Cahiers du plurilinguisme européen
ISSN : 2970-989X

2 | 2010
Outils de traduction – outils du traducteur ?

La préparation des corpus parallèles et comparables — nouvelles bases pour la traduction entre le français et le slovène

Mojca Schlamberger Brezar

https://www.ouvroir.fr/cpe/index.php?id=246

DOI : 10.57086/cpe.246

Electronic reference
Mojca Schlamberger Brezar, « La préparation des corpus parallèles et comparables — nouvelles bases pour la traduction entre le français et le slovène », *Cahiers du plurilinguisme européen* [Online], 2 | 2010, Online since 01 January 2010, connection on 06 November 2026. URL : https://www.ouvroir.fr/cpe/index.php?id=246

Copyright
Licence Creative Commons – Attribution - Utilisation non commerciale - Partage dans les mêmes conditions 4.0 (CC-BY-NC-SA)

La préparation des corpus parallèles et comparables — nouvelles bases pour la traduction entre le français et le slovène

Mojca Schlamberger Brezar

OUTLINE

1. L'origine des corpus
2. Les corpus en traductologie
3. L'utilité des corpus dans le travail du traducteur et dans la recherche en traductologie
   3.1. Les relations interlinguistiques et interculturelles entre le français et le slovène
   3.2. Les domaines de la traduction favorisant la création des corpus entre le français et le slovène
   3.3. La composition du corpus des textes parallèles et comparables français-slovène
   3.4. Les droits d'auteurs
Conclusion

TEXT

1 La traduction entre le français et le slovène est présente dans l'espace culturel slovène quasiment depuis toujours, au moins depuis l'époque où la Slovénie faisait partie des Provinces illyriennes[1]. Malgré cette longue tradition, surtout en comparaison avec l'anglais avec lequel il n'y avait pas de contact historique, dans le cadre de la préparation des ressources linguistiques, la paire linguistique français-slovène reste le parent pauvre. La question se pose de savoir comment combler cette lacune.

2 Une partie de la réponse se cache dans le développement du domaine de la linguistique de corpus, de tradition anglo-saxonne. Nous en donnerons un bref historique avant de présenter par la suite les possibilités offertes par les corpus pour les recherches contemporaines en linguistique et en traduction, en prenant l'exemple du français et du slovène. Pour cela, il faudra entre autres se demander : quels sont les échanges entre le français et le slovène et quels

domaines de la traduction sont visés ? La réponse à cette question nous permettra de trouver la typologie des textes qui devraient être inclus dans un corpus de référence pour le français et le slovène, destiné à des recherches en traductologie. Il est inutile d'en souligner la nécessité. Nous présenterons les domaines où l'initiation à la compilation a déjà été faite et les bénéfices que nous pourrons en tirer.

1. L'origine des corpus

3 En linguistique, les corpus ne constituent pas une nouveauté. Depuis toujours, les linguistes en ont construit pour valider leurs hypothèses. Dans l'espace culturel slovène, les grammaires des années 70 ont été rédigées de cette manière, par exemple la grammaire slovène de Jože Toporišič (1976, 2000) et le précis de grammaire slovène en français de C. Vincenot (1975) ainsi que le dictionnaire monolingue du slovène SSKJ (*Slovar slovenskega knjižnega jezika* – dictionnaire du slovène standard dont la composition a débuté dans les années 50) qui, lui aussi, est basé sur un corpus, en majorité littéraire. Selon un de ses éditeurs F. Jakopin (Bajec *et al.*, 1994 : XVII), les corpus étaient écrits et triés manuellement sur les *cartes* ou *fiches* — le SSKJ est basé sur 6 millions de fiches et 300 000 mots-clés issus des textes littéraires, scientifiques, techniques, journalistiques et autres. Toutefois, la linguistique chomskyenne des années 60 et 70 a prétendu que les corpus étaient inutiles et qu'un locuteur natif peut produire bien plus d'exemples pertinents pour un linguiste qu'une grande quantité de données textuelles. Tout a été misé sur l'intuition du locuteur naturel d'une langue.

4 Malgré ce fait, la linguistique de corpus au sens actuel, poussée par le développement des ordinateurs, a pris son essor dès le début des années 80. Sa nouveauté par rapport aux travaux basés sur les *cartes* et *fiches*, selon les mots de B. Habert (Habert *et al.* 1997 : 7), « réside dans l'enrichissement des corpus, l'accroissement de leur taille et dans l'accessibilité effective des corpus et des outils ». De nos jours, les corpus en linguistique sont qualifiés de bases de données annotées, textuelles et écrites, ou énoncés oraux transcrits et stockés dans l'ordinateur, qui se prêtent à l'analyse terminologique ou linguis-t-

ique et à la description (Kennedy 1998 : 1, Olohan 2004 : 11-14, Habert *et al.* 1997 : 11).

5 Cette linguistique est née dans le monde anglo-saxon. Avec les corpus de la 1e et 2e génération — la différence entre les deux concerne surtout la quantité des données élaborées, il faut mentionner au moins le *British National Corpus* (BNC), avec ses 100 millions de mots, dont 90 % viennent de textes écrits et 10 % de l'oral. En 1994, lors de sa phase finale, il a représenté de nouveaux standards pour la composition des corpus (Kennedy, 1998 : 12, 45).

6 Les langues qui ont suivi, notamment le norvégien, le suédois, le danois, le néerlandais ou le flamand, ont d'abord composé leurs corpus de préférence en combinaison avec l'anglais. Ces langues étaient fortement liées à l'apprentissage, à l'enseignement, à la traduction et à la lexicographie de l'anglais. Le monde latin, à l'époque, avait du retard par rapport au monde germanique. Pour sa part, le français présente une exception avec la base textuelle *Frantext*dont la composition a commencé à l'Institut national de la langue française (InaLF) dans les années 60 (Leech, 1992, dans Gorjanc, Krek 2005, Habert et *al.*, 1997 : 7). Avec ses 160 millions de mots, il est accessible par abonnement électronique (TLF : http://atilf.atilf.fr/).

7 En France et dans les pays francophones, on peut constater une émergence des corpus pour le français, accessibles sur internet, depuis les années 90 : French Collection at University of Virginia (http://etext.lib.virginia.edu/collections/languages/french/), Association des Bibliophiles Universels (http://abu.cnam.fr/), corpus Athena de l'Université de Genève (http://un2sg4.unige.ch/athena/html/francaut.html) comprenant des textes littéraires, le recueil belge (http://bach.arts.kuleuven.ac.be/lancom/), le corpus de l'Université Paris VII — Diderot Jussieu (http://www.talana.linguist.jussieu.fr/Presentation/corpus.html), le corpus de référence français COFRANS destiné surtout à la recherche en syntaxe (www-rali.iro.unmontreal.ca), et le corpus du français parlé réalisé par le Groupe Axois de recherche scientifique (GARS) sous la tutelle de Claire Blanche-Benveniste. Dans le domaine des corpus du français parlé annoté, nous devons mentionner également celui de M. A. Morel de l'Université Paris III — Sorbonne nouvelle. Le groupe CRIM (Centre de recherche en ingén-

ierie multilingue, Inalco : www.inalco.fr/) est en train de créer des bases de données pour le français et les langues enseignées à l'Institut national des langues et civilisations orientales (INALCO)[2].

8 Quant à la langue slovène, les corpus monolingues qui se veulent de référence sont *Fida*, *FidaPLUS*, *Beseda*, *nova Beseda*. Le premier corpus de référencedu slovène *Fida* a vu le jour entre 1997 et 2000, comme le fruit de la collaboration entre la Faculté des lettres de l'Université de Ljubljana, l'Institut des recherches technologiques Jožef Stefan, la maison d'édition DZS et l'entreprise Amebis (www.fida.net). À partir de 2007, il est complété par le mégacorpus *FidaPlus* (www.fidaplus.net) de plus de 600 millions de mots.

9 Le corpus slovène *Beseda* a été élaboré dans le cadre du groupe de recherche de l'Académie slovène des arts et des sciences ZRC SAZU (http://bos.zrc-sazu.si/s_beseda.html) et comprend des œuvres littéraires ainsi que des articles de presse (Gorjanc, Krek, 2005 : 186-187). Aujourd'hui, sa forme a évolué et, avec 240 millions de mots, il englobe la littérature slovène, les traductions des textes littéraires et non-littéraires, les textes scientifiques et techniques, les articles de journaux, les textes juridiques, etc.

10 En tant que données textuelles, les corpus du slovène mentionnés ci-dessus fonctionnent comme représentations de la réalité linguistique soit pour les descriptions textuelles et grammaticales, soit pour la recherche et la fréquence des mots en slovène standard, ce qui peut parfois désambiguïser les solutions proposées et combler les lacunes lorsque les dictionnaires n'offrent pas de solutions. Leurs annotations morphologiques, syntaxiques, sémantiques, prosodiques et autres constituent une valeur ajoutée. Mais ces corpus ne peuvent pas être inclus dans la mémoire de traduction des logiciels de traduction assistée car, à part le fait qu'ils sont traités de manière à effacer le texte de départ, ils sont monolingues.

2. Les corpus en traductologie

11 Les corpus en traductologie sont aussi importants pour la recherche que pour l'utilisation commerciale puisqu'ils représentent la base de la traduction assistée par ordinateurs. Leur emploi en traductologie est relativement tardif, les linguistes de corpus n'ont d'ailleurs pas

prévu leur utilisation dans le but de traduire, à l'exception du livre *Corpora in Applied Linguistics* de Susan Hunston (2001 : 123-128) et B. Habert (1997 : 135-143). Toutefois, la collaboration des traducteurs / traductologues et des linguistes allait dans un sens seulement — les traducteurs et les chercheurs en traductologie s'inspiraient en effet des corpus établis par les linguistes tandis que la pratique inverse n'existait pas.

12 La composition du premier corpus bilingue a commencé avec Mona Baker dans les années 90. Il s'agissait d'un corpus parallèle anglais-français contenant les textes du parlement canadien *Canadian Hansard* (Malmkjaer 2004 : 91), accessible sur http://www.tsrali.com, incluant 60 millions de mots et utilisé à titre d'essai dans le but du développement de nouvelles techniques de traduction. La composition du corpus continue depuis (Laviosa, 1998).

13 Il est intéressant de noter que les chercheurs qui ont créé le corpus de référence *British National Corpus* (BNC, 1994) en ont exclu toutes les traductions, ce qui signifie que, selon des données récentes, les traductions ne font pas partie de l'usage représentatif de la langue dans l'espace linguistique anglais (Olohan, 2004 : 13). En revanche, pour la langue slovène, on rencontre une autre conception de la représentativité. Les corpus slovènes *Beseda* et *Nova Beseda* (http://bos.zrc-sazu.si/s_beseda.html) contiennent non seulement des textes de différents types, rédigés en slovène, mais aussi les traductions des textes littéraires et non-littéraires qui font partie du paysage textuel slovène. Il s'agit des traductions de la littérature contemporaine ainsi que des manuels ou des ouvrages de vulgarisation et journaux ou revues (par exemple Le *National Geographics* dans le corpus *Nova Beseda*) parus depuis le début de la création du corpus *Fidaplus* en 1990 (Gorjanc 2005 : 47).

14 Aujourd'hui, les corpus en traductologie sont destinés aux recherches dans le cadre de la lexicologie, l'analyse contrastive, etc. Le cadre théorique de l'utilisation des corpus dans les recherches en traductologie est vaste (M. Olohan, 2004 : 16) : ils sont inclus dans les études descriptives, les études qui traitent de ce qui est typique dans les traductions ; les corpus sont le support des étudescombinant les méthodes qualitatives et quantitatives qui se concentrent sur le lexique, la syntaxe et les caractéristiques discursives et, en même

temps, permettent l'application de la méthodologie à différents types de traduction, par exemple la traduction dans divers milieux sociologiques, etc.

15 Le choix se pose entre les corpus parallèles et comparables. Les corpus parallèles contiennent l'original et sa traduction vers une langue étrangère. Quant aux corpus des textes comparables, ils constituent selon la définition de B. Habert (Habert et al. 1997 : 144) « des sélections de textes similaires dans plus d'une langue ou plusieurs variétés d'une langue. On peut considérer LOB et Brown comme des corpus comparables ». Cette définition est valable aussi pour la définition en traductologie (Olohan 2004 : 35) où l'auteur cite plusieurs définitions des corpus comparables, parmi lesquelles la plus générale, dénotant ces corpus comme les compilations de textes originaux avec un contenu et une longueur comparables dans deux ou plusieurs langues sources, par exemple la consigne suivante : la recette de cuisine ou l'invitation en français et en slovène. Les corpus comparables sont les corpus destinés à des recherches interlinguistiques, surtout à celles plus longues consacrées à l'étude des stratégies et des procédés de traduction interlinguistiques (M. Olohan, 2004 : 17).

16 Les traducteurs préfèrent parfois les corpus parallèles ou alignés aux corpus comparables. Les raisons en sont la facilité de la recherche des concordances, la possibilité de les inclure dans le mémoire de traduction. Cependant, selon M. Doherty (1998 : 235), il existe un risque que les résultats de la recherche des traductions aient un impact sur la norme linguistique de la langue cible, la multitude et la complexité des données pouvant conduire à une confusion des données. En outre, pour éviter le risque de contamination des structures du texte cible par les structures du texte source, il est à notre avis souhaitable d'inclure aussi dans la recherche les éléments comparables, par exemple les originaux en plus des traductions. Une telle démarche s'est révélée utile dans notre recherche concernant la traduction des textes politiques européens que nous avons comparés avec les originaux slovènes (Schlamberger Brezar 2005a).

17 Les corpus parallèles et leur capacité dépendent surtout de l'importance de la combinaison linguistique dans un domaine donné et des textes qui font partie de l'échange traduisant pour la combinaison

linguistique en question. Les corpus bilingues en slovène existent surtout pour la combinaison slovène-anglais *ELAN* (http://nl.ijs.si/elan/), le résultat d'un projet européen en collaboration avec l'Université de Ljubljana et l'Institut Jožef Stefan comprenant 15 textes alignés anglais et slovènes et 1 million de mots. Le corpus intitulé *Evrokorpus*, accessible en ligne (http://evroterm.gov.si/index.php, http://evrokorpus.gov.si/index.php), est spécialement élaboré pour les traducteurs. Dans *Evrokorpus*, à part l'anglais, les données pour le français et l'allemand apparaissent, mais sont moins nombreuses. Le corpus anglais-français contient 56 millions de mots et 1 400 000 unités de traduction ; le corpus français-slovène, 24 millions de mots et 540 000 unités de traduction ; le corpus italien-slovène, 11 millions de mots et 270 000 unités de traduction ; le corpus allemand-slovène, 12 millions de mots et 320 000 unités de traduction ; le corpus espagnol-slovène, 10 millions de mots et 230 000 unités de traduction ; enfin, le corpus multilingue comprenant les 22 langues officielles de l'Union européenne, 98 millions de mots et 610 000 unités de traduction (http://evrokorpus.gov.si/index.php). Un autre corpus slovène-anglais parallèle *TRANS*, comprenant 45 textes en deux langues (http://www-ai.ijs.si/~spela/trans-index.html) a été élaboré dans le cadre du séminaire de TAO de Spela Vintar au Département de traduction et d'interprétation de l'Université de Ljubljana. Le corpus parallèle slovène-italien a été composé dans le cadre du doctorat de T. Mikolič Južnič (Mikolič Južnič 2007). Ces corpus, trop petits pour pouvoir constituer un corpus de référence, puisent les textes parallèles dans différents domaines. Le corpus ELAN (http://nl.ijs.si/elan/headers-v2.html comprend un million de mots provenant de 15 textes parallèles anglais-slovène portant sur la législation européenne, quelques textes spécialisés et un texte littéraire. Le corpus slovène-italien ISPAC (Mikolič Južnič 2007) est constitué d'une partie littéraire et d'une partie non-littéraire, comprenant surtout la législation européenne en italien et en slovène. Leur utilisation est variée et va de l'étude de l'impact de la traduction sur le texte cible et de la comparaison des originaux et des traductions aux études terminologiques, lexicologiques, contrastives et culturelles.

3. L'utilité des corpus dans le travail du traducteur et dans la recherche en traductologie

18 À l'aide des corpus parallèles ou comparables, nous pourrions répondre à l'éternelle question de savoir « comment mesurer la qualité de la traduction ». À part les manuels normatifs que sont le dictionnaire, la grammaire et qui peuvent nous aider à juger si la tournure employée est grammaticalement correcte, il nous reste à déterminer si elle est adéquate. La réponse à cette question apparaît le plus probablement dans l'approche textuelle. Il existe plusieurs outils sur le web qui peuvent en témoigner. Pour le slovène, le corpus de référence monolingue — *FidaPLUS* — est accessible en ligne et gratuit. D'ailleurs, pour les traductions d'un certain type de textes, notamment les textes européens (politique, législation européenne…), il existe déjà une standardisation pour la langue cible qui est le slovène. Pour d'autres types de textes, le traducteur est obligé de créer ses propres outils, notamment les corpus monolingues et bilingues, sous la forme de textes modèles qui pourraient être des outils pour le type de traduction en question. Nous laisserons de côté la constitution des mémoires de traduction.

19 En vue d'améliorer la performance des traducteurs et l'accessibilité des outils, nous avons lancé le projet de préparation d'un corpus parallèle, constitué de textes français et de leurs traductions vers le slovène, et vice-versa, ainsi que d'un corpus comparable, comprenant les originaux de type textuel en français et en slovène. Ce projet, qui a débuté dans le cadre d'un séminaire en 4^{e} année au Département de traduction et d'interprétation de Ljubljana, est basé sur l'identification des types (ou des prototypes) des textes selon la classification de J. M. Adam (1997), parmi lesquels figurent le texte narratif, le texte descriptif, le texte explicatif et le texte argumentatif. Ces types, dans le cadre desquels nous identifions plusieurs genres de textes, nous servent de base textologique pour la recherche. Vu l'analyse des domaines dans lesquels se fait la traduction entre le français et le slovène que nous présentons ci-dessous, nous supposons que ce corpus devra être enrichi des corpus comparables pour les domaines

qui n'existent pas dans la traduction entre les deux langues, là où les contacts linguistiques sont limités.

3.1. Les relations interlinguistiques et interculturelles entre le français et le slovène

20 La réalité franco-slovène est en corrélation avec l'importance des deux langues en contact. Les liens entre le français et le slovène sont asymétriques ; il existe un nombre considérable de traductions du français vers le slovène et non l'inverse, résultat de l'importance des langues (2 millions de locuteurs slovènes) et du rayonnement culturel.

21 Les traductions du français vers le slovène se répertorient autour de la littérature française, assez bien représentée depuis toujours dans les traductions, suivent les ouvrages scientifiques du domaine des humanités, surtout dans le cadre de trois maisons d'édition spécialisées (Studia Humanitatis, /*cf. et Krtina).

22 Les textes scientifiques du domaine des sciences naturelles et techniques sont moins nombreux et limités aux entreprises qui sont liées à la France, ce qui est le cas de l'industrie automobile (Renault chez Revoz, Peugeot), ou présentes globalement dans le monde comme les entreprises de l'industrie pharmaceutique et cosmétique représentées par Krka ou Lek. On connaît également la traduction dictée par les besoins internes des Français travaillant dans un milieu slovène — les articles de presse contenant les commentaires sur le rôle de leur entreprise, les comptes-rendus des réunions, les consignes internes (ces données nous ont été confiées lors d'un séminaire organisé pour les traducteurs chez Revoz).

23 Les articles de la presse française en slovène apparaissent soit en tant que traductions comme c'est le cas du *Monde diplomatique*, soit comme adaptations d'articles des journaux *Le Monde*, l'*Express*, le *Nouvel Observateur* — dans le journal slovène *Delo*, à la rubrique « Les gens et le monde » (*Svet in ljudje*) et « Panorama ». La revue *Življenje in tehnika* publie les adaptations d'articles de la revue *Science et vie*.

24 En outre, une des sources possibles vise les textes institutionnels — politiques et juridiques — de l'Union européenne, ce qui a déjà été

exploré par l'*Evrokorpus*. Mais, dans ce cas-là, il est difficile de déterminer la vraie langue source, très souvent inconnue dans les textes européens, car ils peuvent être produits en plusieurs langues à la fois. La pratique traduisante en Slovénie veut que la plupart des traductions se fassent à partir de l'anglais ; le recours au texte français s'effectue le plus souvent là où la traduction à partir de l'anglais risque d'être ambigüe.

25 Les traductions du slovène vers le français sont plus rares. Nous pouvons mentionner la traduction de la littérature slovène — les œuvres fondamentales des classiques comme France Prešeren, Josip Jurčič, Ivan Cankar et les contemporains Drago Jančar, Boris Pahor, Boris A. Novak. Le choix des auteurs est favorisé par les traducteurs du slovène vers le français qui choisissent leurs auteurs préférés et l'activité des auteurs slovènes sollicitant eux-mêmes les traducteurs pour traduire leur livre, sans un éventuel programme culturel du Ministère slovène ou des éditeurs (Klinar 2008). Cependant, pour des raisons commerciales, la Slovénie trouve par exemple sa présentation écrite directement en français dans le *Guide du Routard*, (http://www.routard.com/guide/code_dest/slovenie.htm).

26 En parcourant les relations économiques et les traductions existantes entre le français et le slovène, il en résulte que tous les domaines ne sont pas représentés dans la traduction entre les deux langues à la même échelle que pour l'anglais ou l'allemand et le slovène. Les corpus parallèles ne peuvent pas couvrir tous les domaines — pour un corpus référentiel français-slovène, il faudrait inclure les textes comparables qui combleraient les lacunes.

3.2. Les domaines de la traduction favorisant la création des corpus entre le français et le slovène

27 Les études de traduction et d'interprétation à l'Université de Ljubljana ont débuté en 1997/98 grâce à un projet *Tempus* qui impliquait plusieurs partenaires parmi lesquels l'ITI-RI de Strasbourg, les Instituts de traduction et d'interprétation de Karl-Franzes Universität de Graz, Trieste, Mannheim, dont les bases ont été posées dès 1995. Les premiers diplômés en sont sortis en 2001, et maintenant nous nous

préparons à accueillir les premiers diplômés de la réforme LMD en médiation interlinguistique qui pourront poursuivre leurs études en master de traduction ou d'interprétation. Les études doctorales se font depuis 2005.

28 Les études dans le cadre du Département de traduction et d'interprétation à l'Université de Ljubljana développent les capacités des apprenants dans deux sens. Premièrement, c'est dans le sens d'une pratique traduisante assurant une qualité qui leur permettra de satisfaire les besoins du marché, et deuxièmement, dans le sens du développement de la capacité de la recherche en traduction. L'épreuve finale consiste en une traduction pour chaque combinaison linguistique qui doit être faite sur place et défendue. Suivent la composition et la soutenance d'un mémoire de traduction où l'étudiant se sert de la méthodologie appropriée pour la recherche en traductologie. Les recherches relatives à ces mémoires se fondent en grande majorité sur une compilation du corpus quel que soit le sujet. Les corpussont ensuite analysés manuellement ou à l'aide d'outils de traitement automatique des langues.

29 Par ailleurs, dès le début de leurs études, les étudiants travaillent sur des corpus monolingues en tant que ressources linguistiques. Leur emploi est obligatoire dans le cadre des études de linguistique appliquée et constitue une base pour les études contrastives ou lexicologiques. Plus tard, dans le cadre des outils de traduction assistée qui figurent au curriculum des études, les étudiants composent eux-mêmes les corpus destinés à la recherche, dont l'objectif est d'étudier un problème de ce domaine — soit un problème de traductologie descriptive, soit de linguistique contrastive ou de lexicologie, mais aussi en vue d'une comparaison des traductions et des originaux en une langue afin de repérer les éventuelles interférences entre l'original et la traduction. Ces travaux de recherche se font sur toutes les langues de travail disponibles au Département de traduction et d'interprétation, notamment l'anglais, l'allemand, le français et l'italien. Dans la suite de cette communication, nous nous consacrerons à la combinaison linguistique français-slovène.

30 Dans le cadre du séminaire de textologie française (en 4^{e} année d'études), les étudiants cherchent les textes qui seraient utiles dans le cadre de ce séminaire destinés à l'étude des quatre prototypes

textuels, notamment les textes narratifs, descriptifs, explicatifs et argumentatifs. D'autre part, nous constatons que de plus en plus de mémoires et de thèses de doctorat visent à explorer les données des corpus. Or il faut d'abord les créer. Ces corpus, annotés ou pas, restent à disposition pour la compilation d'un corpus bilingue plus large. Nous présenterons les types d'études qui ont été menés dans différents domaines de traduction entre le français et le slovène. Ici, un travail important est effectué aussi par les étudiants de traduction dans le cadre de leurs mémoires ou par les doctorants.

31 Parmi les mémoires de diplôme basés sur les corpus créés dans un but lexicologique, il faut mentionner les compositions de glossaires, dont un sur l'industrie automobile (Dejak 2007) et un sur l'œnologie (Gorjup 2008). Les corpus ont été élaborés sur la base de textes techniques représentant les domaines concernés.

32 Pour ce qui est des travaux orientés vers une analyse grammaticale contrastive, A. Mezeg (2007) a étudié les verbes introducteurs dans l'œuvre de Flaubert *Madame Bovary* et ses traductions vers le slovène. J. Zidar (2007) a étudié le conditionnel français et ses variantes dans les traductions vers le slovène. Ces recherches ont exploité les corpus littéraires les plus facilement accessibles. Deux mémoires ont été consacrés à l'oral transcrit venant des discours et leurs interprétations dans le Parlement européen (M. Miljančič 2008, L. Rogič 2006), pour n'en mentionner que quelques-uns.

33 L'analyse d'un corpus comparable constitué de 200 articles de journaux originaux slovènes et français a servi de base pour une étude sur l'ordre des mots et la variation thème-rhème en français et en slovène (Schlamberger Brezar 2005b). Ces données ont été élaborées comme un recueil pré-corporel, mais pourront faire partie d'un corpus plus vaste si l'occasion se présente.

34 Dans le cadre des études doctorales, deux personnes s'occupent de la compilation des corpus électroniques parallèles pour la combinaison français-slovène. Adriana Mezeg prépare un corpus parallèle où le français est la langue source et le slovène la langue cible. Ses sources sont le *Monde diplomatique* et ses traductions pour l'édition slovène et quelques textes littéraires. Elle cherche la définition qui pourrait constituer une base pour une recherche informatisée des constructions détachées dans son corpus.

35 Le corpus de Simona Šumrada (2009) va du slovène vers le français et comprend les textes touristiques élaborés en Slovénie à l'intention des Français. Elle cible les relations entre l'implicitation et l'explicitation dans le texte de la langue source et le texte de la langue cible.

36 Des recherches doctorales se font aussi dans le cadre de *WordNet*, un thésaurus électronique, exemple d'une base lexicale conçue pour le support électronique (Habert, 1997 : 85). Son objectif est de décrire les sens de mots ou concepts. Il existe pour l'anglais, l'allemand, l'italien, l'espagnol, mais Darja Fišer vient de l'élaborer pour le slovène dans le cadre de ses études doctorales. Elle poursuit son travail sur l'élaboration du *WordNet* français avec B. Sagot (Sagot, Fišer, 2008).

3.3. La composition du corpus des textes parallèles et comparables français-slovène

37 La constitution du corpus slovène-français débutera à une grande échelle dans le cadre d'un projet du Département de traduction et d'interprétation avec l'aide du Ministère de la science et de la technologie. La composition d'un corpus bilingue français-slovène semble urgente aussi bien pour les recherches lexicologiques que textologiques dans le cadre de la traductologie.

38 Ayant en vue les relations interlinguistiques entre le français et le slovène, nous constatons que la recherche devra se dérouler dans les genres suivants : les textes littéraires, les articles de presse, la législation européenne, les textes du domaine de la technologie et du tourisme. Malgré les domaines relativement restreints d'échange entre les langues, nous pouvons parler de la programmation de ce corpus dans lequel, si le besoin s'en ressent, nous inclurons des textes comparables qui combleraient les lacunes.

39 Le futur corpus voudrait traiter les prototypes textologiques selon la classification de J. M. Adam (1997) dans le cadre typologique suivant : narratif, descriptif, argumentatif, explicatif et dialogique. La base des prototypes de textes nous garantit la variété textuelle — dans le meilleur des cas — aussi bien pour les corpus comparables que pour les corpus parallèles. Dans le cadre de ces quatre prototypes, il faut également veiller à une diversité thématique. Il faudra ajouter des

textes à schéma culturel typique comme les invitations, les faire-part, la correspondance commerciale, ces textes pouvant être analysés de manière contrastive à l'aide des corpus comparables composés des originaux en deux langues. Il ne faut cependant pas oublier l'aspect pratique et considérer surtout les textes qui représentent un enjeu pour les traducteurs de cette combinaison linguistique actuellement mais aussi dans le futur.

40 Les textes sont unis dans un corpus parallèle et ils sont déjà alignés, la phase suivante est l'annotation qui associera aussi les informaticiens. L'annotation sera faite de manière à inclure aussi les données textologiques, concernant le thème et le rhèmece qui permettra aussi d'explorer les différences d'ordre des mots dans les deux langues concernées. Nous pourrons inclure dans la recherche l'exploitation du corpus référentiel slovène monolingue *FidaPlus*, accessible librement par abonnement. Un tel corpus, malheureusement, n'existe pas pour le français.

3.4. Les droits d'auteurs

41 Finalement, il nous reste à discuter la question de savoir s'il est encore possible de constituer un corpus librement. Aujourd'hui, un grand nombre de corpus de première et de deuxième génération ne peuvent plus être utilisés parce que la question des droits d'auteurs n'a pas été résolue (Kennedy, 1998). Le domaine de la littérature jouit d'une certaine liberté, au moins en Slovénie puisque les oeuvres littéraires font partie du domaine public si plus de 70 ans se sont écoulés depuis la mort de l'auteur. Mais les enjeux se situent le plus souvent dans la langue moderne. Ainsi, la possibilité de négociation des droits visant des objectifs scientifiques, non commerciaux, existe toujours.

42 En général, les auteurs permettent habituellement l'utilisation de leurs oeuvres en vue d'une exploitation non commerciale (Malmkjaer, 2004 : 89). Telle a été aussi notre expérience avec les auteurs ainsi que les maisons d'édition slovènes. Les éditeurs slovènes du journal *Le Monde diplomatique* ainsi que quelques maisons d'édition littéraires (Učila, Beletrina) ont permis l'utilisation des traductions vers le slovène. Si les éditeurs sont sensibles à l'intérêt scientifique d'une telle démarche, les problèmes des droits d'auteurs ne présentent pas un obstacle. Mais il faut être sensible à cet aspect et y sensibiliser

aussi les étudiants dans leur travail pour les protéger des conséquences néfastes de la publication ou l'exploitation des contrats ou des documents industriels confidentiels qu'ils ont recueillis quelque part et exploités par la suite.

Conclusion

43 Dans la recherche linguistique de nos jours, il n'y a plus de place pour la fameuse « intuition linguistique » prêchée par Noam Chomsky et ses successeurs. Les corpus électroniques en linguistique ainsi qu'en traductologie deviennent des outils de description indéniables. Ainsi, la compilation des corpus comme base de recherche est inévitable pour la crédibilité scientifique.

44 Les corpus en traductologie aujourd'hui sont une nécessité, aussi bien du point de vue commercial comme les mémoires de traduction pour la traduction assistée par ordinateurs, que du point de vue de la recherche linguistique et traductologique. Les étudiants apprennent les notions de compilation de corpus au moins pour leurs présentations et la recherche manuelle peut être une bonne base en vue d'une future pratique dans la composition des corpus et leur annotation.

45 Pour la paire linguistique slovène-français, il est nécessaire de créer des matériaux qui rendraient possibles des recherches dans le domaine de la lexicologie et de la linguistique contrastive basées sur les descriptions faites à partir des données modernes. Ces recherches pourraient mener à une meilleure pratique traduisante fondée sur la réalité linguistique. C'est pourquoi la composition d'un corpus bilingue parallèle ou comparable est nécessaire pour assurer un avenir aux traducteurs disposant de cette combinaison linguistique.

BIBLIOGRAPHY

ADAM Jean-Michel, 1997, *Les textes : types et prototypes. Récit, description, argumentation, explication et dialogue*, Paris, Nathan.

BAJEC Anton *et al.*, 1994, *Slovar slovenskega knjižnega jezika*, Ljubljana, DZS.

BAKER Mona, 1993, «Corpus linguistics and translation studies. Implications and Applications », dans BAKER Mona, FRANCIS G., TOGNINI-BONELLI E. (éd.) : *Text and Technology : In honour of John Sinclair*, Amsterdam, John Benjamins, p. 17-45.

BAKER Mona, 1995, « Corpora in translation studies. An overview and some suggestions for future research », dans *Target* 7, 1995, p. 223-243.

BIBER Douglas, CONRAD S., REPPEN R., 1998, *Corpus linguistics. Investigating language structure and use*, Cambridge, Cambridge University Press.

CHOMSKY Noam, 1957, *Syntactic Structures*, Berlin, New York, Mouton de Gruyter.

DEJAK, Špela, 2007, *Francosko-slovenski glosar s področja notranje opreme avtomobilov*, mémoire de diplôme, Ljubljana, Filozofska fakulteta, Oddelek za prevajalstvo.

GORJANC Vojko, 1999, « Korpusi v jezikoslovju in korpus slovenskega jezika FIDA », dans KRZISNIK Erika, LOKAR Meta (éd.) *Zbornik predavanj*. Ljubljana : Center za slovenščino kot drugi/tuji jezik pri Oddelku za slovanske jezike in književnosti Filozofske fakultete, p. 47-59.

GORJANC Vojko, 2003, « Korpusi in jezikoslovje », dans *Jezik in slovstvo*, 48, nº 3-4, p. 19-27.

GORJANC Vojko, 2005, *Uvod v korpusno jezikoslovje*, Domžale, Izolit.

GORJANC Vojko, KREK, Simon. (éd.), 2005, *Študije o korpusnem jezikoslovju*, Ljubljana, Krt.

GORJUP Polona, 2008, *Francosko-slovenski razlagalni glosar enoloških izrazov v programu Trados MultiTerm*, mémoire de diplôme, Ljubljana, Filozofska fakulteta, Oddelek za prevajalstvo.

HABERT Benoît, NAZARENKO Adeline, SALEM André, 1997, *Les linguistiques de corpus*, Paris, Armand Colin.

HUNSTON Susan, 2002, *Corpora in Applied Linguistics*, Cambridge, Cambridge University Press.

KENNEDY Greame, 1998, *An introduction to Corpus Linguistics*, Amsterdam, Rodopi.

KOCIJANCIC POKORN Nike, 2008, « Translation and TS research in a culture using a language of limited diffusion : the case of Slovenia », dans *The journal of specialised translation*, 10, p. 2-9, en ligne [http://www.jostrans.org/issue10/art_pokorn.pdf (2008-08)], consulté le 9 août 2008.

KLINAR Meta, 2008, *Slovensko-francoske književne izmenjave*, mémoire de maîtrise, Ljubljana, Filozofska fakulteta, Oddelek za slovenistiko.

LAVIOSA Sara (ed.), 1998, « L'approche basée sur le corpus / The corpus based approach », dans *Meta* 43 (4), 1998, p. 474-479, en ligne [http://www.erudit.org/revue/meta] [http://id.erudit.org/iderudit/003424ar], consultés le 9 août 2008.

LAVIOSA, Sara, 1998, « Core patterns of Lexical Use in a Comparable Corpus of English Narrative Prose », dans *Meta* 43 (4), 1998, p. 557-570, consulté le 9 août 2008, en ligne [http://www.erud-

it.org/revue/meta, http://id.erudit.org/iderudit/003425ar].

MALMKJAER Kirstin (éd.), 2004, *The Linguistic encyclopaedia*, London, Routledge.

MEZEG Adriana, 2007, « Načini prevajanja francoskih glagolov iz spremnih stavkov premega govora v treh slovenskih prevodih Flaubertove Madame Bovary », dans *Jezik in slovstvo*, 52, št. 2, p. 81-93.

MIKOLIC JUZNIC Tamara, 2007, *Nominalne strukture v italijanščini in slovenščini : pogostnost, tipi, in prevodne ustreznice*, thèse de doctorat du III[e] cycle, Ljubljana : Univerza v Ljubljani, Filozofska fakulteta.

MILJANCIC Marijana, 2008, *Analiza argumentacije pri simultanem tolmačenju iz francoščine v slovenščino*, mémoire de diplôme, Ljubljana, Filozofska fakulteta, Oddelek za prevajalstvo.

OLOHAN Maeve, 2004, *Introducing Corpora in Translation Studies*, London, New York, Routledge.

PAGON Neda *et al.*, 2005, *Studia Humanitatis : Dvajset let*, Ljubljana, Studia Humanitatis.

QUIRK Randolph, GREENBAUM Sidney, LEECH Geoffrey, SVARTVIK Jan, 1985, *A Comprehensive Grammar of the English Language*, London, Longman.

ROGIC Luka, 2006, *Prevajanje povezovalcev v simultanem tolmačenju iz francoščine v slovenščino*, mémoire de diplôme, Ljubljana, Filozofska fakulteta, Oddelek za prevajalstvo.

SAGOT, Benoît, FISER Darja, 2008, « Building a free French wordnet from multilingual resources » (http://lojze.lugos.si/~darja/papers/ontolex08-sagot-fiser.pdf), dans *Proceedings of the OntoLex workshop*OntoLex'08 (http://www.loa-cnr.it/ontolex08), 31 May 2008, Marrakech, Maroc, en ligne [http://lojze.lugos.si/~darja/papers/ontolex08-sagot-fiser.pdf], consulté le 9 août 2008.

SAGOT Benoît, FISER Darja, 2008, « Construction d'un wordnet libre du français à partir de ressources multilingues (http://lojze.lugos.si/~darja/papers/taln08-sagot-fiser.pdf)», dans *Proceedings of the conference on automatic processing of natural languages* TALN'08 (http://www.lia.univ-avignon.fr/index.php?id=77), 9-13 June 2008, Avignon, France, consulté le 9 août 2008.

SCHLAMBERGER BREZAR Mojca, 2005a, « Politična besedila kot tip besedil in postopki prevajanja », dans KOCIJANCIC POKORN Nike, PRUNC Erich, RICCARDI Alessandra (éd.), *Beyond equivalence = Jenseits der Aequivlenz = Oltre l'equivalenza = Onkraj ekvivalence*, Graz, Institut für Translationwissenschaft, Karl Franzens-Universität, p. 121-135.

SCHLAMBERGER BREZAR Mojca, 2005b, « Problem of information in French and Slovene short newspapers articles », dans *Formal, functional and typological perspectives on discourse and grammar*, Valencia, SLE, 2005, p. 214-215.

STABEJ Marko, 2003, « Jezikovnetehnologije in jezikovno načrtovanje », dans *Jezik in slovstvo*, 48, n° 3-4, p. 5-18.

ŠUMRADA Simona, 2009, « Eksplicitnost pri prevajanju slovenskih narodnih jedi », dans *Jezik in slovstvo* 54/2, 2009, p. 81-100.

TOPORISIC Jože, 2000, *Slovenska slovnica*, Maribor, Obzorja.

VINCENOT Claude, 1975, *Essai de grammaire slovène*, Ljubljana, Mladinska knjiga.

VINTAR Špela, 2003, « Kaj izvira iz jezikovnih virov », dans *Jezik in slovstvo*, 48, nº 3-4, p. 77-88.

ZIDAR Jana, 2007, *Francoski sedanji pogojnik v prevodih romanov v slovenski jezik*, mémoire de diplôme, Ljubljana, Filozofska fakulteta, Oddelek za prevajalstvo.

Ressources internet

Le Trésor de la langue française informatisé, CNRS, [http://atilf.atilf.fr/], consulté le 29 janvier 2009.

Corpus Fida, [www.fida.net], consulté le 29 janvier 2009.

Corpus FidaPlus, [www.fidaplus.net] consulté le 29 janvier 2009.

French language ressource, University of Virginia, [http://etext.lib.virginia.edu/collections/languages/french/], consulté le 29 janvier 2009.

Abu : la bibliothèque universelle, [http://abu.cnam.fr/], consulté le 29 janvier 2009.

Athéna : textes français, [http://un2sg4.unige.ch/athena/html/francaut.html], consulté le 29 janvier 2009.

Université catholique, Leuven [http://bach.arts.kuleuven.ac.be/lancom/], consulté le 5 août 2008.

Université Paris VII — Diderot, [http://wwwtalana.linguist.jussieu.fr/Presentation/corpus.html], consulté le 5 août 2008.

ELAN, [http://nl.ijs.si/elan/headers-v2.html], consulté le 29 janvier 2009.

Evrokorpus, [http://evrokorpus.gov.si/index.php], consulté le 5 août 2008.

Corpus Nova Beseda, [http://bos.zrc-sazu.si/s_beseda.html], consulté le 5 août 2008.

Corpus Elan, [http://nl.ijs.si/elan/], consulté le 5 août 2008.

Trans search, Canadian Hansard, [http://www.tsrali.com/], consulté le 29 janvier 2009.

The English-Norvegian parallel corpus, [http://www.hf.uio.no/ilos/forskning/forskningsprosjekter/enpc/], consulté le 29 janvier 2009.

NOTES

1 1809-1813

2 Il s'agit de 97 langues parmi lesquelles le slovène, mais la base des données franco-slovène n'existe pas encore.

ABSTRACTS

Français

Dans les années 80, les recherches linguistiques se sont généralement orientées vers l'élaboration des corpus qui ont aussi influencé les recherches en traduction. Pour le slovène, il existe des corpus monolingues référentiels ainsi que bilingues destinés à l'usage des traducteurs, mais surtout en combinaison avec l'anglais. Il est important de créer un corpus slovène-français. Les liens interculturels entre ces deux langues ont toujours été assymétriques, le français ayant plus d'importance pour les Slovènes que vice-versa. Les traductions se faisaient surtout dans le cadre de la littérature et des sciences humaines, et moins en technologie. Dans le cadre du département des études de traduction et d'interprétation à l'université de Ljubljana, nous avons lancé un projet de compilation de corpus parallèle composé de textes français et de leurs traductions vers le slovène, et vice-versa, ainsi que d'un corpus comparable composé de textes originaux du même type en français et en slovène. Ce projet constitue un fondement pour les recherches lexicologiques et textologiques en traduction entre ces deux langues pour le futur.

English

Linguistic research from the end of '80 seeks for empirical confirmation within the corpora, used today for linguistic, lexicological research and in translation studies. The Slovene language is well provided with monolingual referential corpora and with bilingual corpora but only in combination with English. Now we have to take care of a translation corpus for French-Slovene language combination. The intercultural links between Slovene and French are assymmetrical, French being more important to Slovenes than the other way round. The translations were mostly done in the field of literature and humanities and much less in technology. At the Department of Translation Studies of Ljubljana University we begun a project of building a parallel corpus from French originals and their translations into Slovene, and vice versa, supported by a comparable corpus consisting of originals of the same textual type in French and in Slovene. The project presents a basis for lexicological ant textological research in translation between the two languages in the future.

INDEX

Mots-clés

combinaison linguistique slovène-français, corpus comparable, corpus parallèle, textes-types en traduction entre le français et le slovène,

traduction

Keywords
comparable corpus, parallel corpus, Slovene-French linguistic combination, translation, typical texts in translation between French and Slovene

AUTHOR

Mojca Schlamberger Brezar
Professeur de linguistique française et de traduction au département de traduction et d'interprétation de l'université de Ljubljana. Elle a fait sa thèse sur les connecteurs dans le discours oral médiatique. Maître de conférences invitée à l'INALCO à Paris en 2007/2008, elle a publié des ouvrages sur l'apprentissage du slovène-langue étrangère et s'occupe également de traduction.
IDREF : https://www.idref.fr/095761314
ISNI : http://www.isni.org/000000011671695X
BNF : https://data.bnf.fr/fr/15017632