93–113.
- DAM H., ZETHSEN K. (2010). Translator status: Helpers and opponents in the ongoing battle of an emerging profession. *Target, 22*(2), 194–211.
- FREROT C., KARAGOUCHE L. (2016). Outils d'aide à la traduction et formation de traducteurs : vers une adéquation des contenus pédagogiques avec la réalité technologique des traducteurs. *ILCEA Revue de l'Institut des langues et cultures d'Europe, Amérique, Afrique, Asie et Australie, 27*, pp. 120–140.
- HARZING A.-W., PUDELKO M. (2012). Language competencies, policies and practices in multinational corporations: A comprehensive review and comparison of Anglophone, Asian, Continental European and Nordic MNCs. *Journal of World Business, 48*(1).
- HUYNH C.-P. (2010). *Des suites de test pour la TA à un système d'exploitation de corpus alignés de documents et métadocuments multilingues, multiannotés et multimédia*. Thèse de doctorat, Université Joseph Fourier.
- KOSKINEN K. (2008). *Translating Institutions: An Ethnographic Study of EU Translation*. Manchester: St Jerome Publishing.
- LAGOUDAKI, E. (2009b). *Expanding the possibilities of translation memory systems: from the translators wishlist to the developers design*. Thèse de doctorat, Imperial College London.
- NGUYEN H.-T. (2009). *Des systèmes de TA homogènes aux systèmes de TAO hétérogènes. Interface homme-machine U*. Thèse de doctorat, Université Joseph-Fourier - Grenoble I.
- SALDANHA G., O'BRIEN S. (2013). *Research methodologies in translation studies*. London: Routledge.
- SLOCUM J. (1985). A survey of machine translation: its history, current status, and future prospects. *Computational linguistics 11.1 (1985): 1-17.*, 11(1), pp. 1–17.

STURGE K. (2004). *The Alien Within: Translation into German during the Nazi Regime*. Munich: Ludicium.

TENZER H., PUDELKO M., HARZING A.-W. (2014). The impact of language barriers on trust formation in multinational teams. *Journal of International Business Studies*, 45(5), 508–535.

VASCONCELLOS M. (1993). The present state of machine translation usage technology; or: How do I use thee? Let me count the ways. *MT Summit IV, July 20–22*, (pp. 47–62). Kobe, Japan.

YIN R. (2009). *Case Study Research: Design and Methods* (éd. 4e). Los Angeles, London, etc.: Sage.

ZHANG Y. (2016). *Modèles et outils pour des bases lexicales "métier" multilingues et contributives de grande taille, utilisables tant en traduction automatique et automatisée que pour des services dictionnaires variés*. Thèse de doctorat, Université Grenoble Alpes.