



HAL
open science

Acquisition, traduction, usages : des linguistiques énonciatives et cognitives à l'automatisation

Caroline Rossi

► **To cite this version:**

Caroline Rossi. Acquisition, traduction, usages : des linguistiques énonciatives et cognitives à l'automatisation. Linguistique. Université Lumière Lyon 2, 2018. tel-03659740

HAL Id: tel-03659740

<https://hal.science/tel-03659740>

Submitted on 5 May 2022

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.

Acquisition, traduction, usages : des linguistiques énonciatives et cognitives à l'automatisation

Dossier présenté en vue de l'Habilitation à Diriger des Recherches
Soutenance publique le 7 décembre 2018 à Lyon (MILC)

Volume 1. Note de synthèse

Jury :

Pr. Yvon Keromnes, Université de Lorraine et ATILF

Pr. Natalie Kübler, Université Paris Diderot et CLILLAC-Arp

Pr. Élisabeth Lavault-Olléon, Université Grenoble Alpes et ILCEA4

Pr. François Maniez, Université Lumière Lyon 2 et CRTT (garant)

Pr. Christine Michaux, Université de Mons (pré-rapporteur)

Pr. Aliyah Morgenstern, Université Paris 3 Sorbonne Nouvelle et SeSyLiA

Pr. Martine Sekali, Université Paris Nanterre et CREA (rapporteur)

Pr. Jean Soubrier, Université Lumière Lyon 2 et CRTT (pré-rapporteur)

Pour Angèle,

ma tante chérie qui a aimé et transmis le français à des générations d'enfants, et qui doit maintenant reprendre, avec beaucoup de courage, le chemin du langage.

Lascià

a so memoria
nantu una pagina
cum'ellu si lascia
sfattu è apertu
u lettu a matina
à purtelli spalancati
ch'ellu si rinfreschi
d'un odore
di mundu novu.

Laisser

sa mémoire
sur une page
comme on laisse
son lit défait le matin
fenêtre grande ouverte
qu'il se rafraîchisse
d'une odeur
de monde nouveau

Sonia Moretti, traduite par elle-même, *Poesie di a curtalina*, 2009, Ajaccio, Albiana/CCU.

Remerciements

Ce travail a été réalisé grâce à un congé de recherche et de conversion thématique de six mois, qui m'a été accordé par l'Université Grenoble Alpes.

Il a également bénéficié d'un suivi attentif et minutieux, et je souhaite commencer par remercier mon garant, François Maniez, qui a accepté de se rendre disponible en plein cœur de l'été, a répondu à mes innombrables questions et m'a orientée à plusieurs reprises de façon décisive.

La préparation de ce document de synthèse a été à bien des égards un travail solitaire, mais en tant qu'il m'a permis de prendre acte d'un tournant que je n'ai pas amorcée seule, il doit beaucoup à mes collègues.

D'abord ceux, lyonnais, qui m'ont recrutée et fait découvrir la filière LEA en même temps que de nouvelles pistes de recherche : mes premières interrogations sur les composés et leur acquisition et mon travail sur la néologie en sont issus.

Ensuite l'équipe grenobloise qui m'a accueillie à l'automne 2013, et qui m'a fait confiance en me proposant des responsabilités et des enseignements qui auront été des accélérateurs de découvertes. C'est à Grenoble que j'ai commencé à appréhender la traductologie dans toute sa richesse, et que j'ai eu envie de suivre certaines des nombreuses pistes ouvertes par les travaux d'Élisabeth Lavault-Olléon.

C'est aussi à Grenoble qu'il m'a été donné de poursuivre un questionnement psycholinguistique que j'avais amorcé avec ma thèse, tout en l'articulant à mes nouvelles pratiques : je n'y serais jamais arrivée sans le soutien et les conseils de Jean-Pierre Chevrot, qui m'a montré que les horizons se créent tout autant qu'ils s'ouvrent.

Si j'ai pu continuer à travailler sur l'acquisition du langage suite à un recrutement qui ne m'y destinait pas, c'est aussi dans le cadre de projets financés et des dynamiques que leurs porteuses ont créé. Je remercie Sophie Kern de m'avoir permis de retrouver le laboratoire DDL où j'avais passé mes années de thèse, et Aliyah Morgenstern qui m'a intégrée à un groupe de recherche dynamique et passionné au sein duquel j'ai énormément appris.

Un grand merci à Dorothy Kenny pour tout ce que nous avons partagé et partagerons encore : cela va bien au-delà d'un PHC ! Merci à Olga Torres d'avoir relevé avec nous le défi du dépôt Erasmus+ cette année et d'accepter de le poursuivre encore jusqu'au prochain dépôt. Merci aux collègues du bureau du Pôle Grenoble Cognition pour la richesse des échanges.

Je remercie aussi les collègues avec lesquels j'ai partagé un bureau, des cours ou des projets de recherche, pour leurs encouragements et leur bienveillance : Pierre-Alexandre Beylier, Grégory Albisson, Sandrine Tolazzi, Cyril Besson, Camille Biros, Séverine Wozniak, Inesa Sahakyan, Evelyne Chabert, Alice Carré, Aurélien Talbot, Emanuela Nanni, Lionel Karagouch et Élisabeth Lavault-Olléon, pour ne citer que les principaux. Je dois également beaucoup à Shaeda Isani, que je remercie pour son *coaching* estival.

Merci à tous mes amis et proches, qui m'ont supportée et ont compris que je ne serais pas très disponible, merci à mes enfants pour leurs jeux, leurs blagues, leurs histoires et leur incroyable énergie. Merci à mes parents et beaux-parents pour leur patience, leur dévouement, et tout ce qu'ils transmettent à leurs petits-enfants.

Merci, surtout, à mon *Fabulous Fab* qui a tenu bon et qui mériterait plusieurs médailles !

Un immense merci, enfin, aux rapporteurs qui ont bien voulu travailler dans des délais serrés, et à l'ensemble des membres du jury.

Table des matières

Remerciements	2
Table des matières	4
Avant-propos : de la diversité des langues à l'automatisation	7
Première partie. La linguistique appliquée, d'un champ à l'autre	10
1. Méthodes et outils	11
1.1 Corpus oraux, corpus écrits	13
1.2 Variation, variabilité	18
1.3 Contextes	25
1.4 Fréquences et seuils de significativité : réflexion sur les statistiques descriptives et inférentielles.....	31
2. D'un champ à l'autre : principaux points d'ancrage théoriques.....	37
2.1 Points d'ancrage linguistiques.....	37
2.2 De la psycholinguistique développementale à la traductologie	47
2.3 La traductologie à la croisée des chemins	52
2.4 Elargir le champ de la cognition	56
2.5 Les linguistiques appliquées « à travers champs »	58
Deuxième partie. Travaux et projets	60
1. Catégorisation et typologies.....	61
1.1 À l'arrière-plan : la catégorisation en philosophie du langage et en linguistique cognitive.....	61
1.2 Noms et verbes, de la catégorisation innée au déclin des catégories traditionnelles	64
1.3 Catégories situationnelles, catégories en contexte	65
1.4 La catégorisation au prisme de la relativité linguistique	66
2. Mouvement et langage	68
2.1 L'expression de la manière de mouvement	68
2.2 Le futur et son acquisition	70
2.3 Aux confins des schémas de lexicalisation : quelques applications à l'enseignement de la traduction	71
3. Les composés et leur acquisition	73
3.1 Mécanismes de composition et d'amalgamation, néologie et créativité.....	73
3.2 L'alternance du génitif : arguments en faveur d'une approche probabiliste de la grammaire	76
3.3 Applications à la didactique de la traduction (vers une langue seconde).....	80
4. Le traducteur et ses outils.....	81
4.1 La traduction automatique	81
4.2 Le corpus et les mémoires.....	85

4.3	Les applications à l'enseignement	87
Synthèse : vers une conception critique de l'automatisation et de ses retombées en sciences du langage et en traductologie		90
1.	Méthodes mixtes, objets variés.....	90
1.1	Mettre en œuvre des synergies pluridisciplinaire pour analyser le discours sur la justice climatique.....	91
1.2	Revenir aux données qualitatives pour mieux expliquer : l'analyse des usages et perceptions de la TA à la DGT.	94
1.3	Appliquer des modèles probabilistes à l'acquisition L2 : une exploration séquentielle complexe	95
2.	Réflexion épistémologique	98
2.1	Des règles aux statistiques : destins croisés de la linguistique et de la traduction automatique.....	98
2.2	Des modèles connexionnistes aux réseaux de neurones.....	100
3.	Construire une critique de l'automatisation	103
3.1	Repenser le rôle de l'utilisateur : le projet M2TLabs	103
3.2	Mieux articuler cognition individuelle et comportements collectifs : le projet Co3i.....	105
3.3	Développer une théorie du contexte : le projet CL3i.....	106
3.4	Interroger nos pratiques : l'éthique de la recherche.....	108
Bibliographie.....		111
Index.....		134
Annexe : Curriculum Vitae		136
1.	Trajectoire professionnelle	136
2	Expérience pédagogique.....	136
2.1	Enseignements	136
2.2	Implication dans des activités liées au numérique.....	137
2.3	Responsabilités administratives.....	137
3	Activités d'animation et d'encadrement scientifique	137
3.1	Participation à la recherche collective à l'Université Grenoble Alpes	137
3.2	Travail éditorial	138
3.3	Colloques et conférences (organisation et évaluation).....	138
3.4	Bases de données, dispositifs empiriques, documentation.....	140
3.5	Participation à des projets de recherche financés	140
3.6	Blogs (carnets)	143
3.7	Réponse à des appels à projet et participation à la constitution de réseaux internationaux	143
3.8	Encadrement de la recherche	144
3.9	Membre de sociétés savantes	146

4	Publications et travaux.....	146
4.1	Articles dans des revues internationales à comité de lecture	146
4.2	Articles dans des revues nationales à comité de lecture	146
4.3	Directions de numéros de revues ou d'ouvrages	147
4.4	Chapitres d'ouvrages (publiés après évaluation par un comité de lecture).....	147
4.5	Communications avec actes édités (après évaluation par un comité de lecture).....	148
4.6	Diffusion de la recherche (publications à visée de vulgarisation).....	149
4.7	Principales communications orales	149
4.8	Principales conférences à titre d'invitée	152
5.	Formations suivies.....	153
5.1	Tests statistiques dans R Studio	153
5.2	Initiation aux démarches de créativité	153

Avant-propos : de la diversité des langues à l'automatisation

Redécouvrir et expliquer le chemin qui mène d'une thèse appliquant les acquis de la linguistique contrastive et de la typologie talmyenne à l'acquisition du langage, jusqu'à une réflexion sur l'automatisation : tel est l'objet de cette synthèse. Ce chemin, je crois l'avoir frayé avant tout par goût pour l'exploration et l'application de théories diverses, et guidée par la même volonté qui m'animait déjà il y a dix ans, lorsque je préparais ma thèse : celle de circonscrire un peu mieux le mystère du langage et des langues. En revenant sur les détours de mon itinéraire, je découvre cependant que les ponts que j'ai cru devoir imaginer ont déjà été évoqués, parfois même construits et parcourus. Voici par exemple ce qu'écrit Antoine Culioli en 1973, à propos de ce qu'il appelle « quelques contradictions en linguistique » :

Si l'on accepte de broser à gros traits l'évolution des recherches linguistiques, on ne simplifiera pas trop en marquant que les quinze dernières années ont vu la redécouverte du langage en tant qu'activité signifiante ; en outre, l'on commence à se poser avec quelque lucidité le problème de la relation existant entre la faculté universelle de produire et interpréter des textes d'un côté, et de l'autre la diversité des langues naturelles.

Il est vrai que, du même élan, se développe à nouveau un fantasme technologique, indûment assimilé à une théorie, qui prend des apparences variées et parfois idéologiquement contradictoires. Ainsi, on voit ressurgir le behaviorisme sous une forme plus agressive mais toujours aussi mécaniste; des psycholinguistes se remettent à chronométrer des opérations au statut incertain; le logicisme revient en course, aidé par la fascination qu'exerce tout calcul sur le public et par l'emprise croissante des mathématiques, comme nouveau symbole de l'accomplissement social. A lire certains, on apprend que la mathématique est le langage parfait auquel se rapporte toute grammaire, que le langage humain est réductible à la logique, que le calcul des prédicats peut encore servir, sans que l'on se préoccupe de ses fondements extensionalistes. De même, à force de parler, par métaphore, d'automate et de dispositifs, on se réhabitue à concevoir le langage comme une machine à traiter les lexèmes et la sémantique comme une sémantique générale. (Culioli 1973: 83).

Les linguistiques de l'énonciation, puis l'ensemble des théories basées sur l'usage, se développent en marge du paradigme computationnel comme des théories chomskyennes (dont l'inspiration est d'ailleurs en partie semblable même si elles s'en sont détachées dès la fin des années 1970, voir par ex. Pullum 1989: 139). En 1973, les formalismes développés au MIT et l'ensemble des travaux menés en linguistique computationnelle (notamment autour des moteurs de traduction automatique) entretiennent depuis plusieurs décennies déjà ce que Culioli appelle « un fantasme technologique ». L'automatisation et toutes les tentatives réductionnistes qui l'accompagnent sont déjà très bien décrites ici : réduction, par le behaviorisme, du comportement

à un ensemble de mécanismes (réponses à des stimulations), réduction de la complexité à des observables simples (mesures temporelles de processus) ou à des formalismes inadéquats (ensemble restreint de concepts issus de la logique, calcul mathématique).

Culiolienne de formation, je me suis toujours intéressée aux formalismes du point de vue de leur application à différents domaines et questions de recherche. Les propositions qui sont faites dans ce volume de synthèse élaborent ainsi quelque chose comme une linguistique des confins, et interrogent le rôle et la place de la linguistique lorsque ses observables sont construits ou saisis par une démarche qui implique d'autres disciplines ou champs disciplinaires. Née au milieu des montagnes, et probablement influencée par l'environnement insulaire dans lequel j'ai également évolué, j'aime en effet travailler « en plein milieu des confins ». L'expression est de Pierre Dac, elle a été reprise par Nicolas Froeliger (2013: 13) qui en fait justement une métaphore de la traduction comme objet difficile à circonscrire, et comme lieu de passage. Or pour comprendre l'acquisition du langage et des langues (maternelles et secondes), il faut aussi accepter de se situer dans un entre-deux, qui peut s'entendre de deux façons différentes au moins : d'une part, il s'agit en permanence de naviguer entre linguistique, psychologie et neurosciences, et d'autre part, l'analyse de productions enfantines, comme d'ailleurs celle des productions d'apprenants d'une langue seconde, pose une série de problèmes qui se situent aux frontières des catégories d'analyse traditionnelles.

La première partie revient sur les modèles et notions théoriques avec lesquels j'ai travaillé, elle en interroge les fondements épistémologiques et propose une réflexion critique sur un ensemble d'éléments méthodologiques communs à plusieurs champs de la linguistique appliquée. Ces champs sont ceux que j'ai traversés : la psycholinguistique (acquisition d'une langue maternelle ou seconde), la linguistique de corpus et la traductologie.

La seconde partie propose un parcours thématique de mes travaux : elle s'ancre dans une interrogation déjà ancienne sur la catégorisation, présente mon travail sur l'expression de l'espace et du mouvement en français et en anglais, avant d'aborder les deux fils conducteurs de mes recherches actuelles. Ceux-ci concernent d'une part une réflexion sur l'acquisition des composés, qui m'a menée du français langue maternelle à l'anglais langue seconde, et d'autre part, un travail en cours sur le traducteur et ses outils. Ce dernier volet remet en jeu, dans un contexte qu'il m'a été donné de découvrir récemment (celui de la traduction professionnelle), plusieurs questions issues de mes travaux antérieurs :

- Il définit un ensemble d'applications des corpus qui est encore assez peu décrit : il concerne notamment les usages de corpus parallèles intégrés aux outils de traduction assistée par ordinateur sous forme de mémoires de traduction, et utilisés pour entraîner

les moteurs de traduction automatique. Comment caractériser les applications qui ont fait évoluer le poste de travail du traducteur, et expliquer que d'autres soient restées plus marginales ?

- Il permet d'interroger les tout derniers développements de l'automatisation et les nouveaux modèles utilisés (réseaux neuromimétiques). En quoi ces modèles éclairent-ils notre compréhension du fonctionnement du langage ?
- Il montre la nécessité d'une réflexion située, qui tienne compte de l'environnement au sens large (ergonomie organisationnelle, approches ethnographiques) mais aussi du corps et de l'ensemble des compétences et processus à l'œuvre (ergonomie physique et cognitive).

Une discussion synthétique de ces questions est proposée dans la dernière partie, qui porte sur des travaux et projets récents ou en cours.

Première partie. La linguistique appliquée, d'un champ à l'autre

L'objet de cette première partie est de situer mon travail en clarifiant certains préalables méthodologiques. Les éléments de méthode qui sont mis en perspective ici ne sont que rarement abordés dans mes publications mais ils ont été pensés en amont, et parfois retravaillés lors de séminaires ou conférences invitées. Ils reposent avant tout sur une lecture critique des évolutions observées dans les champs où se sont situées mes questions de recherche, c'est-à-dire tour à tour la psycholinguistique, la linguistique de corpus et la traductologie.

Les considérations épistémologiques qui structurent cette partie sont les suivantes : quels sont les modèles utilisés dans chacun des trois champs, comment se sont-ils constitués et quels sont les critères de validité utilisés ? Le fil conducteur correspond à la recherche d'une cohérence, qui permette de définir la linguistique appliquée autrement que par le type de problèmes qu'elle aborde, comme le suggère la définition qui en est donnée sur le site de son association internationale (AILA) :

Applied Linguistics is an interdisciplinary field of research and practice dealing with practical problems of language and communication that can be identified, analysed or solved by applying available theories, methods and results of Linguistics or by developing new theoretical and methodological frameworks in Linguistics to work on these problems. (<http://www.aila.info/en/about.html>, site consulté le 1.08.2017)

Si je fais référence à l'appellation « linguistique appliquée », qui est assez peu revendiquée en France (Carton et al. 2015 : 20), c'est qu'elle délimite un champ dans lequel s'inscrivent l'ensemble de mes travaux. Ce champ n'est pas uniforme : l'unité de la linguistique appliquée est peut-être avant tout le produit de sa « disciplinarisation » relativement récente, qui est à l'évidence une construction discursive (Stegu, 2011: 134) et institutionnelle (Linn et al. 2011: 7). La linguistique appliquée se distingue en ceci, par exemple, de la linguistique de terrain, également récente (Auroux, 1998: 89) et qui semble marquée par une plus grande unité théorique et méthodologique (Blanchet 2012: 28), sans que son statut de discipline soit marqué par l'existence d'associations la représentant, Je pars cependant du principe que les rapprochements qu'opère la linguistique appliquée, à commencer par celui de la traduction automatique et de l'enseignement des langues étrangères, ne sont pas accidentels. La cohérence que je cherche à retrouver au sein de la linguistique appliquée correspond en effet à celle de mon parcours, dont je veux croire qu'il n'est pas uniquement le produit d'accidents, et qu'il reste assez peu tributaire des frontières disciplinaires.

1. Méthodes et outils

Faire de la méthode et des outils un préalable, c'est déjà s'orienter vers une linguistique résolument appliquée. La tradition chomskyenne faisant souvent figure d'archétype de la linguistique théorique, ou « en chaise longue » (Fillmore, 1992, cité et traduit par Poibeau, 2014: 2) se distingue ainsi d'une linguistique outillée, qui désigne en première approximation la ou les linguistiques de corpus. Les méthodes et outils présentés ici ne se limitent cependant pas aux corpus. Ils retracent un cheminement soucieux d'adapter les choix méthodologiques à la diversité des objets étudiés. Les différents points abordés dans la discussion méthodologique qui suit permettent de rendre compte de ces choix et de leur articulation. Chacun de ces points est en effet transversal, et chacun apporte un éclairage particulier sur le chemin qui conduit du suivi longitudinal pour étudier l'acquisition d'une langue maternelle, au recueil de données expérimentales sur l'acquisition d'une langue seconde, et finalement à l'adoption de méthodes d'inspiration ethnographique afin d'analyser les perceptions de la traduction automatique.

Cette partie met aussi et surtout en avant l'un de mes premiers engagements, sur lequel je ne suis jamais revenue : celui de prendre très au sérieux le caractère empirique du langage. Mon goût pour l'observation et l'analyse de productions langagières situées l'explique en grande partie, et rentre en résonance avec le constat, issu de l'histoire des théories linguistiques, de ce qu'« il y a des informations sur le langage que nous ne pouvons recueillir que par le truchement de données externes » (Auroux, 1998 : 90). Même si j'ai fait ce travail il y a une dizaine d'années, je ne saurais à présent aborder autrement le langage de l'enfant qu'en observant, sur des enregistrements qui m'ont été transmis mais aussi, caméra à l'épaule, des interactions et situations variées qui toutes correspondaient au quotidien des dyades filmées. L'observation est pour moi primordiale, parce qu'elle façonne le questionnement ou fait surgir de nouvelles interrogations, mais aussi parce que son analyse convoque des modèles qu'il nous faut parfois aussi révoquer.

Je n'étudierai ici que quelques caractéristiques, choisies et analysées pour les problèmes qu'elles posent, des modèles avec lesquels je travaille. Ce ne sont en effet pas les modèles en tant que tels qui m'intéressent, mais les problèmes rencontrés et les stratégies utilisées pour s'y attaquer. A l'arrière-plan de la diversité d'approches évoquées ici se trouve cependant une diversité de modèles, et peut-être faut-il apporter au préalable quelques éclaircissements sur ce que nous entendons par là.

Qu'est-ce qu'un modèle ? Issue des mathématiques et de la logique formelle, la notion de modèle est assez peu utilisée en linguistique générale, en particulier dans la tradition énonciativiste française (Rotgé, 2002: 1) et plus volontiers rattachée à la tradition générativiste

ou aux formalismes utilisés en traitement automatique des langues. La psycholinguistique, en revanche, développe de nombreux modèles qui se distinguent surtout par leur valeur heuristique, et c'est bien ce que nous voulons mettre en avant ici. Au sens large, et dans la perspective épistémologique qui est la nôtre, on peut donc dire qu'un modèle est « un cadre représentatif, idéalisé et ouvert, reconnu approximatif et schématique mais jugé fécond par rapport à un but donné (prévoir, agir sur la nature, la connaître mieux, etc.) » (Soler, 2000: 45). En linguistique appliquée, le modèle n'est d'ailleurs le plus souvent pas prescriptif, mais plutôt descriptif. Il s'agit alors d'une description systématique (ou systématisée), d'un ensemble d'hypothèses dont la fonction est de représenter au mieux ses objets (voir par ex. Chesterman, 2015 : 9).

Les modèles utilisés pour expliquer l'acquisition du langage peuvent se présenter sous la forme de règles ou d'exemplaires, en fonction du cadre théorique adopté : nos questionnements sont issus d'une approche largement basée sur l'usage, et donc plutôt liés à des modèles exemplaristes. Ces modèles ne sont cependant pas convoqués seuls ni considérés pour eux-mêmes : nous interrogeons par exemple aussi le statut des règles dans l'acquisition des langues secondes (voir ci-dessous : deuxième partie, 3.2). Au-delà des modèles figés, ce sont plutôt les modélisations, comme processus toujours inachevés et perfectibles, qui nous intéressent. Toute théorie linguistique suppose par exemple une modélisation de la chaîne parlée, susceptible de permettre, entre autres, le découpage en unités comme les phonèmes, mais aussi les mots, les syntagmes, et les énoncés au sein desquels ils sont agencés. Ces énoncés, qui sont des « unité[s] de discours attestée[s], ou attestable[s], et située[s] dans un espace énonciatif » (Chuquet et Paillard 1992: 242) sont souvent ramenés lors de leur transcription à l'unité de la phrase (« suite bien formée selon des règles syntaxiques » *ibid.*), et il me semble que la question de savoir pourquoi l'on adopte tel ou tel découpage ne se pose vraiment que lorsque l'on travaille à la transcription de données d'oral enregistrées. Le nécessaire passage par l'écrit, qui conditionne toute analyse (voir par ex. les travaux de Claire Blanche-Benveniste), suppose que l'on prenne un certain nombre de décisions qui ne sont jamais neutres : la transcription est le reflet des modèles qui la rendent possible, et ces modèles passent difficilement inaperçus dans la transcription du langage de l'enfant Ochs (1979).

Dans ce qui suit, je prolonge jusque dans mes travaux actuels, qui portent surtout sur des données textuelles, la réflexion entamée en 2008 à propos des enregistrements filmés et transcrits utilisés pour l'étude de l'acquisition précoce du langage [Rossi & Morgenstern 2008]. Cette première partie méthodologique permet d'aborder un à un les problèmes que j'ai rencontrés en travaillant sur des données langagières depuis une dizaine d'années, et de motiver les réponses que mes travaux ont apportées ou les solutions adoptées.

1.1 Corpus oraux, corpus écrits

Construire des corpus en acquisition du langage, c'est en quelque sorte faire de la linguistique de terrain : le laboratoire dans lequel j'ai fait ma thèse (Dynamique du langage, UMR 5596) rassemblait d'ailleurs une équipe travaillant sur la description des langues en danger, et une autre, à laquelle j'appartenais, et qui s'intéressait au développement du langage chez l'enfant. J'ai appris au cours des séminaires communs comme à travers certains échanges amicaux que nos pratiques n'étaient pas absolument distinctes. C'est en effet à la manière du linguiste de terrain que l'on est amené à s'imprégner du quotidien des familles filmées avant de transcrire et d'analyser les données langagières recueillies. Si j'ai voulu recueillir des données et prendre en charge au moins une partie des transcriptions, c'est parce que je suis convaincue que les meilleures questions de recherche en sont issues. Il existe certes d'autres façons de travailler, notamment en réutilisant des données créées par d'autres chercheurs, ce que j'ai également fait. C'est une pratique courante en raison de la longue tradition de partage des données sur l'acquisition du langage : la base de données sur l'acquisition du langage (Child Language Data Exchange System : CHILDES) créée dans les années 1980 par Brian MacWhinney et Catherine Snow (MacWhinney, 2000), qui a rapidement pris de l'ampleur, rassemble à présent des centaines de suivis longitudinaux¹ documentant l'acquisition monolingue et bilingue, précoce et tardive, ainsi que des données produites dans un cadre expérimental, dans plus de 30 langues. Depuis 2001, ces données ont été progressivement intégrées à TalkBank, qui constitue la plus grande base de données d'oral en accès libre. Pour faciliter l'échange et garantir la fiabilité des données, un format de transcription unique (CHAT) et un ensemble de programmes d'analyse (CLAN) ont été mis en place. Ce partage des données permet de multiplier les analyses des données disponibles et d'enrichir notre compréhension de ce que fait chaque enfant, ainsi que de faire des analyses à plus grande échelle, mais il signifie aussi que le recueil et la transcription des données ne sont plus un passage obligé pour le chercheur. J'ai ainsi, dans le cadre de ma thèse, dû me familiariser avec des données que je n'avais pas recueillies : le corpus de Lyon (Demuth et Tremblay, 2008) qui avait été constitué par des chercheurs de l'équipe à laquelle j'appartenais, et celui de Providence (Demuth et al., 2006), constitué aux Etats-Unis, en lien avec l'équipe de Lyon, en sorte que l'adoption de méthodes de recueil communes garantissait une bonne comparabilité des enregistrements. J'ai cependant consacré le plus grand soin au choix des enfants dont j'allais analyser les productions, et travaillé au moins autant à partir des enregistrements vidéo, qui

¹ Il s'agit d'enregistrements audio ou vidéo réguliers (le plus souvent à raison d'une heure tous les 15 jours) réalisés au domicile de l'enfant, en interaction avec ses parents, pendant une durée définie dans le cadre d'une étude donnée. Nous utiliserons également pour parler de ce type de recueil les adjectifs « naturaliste » et « écologique ».

constituent le point de départ des premières analyses (basées sur une observation attentive, sans catégories préétablies), que de l'analyse en partie automatisée du corpus de transcriptions. Cette méthode mixte, plus gourmande en temps, permet d'aboutir à des analyses plus fiables : « Un bon travail implique d'avoir visionné au moins une fois (et même plusieurs fois) la vidéo avant et en même temps que l'analyse des données transcrites » (Morgenstern et Parisse, 2007 : 71).

Ainsi, la réflexion méthodologique qui a accompagné mon travail de thèse était directement issue d'un questionnement de jeune chercheuse aux prises avec des données complexes. A la difficulté, souvent décrite, d'aborder des corpus de langue parlée en interaction spontanée, objets complexes se rapportant à des phénomènes dont ils ne permettent pas toujours de saisir la complexité (Mondada, 2000: §5 ; Traverso, 2011: 315), s'ajoutait en effet la dimension développementale. Pour en tenir compte, la transcription doit s'adapter aux productions langagières de l'enfant, constituées de catégories « émergentes » (Clark, 2001) et d'usages qui ne correspondent pas à ceux de l'adulte, qu'une transcription normalisée ferait disparaître. Pour tenir compte de ce problème, nous avons proposé la mise en place d'un codage explicite, dans les transcriptions, des syllabes « fillers », ces productions vocaliques du jeune enfant, qui peuvent remplacer un déterminant, un auxiliaire, un clitique ou une préposition [Rossi et Morgenstern 2008]. Pour garder la trace des évolutions développementales, nous avons aussi utilisé un codage des mots dont la transcription n'était pas évidente, parce qu'ils ne ressemblaient pas entièrement à la cible adulte, en indiquant le degré de certitude de la personne qui l'avait transcrit.

Une grande partie de cette réflexion a été menée dans le cadre de deux projets ANR successifs auxquels j'ai participé, tous deux portés par Aliyah Morgenstern : le projet Léonard (<http://anr-leonard.ens-lsh.fr/>) puis le projet CoLaJe qui a permis de suivre les mêmes enfants jusqu'à l'âge de 7 ans (<http://colaje.scicog.fr/index.php/le-projet>). Mon intérêt pour les questions méthodologiques en même temps que les connaissances que j'avais développées en réalisant des transcriptions au format CHAT m'ont permis de réaliser la première version d'un guide qui a été utilisé par l'ensemble de l'équipe pendant toute la durée du premier projet et enrichi à mesure que nous avançons. Ce guide a ensuite été mis à disposition de la communauté via CHILDES et il peut être téléchargé depuis la page de présentation suivante : <http://childes.talkbank.org/intro/>. En plus des informations nécessaires à la prise en main et des fonctions de base, le guide présente certaines décisions méthodologiques propres au projet, comme le codage sur une ligne dépendante (%int) d'une ou plusieurs interprétations de ce que dit l'enfant, en faisant figurer à sa suite le degré de certitude.

Il faut ajouter au travail réalisé pour ces deux projets financés par l'ANR la réflexion plus large menée au sein du laboratoire junior sur les « Approches linguistiques de corpus d'oral

spontané » (ALCOS) que j'ai porté à partir de 2008. Il rassemblait une équipe de jeunes chercheurs et de chercheurs confirmés, financée pendant deux ans à l'ENS de Lyon. L'intention était de mieux appréhender la diversité d'approches et de mettre au jour des points de rencontre. Le travail a été rythmé par une série de conférences que nous animions, en tout début d'après-midi (les « conf'Expresso »). Il n'a pas donné lieu à une publication spécifique, mais il constitue le point de départ d'une réflexion sur les caractéristiques des données orales, allant du travail de terrain (conférence d'Alexis Michaud sur l'élicitation, le recueil et l'analyse prosodique de données dans une langue sino-tibétaine, le naxi) à l'analyse multimodale de données d'acquisition du langage (conférence de Mats Andren sur les gestes liés à la manipulation d'objets). Parmi les problèmes repérés alors, certains restent à explorer : je n'ai, par exemple, pas eu l'occasion de travailler sur le rôle de la multimodalité dans les premières formes d'expression du mouvement acquises par l'enfant comme je l'aurais souhaité : mes travaux sur la gestualité se sont pour l'instant développés uniquement dans le cadre d'un groupe de réflexion alliant psychiatres, psychanalystes et linguistes ([Rossi 2012] et [Rossi & Nassikas 2019]). Ceci s'explique en partie au moins par une réorientation de mes travaux de recherche, en accord avec le profil du poste de Maître de conférences sur lequel j'ai été recruté.

Mes travaux récents m'ont ainsi fait passer de corpus oraux d'interactions spontanées mère-enfant, à des corpus de textes qui concernent principalement la langue de spécialité. Le changement d'outils en même temps que la nature différente des données appelait nécessairement une réflexion méthodologique, qui a été menée en lien avec un projet de constitution de corpus. Il s'agit d'un corpus sur la justice environnementale, climatique et énergétique, constitué de documents produits par différents types d'acteurs (Nations-Unies, organisations non-Gouvernementales environnementales, entreprises du secteur énergétique, auteurs d'articles scientifiques, et journalistes). Nous considérons le discours sur l'environnement comme un « discours spécialisé émergent, dans le sens où l'on voit des normes de communication et une terminologie stabilisée se mettre en place progressivement dans le domaine. » (Biros, 2014 : 45). Pour procéder à la collecte de données dans un domaine de spécialité encore peu décrit, nous avons choisi une thématique permettant de mieux cerner les développements contemporains de ce discours « traversé par des lignes idéologiques multiples » (*ibid.*). La justice climatique fait référence aux conséquences humaines du changement climatique et des mesures mises en place pour y faire face. Elle est liée à un questionnement sur le caractère équitable de la distribution des émissions, des impacts et des efforts, entre différents pays d'une part, et entre différentes communautés au sein d'une même nation d'autre part. La justice énergétique, elle, désigne la nécessité d'assurer l'accès à l'énergie pour tous, en tenant compte des conséquences à long terme des politiques énergétiques et en valorisant le recours aux énergies renouvelables. La notion est d'abord et surtout utilisée dans un contexte militant (Heffron & MacCauley 2017: 2),

comme déjà celle de justice environnementale, lorsqu'elle est apparue aux Etats-Unis, en référence au préjudice subi par les minorités raciales du fait de certaines dégradations environnementales (Bullard 1990), mais elle s'est ensuite diffusée dans de nombreux pays européens, ainsi qu'en Afrique du Sud, ou encore en Amérique latine (voir par ex. Beretta, 2012). Le corpus a été construit pour observer la diffusion de telles questions dans des communautés de discours que nous avons identifiées, c'est-à-dire au sein de groupes sociaux définis par « un ensemble d'objectifs publics communs faisant l'unanimité, ou presque [...] des mécanismes de communication entre ses membres » et le recours à « un ou plusieurs genres discursifs, ainsi qu'à un lexique spécifique, pour mettre en œuvre une communication qui réponde à [leurs] objectifs » (Swales, 1990: 24 et sq, notre traduction²). Une réflexion sur l'importance de l'appartenance de l'énonciateur à une communauté de discours a donc présidé au choix des sources, qui ont été triées en fonction de l'organisation concernée. Les organisations dont les rapports figurent dans le corpus ont été recensées et leur nom lié au fichier de métadonnées qui permettent un tri et la création de sous-corpus, en utilisant par exemple le logiciel libre TXM (Heiden et al., 2010). Une partie de ce corpus est également consultable dans le concordancier en ligne ScienQuest³ (Falaise et al, 2011), qui permet d'utiliser toute la richesse de ces métadonnées (en particulier la date de publication, le type de communauté de discours, la langue et le type de texte) pour interroger les données.

Cette attention prêtée au sujet de l'énonciation est directement issue de ma formation à la théorie culiolienne des opérations énonciatives et prédicatives, qui considère toute assertion, affirmative ou négative, comme résultant d'une opération de prise en charge par un énonciateur, et s'attache à définir l'ensemble des repères énonciatifs nécessaires à l'analyse. Elle témoigne également d'une influence de mes premiers travaux, portant sur des données d'oral (qui contrairement aux données textuelles ne font jamais disparaître complètement l'énonciateur) sur les analyses ultérieures. C'est ainsi qu'un travail non encore publié car commencé récemment sur ce corpus [Rossi et al. 2018c-oral], m'a permis de montrer qu'il est important de définir le niveau de granularité souhaité lors de la constitution et de l'analyse de corpus écrits. Ce travail part du constat suivant : le plus souvent, l'individu représente en effet un niveau de granularité qui n'est pas conservé dans les grands corpus de référence, qu'il soit locuteur dans le cas de corpus oraux ou scripteur, auteur d'un texte intégré au corpus (voir par ex. Gries, 2015 : 99). A la différence des

² Discourse communities are defined by “a broadly agreed set of common public goals [...] mechanisms of intercommunication among [their] members” and resort to “one or more genres in the communicative furtherance of [their] aims” as well as “some specific lexis”

³ L'accès est restreint car les droits n'ont pas encore tous été obtenus, il faut se connecter à l'adresse suivante <https://corpora.aiakide.net/scientext19/?do=SQ.setView&view=login> (login : cop21 et mot de passe : Kefei1ie)

corpus oraux, les corpus écrits sur lesquels nous travaillons se composent de textes le plus souvent produits par plusieurs auteurs : en distinguant des communautés de discours, nous avons cerné un niveau de granularité collectif, mais unifié du point de vue des pratiques discursives et des objectifs qui ont présidé à l'élaboration de ces écrits. La question posée dans ce travail en cours est celle de savoir si ces communautés constituent un niveau de granularité pertinent pour l'analyse du discours spécialisé, que l'on définit généralement plutôt par sa capacité à représenter un genre ou un domaine (Kübler, 2014). Notre analyse de termes se rapportant au traitement des questions de justice, d'équité et de droits de l'homme au sein de trois communautés de discours distinctes (ONG, ONU et entreprises du secteur énergétique) a permis de montrer que les indices de spécificité (calculés dans le logiciel TXM) étaient fortement liés à deux indices de répartition : la répartition par organisation (nombre d'organisations utilisant le terme / nombre d'organisations totales dans le sous-corpus), mais aussi la répartition au sein des documents produits par chaque communauté. Le graphique ci-dessous⁴ montre bien que les différences manifestes d'une communauté à l'autre, rendues visibles par le calcul des spécificités, correspondent presque toujours à une répartition différente des termes dans chacun des sous-corpus. Ces premiers résultats montrent l'importance de constituer des bases de données aux unités « atomisables » (Habert, 2000), parfois appelées aussi « corpus distinguables » (Reffay et al. 2008), et ce jusqu'au plus fin niveau de granularité.

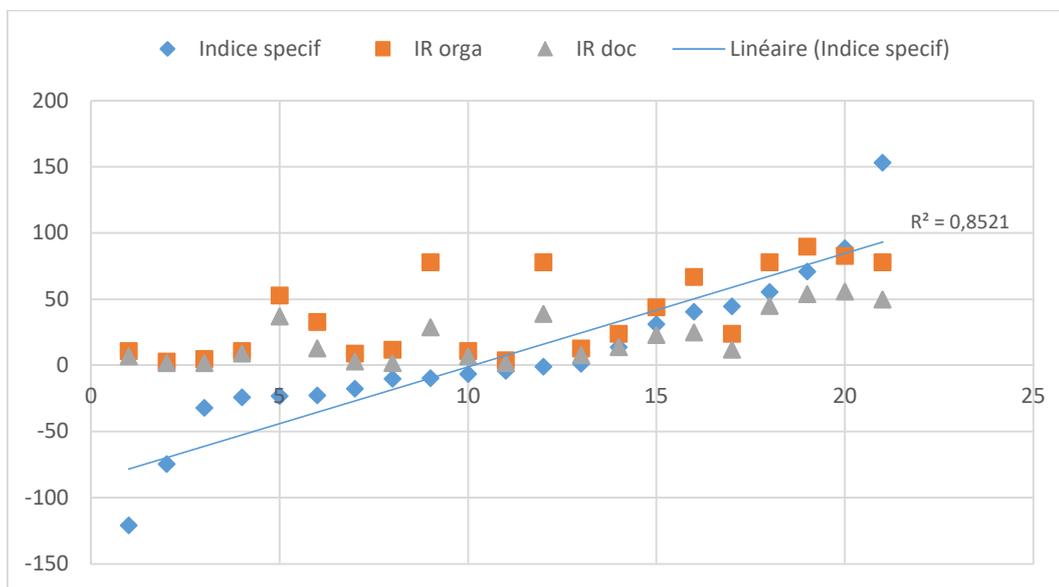


Figure 1. Caractérisation de la variation terminologique par communauté de discours

⁴ On considère ici 7 termes pour lesquels les indices de spécificités diffèrent fortement au sein de trois communautés : adaptation, investment, poor, vulnerability, migration, efficiency, communities.

Avant d'aborder un ensemble de questions liées à l'étude de la variation, et pour conclure sur le transfert de connaissances qui se produit lorsque l'on passe de corpus oraux à des corpus écrits, j'évoquerai brièvement la question du partage des données. C'est en participant à une école thématique du CNRS dès la première année de ma thèse⁵ que j'ai été sensibilisée à ces questions, au sein d'une communauté de chercheurs en acquisition du langage, où l'habitude du partage s'est rapidement installée, comme en témoigne l'étendue de la base de données TalkBank. En pratiquant ensuite ce partage, j'en ai éprouvé les limites (comme expliqué ci-dessus) mais j'ai aussi remarqué qu'il constituait indéniablement un accélérateur de découvertes, et permettait de vérifier ou de répliquer des résultats. C'est cependant en travaillant dans d'autres champs que j'ai découvert que le partage des données de recherche n'allait pas de soi : il se met en place plus lentement dans la communauté des chercheurs travaillant sur l'acquisition d'une langue seconde (L2), et il n'est pas encore vraiment pratiqué en traductologie (il n'existe pas à ma connaissance de base de données issues de l'étude des processus traductionnels). En acquisition L2, la constitution depuis 2007 d'un complément à TalkBank, intitulé SLAWeb, ainsi que la création de la base de données IRIS en 2011, puis de l'infrastructure CLARIN en 2012 sont des tentatives remarquables mais non encore pleinement abouties. Les corpus écrits, quant à eux, se heurtent en France à la question épineuse des droits d'auteur, qui trop souvent découragent tout partage. Il est toutefois non seulement possible mais également souhaitable que les corpus soient partagés entre chercheurs, ne serait-ce que pour favoriser le dialogue et multiplier les points de vue sur les données. C'est à cette fin que la mise à disposition de données étiquetées dans des concordanciers en ligne, sans accès direct au texte, a été largement adoptée depuis plus d'une décennie (voir par ex. Baroni et Kilgarriff 2006 : 89), mais même cet accès aux seuls extraits du texte affichés dans un concordancier pose problème dans le cadre législatif français. Pour les données sur lesquelles nous avons travaillé récemment, l'obtention des droits est en cours et a considérablement retardé le partage sur l'infrastructure ORTOLANG (<https://www.ortolang.fr/>). Nous disposons toutefois déjà de plusieurs accords et espérons que l'ensemble des organisations onusiennes et ONG dont nous avons rassemblé des rapports permettront le partage sous licence Creative Commons.

1.2 Variation, variabilité

Mes publications sur l'acquisition du langage s'inscrivent dans la lignée des théories basées sur l'usage (UB), lesquelles ont été adoptées assez tôt dans un ensemble de travaux, en partie menés en réaction à la grammaire générative et transformationnelle de Chomsky. La

⁵ « Acquisition du langage & corpus en linguistique et en psychologie : une mutualisation des ressources, outils et données » Moissac 14-19 mai 2006, cf. <http://acquisition.scicog.fr/> (dernière consultation le 11 juillet 2018)

construction de bases de données partagées en a été l'un des fondements, contre la conception du linguiste « en chaise longue » évoqué plus haut. Les théories UB ne s'opposent pas seulement à l'innéisme chomskyen, mais elles s'inscrivent aussi en faux contre les « principales hypothèses de travail des structuralismes européens et américains » (Chevrot, 2012 : 50) et proposent une vision émergentiste des compétences langagières. Ainsi, l'idée que toute connaissance langagière soit construite en contexte, en s'appuyant sur des facultés diverses et non spécialisées, et se structure en réseaux dynamiques et changeants est parfaitement incompatible avec une vision modulaire de l'esprit, comme avec une conception classique de la catégorisation. C'est le retour à l'analyse du discours, et donc aussi à la prise en compte de données authentiques, plutôt que d'énoncés repérés ou reconstruits par l'intuition du linguiste, qui produit une conception au sein de laquelle les unités, tout comme les règles permettant de rendre compte de leur combinaison, sont émergentes et non définies a priori. Cette conception est essentiellement empiriste, et propose qu'il y ait « un ensemble de connaissances tirées de l'expérience qu'ont les locuteurs des formes linguistiques et de leurs contextes d'usage, qui permettent l'émergence progressive d'une sensibilité et d'une adaptabilité aux informations générales et spécifiques dont ces formes sont porteuses » (Bybee & McClelland, 2005 : 405, notre traduction⁶)

Le statut de la variation dépend fortement de ces deux conceptions : les modèles UB permettent de donner plus de poids à l'analyse de la variation, qui ne correspond plus à une multitude de formes incomplètes ou imparfaites, mais à des usages motivés et toujours susceptibles de façonner les connaissances langagières. Observer la variation suppose aussi que l'on dispose de données riches et variées, et non d'exemples choisis. C'est ainsi que dès 1973, Katherine Nelson réussit à mettre en évidence, en utilisant des suivis longitudinaux réalisés au domicile de 18 enfants américains, l'existence d'une grande variabilité dans les premiers mots produits par les enfants (Nelson, 1973: 8). Depuis, les grandes bases de données rassemblées grâce aux inventaires du développement communicatif⁷ (voir Kern & Gayraud, 2010 pour l'adaptation française) ont permis de réaliser des études à très grande échelle de cette variabilité. A l'université de Stanford, un outil permettant d'interroger librement les données issues de plus de 80 000 inventaires administrés dans 29 langues différentes et de visualiser les résultats a été développé et mis en ligne récemment (Frank et al. 2017)⁸. Schneider et al (2015) l'ont utilisé pour

⁶ La citation originale est la suivante : « Instead there is knowledge derived from experience with language forms and their contexts of use that allows sensitivity to general and specific information to arise in a graded and malleable way from experience »

⁷ Il s'agit de questionnaires standardisés et adaptés dans de nombreuses langues, permettant de recueillir une évaluation, faite par les parents, des premiers développements langagiers de leur enfant.

⁸ C'est la base Wordbank, qui est accessible à l'adresse suivante : <http://wordbank.stanford.edu/> (dernière consultation le 11 juillet 2018).

montrer que la variation observable au sein des premiers mots produits par les enfants était également plus structurée qu'il n'y paraissait : l'analyse fait ressortir une préférence pour les mots fréquents dans l'input et de faible complexité phonologique, mais aussi (et ce résultat est moins attendu) pour un type de référents concrets et très ancrés dans le quotidien de l'enfant, à savoir les noms d'animaux, de jeux et de routines. On voit bien ici comment l'observation structurée et contrôlée permet de comprendre la variation, c'est-à-dire de retrouver des régularités qui n'apparaissent pas nécessairement dans d'autres conditions.

Or si mon questionnement sur l'acquisition du langage a immédiatement porté sur la variation, mes premiers travaux ne m'ont probablement pas permis d'éviter ce que Labov nomme le paradoxe de l'observateur (Labov, 1972: 209), et il m'est apparu assez tôt que le type d'observations sur lequel je travaillais ne permettrait pas de rendre vraiment compte de la variation. D'abord parce que les enregistrements, réalisés au domicile, se distinguaient dans certains cas au moins de ce qui se produisait naturellement et spontanément dans les interactions avec le jeune enfant. J'ai en effet constaté en le pratiquant que le recueil de données créait une situation particulière, susceptible de modifier les observables : la petite fille bilingue que j'ai filmée pendant ma thèse se montrait souvent réticente ou mécontente d'être filmée, tant et si bien que les données recueillies comportaient tout un éventail de formes négatives que nous avons décidé d'étudier [Rossi & Shimanek 2007-oral]. La seconde raison pour laquelle mes premiers travaux sur la variation n'étaient pas entièrement satisfaisants concerne la quantité de données étudiées. Le point de départ de mon travail de thèse est un point de variation inter-langue souvent décrit : l'expression du mouvement, qui se structure différemment en français (langue à cadre verbal dans la typologie talmyenne) et en anglais (langue à cadre satellitaire). Cependant, en m'intéressant aux premières productions langagières, et en me focalisant sur un petit nombre de suivis longitudinaux, c'est en définitive surtout la variation individuelle que j'ai pu observer. Les données que j'ai étudiées suggèrent par exemple que les différences dans l'expression de la manière de mouvement ne trouvent pas leur source dans des facteurs culturels ou sociaux, mais sont plutôt le reflet d'idiosyncrasies, que l'on peut au mieux rapprocher de certaines pratiques et habitudes des dyades observées. C'est probablement le type de variation qui a été le plus souvent observé et analysé à partir de telles données, comme le faisaient déjà remarquer Bloom et ses collaborateurs en 1975 :

L'origine de la variation linguistique que l'on trouve décrite dans la littérature (et qui concerne des enfants de classe moyenne, dont les parents ont généralement fait des études supérieures) n'est ni culturelle ni sociale. Bien qu'un conditionnement environnemental puisse se produire, dans la mesure où la variation est liée à des différences de style interactionnel des parents (Nelson 1973), il est plus probable que la variation des productions langagières de l'enfant dépende de son

développement cognitif individuel et de l'interaction de celui-ci avec certaines caractéristiques du code linguistique. (Bloom et al. 1975 : 4, notre traduction⁹)

Il est indéniable que les parents ont tous des façons différentes d'interagir avec leur enfant, mais la variation qui est liée au contexte interactionnel est peut-être également fonction de différences culturelles. J'avais été sensibilisée à la possibilité de telles différences par les discussions et choix méthodologiques minutieux et prudents réalisés dans le cadre du projet Léonard [P1]¹⁰, où il est arrivé que deux personnes filment le même enfant pour multiplier les points de vue. J'ai ensuite commencé à repérer certaines différences dans mon travail de thèse, parce que le recueil des données, qui avait été fait aux Etats-Unis comme en France en suivant le même protocole, présentait des différences situationnelles remarquables. Je n'ai cependant pas quantifié ces différences, qui n'étaient pas l'objet de ma thèse, mais elles constituaient un point de variation sur lequel j'ai voulu travailler ensuite. C'est dans le cadre du projet PREMS [P3] que j'ai engagé la réflexion, avec une jeune doctorante, afin de mettre en évidence les contrastes liés à la présence de l'observateur dans de nouvelles séries d'enregistrements, également réalisés en situation « naturaliste » (c'est-à-dire au domicile sans intervention de l'observateur). Notre étude des activités filmées [Rossi et al. 2019a], qui prend appui sur les suivis longitudinaux de deux enfants francophones (corpus PREMS), de deux anglophones (corpus de Providence, qui correspond aux données utilisées pour ma thèse) et de deux arabophones (corpus PREMS), a notamment permis de montrer une répartition différente des activités lorsque l'observateur était présent, et la quasi disparition des activités de maintenance (activités liées à l'hygiène et à l'alimentation) lorsque l'observateur ne pouvait être présent et que la caméra était laissée sur un pied. Dans ce travail, où nous lions la variation linguistique observée à des caractéristiques de la situation d'interlocution, les résultats sont attestés quelle que soit la langue ou la culture. Une nouvelle fois, les données analysées n'étaient pas en quantité suffisante pour nous permettre de mettre en évidence des différences linguistiques et culturelles, mais la méthodologie proposée pourrait être étendue à des échantillons plus larges afin d'y parvenir.

La socialisation de l'enfant dans la dyade est un microcosme intéressant à plus d'un titre, mais les données recueillies sont également le plus souvent socialement homogènes, et elles ne

⁹ « The source of linguistic variation in the child speech data that have been described so far in the literature (from children of middle-class and generally college-educated parents) is neither cultural nor social. Although it may be environmentally conditioned to the extent that it reflects differences in parent interaction styles (Nelson 1973), it is more likely that variation in child speech is a function of individual cognitive development in interaction with different aspects of the linguistic code. »

¹⁰ Les projets auxquels j'ai participé sont recensés et décrits en Annexe (3.5)

permettent pas d'aborder les productions linguistiques dans leur dimension collective, sociale. Mes travaux récents m'ont permis de commencer à aborder cette dimension collective, en suivant certaines des pistes ouvertes par la linguistique variationniste basée sur corpus (« corpus-based variationist linguistic » ou CVL) telle que la définit Benedikt Szmrecsanyi dans un article récent qui décrit et analyse « le chevauchement entre la linguistique de corpus et la sociolinguistique variationniste » (Szmrecsanyi 2017 : 685-686). Le premier point est l'existence d'un lien entre les approches variationnistes et l'étude des discours spécialisés. Les approches variationnistes se sont d'emblée appuyées sur des entretiens transcrits ou sur des données recueillies en situation, elles ont donc toujours travaillé avec des corpus spécialisés (*ibid.*). La ressemblance n'est certainement pas fortuite, et il faut remarquer que les travaux qui portent sur les langues et cultures de spécialité s'intéressent à la variation sous toutes ses formes. Certains auteurs appellent d'ailleurs de leurs vœux une convergence avec les approches variationnistes : c'est ce que fait Isabel Desmet en proposant dans un article de 2007 des « [é]léments pour une théorie variationniste de la terminologie et des langues de spécialité » :

Aujourd'hui, nous proposons une approche des langues de spécialité fondée pour l'essentiel sur les principes de la linguistique descriptive, intégrant la variation à plusieurs niveaux: variation des langues, des discours et des textes de spécialité ; variation dans le temps, dans l'espace, en fonction des interlocuteurs et des situations de communication ; variation des unités linguistiques de spécialité sur l'axe syntagmatique et sur l'axe paradigmatic ; variation sur le plan lexical, sur le plan phrastique, sur le plan textuel et sur le plan discursif. Une théorie variationniste pour la description des terminologies et des langues de spécialité nous semble être nécessaire pour toutes les langues naturelles. (Desmet, 2007: 3-4, in Dury, 2013: 3, note 1).

C'est afin de mieux cerner la variation diachronique (de la COP15 à la COP21) et la variation d'une communauté de discours à l'autre que nous¹¹ avons constitué le corpus COP21 [BD1]. Nous avons travaillé, en amont de la constitution du corpus, avec des collègues spécialistes d'histoire des idées et procédé ensemble à la lecture critique d'un premier ensemble de textes, afin de parvenir à identifier « les contextes sociaux, historiques, politiques et intertextuels » (Baker et al. 2008 : 273, notre traduction) pertinents. Les communautés de discours que nous avons choisies se distinguent des communautés de locuteurs (« speech communities ») de la sociolinguistique variationniste parce qu'elles sont construites autour de pratiques écrites et codifiées, le plus souvent en lien avec des objectifs professionnels. Dans le cadre du travail que nous avons mené sur ces données, nous avons mis en avant plusieurs options méthodologiques pour analyser la variation. Certaines sont assez peu quantitatives : nous avons par exemple étudié

¹¹ Le pronom renvoie à un groupe de travail sur ces questions, que je co-anime depuis deux ans avec ma collègue Camille Biros (cf. *infra*, Curriculum Vitae, 3.1).

les énoncés dans lesquels une définition de la notion recherchée était donnée, afin de découvrir d'autres options expressives, ou d'autres concepts qui pourraient être utilisés pour désigner une autre approche d'un même problème. Il apparaît ainsi que les organisations onusiennes ne reprennent jamais à leur compte le discours (essentiellement militant) sur la justice climatique, et lui préfèrent des expressions plus neutres comme « climate-informed development », ou « pro-poor climate policies » [Rossi et al. 2019b]. D'autres sont issues du traitement statistique de données de corpus, comme le calcul de mots clés qui utilise un test de significativité : les mots clés, ainsi que les concepts clés calculés à l'aide du logiciel WMatrix nous ont notamment permis de faire ressortir des caractéristiques de chaque sous-corpus et d'identifier de nouveaux ensembles de variantes (cf. [Rossi et al. 2018] et [Rossi et al. 2019b]).

Le second point concerne les apports d'une linguistique variationniste basée sur corpus à notre compréhension de la grammaire, ou en tout cas des contraintes qui permettent de rendre compte du choix d'une forme plutôt qu'une autre. De nombreux cas d'alternance ont été étudiés : je me suis intéressée à l'alternance du génitif en anglais, précisément parce que la multiplicité des contraintes, qui entrent parfois en conflit les unes avec les autres, en complique l'enseignement auprès d'étudiants généralement francophones et pour qui l'anglais est une langue seconde (L2). C'est le cas de la contrainte de persistance citée par Rosenbach (2014: 232) parmi celles qui ont été étudiées et dont l'influence a été démontrée. Cette contrainte veut que l'occurrence d'une des variantes de l'alternance dans un texte encourage le choix de cette même variante dans la suite du même texte (ou du même échange) : elle peut donc, et vraisemblablement en de nombreux endroits, contredire la contrainte liée au statut informationnel (préférence pour la construction en *of* pour une première mention et choix d'un génitif en 's ensuite). Deux questionnements sont issus de ce constat : le premier concerne les conditions de possibilité d'un enseignement basé sur des règles, lorsqu'une forte variabilité existe et semble ne pouvoir être décrite que de façon probabiliste, et le second concerne le traitement que font les locuteurs de l'anglais L2 de telles variations. J'ai donc travaillé d'une part à évaluer l'impact d'un enseignement basé sur des règles, en l'opposant à la seule présentation d'exemples choisis et rendus saillants par des techniques connues et robustes (mise en évidence par le soulignement ou l'usage de caractères gras, et présentation resserrée ou intensive, c'est-à-dire suppression ou reformulation des éléments du texte qui ne comprennent pas la construction en question¹²) : les résultats d'une enquête préliminaire présentée lors de la conférence « Thinking, Doing, Learning » à Munich en 2017 [Rossi, Chevrot, Sahakyan et Batailler 2017-oral] montrent que la compréhension d'un ensemble de règles complexes et en partie contradictoires dépend du niveau de langue des étudiants, et que

¹² Les termes anglais sont : “visual enhancement” et “input flooding”

les étudiants les plus faibles produisent davantage d'erreurs du fait de l'exposition aux règles. Ces résultats montrent qu'il est important de mieux cerner le traitement de la variation par des locuteurs non natifs de l'anglais : une étude à grande échelle est en cours, que nous présentons dans la deuxième partie de cette synthèse (voir ci-dessous, point 3.b de la Deuxième partie)

C'est aussi et enfin pour tenter de décrire et d'expliquer la variation que j'ai souhaité utiliser des méthodologies et des dispositifs variés, allant du recueil de données dans une situation que l'on appelle parfois aussi « écologique », au domicile, dans le cadre de ma thèse, à l'utilisation de dispositifs expérimentaux que je décrirai dans la seconde partie de ce travail, et jusqu'à l'utilisation de méthodes ethnographiques pour recueillir des données sur les usages et perceptions de la traduction automatique auprès des traducteurs de la Commission européenne [Rossi & Chevrot. 2019] et auprès de mes étudiants [Rossi 2017a]. Ces méthodes correspondent aussi à une variété d'objets, des interactions naturalistes à la production contrôlée et à la traduction, que je commence à envisager sous l'angle des processus complexes qu'elle met en l'œuvre. Au total, et même si c'est aussi mon parcours qui m'a amenée à faire varier méthodes et objets, ma volonté de croiser et de combiner plusieurs analyses et plusieurs regards sur les phénomènes correspond à une conception que j'ai exprimée dans une de mes premières publications [Rossi & Morgenstern 2008] : les observables ne sont jamais neutres, ce sont des productions situées, et en partie au moins construites ou reconstruites par le chercheur. Anne Trevisse (1996) l'explique si bien qu'elle semble décrire une évidence, alors qu'il n'est pas certain que la conception des données qu'elle énonce soit consensuelle : « On sait que les données sont forcément du construit : elles ne peuvent être observables que dans un va-et-vient avec une théorie de l'observation, et donc une métalangue d'observation qui les catégorise, en fait ressortir les aspects discriminants. » (Trevisse, 1996 : §23). Une théorie de l'observation située est peut-être d'abord une théorie du contexte : c'est en tout cas cette notion « protéiforme, flexible » (Kleiber, 2009: 17) que j'ai rencontrée à plusieurs reprises dans mes travaux et qu'il importe de définir ici car elle continue de m'occuper, ainsi que je l'expliquerai dans la troisième partie de ce travail (voir ci-dessous, Synthèse, point 3.c).

1.3 Contextes

L'ensemble des travaux présentés ici témoigne d'une attention particulière au contexte. Ils se situent dans la lignée des linguistiques contextualistes, qui analysent des productions langagières situées, dans leur environnement langagier. Appliquée à la linguistique textuelle, cette conception s'ancre dans le célèbre principe énoncé par Firth à la suite de Wittgenstein : « You shall know a word by the company it keeps » (Firth 1957 : 11, in Kleiber 2009: 24 ; voir aussi Poibeau 2014: 3), qui crée une première théorie des collocations. Elle intègre aussi l'ensemble des éléments que regroupe la notion firthienne de « contexte de situation », c'est-à-dire (et nous reprenons également cette liste à Kleiber *ibid.*) :

A. les traits pertinents des participants : individus, personnalités

(i) l'action verbale des participants,

(ii) l'action non verbale des participants ;

B. les objets pertinents ;

C. les effets de l'action verbale.

Ces éléments, qui ne correspondent pas à une simple description mais bien à une construction théorique ancrée dans l'expérience authentique des interactions verbales (Kenny 2001 : 9) permettent de situer mes travaux : mes publications sur l'acquisition précoce d'une langue maternelle se situent dans le cadre d'une linguistique interactionniste, qui est aussi et peut-être d'abord une linguistique de l'énonciation. Ainsi, je ne me suis pas intéressée aux seules formes linguistiques (qu'il s'agisse des verbes de mouvement ou de l'alternance *will/gonna*, par exemple) et à leur combinaison dans la phrase, mais à leur prise en charge par un ou plusieurs énonciateurs présentant des caractéristiques individuelles, sociales, culturelles. La notion de prise en charge, très utilisée par Culioli qui la définit cependant rarement (Coltier et al. 2009: 7), permet d'associer une relation prédicative au sujet qui l'énonce. Pour Culioli, « toute énonciation suppose une prise en charge de l'énoncé par un énonciateur » (1971: 4031, in Coltier et al. 2009 *ibid.*). Pour comprendre ce qui se joue dans les premières interactions avec l'enfant, il faut analyser le travail énonciatif à l'œuvre dans celles-ci, celui qui permet à l'enfant de « construi[re] sa place d'énonciateur » ([Rossi et al. 2013], cf. page 265), en mobilisant bien souvent les éléments listés en A, B et C. Pour ne donner qu'un exemple, dans l'interaction transcrite ci-dessous¹³ entre

¹³ Cet extrait peut être visionné en suivant le lien : <http://ct3.ortolang.fr/devlang/5-7-theophile-vroumvroum-1-10.mp4>. Je propose ici une nouvelle analyse d'un passage étudié dans une de mes publications [16] où l'analyse porte sur l'expression du mouvement.

Théophile (1 an et 10 mois) et sa mère, on ne peut comprendre l'usage des onomatopées sans tenir compte des caractéristiques de la dyade : la mère, qui est violoniste et dont Théophile est le premier enfant, est très attentive à l'ensemble de ses conduites, verbales et non verbales, et s'y adapte constamment¹⁴. On voit ici qu'elle construit des syntagmes nominaux en reprenant les onomatopées qu'il produit pour désigner sa petite voiture. Le non verbal permet à l'enfant de préciser sa demande, d'abord avec un geste de pointage, puis en agitant la main, toujours en direction de la voiture, pour dire son impatience. Parmi les éléments du contexte situationnel qui sont au cœur de cette séquence (B dans la liste ci-dessus) se trouve le pyjama de l'enfant, que la mère est en train de boutonner : il ne sera convoqué que plus tard, et sans être nommé, lorsque la mère demandera à Théophile de lui laisser le temps de terminer (« j'ai presque fini Théophile mais (il) y a plein de boutons »). Quant aux effets de l'action verbale (C, ci-dessus), ils sont immédiatement visibles ici puisque la mère se dépêche d'accrocher les boutons, et ils conditionnent aussi le déroulement du dialogue : on le voit lorsque la mère annonce à l'enfant, pour le faire patienter, qu'elle va répondre à sa demande. Enfin, le pointage de l'enfant et le jeu des pronoms personnels dans les énoncés que la mère adresse à l'enfant (« tu » mais aussi « on » qui renvoie bien ici aux deux personnes filmées) sont autant d'éléments déictiques, ancrés dans la situation d'énonciation.

THÉOPHILE a vroom vroom

pho [oe vʁvʁvʁ]

Théophile pointe vers son petit camion, hors caméra, main ouverte.

MÈRE tu veux ta vroom-vroom oui

THÉOPHILE papa papa !

pho [pupwa epupwoe:]

MÈRE on va aller voir papa !

Séquence 1. Théophile demande sa petite voiture

L'analyse de données langagières situées, comme celles que nous venons de citer, convoque nécessairement un ensemble de données contextuelles, dont certaines font référence à la situation d'énonciation. Le contexte de situation ne se situe donc jamais vraiment dans un hors-

¹⁴ C'est l'une des caractéristiques fortes du projet CoLaJe que d'avoir systématiquement prêté attention à ces éléments, qui sont trop souvent gommés des collectes. Les familles ont accepté que les vidéos ainsi que de brèves descriptions figurent sur le site du projet. Voici, à titre d'exemple, le début de la description qui est donnée pour le corpus en question : « Théophile est un petit garçon français né le 04 juillet 2005. Son père est cadre supérieur dans une entreprise et sa mère est violoniste. Il est bricoleur, curieux et constamment en quête de nouvelles expériences. Au moment des premiers enregistrements, Théophile était encore enfant unique. » (<http://colaje.scicog.fr/index.php/les-enfants/theophile>)

langue, il désigne plutôt les connaissances qui sont nécessaires au bon fonctionnement des interactions langagières, et que l'analyste va reconstruire. La distinction entre un contexte interne, langagier, et un contexte de situation, externe à ces données langagières, est donc surtout pertinente pour l'analyste, et je rejoins les conclusions de C. Kerbrat-Orecchioni (2012: 20):

Si donc la distinction entre données « internes » et « externes » n'est pas une vue de l'esprit (il y a bien des énoncés, et des choses autour), d'un point de vue descriptif la question pertinente se pose en ces termes : *quels sont les types de savoirs nécessaires à l'interprétation du segment discursif soumis à l'analyse* (savoirs linguistiques, pragmatolinguistiques et socio-pragmatiques, génériques, cotextuels et intertextuels, situationnels, encyclopédiques...) ? S'agit-il de savoirs préalables ou construits au cours de l'événement communicatif lui-même, et sont-ils ou non partagés par les différents participants à l'interaction ?

Il me semble cependant que répondre à de telles questions revient en fait à s'inscrire assez fortement dans un cadre théorique. Il importe donc pour moi de tâcher de définir au mieux celui-ci. Le détour épistémologique que je propose ici est une analyse à rebours : je commence par analyser les présupposés de mes travaux récents sur des corpus écrits, avant de revenir sur l'analyse de données d'interaction en acquisition du langage.

Le choix de travailler sur des corpus écrits en distinguant plusieurs communautés de discours s'est imposé à l'issue d'une démarche collective et pluridisciplinaire, mais si j'ai cherché à rendre compte de la validité d'un tel choix dans mes derniers travaux (cf. supra) c'est parce que ce choix n'a rien d'évident. Du point de vue de l'histoire des idées, c'est également à Firth qu'est souvent rattachée la notion de *communauté de locuteurs* (« speech community »), se rapportant aux participants du contexte de situation. Très utilisée en sociolinguistique, elle n'a à ma connaissance presque pas été reprise par les linguistiques de corpus¹⁵. Il semble en effet que les linguistiques de corpus, pourtant issues de cette tradition contextualiste, aient été amenées à délaisser la notion firthienne de *contexte de situation* et ses dérivés, en particulier avec l'avènement de bases de données de plus en plus volumineuses et l'établissement de conventions stables pour l'élaboration de corpus de référence, c'est-à-dire de « bases de données autonomes et générales », qui ne sont plus construites en lien avec un projet linguistique particulier (Léon, 2008 : 28), mais guidées par le projet global d'échantillonner pour représenter la langue générale. Ces notions gardent cependant une grande pertinence pour l'analyse des discours spécialisés, comme en témoignent les nombreuses références aux *communautés de discours* de Swales dans la littérature récente : l'auteur du concept s'est d'ailleurs interrogé récemment sur ces usages (Swales 2016 : 8). En revanche, la notion de *contexte de situation* semble connaître un relatif

¹⁵ On en trouve quelques emplois chez Stubbs (1996)

désintérêt, si bien que l'on peut se demander s'il est encore légitime de vouloir intégrer ces éléments à la réflexion linguistique. Un calcul des occurrences du terme et de la tournure adjectivale équivalente dans l'ensemble des livres de la collection Google Books donne un aperçu de cette tendance dans les publications de langue française et anglaise¹⁶.

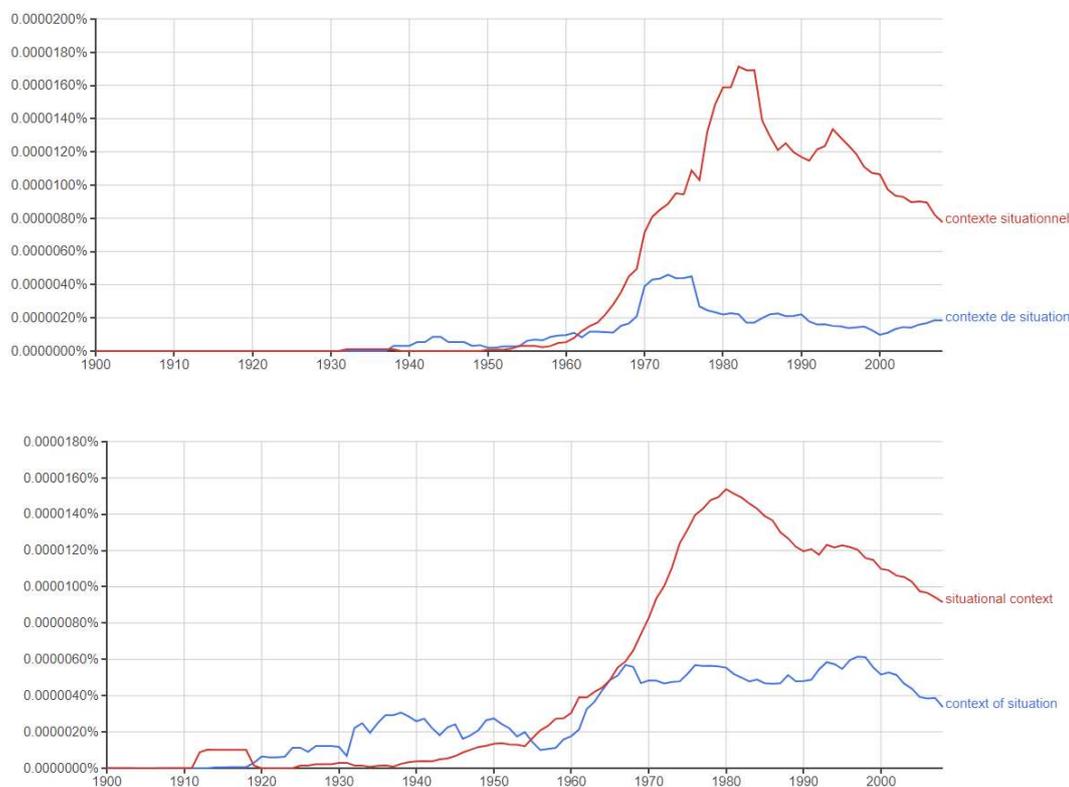


Figure 2. Evolution des occurrences de « contexte situationnel » et de « contexte de situation » (FR) / « situational context », « context of situation » (EN) dans Google Books

Si la majorité des usages recensés se situe dans les années 1980 et 1990, la tendance à la baisse n'est pas toujours franche et il serait intéressant de pouvoir observer aussi l'évolution de ces dix dernières années dans ce même corpus. En interrogeant la bibliographie en ligne des travaux de linguistique de Brill, il apparaît que la grande majorité des occurrences recensées figurent dans des travaux parus entre 2010 et 2017. Les chiffres suggèrent aussi que la variante adjectivale soit bien moins utilisée dans les travaux de linguistique, qui conservent plutôt le terme de Firth¹⁷.

¹⁶ Il est probable que les calculs intègrent, du moins en français, des énoncés non pertinents (du type « dans le contexte de situations difficiles »).

¹⁷ C'est en fait un terme que Firth reprend à Wegener, Malinowski et Gardiner (Léon 2008 : 15), ce qui explique que l'on en trouve des occurrences dès les années 1920.

	situational context	context of situation
Depuis 2017	10	28
Depuis 2016	17	77
Depuis 2015	37	144
Depuis 2014	54	198
Depuis 2013	64	252
Depuis 2012	76	299
Depuis 2011	82	333
Depuis 2010	86	347
Depuis 2000	97	384

Tableau 1. Evolution des occurrences de « situational context », et de « context of situation » (EN) dans la bibliographie en ligne des travaux de linguistique de Brill

Il faut donc relativiser ce désintérêt, et l'on peut supposer qu'il concerne certainement plusieurs courants de la linguistique, alors que le contexte situationnel continue de jouer un rôle important dans d'autres. Cette hypothèse correspond à celle de F. Rastier, qui explique qu'en intégrant à la définition du contexte une conception non réductionniste de la situation comme « une occurrence d'une pratique sociale » on définit « deux corps disciplinaires bien différents » (Rastier, 1998 : 99) : la grammaire et la logique d'une part, la rhétorique et l'herméneutique de l'autre. Dans le premier de ces corps disciplinaires, que Rastier nomme aussi « paradigme logico-grammatical » (*ibid.*, p.100), le contexte ne sera défini que par la taille et le voisinage, et en fonction d'unités discrètes. Pour les positivistes du Cercle de Vienne, à chaque mot doit correspondre une signification dans un contexte donné. En revanche, dans ce que Rastier nomme la « problématique rhétorique/herméneutique » (*ibid.*), se développe une conception non-compositionnelle du contexte, ouvrant la possibilité de tenir compte de l'existence de parcours interprétatifs et de variations individuelles. Alors que les linguistiques formelles se situent dans le premier ensemble, on retrouve dans le second les linguistiques de l'énonciation, mais aussi certains courants d'analyse conversationnelle, de sociolinguistique et d'ethnolinguistique. Il n'est cependant pas évident de situer les théories basées sur l'usage.

En acquisition du langage, les approches contextualistes se sont développées très tôt. L'importance du contexte est bien analysée dès la fin des années 1970, comme en témoigne par exemple l'article de D. Keller-Kohen (1978) qui explique que le contexte, défini par « toute la gamme d'éléments sur lesquels la compréhension et la production langagière de l'enfant pourrait s'appuyer »¹⁸ cristallise les interrogations de cette période (*ibid.*, p.453, notre traduction). Ce questionnement se poursuit actuellement dans le cadre de ce que M. Tomasello a appelé les « nouvelles perspectives enthousiasmantes » ouvertes par les modèles UB (Tomasello, 2001 : 77,

¹⁸ "I believe that the current period in child language study has been characterized by attempts to identify the wide range of factors that children seem to use in understanding and producing speech. The evidence suggests that these factors, often referred to as context, must be taken into account in the investigation of language acquisition."

notre traduction). C'est bien d'ailleurs en faisant appel à la notion de contexte qu'E. Lieven (2014: 48) définit l'approche UB du développement d'une langue maternelle :

Les propositions fondatrices de l'approche [UB] sont les suivantes : les enfants acquièrent le langage sur la base d'énoncés spécifiques produits dans des contextes spécifiques, et c'est à partir de ces premières acquisitions que des représentations linguistiques de plus en plus complexes et abstraites se construisent (notre traduction¹⁹).

Pour la psycholinguistique développementale, le contexte est donc le socle et le point d'ancrage de l'ensemble des développements ultérieurs : l'abstraction se construit petit à petit sur la base d'énoncés compris et appris en contexte. Mais que veut dire contexte ici ? Les théories UB de l'acquisition du langage s'appuient sur une analyse quantitative du discours adressé à l'enfant, désormais souvent appelé « input », trace de l'importance de la métaphore computationnelle dans les développements récents. Ceux-ci s'appuient en effet sur de grandes bases de données et c'est la modélisation computationnelle de celles-ci qui a permis d'en analyser la structure et d'évaluer la quantité d'informations entendues par l'enfant (Lieven, 2014: 48-49). Dès la fin des années 1990, des travaux ont ainsi proposé des modélisations du contexte linguistique présent dans les données transcrites. Au-delà du contexte linguistique (ou co-texte), ces théories qui définissent avant tout le langage comme instrument de communication (voir par ex. Diessel, 2014), s'intéressent au contexte situationnel dans lequel se mettent en place les dimensions fonctionnelles et communicationnelles du langage. Les travaux pionniers de l'équipe du MIT dirigée par Deb Roy et le travail de thèse de Brandon C. Roy (D. Roy, 2009, B. C. 2014) ont permis, à partir d'un corpus d'une densité inédite (celui qui est issu du projet « Human Speechome ») de modéliser l'usage des mots en tenant compte du contexte situationnel, défini comme un ensemble d'activités observées et catégorisées en s'appuyant sur leur distribution spatiale et temporelle. Cette modélisation, qui intègre le contexte linguistique et situationnel, fait ressortir des mots qui sont utilisés dans de nombreux contextes, et d'autres se produisant dans des contextes particuliers : « les mots qui sont utilisés dans des contextes distinctifs, sur le plan linguistique, spatial et temporel, sont ceux que l'enfant produit en premier, ce qui suggère qu'ils soient plus faciles à appréhender » (Roy et al. 2015 : 12663, notre traduction²⁰).

Ces développements suggèrent que la compréhension du langage en contexte, ou compréhension située, qui est au cœur des interactions dialoguées mais aussi de tâches comme la

¹⁹ "The fundamental proposals of [the usage-based] approach are that children learn language from particular utterances in particular contexts and that increasingly complex and abstract linguistic representations are built from this learning."

²⁰ "words used in distinctive spatial, temporal, and linguistic contexts are produced earlier, suggesting they are easier to learn."

traduction, puisse faire l'objet d'une modélisation qui améliore notre compréhension de ces phénomènes. Il faut cependant remarquer que la compréhension située constitue l'un des défis que rencontrent actuellement les chercheurs travaillant au développement d'intelligences artificielles²¹. Les nouveaux modèles d'apprentissage profond s'appuient sur une catégorisation toujours plus fine, mais contextualiser ne revient pas uniquement à discriminer : il s'agit bien aussi de juger, en s'appuyant sur des informations indexicales que chaque énonciateur/co-énonciateur reconstruit. A trop vouloir circonscrire le contexte en un ensemble fini de catégories, on court le risque de le réifier, en le réduisant à la somme des unités ainsi identifiées (Rastier 1998 : 101). A l'inverse, une lecture riche et ouverte du contexte est susceptible d'éclairer et d'orienter les analyses quantitatives, pour tenir compte des particularités du langage et des interactions humaines. Le traitement statistique de données langagières est également susceptible de nous faire perdre de vue les observables : ce questionnement auquel je suis revenue à plusieurs reprises et sous des angles différents est abordé dans la partie suivante.

1.4 Fréquences et seuils de significativité : réflexion sur les statistiques descriptives et inférentielles

L'étude des effets de fréquence, c'est-à-dire l'ensemble des travaux qui interrogent les effets de l'usage sur la cognition, se situe à l'interface de la linguistique cognitive de corpus et de la psycholinguistique, qui initialement au moins n'ont pas porté le même regard sur cet objet (Divjak et Gries 2012: 3). En psycholinguistique, la fréquence calculée est avant tout celle des mots, comme en témoignent d'ailleurs les principales commandes disponibles dans la suite de programmes CLAN (MacWhinney 2018: 16 et sq.). Dès la fin des années 1970, la fréquence d'occurrence des mots a été repérée comme organisatrice du « lexique mental » à partir des premières tâches de décision lexicale. Cette première définition de la fréquence comme décompte d'occurrences d'éléments lexicaux rejoint un présupposé qui a présidé à l'élaboration de la plupart des modèles classiques de perception et de production du langage : c'est le mot qui en constitue l'unité de base, qu'il s'agisse des modèles interactifs de lecture, d'accès lexical ou de production orale (Baayen 2010: 436-437). Elle témoigne aussi d'une influence des modèles de issus de l'analyse de la langue écrite, et d'une prise en compte insuffisante des caractéristiques propres à l'oral (Parisse, 2002: 479). Nous avons vu cependant que dès les premiers enregistrements d'interactions langagières documentant l'acquisition précoce du langage, les chercheurs ont cherché à tenir compte des situations de production et ont favorisé une approche du mot en contexte : ces approches, devenues quantitatives avec l'avènement de grandes bases de données

²¹ Voir par exemple le projet ANR Atlantis porté par T. Poibeau (ArTificial Language uNdersTanding In robotS, <http://www.agence-nationale-recherche.fr/Project-ANR-15-CHR2-0002>)

comme CHILDES, ont commencé à s'intéresser à la fréquence en tant qu'indice complexe au tournant des années 1980 (Chenu et Jisa 2009b) rejoignant ainsi la tradition contextualiste qui caractérise l'émergence des linguistiques de corpus. Elles ont intégré de nouveaux outils d'analyse pour articuler la fréquence des mots à leur environnement textuel, qu'il s'agisse des séquences que la phraséologie analyse comme préconstruites et qu'elle considère dans leur ensemble, le plus souvent de façon non compositionnelle (Legallois et Tutin, 2013) ou simplement de segments répétés qui peuvent contenir une ou plusieurs places susceptibles d'être instanciées différemment : c'est ce que Theakston et Lieven (2017) nomment des chaînes à plusieurs unités (« multi-unit strings »), dans un article qui les situe au cœur de la compréhension UB de l'acquisition du langage.

En linguistique de corpus, la nécessité de disposer d'échantillons susceptibles de représenter la diversité des langues et des genres discursifs a été prise en compte très tôt. Or la représentativité, qui définit « l'intégration plus ou moins exhaustive de la variabilité observable dans un échantillon » (Biber 1993: 244, notre traduction²²) est liée à la fréquence d'au moins deux façons. D'abord parce qu'il faut faire des hypothèses sur la fréquence afin de procéder à l'échantillonnage des données, mais aussi et surtout parce que la constitution des échantillons varie en fonction de la fréquence des phénomènes langagiers à étudier. Alors que les faits de langue fréquents ont une distribution relativement régulière qui permet de s'en tenir à des fragments de textes, les phénomènes moins fréquents ne seront observables que dans des données plus denses (*ibid.* p. 252). Dans l'ensemble, si les grands corpus permettent d'étudier de façon approfondie et sûre des constructions ou termes fréquents, l'étude d'éléments de langage de basse fréquence est plus ardue et doit faire l'objet d'une réflexion méthodologique. Dans notre corpus COP21, pour observer la diffusion (progressive et souvent très prudente) des termes liés à la justice climatique dans des communautés dont le discours n'est pas militant, nous avons ainsi fait le choix d'intégrer des rapports entiers plutôt que des extraits de ceux-ci. Ce choix entraîne inévitablement des déséquilibres : les rapports sont de taille inégale, et ceux des organisations onusiennes sont toujours beaucoup plus fournis que ceux des ONG, par exemple. De plus, nous avons choisi d'intégrer les organisations dont les publications sur ces questions étaient connues et bien répertoriées, plutôt que de vouloir à tout prix rassembler le même nombre de documents dans chaque sous-corpus. Ce déséquilibre peut en effet être compensé, en travaillant sur des fréquences normalisées²³ et en utilisant des tests statistiques peu sensibles à la taille des

²² "Representativeness refers to the extent to which a sample includes the full range of variability in a population."

²³ Il s'agit des fréquences brutes divisées par la taille du corpus ou du sous-corpus, et multipliées par un facteur de normalisation (en général 1000 ou 1 000 000).

échantillons, alors que certaines informations présentes uniquement dans certaines parties des rapports auraient été perdues si nous avions choisi de n'intégrer que des extraits de ceux-ci. Voici l'une des stratégies mises en œuvre afin de compenser le déséquilibre : pour produire des listes de mots clés [Rossi et al. 2019b], nous avons considéré conjointement les seuils de significativité donnés par la mesure de LL (*log-likelihood ratio*, ou rapport de vraisemblance logarithmique), dont le calcul est basé sur les différences de fréquences observées (Rayson et Garside 2000 : 3), et la taille d'effet donnée par le logarithme binaire du ratio des fréquences relatives, ou « log ratio » (Hardie 2014). Dans les deux cas, l'utilisation de logarithmes permet d'obtenir des séries de mesures faciles à comparer, mais alors que LL estime la probabilité d'avoir une différence significative entre deux fréquences observées, en tenant compte de la taille des corpus comparés, le *log ratio* se base sur le ratio des fréquences relatives observées et permet grâce à la transformation logarithmique de savoir que chaque point supplémentaire correspond à une différence de fréquence multipliée par deux (*ibid.*). L'ajout de cette seconde mesure ne va pas de soi, et de nombreux travaux se contentent d'un test statistique comme LL. Cependant, l'inférence statistique contenue dans une mesure de significativité n'est pas toujours suffisante, notamment parce que l'hypothèse sous-jacente est celle d'une répartition égale, qui ne devrait que très rarement correspondre à un observable dans des données langagières, sauf à accepter que le hasard y joue un rôle. C'est le sens d'une déclaration de Sinclair et al. (2004[1970]) qui disent avoir peu à peu cessé de se fier aux tests statistiques, et expliquent que les mesures de significativité nous révèlent en réalité « qu'il est peu probable que les répartitions observées soient liées au hasard, alors même que [le linguiste] ne s'attend en aucun cas à ce qu'elles le soient. » (Sinclair 2004: xxi, in Léon 2008: 27 note 21, notre traduction²⁴). C'est aussi ce qu'a bien analysé Kilgarriff (2005) dans un article entièrement consacré à la réfutation du rôle du hasard dans les productions langagières.

Les grammaires cognitives et les grammaires de construction ont aussi développé des méthodes plus fines et plus ciblées pour analyser les différences de fréquence, notamment dans le cas d'alternances syntaxiques. Il se pourrait que l'analyse statistique soit ici moins risquée, ainsi que le suggèrent Tummers et al. (2005: 234²⁵) : si la grammaire prend la forme d'un réseau de

²⁴ Voici la citation originale, que je n'ai pas traduite intégralement : “[O]ver the years [...] I have lost most of my original confidence in [the t-score] and in other statistical procedures. If something like the co-occurrence of two or more words is statistically significant, this tells me that there is but a small chance of it being accidental. But I don't expect it to be accidental anyway. One of the worrying aspects from the beginning, what really made me suspicious, was the frequent finding that the actual co-occurrence of words in texts is many times the prediction that is made on a statistical basis. Not just slightly over the estimate, but hundreds or thousands of times more frequent than the expected. Statistical prediction based on chance seems just irrelevant.”

²⁵ “Grammatical models like Cognitive Linguistics and Construction Grammar in particular, which accept the existence of networks in which elements or levels may have different degrees of salience, should

constructions au sein desquelles des différences de saillance se font jour, alors calculer la force d'attraction d'un item au sein d'une construction revient à donner une définition plus précise du lien entre fréquence et ancrage (*entrenchment*) en permettant de mesurer quelque chose comme le caractère plus ou moins prototypique d'un item au sein d'une construction. Il existe cependant, ici aussi, une différence entre décrire et prédire, et la plupart des mesures d'association utilisées dans l'analyse colostruccionnelle (voir Desagulier 2015: 104) intègrent un test de significativité (souvent basé sur un rapport de vraisemblance logarithmique d'ailleurs) qui suggère que l'occurrence d'un élément plutôt qu'un autre dans une construction puisse être le fruit du hasard. Or il me semble particulièrement difficile d'imaginer à quoi ce hasard pourrait correspondre pour un locuteur. Même la propension de certains locuteurs à déformer ce que d'autres reconnaissent comme des expressions figées n'est certainement pas le fruit du hasard. Sans remonter au personnage de Sheridan, on trouve dans le roman de Philip Roth intitulé *Sabbath's theater* quelques exemples de ce qu'un académicien a récemment proposé de nommer « malapropisme » en français, remarquant que nous manquions d'un équivalent pour ce terme pourtant issu de l'expression française « mal à propos »²⁶ :

Remembering the confidence with which she believed herself to be smoothly idiomatic, lovingly recalling from over all the years as many as he could of Drenka's malapropisms stripped him now of every defense, and once again he descended to the very pit of his sorrow: bear and grin it . . . his days are counted . . . a roof under my head . . . when the shithouse hit the fan . . . you can't compare apples and apples . . . the boy who cried "Woof!" . . . easy as a log . . . alive and cooking . . . you're pulling my leg out . . . I've got to get quacking . . . talk for yourself, Johnny . . . a closed and shut case . . . don't keep me in suspension . . . beating a dead whore . . . a little salt goes a long way . . . he thinks I'm a bottomless piss . . . let him eat his own medicine . . . the early bird is never late . . . his bark is worse than your cry . . . it took me for loop . . . it's like bringing coals to the fireplace . . . I feel as though I've been run over by a ringer . . . I have a bone to grind with you . . . crime doesn't pay off . . . you can't teach an old dog to sit . . . (Roth 1996: 71)

Toutes ces expressions loufoques sont probablement en partie au moins le fruit de l'imagination du romancier, mais elles correspondent aussi à des conduites langagières réelles. Il s'agit de créations inconscientes, susceptibles de nous renseigner sur le traitement et la production du langage (voir notamment MacWhinney et Anderson 1986 : 13, qui distinguent les extensions sémantiques d'expressions idiomatiques des « malapropismes »). Or les grands corpus utilisés pour l'analyse colostruccionnelle en gardent en général peu de traces. Le tableau 2 montre

be particularly sensitive to differences in the frequency of alternative expressions. That is to say, a usage-based linguistics needs quantification and statistical analysis."

²⁶ <http://www.academie-francaise.fr/malapropisme>

que, pour les trois premiers exemples ci-dessus, les expressions figées apparaissent bien comme telles dans le Corpus of Contemporary American English (COCA²⁷) mais que l'on ne trouve pas de variation du même type que dans l'extrait ci-dessus.

Requête dans le COCA	Fréquence (brute)	Variantes (fréquence brute)	Remarques
grin and bear	64	bear and _v* (139)	Une centaine de formes verbales différentes, pluralité d'effets de sens. Aucune occurrence de « grin »
his days are (numbered)	42 (17)	His days are short (2) his days are limited (2)	Aucune occurrence de « counted », et d'autres constructions fréquentes (« his days are full »).
a roof over my head	50	a roof over my children's head (1) a roof over my family's head (1)	

TABEAU 2 : EXPRESSIONS IDIOMATIQUES DANS LE CORPUS OF CONTEMPORARY AMERICAN (COCA)

Ces « malapropismes », ou emplois mal à propos, sont traditionnellement basés sur une ressemblance phonétique. La liste ci-dessus fournit également des exemples de glissements dans lesquels le sémantisme est conservé en totalité (*counted* pour *numbered*) ou partiellement (*bringing coals to the fireplace* n'est peut-être pas aussi inutile que ne peut l'être le transport de charbon jusque dans la ville minière de Newcastle, comme le suggère l'expression originale), mais dans tous les cas, ces exemples montrent que le choix d'un élément de substitution, si maladroit soit-il, n'est jamais lié au hasard.

Ainsi, même si le traitement probabiliste fonctionne bien sur de grandes bases de données langagières, il produit une description des faits de langue qui nous éloigne de l'énonciation, parce que le calcul statistique ne correspond pas à l'intuition du locuteur. Si l'assimilation perdure, c'est peut-être parce qu'elle fait fond sur une compréhension simple et relativement convaincante du traitement de la fréquence par le récepteur humain, qui procéderait à la façon d'un « compteur mental » (Baayen 2010). J'espère avoir rassemblé ici quelques éléments qui permettent de montrer que les données de fréquence recouvrent une réalité beaucoup plus complexe qu'il n'y paraît. Mais se défaire de cette analogie trompeuse ne revient pas à dire que les tests de significativité ne sont pas valables dès lors qu'ils sont appliqués à des données langagières. Il s'agit plutôt de situer le traitement probabiliste, qui peut constituer une étape dans un ensemble d'analyses, et d'accepter qu'il ne puisse être qu'un outil heuristique (Kilgarriff 2005 : 264). Cette proposition, qui rejoint l'idée de va et vient entre analyses quantitatives et qualitatives déjà discutée dans mes premiers travaux et en constitue une version plus aboutie, sera illustrée à plusieurs reprises dans ce qui suit. Je me contenterai donc ici d'aborder deux conséquences majeure de la réflexion esquissée ici. La première est que mes travaux sont guidés par une grande

²⁷ Toutes les requêtes ont été faites sur <https://corpus.byu.edu/coca/>

prudence quant à l'usage des statistiques inférentielles : il en résulte notamment une préférence pour les statistiques descriptives, face au risque de proposer des inférences trompeuses quand les données ne sont pas en quantité suffisante [Rossi et al. 2019a]. La seconde correspond à une conception de l'analyse statistique comme point de vue sur les données, ce dernier pouvant être plus ou moins éclairant : c'est ce que montre le passage, dans le travail que j'ai mené avec C. Parisse sur le rôle d'informations sémantiques dans l'acquisition des catégories nominales et verbales chez l'enfant, d'une recherche infructueuse basée sur le calcul d'un indice de significativité [Rossi & Parisse 2012] à une analyse des correspondances qui permet de faire ressortir des tendances en considérant les interactions entre plusieurs catégories sémantiques, tout en restant au plus près des données ([Parisse et Rossi 2017] cf. page 426 et sq.)

Le fil conducteur de l'interrogation qui m'a conduite à me former aux méthodes statistiques et à les utiliser, avec beaucoup de précautions, a toujours été pour moi la question suivante : quelle est la pertinence des statistiques pour l'analyse de données langagières ? D'une réflexion sur la fréquence à l'acquisition d'une meilleure connaissance des modèles statistiques, mon cheminement est ensuite arrivé, récemment, jusqu'aux algorithmes et à leur utilisation dans les modèles d'apprentissage profond, qui ont permis à de nouveaux moteurs de traduction automatique de donner des résultats jamais atteints auparavant, et dont il n'est pas exclu qu'ils permettent de mieux comprendre l'acquisition du langage. J'expliquerai, dans la troisième partie de cette synthèse, mes travaux actuels et en cours sur ces questions. Je souhaite que ces travaux permettent d'encourager une réflexion sur ces modèles, et de réfléchir à une analogie dont je viens de suggérer qu'elle est au moins en partie trompeuse, ainsi que l'a très bien expliqué Sylvain Auroux il y a vingt ans déjà :

« La mécanisation du langage ne doit pas nous entraîner à des thèses nouvelles sur la nature du langage et, en particulier, ne doit pas nous entraîner à penser que, si nous parlons, c'est parce que nous avons des algorithmes dans notre tête. Au contraire, il y a deux thèses parfaitement compatibles et qui suffisent pour comprendre comment fonctionne le langage. C'est de supposer que d'un côté, il y a des formes de vie, et de l'autre des outils linguistiques plus ou moins compliqués. Et que jamais on ne pourra faire que l'un soit identique à l'autre. Cela ne veut pas dire qu'on ne peut pas remplacer l'un par l'autre si l'on entend par là que nous allons vers une standardisation et que l'on assiste peu à peu à une diminution du caractère variable de ce qui fait la vie et l'histoire. » (Auroux, 1998 : 95).

Statistiques et calculs probabilistes seraient alors autant d'outils heuristiques, au même titre que les autres outils linguistiques que sont par exemple les formalismes. Ils s'en distinguent cependant, car ils participent à un mouvement de sortie de la « sphère du langage » que Steiner fait remonter au 19^{ème} siècle, dans un texte qui évoque le rétrécissement du « monde des mots » (Steiner 2010 : 35-36 in Rey 2016 : 116). Ils sont liés non seulement au traitement de données

riches et de plus en plus massives, mais aussi à l'essor des modèles computationnels, dont je pense qu'il ne devrait pas se faire au détriment d'autres perspectives tout aussi riches, notamment celles de l'analyse linguistique et traductologique.

2. D'un champ à l'autre : principaux points d'ancrage théoriques

Pour suivre le fil de ma formation intellectuelle, je commencerai par aborder ce que j'appelle les points d'ancrage linguistique de mes travaux. Ce choix ne revient cependant pas à leur donner une place primordiale et fondatrice. La recherche d'applications et donc de contextes dans lesquels les théories fonctionnent bien est au contraire porteuse d'interrogations qui ont été le moteur de mon cheminement, et ne permettent pas d'inscrire mes travaux dans l'un des deux ensembles théoriques évoqués dans ce qui suit, du moins pas si cette inscription se fait à l'exclusion des autres. Pour chacun des points d'ancrage que j'identifie dans ce qui suit, je propose un ou plusieurs points de passage de l'analyse linguistique à la traductologie. Consciente de la complexité des relations qui unissent linguistique et traductologie, j'ai souhaité apporter un éclairage personnel qui, si partiel soit-il, ne rejoint pas les antagonismes souvent décrits.

Ce cheminement, que je place plus volontiers du côté des linguistiques appliquées et empiriques, est également très ancré dans la psycholinguistique développementale dont je montrerai qu'elle a constitué une voie d'accès privilégiée vers la traductologie. Enfin, mes travaux en cours s'inscrivent dans le champ d'une traductologie empirique émergente, que je m'attacherai à caractériser et dont je décrirai quelques évolutions récentes. Pour conclure, je reprendrai les principales interrogations sur la cognition qui traversent mes travaux, afin de les inscrire dans le cadre des perspectives très larges et pluridisciplinaires ouvertes par les sciences cognitives.

2.1 Points d'ancrage linguistiques

La linguistique énonciative

C'est principalement sur la question de l'invariant que mes travaux ont rencontré les linguistiques énonciatives. Issue des premières linguistiques énonciatives (guillaumiennes et adamczewskiennes) la notion a évolué pour devenir « forme schématique » chez Culioli, mais elle désigne toujours la recherche du « plus petit dénominateur commun » (Lowrey et Toupin 2010: §12). Elle fait partie de la formation de tout angliciste (voir notamment Lapaire et Rotgé 2002 : 39 et sq.) parce qu'elle doit permettre d'unifier les descriptions linguistiques en dépassant les taxinomies, et par conséquent de simplifier l'enseignement de l'anglais. Dans la théorie des

opérations énonciatives (TOE), la recherche d'invariant se fait à partir d'un marqueur²⁸, et c'est surtout en travaillant sur les prépositions dans une perspective contrastive que j'y ai été confrontée. L'invariant sémantique des prépositions repose souvent sur l'hypothèse de la primarité du spatial (Groussier 1997), mais cette hypothèse n'implique pas que les enfants apprennent d'abord le sens spatial des prépositions, ou que les usages spatiaux de ces marqueurs soient les plus fréquents : « primarité n'est pas primauté » (*ibid.* p. 221), et les formes schématiques de la TOE comprennent des dimensions configurationnelles ou topologiques qui ne correspondent pas à un décalque des relations spatiales mais se situent à un degré d'abstraction supérieur. C'est ainsi qu'elles doivent permettre de rendre compte conjointement des usages spatiaux et non spatiaux des prépositions comme des particules adverbiales (voir par ex. Dufaye 2012, ou Ranger 2014). Une telle recherche d'invariants à partir de marqueurs ne revient cependant pas à suggérer que la comparaison d'une langue à l'autre puisse se faire sur la base des marqueurs : c'est plutôt la compréhension du fonctionnement d'un marqueur dans une langue donnée qui peut éclairer la recherche d'équivalents dans une autre langue. Mon travail sur les premières prépositions dans la langue de l'enfant [Rossi et al. 2007] est ainsi soucieux de tenir compte de la pluralité d'effets de sens observables, et il ne cherche pas à comparer deux à deux des marqueurs anglais et français dont le fonctionnement est à l'évidence différent. Ce choix méthodologique que j'explique ici beaucoup plus longuement que je ne l'avais fait dans la publication en question, se trouve conforté par la comparaison de grandes bases de données et la visualisation rapide que permet désormais l'outil *Wordbank* (cf. supra 1.2, note 8). Je n'ai sélectionné ci-dessous que deux grands ensembles qui correspondent aux données analysées dans mes publications sur l'acquisition du français et de l'anglais : les données issues des inventaires parentaux présents dans la base pour l'anglais américain (Fenson et al. 2007 ; Thal et al. 2013 ; Fernald et al. 2013) et pour le français (Kern et Gayraud 2010 ; Von Holzen et al. 2018). Les trois prépositions de la langue anglaise sélectionnées correspondent à celles qui encodent les premières relations spatiales comprises par les enfants, quelle que soit la langue (Hickmann 2012 : 27) et les proportions représentées sur les figures 3 et 4 ci-dessous montrent bien qu'il n'existe pas de correspondance bi-univoque entre les emplois fréquents et précoces des marqueurs *up* et *down* et ceux des marqueurs qui sont proposées comme équivalents pour les enfants francophones²⁹. J'ai restitué deux marqueurs repérés dans mes travaux, qui permettent déjà une meilleure comparabilité.

²⁸ Pour Culioli, la langue est un agencement de « marqueurs » d'opérations énonciatives (unités de taille variable, allant des morphèmes aux constructions syntaxiques) et c'est l'observation analytique de ces agencements qui permet de reconstruire les opérations.

²⁹ La version française des inventaires n'est pas une traduction de la version originale (américaine) mais bien une adaptation au contexte français : elle a été élaborée sur la base de très nombreux

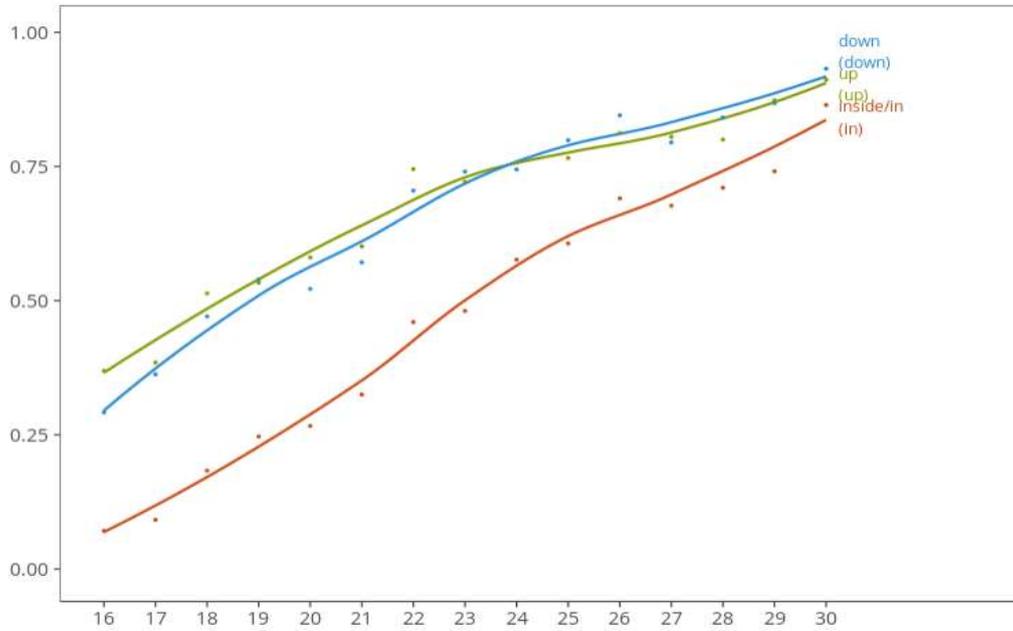


Figure 3 : proportion des enfants américains produisant *down*, *up* et *in* entre 16 et 30 mois (source : wordbank)

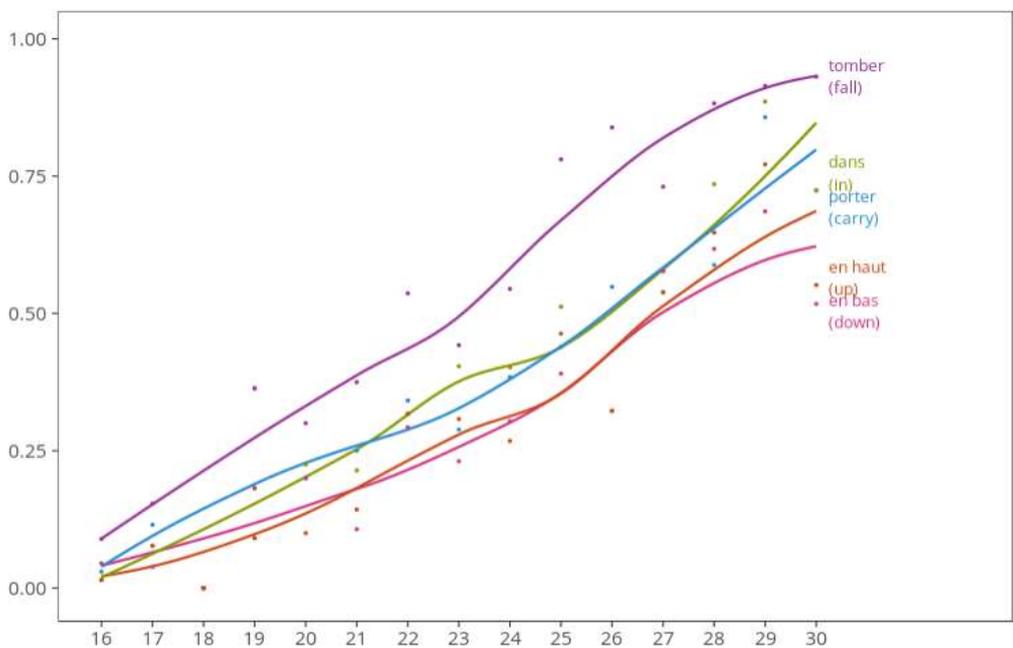


Figure 4 : proportion des enfants français produisant *en bas*, *en haut* et *dans* mais aussi *tomber* et *porter* entre 16 et 30 mois (source : Wordbank)

Sans rentrer dans l'analyse des différents marqueurs, on vérifie ici que les formes schématiques sont conçues pour décrire le fonctionnement de marqueurs dans une langue

enregistrements d'enfants francophones puis testée sur de grandes cohortes. Il n'est donc pas étonnant que la correspondance que suggère ici *Wordbank* soit trompeuse.

donnée, et non pour traiter du passage d'une langue à l'autre. En tant que représentation abstraite reconstruite à partir d'énoncés, et susceptible de déformations afin de rendre compte du travail énonciatif, elles ne sauraient donc constituer quelque chose comme l'*interlangue* que le cognitivisme classique liait aux représentations sémantiques. Si l'analyse contrastive a partie liée avec les linguistiques de l'énonciation, c'est parce qu'elle cherche d'abord et surtout à éclairer la « diversité des langues naturelles » (voir par ex. Culioli 1990: 14 et 1999: 43) et les contrastes que cette diversité permet de saisir. Et si la TOE a si fréquemment rencontré la traduction, ce n'était pas pour y trouver la validation d'un modèle abstrait, au-delà ou en-deçà de la diversité des langues : ce qui intéresse la TOE, ce sont les variations paraphrastiques qui résultent de choix énonciatifs toujours particuliers. Dans ce cadre, la traduction est « une modalité énonciative particulière » (Boisseau 2016 : 170), et « un cas particulier de paraphrase » (Culioli 1971 : 8). Il est d'ailleurs remarquable que la tentative d'unification qui s'intéresse à « rendre compte dans la même théorie de phénomènes en apparence hétérogènes » (*ibid*) intervienne *a posteriori* chez Culioli, pour « permettre au mathématicien et au linguiste de collaborer de façon fructueuse » (*ibid.* page 7). L'analyse contrastive, quant à elle, se déploie d'abord en vue d'applications qui ne sont pas liées à l'analyse mathématique ou à l'unification théorique : les travaux de J. Guillemin-Flescher sont bien ceux d'une linguiste culiolienne « dont la réflexion s'enracine dans une triple expérience du bilinguisme, de la traduction littéraire et de l'enseignement » (Chuquet et Paillard 1992: 243). Ses propositions d'analyse contrastive, comme celles de Chuquet et Paillard (1987) qui se situent dans le prolongement de la *Syntaxe comparée du français et de l'anglais. Problèmes de traduction* (Guillemin-Flescher 1981) correspondent à une étude des problèmes de traduction au prisme de la TOE, c'est-à-dire en termes de choix énonciatifs. Il me semble ainsi que, même si l'énoncé ne constitue pas l'unité la plus pertinente pour la traductologie, les linguistiques énonciatives en général et la TOE en particulier permettent d'interroger l'accès au sens dans une situation d'énonciation donnée. D'autre part, dès lors qu'elle se situe au niveau de l'organisation du discours, l'analyse linguistique est bien plus susceptible de rejoindre la réflexion traductologique. L'analyse conversationnelle, qui partage ses fondements théoriques et épistémologiques avec les linguistiques de l'énonciation et fait la part belle à l'action « située, occasionnée, sensible aux contingences et orientée vers le contexte de sa production et vers sa spécificité [donc] irrémédiablement *indexicale* » (Mondada 2008 : 882), présente d'ailleurs des affinités fortes avec l'analyse traductologique, qui s'intéresse bien davantage à la communication qu'au système linguistique. Ainsi par exemple, la priorité accordée à la communication par Jean-René Ladmiral dans la description des objets langagiers de la traductologie n'est certainement pas fortuite :

S'il est vrai que la traduction n'est pas *qu'une* affaire de linguistique, on ne peut pas dire qu'elle n'est pas *du tout* l'affaire des linguistes. Dans la mesure où la traductologie prend pour objet ces

réalités langagières que sont la communication, les langues, les textes, les mots, etc., il est bien clair qu'il lui aura fallu emprunter à la linguistique les éléments de sa terminologie, qui permettront de conceptualiser cet objet, ainsi que les linéaments de sa méthodologie. (Ladmiral 2004 : §6).

Certains auteurs comme Deslile (1980), ont vu très tôt dans l'analyse du discours une « méthode de traduction », et dans un article plus récent Yves Gambier explique que de nombreux travaux abordent la traduction comme une « communication complexe » (Gambier 2000 : 99) et propose de l'étudier comme « discours, c'est-à-dire énoncé et énonciation, formation discursive (cf. Foucault 1971; Pêcheux 1975), intrication d'un mode de dire et d'un lieu socioculturel déterminés. » (*ibid.* p.104). Or l'importance de l'énonciation pour certains travaux souvent rangés parmi les premières analyses contrastives a rarement été souligné, et la critique des « stylistiques comparées » rassemble des travaux différents : l'ouvrage fondateur de Vinay et Darbelnet (1958) s'inscrit dans le contexte universitaire canadien de l'époque, et se place plutôt sous le patronage linguistique de Charles Bally. La véritable limite de toute démarche contrastive perdure cependant même lorsque l'on considère l'énoncé en contexte, et la variation du sens en fonction des situations d'énonciation³⁰, car le linguiste réfléchit évidemment toujours à partir du « produit fini » (Guillemin-Flescher 1986 : 59) plutôt qu'à partir des processus de traduction. Les linguistiques cognitives, qui se donnent pour objectif de parvenir à une meilleure articulation du langage et de cognition, sont-elles plus à même de rendre compte des processus de traduction ?

Les linguistiques cognitives

Même s'il existe de nombreuses correspondances entre la TOE et les linguistiques cognitives (voir notamment Col et Victorri 2007), ces deux ensembles théoriques ne partagent pas la même conception du sens, et le statut des invariants en linguistique cognitive ne rejoint pas ce que nous venons de décrire. La structure topologique des schémas d'images (*image schemas*, voir notamment Lakoff 1990 : 54) les rapproche des formes schématiques mais ils se caractérisent par un ancrage beaucoup plus net dans l'expérience vécue : ils représentent donc schématiquement des contenants, des trajectoires, des liens ou des forces, pour ne citer que quelques schémas souvent utilisés. Les linguistiques cognitives proposent que ces schémas d'images correspondent à des structures pré-conceptuelles présentes dans notre expérience et qui permettent d'expliquer l'émergence des structures conceptuelles. A la primarité du spatial évoquée ci-dessus s'ajoute donc, en linguistique cognitive, une hypothèse sur la primauté des relations spatiales conçues comme le socle de relations plus abstraites. Ainsi l'espace perceptuel serait avant tout appréhendé

³⁰ Ce que font d'ailleurs déjà, par endroits au moins, Vinay et Darbelnet (1958 : 173 et sq.), comme l'a très justement fait remarquer Isabelle Colombat (2003 : 424)

comme un espace physique : c'est ce dont témoigne, d'après Lakoff (1987 : 509) le rattachement d'un son perçu à un emplacement par l'intermédiaire du déictique *there* dans l'énoncé :

« There's the buzzer! »

D'une manière plus générale, ce sont les processus de conceptualisation qui intéressent la linguistique cognitive, et il importe que les concepts et modèles utilisés par la théorie linguistique leur correspondent. La théorie de la métaphore conceptuelle constitue ainsi l'une des tentatives les plus abouties de lier modèles linguistiques et processus cognitifs. Elle consiste à voir dans la métaphore un mode de conceptualisation. Lakoff et Johnson (1980b : 195) opposent les concepts qui sont directement issus de l'expérience (notamment ceux qui sont liés à l'orientation et au positionnement dans l'espace) aux concepts métaphoriques. Ces derniers correspondent aux concepts « qui sont compris et structurés [...] à partir d'autres concepts » (*ibid.*, notre traduction). Une telle proposition rejoint les préoccupations du traductologue bien plus que ne pourrait le faire l'idée de primitives sémantiques, ou toute autre proposition d'un ensemble de catégories se voulant universelles, distinctes et délimitées une fois pour toutes. Le projet de traductologie intégrée de Mary Snell-Hornby s'annonçait ainsi dès la fin des années 1980 comme une « prototypologie » : système dynamique de relations basées sur des prototypes, qu'elle oppose aux « typologies rigides » (Snell-Hornby 1988 : 31). L'auteure formule dès la fin des années 1980 une critique des formalismes linguistiques, qu'elle propose de dépasser en se fondant sur les acquis de la psychologie cognitive ; ceux-là même qui sont également au cœur des linguistiques cognitives³¹. En l'occurrence, il s'agit surtout de la théorie des prototypes de Rosch (1975) dont Lakoff (1987 : 41 et sq.) a bien montré l'importance. Au-delà de ce point de convergence épistémologique, les métaphores conceptuelles ont été utilisées de deux façons au moins dans le champ de la traductologie : la première est métathéorique et consiste à utiliser la conceptualisation métaphorique comme indice d'évolutions de la théorie, la seconde concerne l'enseignement de la traduction et fait du repérage des métaphores « un outil cognitif puissant permettant de fonder un grand nombre de décisions traductionnelles » (Vandaele et Lubin 2005 : 415). Ces deux types d'utilisation m'ont intéressée et apparaissent dans mes travaux, ainsi que je vais l'expliquer.

L'analyse du « rôle des métaphores en traductologie contemporaine » de D'hulst (1992) est un exemple éclairant d'utilisation métathéorique des métaphores conceptuelles : l'auteur

³¹ Les propositions théoriques de Snell-Hornby ne sont qu'un exemple parmi d'autres. Je les cite parce que je les ai trouvées particulièrement éclairantes. La proposition bien plus récente d'une « approche expérientielle de la traduction » faite par Isabelle Collombat lors du colloque VUAT (Vers une anthropologie de la traduction, 16-17 mai 2017) montrait bien la pertinence et l'actualité des modèles issus de la psychologie cognitive.

s'intéresse au caractère préthéorique des métaphores conceptuelles et montre « le rôle crucial qu'elles ont joué dans le façonnement des théories traductives » (*ibid.* p.38). L'assimilation de la traduction à « un processus de communication » fait partie, selon D'hulst, des « métaphores-clefs de la traductologie contemporaine » (*ibid.* p. 39 et sq). Or le domaine source, que D'hulst appelle la *communication* ne correspond pas à la linguistique conversationnelle que nous évoquions ci-dessus, mais bien plutôt aux modélisations informatiques de celle-ci : les références au *stockage* en mémoire et au *traitement de l'information* sont la trace d'une analogie entre esprit et ordinateur dont D'hulst explique qu'elle est issue des modèles psycholinguistiques qu'utilise la traductologie de l'époque. Qu'en est-il actuellement de cette analogie ? Avec l'essor des outils de traduction assistée par ordinateur (TAO) et l'intégration au poste de travail de traducteur de nouveaux systèmes de traduction automatique (TA) toujours plus performants, les métaphores se sont-elles figées ou transformées ? Sans considérer pour l'instant les évolutions diachroniques, je me suis intéressée au redéploiement de cette métaphore en lien avec les nouveaux systèmes de TA [Rossi 2017-v ; Rossi 2017a-oral]. Les analyses portent sur trois « mini-corpus maison » (je reprends ici la traduction du terme anglais *DIY corpora* proposée par Loock, 2016) : le premier se compose d'articles de recherche proposant une analyse traductologique des changements actuels, le deuxième rassemble des documents accompagnant et décrivant l'usage de la traduction automatique par des traducteurs salariés d'une institution internationale (la DG-Traduction de la Commission européenne, désormais DGT), et le dernier comprend des articles issus de la recherche Google portant sur le développement et l'amélioration du nouveau système neuronal lancé en novembre 2016 . La méthode mise en œuvre pour repérer les domaines source et les métaphores caractéristiques d'un corpus est empruntée à L'Hôte et Lemmens (2009 : §18 et sq.) : j'ai utilisé le logiciel WMatrix (Rayson 2008) pour détecter les mots-clés et les domaines sémantiques en comparant les corpus deux à deux. L'analyse qualitative se construit ensuite à partir des points saillants ainsi repérés, en opérant un retour aux concordances, et souvent au contexte élargi de celles-ci. L'analyse des mots-clés montre notamment que ce qui distingue le corpus Google, c'est la référence à des « modèles », et à « l'encodage » de l'information. J'ai utilisé ce constat comme point de départ pour rechercher ensuite des composés et collocations fréquentes et caractériser deux conceptions antagonistes de la traduction : la traduction comme tâche modélisable, et réductible au calcul d'une part, et d'autre part la traduction comme expérience humaine riche et complexe, irréductible même à l'unicité d'une métaphore [Rossi 2017-v]. Le mot clé anglais *neural*, qui est également repéré dans le corpus Google lorsqu'on le compare aux articles écrits par des traductologie (LL : 160.13 et log ratio : 8.41), m'a permis d'identifier les passages dans lesquels la métaphore de l'esprit ordinateur était susceptible de se déployer. En voici un exemple, où l'on est passé des modèles à leur actualisation en « systèmes »,

qui sont caractérisés comme autant d'agents capables de comportements intentionnels comme l'évitement (*sidestep*) ou l'entraînement (*training*) :

An advantage of Neural Machine Translation is that it sidesteps many brittle design choices in traditional phrase-based machine translation. In practice, however, NMT systems used to be worse in accuracy than phrase-based translation systems, especially when training on very large-scale datasets as used for the very best publicly available translation systems. (Wu et al. 2016).

La métaphore du système informatique devenu « réseau de neurones » complète et renforce celle de la communication modélisée, et s'accompagne d'un effacement tangible de toute présence humaine. L'occasion de réfléchir aux implications de ce constat m'a été donnée dans le cadre du suivi d'un mémoire de traductologie puis d'un article co-écrit avec l'un de nos jeunes diplômés [Rossi & Dirand 2019]³². En complément des données de perception recueillies par B. Dirand grâce à un sondage en ligne, j'ai proposé des analyses basées sur corpus qui ont pour point de départ l'exploration d'une catégorie sémantique particulièrement distinctive dans les données évoquées ci-dessus : les langues, le langage et la grammaire (il s'agit de la catégorie Q3 dans le système USAS qu'utilise le logiciel WMatrix³³). Un examen des éléments ainsi étiquetés montre que cette catégorie rassemble un nombre disparate d'occurrences de *translator(s)* au sein de nos trois corpus. Les occurrences sont particulièrement rares dans le corpus Google, où l'on ne fait référence que trois fois à la qualité et à la précision qui caractérisent le traducteur humain moyen (« average human translators ») : cette moyenne, dont le calcul n'est d'ailleurs jamais expliqué, est une fiction destinée à servir d'étalon pour mesurer les performances du nouveau système neuronal. Dans les deux autres corpus, la référence au traducteur humain est beaucoup plus concrète : on la trouve surtout en position de complément d'agent dans des tournures passives de type « used by [...] translators » ou de bénéficiaire (« for [...] translators ») dans le corpus de la DGT, et en position sujet dans les articles de recherche en traductologie, qui rapportent les propos de traducteurs interrogés ou décrivent leurs activités.

On retrouve ici la même attention à des communautés de locuteurs que dans mes travaux sur la justice climatique, où j'ai mis en œuvre une démarche semblable pour cerner et décrire la pluralité des discours [Rossi et al. 2019a]. Les analyses dessinent le chemin qui va d'une analyse préliminaire automatisée à une lecture minutieuse et située des textes, soucieuse de rendre compte des choix conceptuels (qui sont aussi des choix énonciatifs). En tant qu'il porte sur le discours environnemental, ce travail rejoint la seconde application décrite ci-dessus, selon

³² L'évaluation de cet article est en cours, il n'a donc pas été intégré au volume de publications mais il pourra être mis à disposition du jury.

³³ La catégorisation est expliquée ici : <http://ucrel.lancs.ac.uk/usas/>

laquelle le traducteur doit comprendre la conceptualisation métaphorique pour s'appropriier les domaines de spécialité. La méthodologie mise en œuvre, semblable à celle évoquée ci-dessus, est en effet susceptible de faciliter le repérage de métaphores en vue de leur exploitation dans l'enseignement de la traduction spécialisée. Ainsi par exemple, la catégorie sémantique qui distingue le plus les rapports d'ONG de ceux des organisations onusiennes regroupe différentes expressions d'émotions positives qui vont de l'intérêt au plaisir (il s'agit de la catégorie X5.2+ dans le système USAS). On trouve dans les rapports d'ONG de nombreuses références à l'enthousiasme qui caractérise l'engagement de ces organisations et constituent presque un credo. Ces rapports sont produits localement et leurs traductions ne sont généralement pas mises à disposition du public, ils pourraient donc être utilisés dans un cours de traduction spécialisée. Dans le passage ci-dessous, la traduction du syntagme nominal souligné pose problème, et l'analyse du mode de conceptualisation choisi par les auteurs du rapport peut aider l'enseignant à l'expliquer. On constate en effet que la métaphore fréquente de l'intensité passionnelle comme source de chaleur est remplacée par l'évocation d'une énergie ou d'une force au service d'objectifs ambitieux : la métaphore amoureuse qu'induit le substantif français de passion serait donc à éviter, on pourrait lui préférer l'adjectif pour proposer par exemple « les valeurs et l'engagement passionné de nos équipes ».

We will put our efforts behind ambitious goals that will make a difference on a global scale, [...], while leveraging our unique worldwide Network of locally rooted national organizations, our successful partnerships and collaborative projects, and the values and passion of our people. (WWF International, *Annual Review 2015*).

Parce qu'elle permet de caractériser les modes de conceptualisation, et donc l'accès au sens, l'analyse des métaphores telle qu'elle se pratique en linguistique cognitive (éventuellement élargie à l'ensemble des amalgames conceptuels ou *blends*) peut devenir un outil au service de l'enseignement de la traduction spécialisée. C'est une piste de recherche issue des premières linguistiques cognitives qui reste très actuelle, et constitue un point de rencontre fécond entre linguistique et traductologie, même si elle ne nous donne qu'un accès très indirect aux processus de traduction.

Quant au tournant des grammaires cognitives (Janda 2013), il a permis l'élaboration de nouveaux formalismes comme les grammaires de construction. Bien que basées sur l'usage, ces nouvelles grammaires produisent le même effet que les précédents propositions d'analyse formelle : elles nous éloignent de l'activité traduisante comme de ses produits. Les difficultés qu'elles rencontrent pour intégrer les dimensions pragmatiques et interactionnelles de la construction du sens montrent qu'elles n'ont pour l'instant que peu de prise sur ce qui intéresse vraiment la traductologie :

Tout en reconnaissant l'existence de phénomènes strictement contextuels et interactionnels, les grammaires de constructions s'astreignent à faire passer le maximum d'information pragmatique sous les fourches caudines de l'interface syntaxe-sémantique. [...] Quelles que soient les raisons qui justifient ou non une séparation (théorique ou empirique) de la sémantique et de la pragmatique, il n'en demeure pas moins que la composante pragmatique des grammaires de constructions pose problème. (Desagulier 2011 : 105)

Mes usages de tels modèles sont pour l'instant limités, et s'appliquent davantage à l'acquisition de l'anglais L2 qu'à l'enseignement de la traduction. Une proposition d'application à l'enseignement de la traduction spécialisée a été faite [Rossi et al 2016], mais elle concerne la traduction vers l'anglais médical. Ce travail est davantage lié à une réflexion sur la difficulté que peut induire, lors de l'appréhension d'une langue seconde, la présence d'une construction syncrétique dans la langue maternelle : j'y reviendrai donc dans la partie suivante (voir ci-dessous, Deuxième partie, 3.3). Je n'aborde pas non plus ici les linguistiques de corpus, que je considère comme un ensemble de méthodes, c'est-à-dire comme un « paradigme méthodologique » (Gries 2006 : 191) qui s'est développé au sein des théories sur l'usage, plutôt que comme une nouvelle « approche théorique » (Teubert, 2005 : 2). La rencontre des méthodes de la linguistique de corpus avec la traductologie seront donc présentées en lien avec le caractère ouvert et interdisciplinaire de cette dernière, et permettront de caractériser les évolutions récentes de la discipline (voir ci-dessous, 2.3).

Au terme de ce parcours des deux points d'ancrage linguistiques de mes travaux, j'espère avoir montré que les affinités entre ces deux grands ensembles théoriques et la traductologie ont conduit à des recherches fécondes et constituent encore des pistes prometteuses. Ces affinités sont principalement basées sur l'analyse du sens, que la TOE et les linguistiques cognitives abordent de façon différente. Cependant, qu'il s'agisse de l'invariant, des métaphores ou des amalgames conceptuels, les modèles proposés ne sont que rarement utilisés pour comprendre les processus et compétences à l'œuvre dans la traduction. Mon intérêt pour ces deux aspects est issu de ma formation de psycholinguiste, et je souligne dans ce qui suit quelques-uns des nombreux points de rencontre entre psycholinguistique et traductologie.

2.2 De la psycholinguistique développementale à la traductologie

On trouve dans un article déjà ancien du psychosémioticien et traducteur italien Giuseppe Mininni³⁴ une discussion des liens qui unissent « les intérêts d'une science psycholinguistique et ceux d'une théorie de la traduction » (Mininni 2007 : 19), et l'auteur y défend l'idée que la psycholinguistique « finira par occuper un poste de premier plan dans le projet de construction d'une « traductologie ». » (*ibid.*, p. 20). L'argumentation développée dans cet article est sinieuse, elle décrit bien la diversité des approches psycholinguistiques et surtout l'emprise, encore forte à l'époque, du cognitivisme. C'est un point important dès les débuts de la psycholinguistique, qui s'intéresse à la traduction comme problème posé aux théories de l'information. Les modèles issus de ces théories veulent atteindre un haut niveau de généralisation pour s'appliquer à l'homme comme aux machines. Ils ne s'intéressent donc pas directement aux spécificités de la traduction humaine, comme le montrent bien ces quelques lignes extraites de l'ouvrage fondateur d'Osgood et Sebeok :

On considère généralement la traduction comme un outil, et ses améliorations comme le fruit de l'expérience et du discernement des traducteurs. Avec le développement de techniques issues des théories de l'information, et celui des calculateurs électroniques³⁵, on s'intéresse de plus en plus aux problèmes théoriques et pratiques que pose la traduction, et il apparaît que l'analyse de la traduction comme processus psycholinguistique pourra être utile [...]. La théorie de l'information nous fournit un modèle général où le traducteur, qu'il soit homme ou machine, est assimilé à un canal, la langue source à l'input, et la traduction, ou langue cible, à l'output. (Osgood et Sebeok 1965 : 185, notre traduction³⁶).

La métaphore du conduit (voir Reddy 1979 pour une description détaillée, que reprennent Lakoff et Johnson dans leur ouvrage de 1980) laisse peu de place à la description de la traduction comme activité humaine complexe, et suggère qu'elle soit réductible à un processus d'encodage-décodage dont le traducteur (homme ou machine) serait le siège. Même si le succès actuel de la TA

³⁴ Mininni, G. (1981). On the need of psycholinguistic research on translation. *International Journal of Psycholinguistics*, 8(2, 22), 79-107. Cet article a été traduit plus récemment par Claudia Cortesi (Mininni 2007), et c'est la version française que j'ai consultée.

³⁵ J'ai préféré ce terme à celui d'« ordinateur », pourtant introduit en 1955 dans la langue française, pour conserver la perspective historique : il me semble que l'on évoque ici les progrès d'un type d'objet technologique, et qu'il importe de conserver la modification adjectivale qui précise le domaine d'application.

³⁶ "Translation is usually regarded as a tool, which is improved only by experience and good judgment on the part of translators. The development of information theory techniques and electronic computers has created increasing interest in both theoretical and practical problems of translation, and it may be worthwhile to analyse translation as a psycholinguistic process (see also section 6.2). Information theory provides a general model in which the translator, whether human or machine, is treated as a channel, the foreign or 'from' language (FL) is the input, and the translation or 'to' language (TL) is the output." (Osgood et Sebeok 1965 : 185).

neuronale remet certains de ces modèles au goût du jour, il me semble que pour comprendre la convergence récente de travaux en psycholinguistique et en traductologie, il faut tenir compte d'une part des évolutions historiques des modèles psycholinguistiques (qui ont fait la part belle au behaviorisme de Skinner puis au cognitivisme sous influence chomskienne, avant de se diversifier), et d'autre part des spécificités de la psycholinguistique développementale que je vais tenter de caractériser brièvement.

Situation(s) de la psycholinguistique développementale

La psycholinguistique développementale, dont le but est de décrire et d'expliquer le développement du langage chez l'enfant, définit le domaine d'inscription principal de mes travaux sur l'acquisition du langage (rassemblés dans la première partie de mon volume de publications). Elle intègre les acquis de la psychologie du développement et se situe dans une tradition interactionniste, du moins dans son acception la plus courante³⁷, et donc hors du champ d'application d'une grammaire universelle. C'est un premier point de convergence avec une traductologie soucieuse d'expliquer les processus à l'œuvre dans tout échange ou communication réussis, et Mininni ne s'y était pas trompé puisqu'il situait déjà la réflexion dans une « psycholinguistique post-chomskienne » (2007 : 42). La question du statut des règles et de ce que Mininni appelle la « réalité psychologique de la grammaire » (*ibid.*) n'est pas pour autant exclue de ce champ d'investigation : on peut notamment supposer que les règles rendent tangibles des tendances ou permettent de cerner des invariants qui à leur tour, pourraient permettre d'éclairer certains aspects des processus de traduction. Je reviendrai sur cette question (voir ci-dessous, Deuxième partie, 3.2), qui concerne davantage l'enseignement que la pratique de la traduction, et probablement aussi davantage l'acquisition de langues secondes que celle de langues maternelles.

La psycholinguistique développementale, en tant qu'elle s'intéresse à l'acquisition d'une ou plusieurs langues maternelles, occupe une position périphérique dans la théorie psycholinguistique qui s'intéresse surtout aux processus de production et de compréhension du langage chez l'adulte sain (François et Cordier 2006 : §37 et figure 9). Cette différence d'objet, ou de sujet(s) étudiés s'accompagne d'une différence dans les méthodes mises en œuvre, qui ne sont pas d'abord expérimentales, mais privilégient l'étude de cas à partir de journaux de bord tenus par des scientifiques qui souvent observaient leurs propres enfants (Morgenstern et Parisse 2010), puis de suivis longitudinaux réalisés au domicile de l'enfant. L'accent mis sur les dimensions fonctionnelles et interactionnelles du langage découle donc probablement au moins

³⁷ C'est-à-dire aussi dans l'enseignement et la recherche universitaire de langue française

autant de la nature mouvante de l'objet étudié que des méthodes adoptées. La mise en place de dispositifs expérimentaux toujours moins invasifs et de plus en plus adaptés aux nourrissons et aux jeunes enfants a transformé les procédures de recueil de données, mais on y retrouve toujours une attention au contexte et aux interactions (cf. supra, 1.3). Les approches fonctionnelles qui s'illustrent notamment dans les travaux d'E. Bates et de B. MacWhinney ont beaucoup bénéficié de ces nouveaux dispositifs (MacWhinney 2014 : 129) et ont mis en évidence un ensemble de principes d'apprentissage (apprentissage basé sur des items avec des effets de fréquence et de prototypes, par exemple) et de traitement (notamment la compétition, que l'on observe dans le traitement monolingue comme multilingue) en lien avec des principes neurolinguistiques comme la plasticité ou la spécialisation, et avec des principes interactionnels comme l'étaiyage ou le partage d'attention (ibid., p. 124-127).

Le rôle central des approches fonctionnelles rapproche la psycholinguistique développementale d'un ensemble de théories traductologiques qui fondent leur compréhension de l'activité traduisante sur la fonction de celle-ci, c'est-à-dire aussi le contexte situationnel dans lequel elle se déroule (voir ci-dessous, 2.3). Le courant fonctionnaliste en linguistique appliquée, comme d'ailleurs les linguistiques cognitives qui en sont issues, a aussi mis l'accent sur la variation d'une langue à l'autre (Kail 2002 : §7) : le psycholinguiste américain Dan I. Slobin a ainsi rassemblé dans cinq volumes successifs des comparaisons inter-langues documentant l'acquisition de presque 40 langues différentes³⁸. Pourtant, dans l'introduction du volume qu'il édite en 1998 pour rendre plus visible ce qu'il appelle « la nouvelle psychologie du langage », rassemblant les approches cognitives et fonctionnelles, M. Tomasello s'étonne de la diffusion encore faible de ces théories en psycholinguistique développementale, alors même qu'elles constituent « une approche psychologique de la langue » (Tomasello 1998 : xiii, notre traduction). C'est probablement ce que les travaux de ces quinze dernières années ont permis d'accomplir : le bilan publié à l'occasion des quarante ans de la revue *Journal of Child Language* (vol. 41(S1) 2014) montre que l'intégration des théories fonctionnelles et cognitives est non seulement effective, mais qu'elle a permis de consolider les découvertes.

De l'acquisition du langage aux approches cognitives de la traduction

Alors que les travaux sur l'acquisition précoce d'une langue maternelle (L1) appliquent principalement les approches cognitives à l'étude de la conceptualisation et de la catégorisation (voir par ex. Tomasello 2007 : 1093 et sq.), l'étude des processus cognitifs à l'œuvre dans la compréhension et dans la production des langues secondes (L2) regroupe un ensemble très large

³⁸ La série s'intitule *The cross-linguistic study of language acquisition*, avec un premier volume publié en 1985 et le dernier, au sous-titre évocateur (*Expanding the contexts*) paru en 1997, voir aussi Kail, *ibid.*

de travaux qui rejoignent les approches cognitives de la traduction. Si j'ai depuis quelques années orienté mes travaux vers l'acquisition d'une langue seconde, c'est non seulement parce qu'il m'est donné d'observer de très nombreux apprenants de l'anglais dans le cadre de mon enseignement, mais aussi et surtout parce qu'il existe une complémentarité forte entre l'étude de l'acquisition des langues maternelles et secondes, et que la comparaison est toujours éclairante. Cette comparaison est implicite dans mes travaux, mais un volume de la revue *Acquisition et interaction en langue étrangère*³⁹ lui a été consacré et les contributions montrent bien que c'est en mettant en regard différents types d'apprenants et d'apprentissages que l'on peut découvrir et analyser « des invariants linguistiques et/ou cognitifs dans les processus rapprochés » (Benazzo 2009 : §4). Outre les similitudes et l'intérêt certain que présente l'observation de locuteurs en classe de langue pour s'interroger sur l'importance de l'usage et mettre à l'épreuve les modèles issus des théories UB (ainsi que je l'expliquerai dans la deuxième partie), je distingue en particulier deux objets théoriques que l'acquisition des langues secondes permet d'étudier de façon privilégiée. Il s'agit, d'abord, de la complexité syntaxique : les phrases complexes ne constituent pas un objet d'étude fréquent en acquisition L1, et c'est surtout leur compréhension qui a intéressé les chercheurs (avec des études en situation expérimentale), si bien que très peu de travaux analysent des données longitudinales (Diessel 2004 : 2 ; Sekali 2012). D'autre part, c'est en étudiant l'acquisition des langues secondes que l'on peut espérer circonscrire l'impact de la langue maternelle : celle-ci constitue un « avantage cognitif » pour les apprenants qui disposent d'un premier ensemble de correspondances forme-sens (Chenu et Jisa 2009a : 37), mais aussi un obstacle à surmonter pour appréhender les différences inter-langues.

Cette question n'est évidemment pas sans lien avec la traduction, mais la place que pourrait occuper une réflexion sur l'acquisition des langues secondes en traductologie n'est pas bien définie, et a été amplement débattue sur deux fronts au moins. Le premier concerne l'usage de la traduction en classe de langue : longtemps polémique⁴⁰, notamment en raison du rejet de la langue maternelle hors du cours de langue dans les méthodes audio-orales et communicatives, il s'agit d'un usage qui reste important dans le contexte universitaire français. Cette permanence est en partie liée à la préparation des concours de recrutement des enseignants qui font de l'enseignement du thème et de la version un passage obligé. Les arguments des défenseurs de l'utilisation de la traduction dans l'enseignement des langues étaient cependant tout autres (voir par ex. Ladmiral 1979, ou Lavault 1985), et les chercheurs qui recommencent, depuis peu, à s'intéresser à ces questions leur donnent raison. En effet, s'il est indéniable que les travaux qui ont

³⁹ 2009(1) : il s'agissait du premier numéro de la revue AILE/LIA qui a succédé à AILE cette année-là.

⁴⁰ Cet usage était officiellement interdit en France par des directives imposant la « méthode directe », mais il semble qu'elles n'aient pas fait l'objet d'une application stricte (voir notamment M. Harvey 1996 : 46)

permis de mieux comprendre le bilinguisme, et de décrire le cerveau bilingue en évitant la métaphore du double contenant étanche, ont joué un rôle dans ce nouveau revirement, il s'agit avant tout d'un « tournant traductologique » (Carreres et al. 2017) qui nous éloigne de la très traditionnelle méthode de grammaire-traduction en proposant de faire de la traduction une pratique située, contextualisée, c'est-à-dire en tenant compte de l'ensemble des dimensions communicatives à l'œuvre dans la traduction. De ce point de vue, il me semble que nous commençons tout juste à mesurer l'apport d'une analyse des pratiques professionnelles de traduction, et des théories s'y rapportant⁴¹, à l'enseignement des langues secondes.

Le second front se situe au sein de la traductologie elle-même, et sera abordé dans la partie suivante : il concerne en effet la place des langues secondes dans l'enseignement de la traduction et dans la formation des traducteurs.

Je n'ai fait ici qu'illustrer quelques-uns des points de passage entre psycholinguistique et traductologie, mais on constate que dans l'ensemble, ils s'articulent autour de deux aspects complémentaires de la recherche en psycholinguistique : une attention aux processus cognitifs d'une part, et de l'autre une préférence pour les théories fonctionnelles et basées sur l'usage. Ces deux aspects sont indissociables et trouvent leur origine dans la contestation de l'innéisme chomskyen. Plutôt que de postuler des règles et structures innées, le fonctionnalisme, les linguistiques cognitives et l'ensemble des théories basées sur l'usage cherchent en effet à proposer des modèles qui tiennent compte du fonctionnement de la cognition humaine. Il n'est donc pas étonnant que la traductologie actuelle rencontre à nouveau l'enseignement des langues, et que cette rencontre fasse fond sur l'étude de la traduction comme pratique régie par un ensemble de dimensions fonctionnelles, en même temps que comme activité mettant en jeu des processus cognitifs complexes.

⁴¹ Je pense évidemment aux travaux pionniers d'E. Lavault-Olléon, qui a proposé d'appliquer la théorie interprétative de la traduction à l'enseignement, et dont les propositions concrètes m'ont guidée dans ma propre pratique (voir par ex. Lavault-Olléon 1994).

2.3 La traductologie à la croisée des chemins

Il ne s'agit certes pas de proposer une liste complète des affinités et autres ressemblances entre psycholinguistique et traductologie, mais s'il est un aspect de la traductologie qui la rapproche de la psycholinguistique, c'est sa nature « polydisciplinaire » (Edgar Morin 1986, cité par Yves Gambier in Duarte et al. 2006 : 36). Snell-Hornby, qui a proposé dès la fin des années 1980 une approche intégrée de la traductologie (Snell-Hornby 1988) et fait de l'interdisciplinarité une des caractéristiques fondatrices de celle-ci. La notion d'« interdiscipline » a été maintes fois reprise depuis, mais elle a probablement suscité autant de craintes que d'enthousiasme : absence d'une identité disciplinaire suffisamment forte, craintes de morcellement du fait du « socle épistémologique relativement pauvre » (Gambier et van Doorslaer 2016 : 4, notre traduction) qui caractériserait la traductologie. Je lui préfère l'optimisme de Snell-Hornby et sa conception holistique d'un tout traductologique qui constitue bien plus que la somme de ses parties :

Je ne conçois pas une "interdiscipline" comme une sorte de *no-man's land* que seules définiraient les frontières bien arrêtées des territoires qui l'entourent. C'est bien plutôt grâce à la variété de ses thématiques, de ses méthodes et de leurs alignements successifs, qu'elle évolue pour devenir une entité dont les qualités ne correspondent pas à celles de constituants d'origine. (Snell-Hornby 2006 : 72, notre traduction⁴²).

Quelles sont les implications de cette inter- ou poly-disciplinarité qui situe la traductologie à la croisée des chemins ? Je ne considère ici que les évolutions récentes que j'ai rencontrées, qui ne constituent que quelques pièces du puzzle, mais où je retrouve plusieurs exemples d'une interdisciplinarité féconde. A l'inverse, il me semble qu'en voulant à toute force se faire discipline, la traductologie risque de se priver des perspectives interdisciplinaires qui pourtant la définissent si bien. Le premier « Congrès Mondial » de traductologie qui a eu lieu à Paris en avril 2017 était ainsi un événement d'une grande richesse, mais il a aussi montré la difficulté de faire dialoguer les approches au sein d'une communauté si nombreuse et diverse. J'espère fournir dans ce qui suit quelques éléments justifiant que l'on continue, en traductologie comme en linguistique, de prendre le « risque de l'interdisciplinarité » (Chevrot 2016a).

⁴² [A]n "interdiscipline" is in my view not merely a kind of no-man's land between other clearly defined territories. We could rather say that through the varying constellations of its topics and methods, it evolves into something qualitatively different from the ingredients of which it originally consisted.

Partons d'un constat sans appel, et d'un tournant de la traductologie qui est tout à fait central pour la réflexion que je présente ici : « l'époque du 'tout linguistique' de la traductologie est bel et bien révolue » (Lejeune, in Boisseau et al. 2016 : 71). La linguistique n'est effectivement plus à l'honneur en traductologie, pas plus d'ailleurs que ses modèles ne s'intègrent aux nouveaux systèmes de traduction automatique neuronaux qui ont tout à fait renoncé à l'idée d'ajouter des règles au traitement issu de l'apprentissage profond (*deep learning*). Cependant, si les nouveaux modèles traductologiques sont davantage basés sur la communication et sur la cognition (Lejeune, *ibid.*), ils ne se situent pas pour autant aux antipodes des théories linguistiques contemporaines. Les modèles communicationnels constituent plutôt ce que Pergnier (1980) a nommé les « fondements sociolinguistiques de la traduction »⁴³, et les modèles cognitifs sont au cœur d'une traductologie empirique naissante, qui se situe au carrefour de la psycholinguistique, des sciences cognitives (Gries et Wulff 2012), des sciences de la communication, pour ne citer que les principales (Orero et al. 2018 : 106).

L'avènement récent d'une traductologie de corpus, en agrégeant à la traductologie l'objet corpus, opère un nouveau rapprochement puisqu'elle est « le fruit de la rencontre entre la linguistique de corpus et la traductologie descriptive » (Loock 2016 : 47). Il s'agit bien, en utilisant les « outils et méthodes de la linguistique de corpus [...] de décrire et d'analyser les textes traduits » (*ibid.*). Tout en remettant la linguistique textuelle au cœur des préoccupations du traductologue, ce courant promet, depuis le texte fondateur de M. Baker, de « nous permettre d'identifier les caractéristiques des textes traduits qui nous aideront à comprendre la nature du traduire⁴⁴ et son fonctionnement » (Baker 1993 : 242-243, notre traduction⁴⁵). Le fantasme que l'on voit poindre ici est bien celui d'atteindre, grâce à l'accumulation des données, un très haut niveau de généralité qui confine à l'universel : les caractéristiques évoquées ci-dessus sont ainsi des « caractéristiques universelles de traduction » (*universal features of translation*), ou universaux de traduction. Sans revenir sur les critiques nombreuses qui ont été adressées à une telle proposition, on constatera simplement que les critères d'objectivité et de scientificité qu'elle véhicule sont assez peu adaptés à l'analyse du traduire, qui aurait beaucoup à perdre à faire disparaître la subjectivité du traducteur. Il n'est d'ailleurs pas surprenant que la traductologie de corpus se trouve confrontée à l'un des paradoxes des linguistiques UB, dont elle utilise les outils

⁴³ L'actualité de cet ouvrage, qui a été réédité en 2017, ne fait aucun doute.

⁴⁴ Je reprends à dessein une expression de Lance Hewson (2016), pour éviter le substantif qui désignerait plus volontiers le produit que le processus.

⁴⁵ "The profound effect that corpora will have on translation studies, in my view, will be a consequence of their enabling us to identify features of translated texts which will help us understand what translation is and how it works."

et méthodes : « Pour que la linguistique soit véritablement fondée sur l'usage, il faut procéder à un aller-retour fécond entre analyse intuitive et méthodes quantitatives en ne perdant pas de vue que l'intuition qui nous intéresse est en définitive celle du locuteur » (Desagulier 2015 : 124). Ce que nous font perdre de vue les tendances générales qu'une traductologie de corpus cherche à cerner⁴⁶, c'est le point de vue et l'intuition du locuteur, qui n'est autre que le traducteur. Or la créativité du traducteur est justement ce qui échappe aux données de corpus : elle se situe précisément au-delà des « limites externes » de tels outils (Kübler 2014). Les tendances observées, si elles sont présentées hors contexte, nous informent sur la langue, qui peut être l'objet d'une théorie linguistique ou permettre de formuler des propositions didactiques, mais elles ne servent finalement qu'à « déblayer les textes » afin de permettre au traducteur « de concentrer [son] attention [...] sur les points véritablement cruciaux. » (Froeliger 2013 : 123-124).

La question de ce qu'une traductologie de corpus peut apporter à l'enseignement est sensiblement différente. Pourtant, elle nous ramène inévitablement à la langue. Les corpus sont autant d'outils qui donnent accès à des régularités observables, et ces régularités, si elles sont présentées avec suffisamment de contexte, sont susceptibles d'être utilisées dans l'enseignement des L2 et de la traduction inverse (ou traduction vers la langue B). C'est ce que j'ai proposé sur la base d'une étude contrastive et basée sur corpus de certains syntagmes nominaux complexes en anglais médical [Rossi et al. 2016]. On n'y trouve cependant que des propositions d'application élaborées collectivement et que je n'ai pas mises en œuvre, et je n'y aborde pas les problèmes liés au statut et de la place des langues secondes dans l'enseignement de la traduction. Ce sont ces problèmes qui définissent le second front des débats que je n'ai fait qu'évoquer précédemment, et je souhaite les expliquer brièvement ici afin de donner un exemple de croisement incomplet des théories et des observations qui se situent dans des champs distincts.

L'enseignement de la traduction suppose une compétence bilingue, mais les liens entre bilinguisme et compétence de traduction, qui ont été abordés dans des publications déjà anciennes⁴⁷, ne sont pas encore connus et décrits avec précision : « Seules des études expérimentales portant sur des traducteurs en action fourniront des éclaircissements sur la façon dont la traduction dépend du système lexico-sémantique bilingue et le modifie à son tour. » (Shreve 2010, traduction française de D. Rocher, 2016⁴⁸). Si ces questions n'intéressent que partiellement la traductologie, c'est probablement parce que la distinction entre la traduction

⁴⁶ Berman parle ainsi des "tendances déformantes" d'une traduction qui cherche à « acclimater » le texte source à la culture cible (Loock 2016 : 185)

⁴⁷ Il s'agit notamment des travaux de Brian Harris, théoricien de la « traduction naturelle », qui datent de la fin des années 1970.

⁴⁸ Cette traduction a été réalisée par un étudiant du Master TSM de l'Université Grenoble Alpes sous ma direction.

naturelle (celle que peut pratiquer un sujet bilingue spontanément) et la traduction professionnelle définit une ligne de faille qui place la traduction naturelle hors du champ de la traductologie⁴⁹. Au-delà des enjeux disciplinaires, c'est une question importante pour ce que Holmes appelait la traductologie appliquée⁵⁰, car elle est susceptible d'aider les formateurs à mieux définir le rôle et la place de la traduction vers la langue B, mais aussi les interférences possibles avec la TA si, comme beaucoup d'enseignants le craignent actuellement, celle-ci venait effectivement se substituer (ne serait-ce que partiellement) à la compétence de langue B.

Le contexte au cœur de la traductologie pragmatique

Essayons pour finir de caractériser la traductologie dans sa transversalité, en montrant qu'une telle démarche ne nous fait pas nécessairement perdre de vue ses objets. La notion de contexte peut nous y aider. J'en ai montré l'importance centrale pour les approches fonctionnelles (1.3), et j'ai esquissé ci-dessus (2.2) un parallèle quelque peu hâtif entre le fonctionnalisme tel qu'il s'est illustré en psycholinguistique développementale et les courants fonctionnalistes de la traductologie. Encore faut-il préciser que la traduction se situe toujours dans un après-coup (ses circonstances ne coïncident pas avec celles de la production du texte), et en retrait aussi vis-à-vis de la réception du texte traduit. Les dimensions spatiales, mais surtout temporelles de la traduction la placent hors du contexte situationnel, et la traduction opère une re-contextualisation (House 2006 : 343). Ce décalage n'empêche pas que s'instaure une relation dynamique entre le contexte et le langage, car si le texte à traduire fige cette relation, il appartient au traducteur de la remettre en jeu. House ancre ces remarques dans la perspective de la linguistique systémique fonctionnelle de Halliday, qu'elle adapte à la traduction en empruntant à l'analyse du discours et à la pragmatique (ibid. p. 344). Dans l'approche fonctionnaliste de Vermeer, c'est le *skopos*, c'est-à-dire le but de la traduction, et donc aussi la définition de la « mission » du traducteur (Lavault-Olléon 2006 : 511) qui permet de mesurer l'éventuel décalage lié à cette re-contextualisation, en définissant l'écart qui peut exister entre le contexte de production du texte source et le *skopos* qui régira la production du texte cible. Il est remarquable en tout cas que les travaux qui mettent en œuvre une démarche fonctionnaliste se penchent sur des cas concrets et proposent l'analyse détaillée d'un projet de traduction en contexte, ou mettent en regard deux projets distincts : les corpus parallèles qu'elles utilisent sont donc, du point de vue de la linguistique de corpus, de petite taille. R. Krüger (2016) étudie par exemple le recours à l'explicitation en traduction scientifique et technique dans un corpus de 27 000 mots traduits depuis l'allemand, répartis en deux sous-

⁴⁹ Cette distinction recouvre d'ailleurs partiellement celle entre les approches que l'on pourrait appeler purement linguistiques, ou formelles, et celles qui se basent sur le sens.

⁵⁰ En l'occurrence, il s'agit de la traductologie appliquée à la formation des traducteurs (voir Lavault-Olléon 2018, pour une adaptation française expliquée de la cartographie de Holmes)

corpus distincts. Ils procèdent donc exactement comme les linguistiques contextualistes évoquées plus haut (1.3).

Outre les approches fonctionnalistes, la théorie interprétative de la traduction, dont on connaît le rejet de l'analyse de la « texture dont est faite la langue »⁵¹, retrouve en de nombreux endroits la notion de contexte en raison de son intérêt pour l'implicite (Froeliger 2017 : 37) -- preuve s'il en fallait de ce que l'opposition entre linguistique et traductologie s'apparente plutôt à un refus des formalismes linguistiques. Si l'on tient compte aussi de l'importance du contexte au sein des approches culturelles, on peut même esquisser une chronologie très générale au sein de laquelle le contexte permet de définir un tournant de la traductologie : « On est passé d'une conception purement textuelle à des études contextuelles, pour arriver aux travaux actuels qui s'intéressent au rôle agentif du traducteur dans son environnement. » (Alonso et Calvo 2015 : 142, notre traduction⁵²). La traductologie pragmatique dans laquelle s'inscrivent mes travaux récents et en cours (voir notamment [Rossi 2019] et [Rossi et Chevrot 2019] opère une synthèse intéressante de ces tendances, puisqu'elle intègre les acquis de la théorie interprétative et des approches fonctionnalistes, et a su aussi prendre acte des apports de l'ergonomie, qui place le « facteur humain », et donc le traducteur au centre de ses préoccupations (Lavault-Olléon 2011 : §19). Il revient donc à la traductologie pragmatique d'avoir permis, en se fondant sur des points de convergence théoriques pluridisciplinaires, de déplacer le curseur et de montrer l'importance de considérer tout « l'écosystème et l'environnement dans lequel la traduction trouve son origine » (Alonso et Calvo, *ibid.*, notre traduction⁵³). Dans ce cadre, l'étude de la traduction du point de vue de ses processus se fait sur la base d'une compréhension incarnée et située de la cognition (*situated and embodied cognition*), qui ne se réduit plus à l'activité cérébrale d'un sujet isolé. C'est ce que Risku et Windhager (2013) ont appelé « un programme de recherche socio-cognitif » sur la traduction, et il présuppose une conception élargie de la cognition.

2.4 Elargir le champ de la cognition

Une définition très générale de la cognition peut s'énoncer ainsi : « La cognition regroupe un ensemble de phénomènes qui se rapportent à l'esprit humain et à son fonctionnement » (Andler et al. 2018 : 9). La cognition définit notre faculté de connaître au sens large, et la réflexion sur celle-ci s'inscrit aujourd'hui dans un champ d'investigation non seulement étendu mais qui

⁵¹ « L'objet de la traduction n'est pas la texture dont est faite la langue mais le sens qu'y trouve celui auquel elle s'adresse » (Seleskovitch 1976 : 9, in Lavault-Olléon 2018).

⁵² "The focus has shifted from purely textual studies to studies of context and now studies that focus on the translator as an active agent in their environment." (Alonso et Calvo 2015 : 142).

⁵³ "However, the most pragmatic theories of recent decades are starting to pay attention to the ecosystem and environment in which the translation originates." (Alonso et Calvo 2015 : 142).

semble encore appelé à croître. Les sciences cognitives convoquent en effet un ensemble de synergies pluridisciplinaires : elles sont au carrefour de traditions épistémologiques différentes, qui vont de l'informatique à l'anthropologie, ou encore des neurosciences à la linguistique et à la philosophie. Elles constituent ainsi un observatoire privilégié pour toutes les questions de méthodologie de la recherche comme pour la réflexion sur les impacts de celle-ci (cf. ci-dessous Synthèse, 3.4).

Si je m'intéresse à la cognition, ce n'est donc pas au sens où l'entendaient les « cognitivistes » classiques (une conception qu'illustre bien la théorie computationnelle de l'esprit de Fodor, laquelle se fonde sur une vision chomskyenne de l'acquisition du langage) et peut-être faut-il considérer quelques-uns des élargissements successifs pour expliquer ce que cela implique. Ce sont d'abord les divergences de vue qu'ont exprimé Lakoff ou Langacker en s'opposant aux présupposés objectivistes du « programme cognitiviste » chomskyen, qui ont permis de créer la linguistique cognitive telle que nous la connaissons. Ce tournant se caractérise par la sortie de l'analogie cerveau-ordinateur comme de l'ensemble des modèles cognitivistes classiques, et par la volonté d'ouvrir la cognition à l'environnement physique et social. Le défi que souhaitent relever les premières linguistiques cognitives correspond à celui que rencontrent les sciences cognitives de l'époque, comme le montrent les douze problèmes relevés par D. A. Norman (1980) :

L'être humain est un organisme vivant, doué de fondements biologiques et d'une histoire évolutionniste et culturelle. C'est aussi un animal social, qui interagit avec ses semblables, avec son environnement et avec lui-même. Les disciplines fondatrices des sciences cognitives ont généralement laissé de côté ces aspects comportementaux. Il en résulte une théorie dans l'ensemble stérile, en dépit des progrès considérables qui ont pu être accomplis dans certains domaines [...] (Norman 1980 : 4, notre traduction⁵⁴).

Les linguistiques cognitives cherchent ainsi à cerner des processus, ou modes de fonctionnement de l'esprit humain à même de rendre compte de l'organisation de nos connaissances, telle qu'elle s'illustre dans la communication langagière. Ce que Lakoff appelle des « modèles cognitifs idéalisés » (*Idealized Cognitive Models*) correspond ainsi à des modes de structuration holistiques de notre expérience (également appelés *gestalts expérientielles*), qui

⁵⁴ Je restitue ici la citation complète : « The human is an animate organism, with a biological basis and an evolutionary and cultural history. Moreover, the human is a social animal, interacting with others, with the environment, and with itself. The core disciplines of cognitive science have tended to ignore these aspects of behavior. The results have been considerable progress on some fronts, but sterility overall, for the organism we are analyzing is conceived as pure intellect, communicating with one another in logical dialogue, perceiving, remembering, thinking where appropriate, reasoning its way through the well-formed problems that are encountered in the day. Alas, that description does not fit actual behavior. » (Norman 1980 : 4).

lorsqu'elles sont transposées pour permettre la compréhension d'un autre domaine produisent ce que Lakoff et Johnson ont appelé des métaphores conceptuelles. Ce sont ces processus qui intéressent les linguistiques cognitives, bien d'avantage que des règles abstraites susceptibles de régir des unités symboliques :

Pour ces nouveaux courants de linguistique cognitive, qui s'intéressent principalement à l'aspect dynamique des langues et du langage, le « noyau dur » réside moins dans les règles de grammaire que dans les opérations de construction de la signification : d'où l'importance accordée à des phénomènes comme la polysémie ou la métaphore, par exemple. (Fuchs 2017 : 12)

En traductologie, les approches cognitives sont d'abord attachées à ouvrir la « boîte noire » du traducteur en utilisant les protocoles de verbalisation, l'enregistrement de l'activité (clavier et écran), l'oculométrie puis de nouvelles techniques d'imagerie cérébrale. Les synergies pluridisciplinaires qui existaient déjà à la fin des années 1990 (en témoigne notamment l'ouvrage de Danks, Shreve, Fountain, & McBeath 1997) sont accentuées par la tendance actuelle des travaux empiriques à mettre en œuvre des méthodes plurielles pour « rendre compte de la complexité de leur objet de recherche » (Ehrensberger-Dow et Englund Dimitrova 2018 : 13). Cependant, le tournant majeur de ces dernières années est l'importance qu'acquiert actuellement la cognition sociale, qui apparaît comme le nouveau creuset des approches cognitives. Les auteures de l'ouvrage précité expliquent ainsi qu'elles ont souhaité rééditer les contributions précédemment publiées dans la revue *Translation Spaces* (5.1) avec l'espoir d'encourager ainsi « de nombreux autres chercheurs à explorer les dynamiques qui se situent à l'interface des dimensions cognitives et situationnelles de la traduction et de l'interprétation » (*ibid.*, notre traduction).

2.5 Les linguistiques appliquées « à travers champs »

Les différents points abordés dans cette première partie sont tous transversaux, et les nombreuses traversées que j'ai proposé d'y effectuer empêcheront peut-être la lectrice ou le lecteur de ces lignes de se forger une vision globale du projet qui sous-tend ce parcours. Les propositions théoriques que j'aborde sont effectivement nombreuses, mais chacune élabore à sa façon une conception des faits de langue et de l'énonciation qui n'efface pas ni locuteurs ni l'ensemble des dimensions situationnelles et co(n)textuelles susceptibles d'en rendre compte. Ce sont donc aussi des propositions soucieuses d'intégrer les dimensions sociales du langage et de la cognition, pour mieux expliquer la variabilité observée.

La question que les réflexions épistémologiques proposées dans ce qui précède laissent en suspens est celle de l'impact des explications qui ont été et pourront être produites. De ce point de vue, les linguistiques appliquées ne désignent pas seulement pour moi des linguistiques de terrain, mais elles correspondent aussi à une pratique engagée. Qu'il s'agisse de l'acquisition du

langage ou de ce que Benjamin appelait « la tâche du traducteur », les recherches que je présente ici sont porteuses de ce « désir de faire rupture » qu'Andrew Linn a décrit comme une motivation centrale des linguistiques appliquées :

La leçon que l'on peut tirer de l'histoire des linguistiques appliquées est que les travaux de recherche qui ont fait rupture étaient d'emblée porteurs du désir de le faire, et qu'ils ne séparaient pas nettement la recherche universitaire du monde réel où les locuteurs utilisent le langage et en font l'expérience (Linn 2011: 25, notre traduction⁵⁵).

Il ne faudrait pas en conclure pour autant qu'une recherche appliquée et engagée se soucie peu de ses fondements théoriques. Il me semble au contraire que le va-et-vient constant entre la construction des observables et la réflexion qui permet de formuler de nouvelles questions de recherche ou de préciser les questions initiales conduit régulièrement celle ou celui qui le pratique à interroger la cohérence des théories qu'il adopte ou évalue. L'intégration de différentes perspectives disciplinaires constitue également un puissant moteur de réflexion, et implique un engagement qui n'est pas moindre : une étude bibliométrique récente a en effet montré que l'interdisciplinarité avait un coût à court terme pour les chercheurs, dont les articles sont moins cités que la moyenne dans les trois premières années qui suivent leur publication (Van Noorden 2015, cité par Chevrot 2016b).

C'est parce que la réflexion théorique me semble primordiale que je suis engagée dans des projets pluridisciplinaires, et que j'ai rejoint en novembre 2017 l'équipe de direction du Pôle Grenoble Cognition : ces deux aspects seront présentés dans la troisième partie de ce document de synthèse, et on trouvera dans la présentation thématique de mes travaux et projets qui en constitue la deuxième partie différents exemples d'applications ainsi définies.

⁵⁵ "The lesson from the history of applied linguistics is that research makes a difference when the desire to make a difference is built into the research from the outset and where the boundary between university research and the world where language is actually used and experienced is a thin and porous one".

Deuxième partie. Travaux et projets

La première direction de mes travaux a consisté à interroger conjointement la diversité des langues et leur acquisition : j'ai engagé ce questionnement sous l'angle de la typologie des langues et dans une perspective interactionniste et développementale. Je me suis intéressée à l'acquisition de l'expression de l'espace et du mouvement, ainsi qu'à celle de la temporalité future, chez de jeunes enfants de langue maternelle française et anglaise. Le mouvement a retenu mon attention du fait des différences typologiques importantes entre le français et l'anglais, et j'ai travaillé ensuite sur l'expression du futur en anglais, où l'alternance des marqueurs *gonna* et *will* ne recouvre pas non plus celles qui sont observées en français. Ces travaux constituent le socle empirique d'une réflexion sur la relativité linguistique, en même temps que sur les fondements épistémologiques des théories de l'acquisition du langage. Cette première direction constitue en fait une boucle, puisque les analyses que j'y ai effectuées m'ont finalement ramenée à une réflexion sur la catégorisation que j'avais commencée avant ma thèse, en produisant un mémoire de maîtrise sur la catégorisation. La suite récente de cet ensemble de travaux m'a permis d'approfondir la notion d'usage en proposant une méthodologie d'analyse des activités et routines de la vie familiale où l'acquisition précoce du langage s'effectue, et en examinant leur impact sur l'analyse des interactions langagières précoces à partir de suivis longitudinaux. La dimension contrastive est présente puisque la méthodologie proposée est fondée sur l'étude d'interactions parent-enfant en français, anglais et arabe tunisien.

Dans une deuxième série de travaux, je me suis interrogée sur les représentations cognitives qui sont au cœur de l'acquisition d'une langue maternelle et de l'apprentissage d'une langue seconde. Ces travaux poursuivent la réflexion épistémologique esquissée ci-dessus et s'interrogent d'une part sur la pertinence des catégories linguistiques (avec deux publications successives sur la catégorisation noms/verbes dans l'acquisition d'une langue maternelle) et d'autre part sur les formats cognitifs mobilisés par les apprenants d'une langue seconde, lorsqu'ils sont confrontés à la fois aux régularités grammaticales structurant la langue cible et à la verbalisation de règles par l'enseignant de langues. La question centrale de mes travaux récents sur l'apprentissage de l'anglais est de savoir si des apprenants de différents niveaux généralisent les règles qui leur sont données par l'enseignement, ou bien s'ils construisent des patrons probabilistes sur la base d'un apprentissage statistiques des usages auxquels ils sont confrontés.

Mes travaux actuels sur la traduction automatique rejoignent cette interrogation sur la pertinence de l'analyse statistique du langage et le pouvoir explicatif de tels modèles. En construisant un cours sur la traduction automatique pour les étudiants du Master de Traduction

spécialisée multilingue de l'Université Grenoble Alpes, j'ai engagé une réflexion sur le traducteur et ses outils dont la présentation viendra clore ce parcours thématique.

1. Catégorisation et typologies

On compare souvent les catégories à des cases et la métaphore du contenant nous conduit à voir les éléments qui s'y rattachent comme inclus dans un domaine plutôt qu'unis par des liens complexes. Dans un ouvrage qui a fait date, G. Lakoff propose que la catégorisation soit bien plus qu'une opération abstraite, assimilable au rangement de parties du réel dans des cases ; il y voit une caractéristique fondamentale de l'esprit humain :

On aurait tort de considérer que la catégorisation est un problème trivial. Nos pensées, nos perceptions, nos actions et nos paroles ne se fondent pas sur autre chose. A chaque fois que nous percevons quelque chose comme un *type* de chose, par exemple un arbre, nous catégorisons ; et dès lors que nous raisonnons sur des *types* de choses (qu'il s'agisse de chaises, de nations, de maladies, d'émotions ou de quoi que ce soit d'autre) nous nous servons de catégories. (Lakoff 1987 : 5-6, notre traduction⁵⁶).

Semblable en ceci à la schématisation ou à la conceptualisation métaphorique, la catégorisation est constitutive de l'accès au sens comme de sa production parce qu'elle « permet de passer de l'expérience corporelle structurée à l'élaboration de modèles abstraits. » (Vandaele 2007 : 134). Je commencerai par présenter deux conceptions antagonistes de la catégorisation pour donner quelques repères épistémologiques sur celle-ci, puis je discuterai successivement les trois points types d'analyse de la catégorisation que l'on trouve dans mes travaux.

1.1 À l'arrière-plan : la catégorisation en philosophie du langage et en linguistique cognitive

En fait de cases, ce sont en réalité les ensembles mathématiques qui constituent l'archétype de ce que Lakoff appelle la théorie « classique » de la catégorisation (*ibid.*) : l'appartenance est donc définie de façon binaire (si un individu / item n'appartient pas à une catégorie, sa valeur d'appartenance sera 0 ; s'il lui appartient sa valeur sera 1), ce qui revient à dire que les ensembles possèdent des frontières précisément définies, et que les membres ont tous le même statut. Ces catégories classiques dont Lakoff donne une définition quelque peu

⁵⁶ "Categorization is not a matter to be taken lightly. There is nothing more basic than categorization to our thought, perception, action, and speech> Every time we see something as a kind of thing, for example, a tree, we are categorizing. Whenever we reason about kinds of things--chairs, nations, illnesses, emotions, any kind of thing at all--we are employing categories." (Lakoff 1987 : 5-6)

caricaturale ne résistent pas à l'épreuve des faits : en 1975, les travaux pionniers d'E. Rosch⁵⁷ en psychologie cognitive montrent, à partir d'un dispositif empirique évaluant la catégorisation des couleurs chez des locuteurs de langues différentes, l'existence d'asymétries entre membres d'une même catégorie. Ce constat permet de définir des « effets prototypiques » liés à l'existence de « meilleurs exemples ». Les résultats expérimentaux de Rosch viennent renforcer la proposition de R. Brown⁵⁸, qui avait détecté plus de 15 ans auparavant l'existence d'un « bas niveau de catégorisation » à même de rendre compte de l'existence de différences au sein d'un ensemble de synonymes : le bas niveau de Brown correspond aux mots les plus courts, les plus fréquents, et à ceux qui sont appris en premier.

Ces propositions, que j'ai rencontrées dès 2003 en produisant, dans le cadre d'un mémoire de maîtrise de philosophie, une traduction commentée du premier livre de l'ouvrage sur la catégorisation de Lakoff (1987), ne fondent pas uniquement les premières linguistiques cognitives : la recherche d'effets de fréquence, ainsi que l'ensemble des travaux qui s'intéressent aux premières acquisitions langagières comme socle des développements ultérieurs, supposent l'existence d'items de bas niveau et de phénomènes de prototypicalité. Ce sont des questions qui, comme je l'ai expliqué dans la première partie de ce document (1.2 et 1.4), sont abordées en plusieurs endroits dans mes travaux, et qui concernent tant la linguistique textuelle que la réflexion sur l'acquisition des langues maternelles et secondes. Ainsi par exemple, c'est en travaillant sur les effets prototypiques dans la variation lexicale à partir d'un corpus d'articles tirés de différents magazines féminins que Geeraerts, Grondelaers et Bakema (1994) ont mis en évidence des effets de saillance, qui concernent aussi bien des éléments fréquents et peu fréquents, et sont davantage liés à « une situation discursive particulière » (Desagulier 2015 : 103).

D'autre part, comme la catégorisation est également l'un des cadres utilisés dans l'élaboration de modèles théoriques, la théorie lakoffienne a permis à C. Delmas (1993) de proposer une conception plus souple de l'invariant, en sortant du modèle de l'appartenance sur la seule base des propriétés communes :

A un certain niveau, ils [les marqueurs] semblent être dotés d'un noyau dur, nécessaire et suffisant (« l'invariant » chez G. Guillaume et H. Adamczewski), et, à un autre, ils semblent impliquer des propriétés contradictoires avec le noyau (cf. *be+ing* ; *-ed* ; *should* ; *the* ; *to* ; *of*, etc.). D'un côté, nous avons l'impression qu'un même trait (ou ensemble de traits) verrouille (comme dans la théorie classique) telle forme dans une classe d'emplois, d'un autre, que ce n'est qu'un « air de famille » (L.

⁵⁷ Rosch, E. 1975. "Cognitive Reference Points." *Cognitive Psychology* 7 : 532-547.

⁵⁸ Brown, R. 1958. "How shall a thing be called?" *Psychological Review* 65 : 14-21.

Wittgenstein) qui relie les emplois d'une forme. Les valeurs ne seraient plus strictement dérivées directement d'un même invariant mais un cas d'emploi en inspirerait un autre, à la fois différent et ressemblant par quelque propriété du modèle. On aurait une légère différence par rapport au premier emploi. On aurait un effet de « cascade » (*chaining* chez Lakoff). Il y aurait un fil conducteur, non transitif. (Delmas 1993 : 198, cité par Lowrey et Toupin 2010)

Dériver les emplois les uns des autres, en « cascade », plutôt que de vouloir les inscrire dans une catégorie uniforme, constitue peut-être une stratégie explicative plus proche des besoins de l'apprenant de l'anglais. C'est un point auquel je m'intéresse actuellement, et le travail en cours, qui porte sur la métaphore de soudure adamczewskienne, a déjà montré les limites d'un enseignement qui utilise une conception rigide de l'invariant [Rossi & Chabert 2019-inv].

Enfin, tourner le dos à une conception objectiviste de la catégorisation revient aussi à introduire une dimension culturelle dans la construction des catégories. La question de savoir s'il existe des catégories conceptuelles universelles, propres à la cognition humaine générale et non à des langues particulières, se pose avec une acuité plus forte dans ce cadre. Les parties du discours, que nous connaissons et utilisons au moins depuis Aristote, sont des candidates plausibles à l'universalité. Pour les linguistiques cognitives et basées sur l'usage, elles ne peuvent cependant pas y prétendre :

Alors qu'elles sont en général considérées comme les principales composantes des théories syntaxiques (ou en tout cas des syntaxes formelles) les catégories formelles de la grammaire et du lexique n'admettent pas d'universaux. En typologie, le caractère universel de la catégorie de *sujet* ne fait plus l'unanimité, mais les parties du discours élémentaires que sont le *nom* et le *verbe* ne sont pas encore remises en cause. (Croft 2010 : §7, notre traduction⁵⁹).

⁵⁹ « First, there are no universals of formal grammatical and lexical categories, which are generally taken to be the building blocks of syntactic theories (formal syntactic theories, at least). This is coming to be widely accepted by typologists with regard to the category of 'subject', but is still resisted for basic parts of speech such as 'noun' and 'verb'. »

1.2 Noms et verbes, de la catégorisation innée au déclin des catégories traditionnelles

Peu discuté jusqu'ici, le statut des catégories verbales et nominales dans l'acquisition du langage a fait l'objet d'un débat récent, initié par un article de Ben Ambridge dans la seconde édition du *Handbook of Categorization in Cognitive Science* (Ambridge 2017 : 567-580), en proposant une conception exemplariste au sein de laquelle les catégories verbales et nominales sont illusoire. En d'autres termes, ces catégories n'auraient aucune réalité psychologique dans la mesure où les locuteurs ne forgeraient pas de telles abstractions.

Mes publications sur ces questions sont toutes deux antérieures à ce débat (trois années s'étant écoulées entre l'écriture et la publication de [Parisse et Rossi 2017]) mais elles partent du principe qui est à l'origine de ce débat : l'idée de catégories émergentes. Les catégories émergentes de Clark sont avant tout des catégories sémantiques. Elles correspondent à des distinctions notionnelles ou conceptuelles que les enfants établissent sans qu'elles soient attestées dans leur langue maternelle : l'exemple le plus abouti concerne la catégorie de « source », que Clark et Carpenter (1989: 3) définissent comme la relation entre un thème, un lieu et une direction. La source est le point de départ d'un mouvement, au sens spatial, temporel, mais aussi en tant qu'origine causale ou agentive. Les enfants anglophones qui rattachent cette catégorie à la préposition *from* produisent ainsi des énoncés comme « I took my temperature from the doctor » ou « he's really scared from Tommy » (Clark 2001 : 391). C'est en confrontant ces énoncés non conventionnels aux premières distinctions qui se mettent en place chez l'enfant que Clark fait l'hypothèse de l'existence d'un ensemble de catégories sémantiques ou conceptuelles « privilégiées » (*ibid.* page 382), qui se construiraient sur une base expérientielle et guideraient les enfants dans la recherche de correspondances forme-sens.

Le travail mené avec Christophe Parisse sur les catégories nominales et verbales prolonge cette hypothèse et prend pour point de départ un ensemble de six dimensions sémantiques ancrées dans la perception, qui pourraient permettre de distinguer les catégories de nom et de verbe : la distance, le nombre, le caractère animé, concret, en mouvement, et défini de l'action ou de l'objet auquel l'enfant fait référence. La correspondance entre ces dimensions et les catégories nominales et verbales que nous avons également codées, dans des productions d'enfants allant des toutes premières combinaisons (énoncés à deux termes) à des énoncés dont la longueur moyenne approchait ou atteignait les trois mots, n'est apparue qu'au terme d'analyses statistiques approfondies [Parisse et Rossi 2017]. Ces résultats, qui donnent plus de poids à l'hypothèse du caractère émergent d'une partie au moins de la catégorisation chez l'enfant, imposent aussi une grande prudence dans l'analyse des premières productions langagières. Nous avons en effet

produit un codage manuel des différentes dimensions, qui supposait une lecture fine du contexte interactionnel, et vérifié l'intégralité du codage automatique des noms et des verbes produit par les programmes d'analyse morphosyntaxique MOR et POST de CLAN. Les données et les transcriptions utilisées étaient également des enregistrements d'enfants que nous connaissions bien (celles qui ont été recueillies dans le cadre des projets ANR Léonard puis CoLaJe), et l'on peut supposer que des données normalisées auraient été moins propices à la découverte de tels phénomènes, qui d'ailleurs ont été observés d'abord sur la base de journaux (Clark 1973).

1.3 Catégories situationnelles, catégories en contexte

L'observation répétée d'interactions filmées avec de jeunes enfants, en même temps que la réalisation d'enregistrements de mon fils aîné dans le cadre du projet ANR PREMS m'ont sensibilisée à l'impact que pourraient avoir les différentes situations d'enregistrement sur les données recueillies. J'ai travaillé sur cet aspect avec une jeune doctorante lyonnaise qui aborde l'acquisition du langage avec les outils de l'éthologie : L. Glas a élaboré une typologie des contextes situationnels observés dans les suivis longitudinaux d'enfants francophones, et le codage (manuel) de données dans différentes langues a ensuite fait l'objet d'un travail collectif. Aux côtés des catégories situationnelles souvent repérées dans la littérature (notamment la lecture d'albums et le jeu), figure une catégorie originale liée à l'alimentation et aux soins apportés à l'enfant (c'est la catégorie de « maintenance »).

En raison du caractère extrêmement chronophage de ce travail de codage, les résultats concernaient à la fois un nombre d'enfants relativement faible (deux pour chaque langue) et un nombre d'heures d'enregistrements également faible (15 heures)⁶⁰. Les statistiques descriptives qui sont présentées dans l'article permettent cependant d'entrevoir des tendances, notamment dans les situations de maintenance qui sont peu décrites dans la littérature : elles constituent des moments d'observation privilégiés, où les énoncés adressés à l'enfant sont d'autant plus complexes que le besoin d'argumenter peut s'y illustrer davantage que dans les interactions sociales ordinaires en situation de jeu ou de lecture.

Enfin, cet article propose une comparaison avec une langue peu étudiée, l'arabe tunisien, et ce choix que le projet ANR PREMS rendait possible nous a également confrontés à une difficulté liée à la transcription. En effet, étant donné que les mots intègrent de nombreux affixes que la

⁶⁰ Pour le français et pour l'arabe tunisien, nous n'avons en effet considéré que les étapes clés identifiées dans le cadre du projet, et pour lesquelles les enregistrements étaient déjà transcrits (la transcription aurait en effet ajouté des dizaines d'heures de travail pour chaque heure à transcrire). Les données d'anglais américain étaient déjà transcrites, mais nous avons retenu les enregistrements correspondant aux mêmes étapes pour obtenir des résultats comparables.

transcription ne permettait pas de segmenter, notre découpage en mots ne correspondait pas à celui qui était effectué dans les deux autres langues (le français et l'anglais). Comme l'autre langue étudiée dans le projet ANR PREMS (le berbère tachelhit) posait des problèmes liés à la syllabe, les discussions sur ces questions ont aussi permis de remettre en perspective ces catégories, et la dépendance dans laquelle elles nous placent vis-à-vis de modèles qui ne sont pas toujours adaptés aux données, dépendance qui est encore accentuée par les outils d'analyse utilisés.

1.4 La catégorisation au prisme de la relativité linguistique

L'hypothèse de la relativité linguistique, c'est-à-dire « la proposition qui veut que la langue (au niveau sémiotique, structurel et/ou fonctionnel) puisse influencer la pensée » [Rossi 2012-v] constitue peut-être le cadre qui permet de prendre un recul décisif sur les catégories que nous utilisons et sur le processus de catégorisation. L'évaluation d'une telle hypothèse est cependant périlleuse et, si de nombreux travaux y font référence, ils n'intègrent le plus souvent que la version dite faible de l'hypothèse, ce qui pour John Lucy revient à renoncer à observer la relativité linguistique et ses effets.

J'ai montré dans la première partie de ce document l'importance de la comparaison inter-langues en psycholinguistique développementale (cf. supra, Première partie, 2.2). Ces travaux ont indéniablement renouvelé notre conception des universaux en linguistique, ainsi que l'explique bien M. Bowerman :

Les universaux ne sont pas les hypothèses de départ que nous évaluons et à partir desquelles nous produisons des déductions. Elles constituent plutôt les conclusions que nous espérons atteindre au terme de l'analyse méticuleuse de données issues d'un échantillon de langues suffisamment grand et varié. (Bowerman 2011 : 598, notre traduction⁶¹)

Partir de plusieurs langues suppose cependant qu'on a déjà postulé un point commun : c'est souvent le « langage spatial » qui intéresse la typologie, et M. Bowerman a beaucoup travaillé sur l'expression des relations spatiales dynamiques (mouvement) ou statiques (emboîtement, relation de contenant-contenu). Ces travaux se situent généralement en deçà d'une réflexion sur la relativité linguistique : ils permettent par exemple de montrer à partir de quel âge les enfants sont sensibles aux différentes catégories spatiales qui sont présentes dans leur langue maternelle (Choi et al. 1999). Partir de l'expression de l'espace, c'est supposer que cette catégorie, qui fait référence à l'extralinguistique, cristallise l'intérêt de tous les locuteurs du monde, c'est-à-dire

⁶¹ « Universals are not what you start out with as hypotheses to support deduction and hypothesis testing; rather, they are what you hope to end up with after careful analysis of data from a suitably large and diverse number of languages. » (Bowerman 2011 : 598)

corresponde à une préoccupation fondamentale de l'esprit humain, et/ou à un aspect central des interactions langagières. C'est en travaillant sur les premières prépositions produites par des enfants francophones et anglophones [Rossi et al. 2007] que j'ai rencontré les limites de tels présupposés : les enfants mobilisent ce que les linguistes appellent des « prépositions spatiales » en réponse à des besoins communicationnels qui ne reviennent jamais seulement à parler de l'espace dans lequel ils évoluent. L'idée même d'un langage spatial est porteuse d'une conception fortement référentielle du langage, qui risque de nous faire perdre de vue toutes les autres dimensions de la construction du sens. Ainsi par exemple, le geste de pointage ne désigne que rarement la seule localisation dans l'espace d'un objet : il dit tout autant la saillance de celui-ci pour le locuteur qui le fait remarquer (c'est une façon de dire « regarde ») et il peut aussi constituer une demande ou une « façon d'agir sur l'objet désigné, à distance » [Rossi 2017 : 107]

La typologie sémantique, qui cherche à définir « comment la langue structure les contenus conceptuels » (Talmy 2000 : 2, notre traduction⁶²), est l'une des préoccupations centrales de la sémantique cognitive et les recherches nombreuses qui se sont inspirées de la typologie sémantique de l'événement spatial de Talmy constituent sans aucun doute l'un des socles de l'actuelle linguistique cognitive (voir par ex. Fortis 2012 : 7). Elles mettent en œuvre une démarche inductive, tout en postulant l'existence de processus invariants susceptibles d'être découverts par comparaison : « l'examen de la diversité permet, s'il est fait selon certaines modalités, de mettre au jour les processus invariants qui sont au cœur du fonctionnement du langage et nous éclairent sur sa nature. » (Croft 2010 : §5, notre traduction⁶³). La perspective typologique rapproche ainsi les linguistiques cognitives de la TOE. Elle montre aussi et surtout que les typologies sémantiques sont des universalismes, qui supposent que les patrons de variation observés dans chaque langue particulière constituent des variations sur un schéma universel. Pour observer les effets de la relativité linguistique, au sens de ce que l'on appelle la version forte de l'hypothèse, il faut bien plutôt partir d'une observation attentive au fonctionnement de langues très différentes de celles que nous connaissons. C'est le point de départ de travaux comme ceux de J. Lucy, et il explique bien l'importance d'un moment d'analyse

⁶² “[H]ow language structures conceptual content”

⁶³ “[B]y examining diversity, one can use techniques to uncover the enduring processes that underlie language and reveal its nature. » Même si elle rappelle la terminologie culiolienne, il me semble que la traduction de « enduring processes » par « processus invariants » dans ce passage est à la fois plus claire et plus acceptable que le calque « processus durable ». De plus, la notion d'invariant n'est pas étrangère aux linguistiques cognitives (comme expliqué dans la première partie : cf. supra, 2.1) auxquelles les travaux de Croft sont très clairement rattachés : si cette traduction est infidèle, au sens où elle suggère davantage de parenté entre les deux théories que ne le fait l'original, elle n'implique aucunement à mon sens que Croft soit culiolien.

préalable qui consiste à « évaluer l'usage dans un certain nombre de conditions expérimentales » [Rossi 2012-v : 6].

Mes travaux sur la catégorisation comme sur l'expression du mouvement ne mettent pas en œuvre de dispositifs expérimentaux pour évaluer l'impact de la langue sur la cognition, mais mes choix théoriques et méthodologiques tiennent compte des analyses que je viens de présenter.

2. Mouvement et langage

A l'origine de mon intérêt pour l'expression du mouvement, il y a le désir de comparer deux langues (le français et l'anglais) et d'en élucider ainsi le fonctionnement. J'ai choisi le mouvement comme notion extralinguistique susceptible de fournir un point de comparaison intéressant, sans bien mesurer les difficultés que j'allais rencontrer pour l'articuler aux premières productions linguistiques. L'un des constats qui a émergé de mon observation des données était celui de l'omniprésence du mouvement dans les enregistrements sur lesquels j'ai travaillé, dans les deux langues étudiées. Un codage des différentes dimensions sémantiques des premiers noms et verbes que produisent les enfants m'a permis d'en rendre compte : les analyses, qui ont été faites à partir des enregistrements de trois enfants francophones, montrent que les premières productions se rattachent en majorité à des référents mobiles, à des routines ou à des activités [Rossi 2017 : 104].

Les travaux présentés ici sont très liés à la réflexion sur la catégorisation que j'ai abordée dans le premier point : ils appartiennent à un même ensemble qui se rattache avant tout à l'acquisition précoce des langues maternelles. Toutefois, j'ai souhaité distinguer la thématique « mouvement et langage » parce qu'elle caractérise en quelque sorte mon identité de jeune chercheuse. Le chapitre publié récemment dans un ouvrage sur le langage de l'enfant [Rossi 2017] m'a ainsi permis de faire un bilan du chemin parcouru.

2.1 L'expression de la manière de mouvement

Mon travail sur l'expression de la manière de mouvement en français et en anglais poursuit l'une des pistes ouvertes par mes recherches doctorales. Conformément aux orientations théoriques explicitées ci-dessus (1.1 et 1.4) la manière n'y est pas abordée en tant que catégorie ontologique, ou comme primitive sémantique ainsi que le suggérait par exemple Jackendoff (1983). Elle désigne plutôt « un concept composite reposant sur un faisceau de paramètres qui l'activent dans l'interprétation de nombreuses formes linguistiques » (Moline et Stosic 2016 : 49). Une telle définition correspond bien à une conception non-objectiviste de l'extralinguistique : elle ne présuppose pas l'existence d'une catégorie objective de manière, qui existerait

indépendamment de sa perception et de sa conceptualisation par l'esprit humain (Lakoff 1987 : 164).

En comparant des suivis longitudinaux d'enfants anglophones et francophones [Rossi 2013a] j'ai mis l'accent sur la pluralité des marqueurs et de choix énonciatifs dans les deux langues plutôt que contraintes liées à la langue. Le repérage de similitudes plutôt que de différences est probablement lié à la méthodologie mise en œuvre, et il faut remarquer que les différences ne sont pas inexistantes dans mes analyses, même si les proportions calculées (pourcentages) ne les font pas ressortir. Le travail porte sur six suivis longitudinaux, et les analyses quantitatives se focalisent sur deux enfants anglophones et deux francophones : il constitue donc plutôt une étude de cas, basée sur plusieurs sujets, et la comparaison qu'il permet est par conséquent limitée. Cet entre-deux dans lequel se situent mes travaux sur l'acquisition précoce du langage, et le caractère peu satisfaisant des analyses quantitatives qui en résultent, a d'ailleurs été l'un des moteurs de mon questionnement sur la fréquence et les seuils de significativité (voir ci-dessus, Première partie 1.4) et m'a aussi poussée à travailler sur de plus grands échantillons, en recueillant des données auprès d'enfants d'âge scolaire et d'apprenants d'une langue seconde, ne serait-ce que pour faire l'expérience de ce qui m'avait manqué.

Dans cette comparaison, les onomatopées ont fait l'objet d'un traitement particulier, parce qu'elles présentent des caractéristiques intéressantes liées à leur grande iconicité, et qu'elles permettent une mise en mots très précoce de la manière de mouvement. J'ai par exemple constaté que leur usage dans le discours adressé à l'enfant, au sein d'énoncés où elles peuvent soit conserver la relative autonomie syntaxique qui caractérise leur fonctionnement comme interjections, soit se décliner en formes nominales et verbales (c'est que l'on observe dans le discours de la mère de Théophile), avait pour effet de redistribuer les autres composantes du sens⁶⁴ en produisant des énoncés qui ne correspondent pas aux patrons canoniques. Pour reprendre un exemple que je présente dans cet article [Rossi 2013 : 197, exemple 7], où je ne l'explique que très brièvement, considérons cet énoncé produit dans un contexte de jeu où l'enfant manipule de petits personnages : « Woooo there she goes on the bike! ». L'iconicité qui caractérise l'onomatopée (*woooo*) s'applique aussi à la structuration de l'énoncé, qui produit une description très proche du mouvement perçu : la construction en *there* pose le cadre spatial qui prépare l'apparition du sujet (figure du mouvement) puis la caractérisation du mouvement, dans un circonstant situé en fin d'énoncé. C'est à mon sens ce qui constitue l'originalité de ce travail, qui arpente les confins de la typologie sémantique talmyenne pour repérer ce que j'ai appelé des

⁶⁴ pour reprendre l'expression de C. Sinha, qui s'est intéressé à ces phénomènes de distribution de la sémantique spatiale (Sinha et Kuteva 1995)

« formes moins attendues », dans le discours adressé à l'enfant puis dans les premières productions de celui-ci.

2.2 Le futur et son acquisition

On pourrait avancer que c'est parce que j'y voyais l'expression d'un mouvement métaphorique (temporel plutôt que spatial : *TIME IS SPACE*, dirait-on pour restituer l'analogie comme le font Lakoff et Johnson) que j'ai choisi de travailler sur l'expression du futur en anglais. En réalité, et même s'il est vrai que j'avais repéré que de nombreux usages de *going to/gonna* dans les données utilisées pour ma thèse correspondaient à une périphrase à sens futur, c'est surtout la confrontation d'invariants (comme l'opération de préconstruction pour les usages de *going to/gonna*⁶⁵ et « l'entreprise commune » pour *will*) à des données authentiques d'acquisition du langage qui m'a intéressée. Autrement dit, si je ne considère pas que le futur participe de l'expression du mouvement, la démarche que j'ai mise en œuvre dans ce travail est directement issue d'une des principales conclusions de mes analyses sur mouvement et langage. Cette conclusion peut s'énoncer ainsi : les hypothèses sur la primarité du spatial ont un pouvoir explicatif très faible et une pertinence limitée lorsque l'on travaille sur l'acquisition précoce du langage à partir de suivis longitudinaux.

L'article intitulé « Le futur en construction » [Rossi et al. 2013] résulte d'une réflexion collective et de l'application (également collective) d'un codage complexe des formes observées chez deux enfants américains du corpus de Providence (cf. supra, il s'agit d'un corpus librement accessible dans la base de données CHILDES). Les catégories de codage comprennent 8 critères objectifs et 5 critères interprétatifs : pour chacun de ces critères, des valeurs ont été attribuées à chacune des formes repérées. Ce codage nous a permis d'effectuer des rapprochements (que nous appelons corrélations dans l'article sans faire référence à un test statistique) en fonction des ressemblances qu'il rendait observables. Surtout, il a rendu possible un balayage des données en utilisant différents points d'entrée.

Les analyses qualitatives présentées à la fin de l'article sont issues de ce travail de codage minutieux. Elles expliquent les trois ensembles de valeurs repérées pour chaque marqueur et les illustrent en s'appuyant sur des séquences choisies parmi les données codées. Même s'ils ne sont pas décrits ainsi dans l'article, ces trois ensembles s'apparentent à ce que Lakoff (1987) appelle des catégories radiales : structurés autour d'un prototype, ils comprennent des énoncés qui lui ressemblent de diverses façons. On est bien loin ici des primitives sémantiques, même telles que

⁶⁵ Lansari (2009) l'explique sous la forme d'un choix fait en amont de la situation d'énonciation, c'est ce que nous avons appelé la valeur de planification [Rossi et al. 2013 : 273].

les a définies Wierzbicka, qui propose sur cette base une analyse sémantique des modaux via la décomposition en éléments plus simples (Wierzbicka 1987). L'idée d'une analyse compositionnelle du sens sur la base de primitives ne fait d'ailleurs pas l'unanimité en linguistique cognitive, précisément parce qu'elle n'est pas compatible avec la théorie des prototypes (Lakoff 1990 : 46).

2.3 Aux confins des schémas de lexicalisation : quelques applications à l'enseignement de la traduction

Les travaux décrits dans ce deuxième point ont des applications à la traduction qui ne se limitent pas à la reprise des propositions de la linguistique contrastive. J'ai en effet essayé de mettre en œuvre une analyse linguistique qui, si elle se fonde sur les observations qui ont pu être faites en considérant des schémas de lexicalisation, se situe au niveau de l'énonciateur et de la situation d'énonciation. J'ai déjà montré dans ce qui précède l'importance de ce niveau d'analyse pour la traduction, je me contenterai donc ici d'un exemple. Le codage et l'analyse des liens (de continuité, de discontinuité ou de changement) entre les formes verbales produites pour exprimer le futur et la situation d'énonciation [Rossi et al. 2013 : 274] nous ont permis de confirmer l'importance des dimensions modales et aspectuelles en anglais, alors qu'en français, dans des situations comparables, on privilégiera les valeurs temporelles (Morgenstern, Parisse et Sekali 2009). Ce constat, qui ne s'apparente pas à une règle de grammaire mais correspond à une analyse du sens en situation, est de nature à éclairer les décisions d'une traductrice ou d'un traducteur (que nous aurions pu, par exemple, solliciter pour mettre en place des sous-titres sur ces enregistrements). Je reprends la première séquence citée dans cet article [Rossi et al. 2013 : 271] et je prends le rôle de la traductrice (mes propositions sont en gras ci-dessous) pour montrer en quoi peut consister cet éclairage :

Exemple 1: Naima ---- 1 an et 6 mois
*MOT: crayons away, can you put it in the box?
On range les crayons, tu veux bien les mettre dans la boîte ?
*MOT: good job!
Très bien !
*CHI: I'll do it.
C'est moi qui le fais.
%pho: 'nail 'durt
*MOT: crayons away, it's time to put the crayons away.
On range les crayons, c'est l'heure de ranger.

On voit bien qu'il est impossible d'utiliser le futur en français pour traduire « I'll do it » dans ce contexte. On voit aussi que la périphrase « Je vais le faire » aurait été plus adaptée dans le contexte d'un dialogue entre adultes. Quant au positionnement interpersonnel, on peut imaginer que l'enfant choisira de l'exprimer en français par un pronom, souligné par la tournure présentative.

L'application à la traduction de la démarche d'élucidation du sens en situation, et de la recherche de ce que D. Bottineau (2004) a appelé des « écarts typologiques » qui s'illustre dans mes travaux sur l'acquisition du langage, correspond à une piste que je n'ai pas encore beaucoup exploitée. J'ai cependant commencé à pratiquer la chose suivante dans mon cours sur la traduction automatique : à partir d'un repérage préalable, dans un corpus parallèle, de ce que Bottineau (*ibid.*) a appelé « correction des écarts typologiques », je sélectionne de courts passages présentant de telles corrections, pour produire ensuite des sorties de traduction automatique (TA) qui fonctionnent comme révélateurs de ces écarts⁶⁶. Voici un court exemple délibérément choisi dans la presse, qui ne constitue toujours pas un domaine d'application valable de la TA, même neuronale :

Passage repéré	The Northern Hemisphere is sweating through another unusually hot summer. Japan has declared its record temperatures a natural disaster. Europe is baking under prolonged heat, with destructive wildfires in Greece and, unusually, the Arctic. (Source : https://www.nature.com/articles/d41586-018-05849-9)
Traduction existante	Cette année encore, l'hémisphère Nord connaît un été exceptionnellement chaud. Au Japon, la vague de chaleur qui s'est abattue sur l'archipel a été déclarée catastrophe naturelle. En Europe, on cuit à petit feu pendant que des incendies ravagent la Grèce et sévissent même au-delà du cercle arctique. (Source : https://www.courrierinternational.com/article/canicules-secheresses-inondations-qui-la-faute)
Sortie Google Translate	L'hémisphère Nord transpire à travers un autre été exceptionnellement chaud. Le Japon a déclaré ses températures record une catastrophe naturelle. L'Europe est en train de cuire sous une chaleur prolongée, avec des feux de forêt destructeurs en Grèce et, exceptionnellement, dans l'Arctique.
Sortie DeepL	L'hémisphère nord transpire pendant un autre été exceptionnellement chaud. Le Japon a déclaré que ses températures record étaient une catastrophe naturelle. L'Europe cuit sous une chaleur prolongée, avec des incendies destructeurs en Grèce et, ce qui est inhabituel, dans l'Arctique.
Sortie eTranslation	L'hémisphère Nord fait l'objet d'un souffle d'été particulièrement chaud. Le Japon a déclaré des températures record pour une catastrophe naturelle. L'Europe est confrontée à la chaleur prolongée, qui a ruiné les incendies en Grèce et, exceptionnellement, dans l'Arctique.

Tableau 3. Exemple de correction d'écarts typologiques

Les écarts repérés concernent d'abord le mouvement fictif de traversée de l'été, auquel s'ajoute dans une prédication compacte l'expression d'une manifestation corporelle que l'on peut rattacher à la catégorie de manière (dans la classification de Cardini 2008, in [Rossi 2013 : 192], elle correspondrait aux aspects directement perceptibles). En français cette référence très concrète à la transpiration n'a pas pu être conservée. Vient ensuite la tournure ditransitive attributive qui permet de rapporter la déclaration officielle de façon neutre, ce qu'autorise le passif en français.

On retrouve la complémentarité souvent décrite entre linguistique et traductologie : le linguiste prend en charge l'explication des observables (les textes sources et les textes traduits), il y trouve un « puissant révélateur [des] écarts en typologie cognitive » (Bottineau 2004 : 10), tandis que les traducteurs et/ou traductologues que seront nos étudiant.e.s pourront reprendre à leur compte les « fondements cognitifs explicables » (*ibid.*) de leur activité, et (bénéfice

⁶⁶ Je donnerai ci-dessous (4.3) d'autres exemples d'usages de la TA comme outil heuristique.

secondaire non négligeable et trop rarement évoqué) expliquer que celle-ci est bien autre chose qu'un transfert direct comme celui qu'effectue la TA, même neuronale.

3. Les composés et leur acquisition

Alors que je m'étais davantage intéressée, en travaillant sur l'expression du mouvement, aux premiers verbes et aux premières prépositions produites par de jeunes enfants, mes travaux sur les composés et leur acquisition m'ont permis d'aborder des étapes plus tardives du développement du langage : ils concernent à la fois le français langue maternelle telle qu'elle est utilisée de façon ludique et créative chez des enfants d'âge scolaire, et l'anglais langue seconde chez des étudiants de licence LEA. La composition a ceci de commun avec le mouvement qu'elle nous ramène aux origines du langage : vraisemblablement issus d'un principe primitif (Jackendoff 1999, in Malmkjaer 2010 : 318) les composés sont très productifs dans la langue anglaise, et parfois difficiles à appréhender pour les francophones. La comparaison inter-langues est donc aussi un point nodal des présentés ici : implicite dans la première étude ci-dessous, elle constitue le point de départ des deux autres, dont la construction s'ancre dans le constat des difficultés rencontrées par les apprenants de l'anglais face à l'alternance du génitif.

3.1 Mécanismes de composition et d'amalgamation, néologie et créativité

J'ai abordé la production de néologismes dans le cadre d'un travail collectif commencé à la fin de l'année 2013 dans le cadre d'un projet porté par V. Renner (Université Lumière Lyon 2 et CRTT) et financé par la Fondation européenne pour la science⁶⁷. Pour étudier la productivité de certains patrons (principalement les composés nom + nom et les amalgames ou mots valises) dans la création de mots complexes chez des enfants francophones d'âge scolaire, j'ai effectué deux recueils successifs dans des écoles de l'agglomération grenobloise. Le premier concerne des enfants de CE2 (N=17) et de CM1 (N=6) [Rossi 2015], et le second, fait l'année suivante dans une autre école, intègre les productions d'une population plus nombreuse d'enfants de CM1 (N=23) et de CM2 (N=18) [Rossi & Renner 2017-oral]. Le matériel expérimental utilisé, qui se compose de 40 images originales (montages photographiques représentant des entités imaginaires perceptuellement hybrides) est issu des travaux de S. Borgwaldt (voir par ex. Borgwaldt et Benczes 2011) avec qui nous avons collaboré dans le cadre du projet précité. Nous distinguons la post modification d'un nom par un syntagme descriptif adjectival ou prépositionnel (SD), la

⁶⁷ Le réseau NetWordS (Cross-disciplinary Approaches to Understanding Word Structure in the Languages of Europe) a financé la visite de 10 jours d'une collègue allemande (Susanne Borgwaldt) pour nous permettre de préparer l'utilisation du stimulus et de la procédure expérimentale qu'elle avait conçus avec des locuteurs francophones : <http://archives.esf.org/coordinating-research/research-networking-programmes/humanities-hum/networks/scientific-activities.html>

composition (NN) et l'amalgamation (A), et nous utilisons une définition large de l'amalgamation pour intégrer la diversité et la richesse des productions recueillies : il s'agit pour nous de tout construit lexical qui ne juxtapose pas ses lexèmes-sources, et qui est assemblé après la troncation segmentale (au niveau graphique et/ou phonique) d'au moins un des formants (Renner 2015). L'opération morphologique d'amalgamation n'est évidemment pas sans lien avec la composition : « On a dans les deux cas affaire à l'association de deux lexèmes-bases et [...] [l']interprétation sémantique des constructions composées et amalgamées peut de plus être appréhendée de manière identique. » (*ibid.* p.100). On peut ainsi supposer que l'amalgamation sera plus aisée dans une langue comme l'anglais : c'est en tout cas ce que suggèrent les résultats expérimentaux de Manegre (2001) qui constituent à notre connaissance la seule étude développementale de l'amalgamation. Dans les données de Manegre les amalgames apparaissent dès 4 ans, pour laisser place à une préférence pour les composés jusqu'à 8 ans. Ce que nous avons observé chez les enfants francophones qui ont participé à cette étude est une production d'amalgames qui ne démarre véritablement qu'en CM1 et se renforce en CM2. Je restitue ci-dessous les mesures qui ont été faites sur le second échantillon, puisqu'elles correspondent à un travail qui n'a pas encore été publié.

Mesure	CM1 (23 enfants, 920 dénom)	CM2 (18 enfants, 720 dénom.)
âge moyen en mois (AA;M)	119,4 (9;9)	129,7 (10;8)
écart type âge	5,8	3,5
durée moyenne du test en secondes (minutes approx.)	307,5 (5)	381,2* (6)
écart type (ET) durées test	118,3	98,3
durée moyenne du post-test (secondes)	126,5	76**
écart type durées post-test	54	32,6
nombre de NN (% du total)	582 (63%)	271 (38%)
nombre de SD (% du total)	226 (24%)	156 (22%)
nombre d'A	109 (12%)	275 (38%)
nombre moyen de NN / enfant	25,3	15,1
écart type	14,4	13,7
nombre moyen de A / enfant	4,7	15,3
écart type	11,3	15,6
nombre moyen de SD / enfant	9,8	8,7
écart type	12,9	12,3

Tableau 4. Données recueillies en 2016 et présentées lors du colloque Neolex [Rossi & Renner 2017-oral]

Les résultats d'une première analyse statistique font ressortir des tendances développementales remarquables : premièrement, une régression linéaire simple⁶⁸ montre qu'il existe un effet du niveau scolaire (CM1 versus CM2, $p = 0,0166$), mais pas de l'âge (que l'on considère qu'il s'agit d'un prédicteur continu ou que l'on fasse deux groupes d'âge) sur la production d'amalgames et deuxièmement, la durée totale du test (soit le temps passé par l'enfant pour produire les 40 dénominations successives, étant donné que nous n'avons pas imposé de rythme dans la succession des images mais suivi celui de l'enfant) prédit particulièrement bien la production d'amalgames ($p < 0,0001$). Il s'agit donc d'une tâche complexe, mobilisant un ensemble de connaissances et de compétences qui semblent se mettre en place à la charnière du CM1 et du CM2. La variété des patrons attestés dans nos données (qu'il s'agisse de réduction des formants, de composition syllabique ou de jointure segmentale) est aussi le reflet de l'acquisition du procédé dans sa complexité.

Les données dont nous disposons au terme de cette seconde collecte sont extrêmement riches (2560 dénominations) et nous encouragent à poursuivre dans deux directions : la comparaison avec d'autres langues, qui pourrait se faire dans le cadre de la collaboration existante avec l'Allemagne, et l'ajout d'un groupe contrôle d'adultes francophones. De plus, l'exploration fine des patrons utilisés et des tendances susceptibles de s'illustrer chez chacun des locuteurs interrogés est encore en cours. Enfin, l'analyse statistique préliminaire que nous avons réalisée (et qui consiste en une série de régressions linéaires) doit être confirmée par l'intégration des différentes mesures à un modèle linéaire généralisé, qui permettra de tenir compte du fait que nous avons des échantillons de taille différente, mais aussi de mieux modéliser la distribution et la variation des formes produites. En effet, il s'agit de données langagières dont on est en droit de supposer qu'elles n'ont pas été produites au hasard, et dont on a ici présupposé à tort que leur répartition suivrait une loi normale (sur cette question, voir notamment Gries 2010 : 277).

⁶⁸ Que nous avons faite pour chacun des prédicteurs supposés

3.2 L'alternance du génitif : arguments en faveur d'une approche probabiliste de la grammaire

Mon travail sur l'alternance du génitif est très ancré dans ma pratique d'enseignante, et il me permet de mettre à l'épreuve des faits à la fois les méthodes traditionnelles que je pratique depuis plus de 10 ans dans l'enseignement de la grammaire anglaise, et ma sensibilité de psycholinguiste qui m'encouragerait plutôt à renoncer à l'explicitation par des règles. La question fait toujours débat et la littérature ne permet pas, à l'heure actuelle, de trancher (Moeller & Ketsman, 2010). Cependant, la formation universitaire française des spécialistes de langue privilégie largement les savoirs explicites, dans les filières traditionnelles (LLCER) comme en langues étrangères appliquées (LEA). L'efficacité du recours à une formulation explicite des règles a été démontrée dans de nombreux travaux et la méta-analyse de Norris et Ortega (2002) en donne un aperçu intéressant, mais les 77 publications analysées ne comprennent aucune étude conduite au sein d'une formation universitaire française. Il faudrait donc ajouter à la diversité d'approches recensées l'exception française d'un enseignement de la grammaire fortement nourri de théorie(s) linguistique(s), avec lesquelles les étudiants des filières classiques doivent aussi se familiariser. L'autre constat qui a motivé ce travail est celui d'un manque dans la littérature : les travaux portant sur l'acquisition de connaissances approfondies en langue seconde, c'est-à-dire de connaissances qui vont au-delà des règles de base, sont rares (Ellis 2006).

Cette étude, que je mène avec deux collègues anglicistes (I. Sahakyan et E. Chabert) et avec un psycholinguiste (J-P. Chevrot) porte en réalité sur l'alternance de trois constructions : le génitif ('s), les groupes prépositionnels en *of*, et les noms composés. Rarement abordée dans les manuels, cette alternance est souvent enseignée dans les cursus universitaires français car elle correspond à une difficulté particulière pour les étudiants francophones. Les résultats ont fait l'objet d'une présentation dans un colloque international [Rossi, Chevrot, Sahakyan et Batailler 2017-oral], et de deux conférences invitées [Rossi 2017a-inv] et [Rossi & Chabert 2019-inv⁶⁹] mais ne sont pas encore publiés.

Trois dispositifs expérimentaux, qui correspondent aussi à trois étapes de la formation des étudiants en langues, ont été mis en place afin de déterminer quel type d'information aide les étudiants à comprendre le fonctionnement de l'alternance. Je les décris brièvement dans l'ordre qui a été celui de cette recherche. Le premier dispositif concerne deux groupes d'étudiants de L3 dans la filière LEA (N=66) dont les connaissances ont été évaluées afin de créer un test qui se

⁶⁹ Initialement prévue en juin 2018, cette journée du Centre de recherche en linguistique anglaise (CRELINGUA), a été reportée à 2019.

fonde sur leurs erreurs. Ces étudiants ont été répartis en deux groupes équilibrés du point de vue du niveau (évalué sur la base de la moyenne des notes obtenues en thème) et soumis à deux modes d'apprentissage distincts avant de compléter un ensemble de phrases en choisissant la possibilité qui leur semblait la meilleure : pour le groupe A il s'agissait de prendre connaissance d'un ensemble de quatre règles (extraites de Huart et Larreya 2007), pour le groupe B, d'un texte constituant un input enrichi, qui comprenait 20 exemples de l'alternance, et où les tournures à repérer étaient présentées en caractères gras. Une régression multiple à deux prédicteurs (niveau et exposition règles ou input enrichi) montre que le niveau des élèves prédit le nombre de réponses correctes ($p < 0,00001$), mais aussi qu'il existe un effet marginalement significatif de l'interaction entre niveau et mode d'apprentissage ($p = 0,0908$). Autrement dit, la présentation explicite des règles aide les étudiants les plus avancés et induit des erreurs chez les autres, et l'exposition à des exemples produit l'effet inverse. La figure ci-dessous le montre bien : on y voit des différences beaucoup plus marquées entre les étudiants qui ont été exposés aux règles. En effet, si l'on considère le score utilisé pour évaluer leur niveau général, on remarque que pour chaque point supplémentaire, le nombre de réponses correctes augmente de 7 % (contre 1,5 % seulement chez les étudiants qui ont été exposés à un input enrichi).

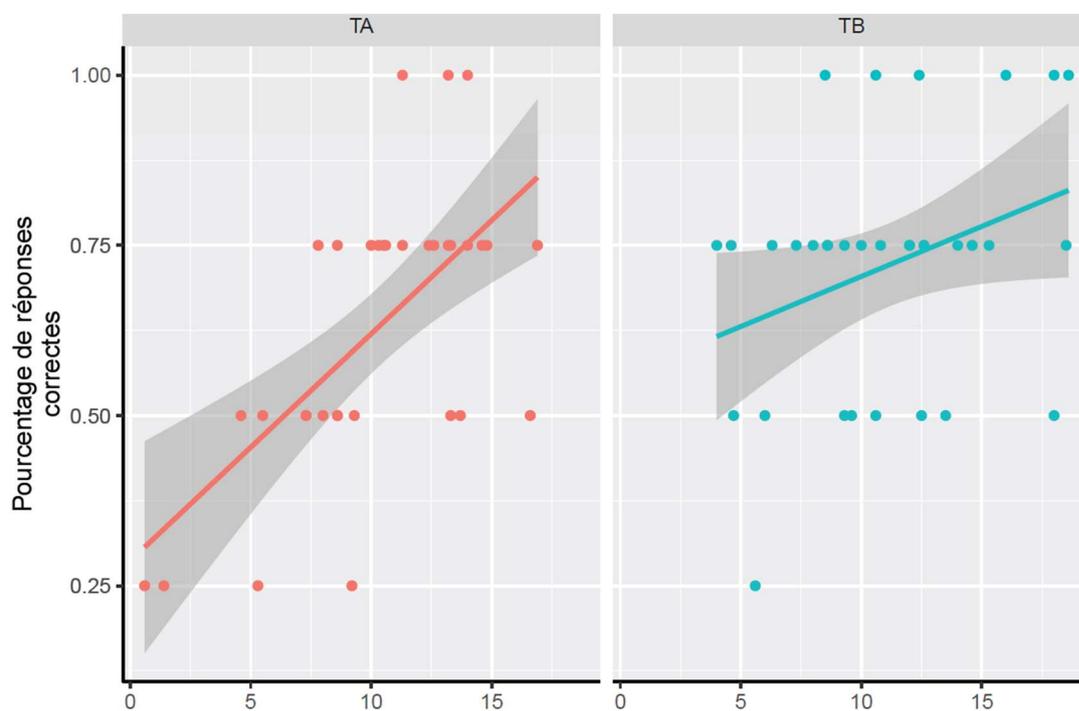


Figure 5. Répartition du pourcentage de réponses correctes au test pour le groupe A (TA : test avec règles) et pour le groupe B (TB : test sans règles) en fonction du niveau (score /20)

Le second dispositif concerne 160 étudiants de L1 et comprend une série de deux questionnaires en ligne. Un premier questionnaire a été proposé aux étudiants afin de déterminer leur sensibilité aux règles et/ou à la fréquence d'occurrence. Ce questionnaire reprend les principes d'un

dispositif élaboré pour évaluer les effets de la fréquence d'occurrence du pluriel des noms masculins commençant par une voyelle sur la production d'une liaison correcte ou incorrecte (Dugua et al. 2009), et présente aux étudiants des cas d'alternance où l'orientation fréquentielle⁷⁰ est favorable à la règle, et d'autres où elle lui est défavorable. Les résultats de plusieurs analyses de la variance indiquent qu'il y a à la fois davantage d'applications réussies des règles chez les étudiants d'un niveau plus avancé, et une plus grande sensibilité à la fréquence.

Notre interprétation globale de ces résultats rejoint les approches constructionnistes de l'acquisition d'une langue seconde (Ellis 2013) : alors que les étudiants les plus avancés élaborent, vraisemblablement sur la base d'une sensibilité assez développée à la fréquence, une connaissance plus abstraite de la grammaire qui leur permet de faire un bon usage des règles, les connaissances des étudiants moins avancés sont très ancrées dans des exemplaires et ils produisent d'avantage d'erreurs en essayant d'appliquer les règles qui leur sont transmises. Nous proposons finalement qu'il puisse être bénéfique d'exposer ces étudiants à un input enrichi, contenant des exemples nombreux et variés, issus de grands corpus, des tournures étudiées, pour les acheminer progressivement vers des connaissances plus abstraites.

Le dernier dispositif interroge ces connaissances plus abstraites, auprès d'un public d'étudiants spécialistes de l'anglais et préparant les concours de recrutement. Il a consisté à évaluer l'application que font les étudiants de l'invariant adamczewskien, suite à un cours qui lui est consacré⁷¹. Le test proposait aux étudiants de choisir la forme la plus appropriée dans plusieurs textes à trous, sélectionnés parce qu'ils fournissent de bons exemples ou contre-exemples de la théorie, et leur demandait à chaque fois de justifier leur choix. Le cadre théorique choisi est délibérément marqué : à la différence des formes schématiques de la TOE, l'invariant adamczewskien n'est en effet qu'assez peu flexible. Nous l'avons précisément considérée comme « une propriété abstraite, à caractère général, qui doit, si le modèle est juste, se retrouver de manière invariante dans chaque occurrence [...] en discours ». (Chabert 2011 : §5). Le même test a également été proposé à un petit groupe d'anglophones. Les résultats montrent que l'invariant

⁷⁰ C'est à dire le ratio, calculé à partir des fréquences observées dans le *Corpus of Global Web-Based English* (GloWBE) de la fréquence de l'item qui correspond à la règle divisée par la somme des fréquences relevées pour les trois constructions qui alternent. Ainsi pour *user's name* qui correspond à la règle d'usage du génitif avec référents animés, mais pour lequel l'orientation fréquentielle ne correspond pas à la règle, on observe les fréquences ci-dessous, et donc un ratio faible :

USER'S NAME	66
NAME OF USER	7
NAME OF * USER	58
USER NAME	2262
RATIO	0.027580443

⁷¹ Les étudiants ont déjà été exposés à cette théorie au cours de leur formation, mais nous avons tout de même consacré une heure à un nouvel exposé théorique.

n'a de valeur prédictive dans aucun des deux groupes. D'autre part, les réponses des anglophones ne manifestent un choix tranché que dans les cas où des contraintes syntaxiques ou d'usage (en l'occurrence l'usage biblique) contraignent fortement l'alternance. Dans tous les autres cas, il s'agit bien pour les natifs comme pour les étudiants, de reconstruire un choix énonciatif sur la base d'indices contextuels, et tous ne l'ont pas fait de la même façon. Même si les résultats de cette première étude sont pour l'instant très descriptifs, ils ne peuvent que nous encourager à présenter aux étudiants d'un niveau avancé plusieurs modèles et plusieurs types d'explications.

Pour conclure en portant le regard sur la méthode plutôt que sur les résultats, on peut remarquer que dans le second dispositif, j'ai adopté une démarche plus franchement quantitative que dans mes précédents travaux, et que les deux premières étapes que je viens de décrire correspondent assez bien à celles qu'évoquent Tummers et al. (2005 : 238 et sq.) dans un état de l'art qui se propose de cerner l'identité méthodologique des linguistiques basées sur l'usage. Le premier dispositif se situe en effet dans une phase descriptive ou exploratoire, et il a été suivi d'une phase explicative, appuyée sur un recueil de données à grande échelle qui nous a permis d'évaluer des hypothèses. En bâtissant un dispositif empirique qui tienne compte des effets de fréquence et permette, le cas échéant, de les observer, nous avons aussi résolument adopté une approche probabiliste. Ainsi la tendance observée chez certains au moins de nos étudiants, qui sont guidés dans leurs réponses par la fréquence, vient confirmer l'un des grands principes de l'approche probabiliste :

La probabilité conditionnelle d'une interprétation ou structure *I* est directement liée à la probabilité antérieure de *i*. Etant donné que la fréquence relative de *i* permet une estimation simple de la probabilité antérieure de *i*, le modèle bayésien prédit que nous devrions rencontrer des effets de fréquence pour de nombreux types de structures. (Bod et al. 2003 : 42, notre traduction⁷²)

Très présente dès les débuts de la sociolinguistique variationniste, qui considère les paroles ou performances des locuteurs comme autant d'« échantillons statistiques issus d'une compétence langagière probabiliste » (Cedergren & Sankoff 1972: 1, in Chevrot 2012: 57, notre traduction⁷³), une telle approche se comprend cependant davantage comme une façon de prendre le contrepied des approches catégorielles que comme une tentative de mathématiser le langage. Il s'agit en effet surtout de dépasser le traitement binaire (vrai/faux, conforme à la règle/erroné) et de placer les

⁷² "This means that the conditional probability of an interpretation or structure *I* is directly related to the prior probability of *i*. Since the relative frequency of *i* provides an easy way to estimate the prior probability of *i*, the Bayesian model predicts that we should find frequency effects for various kinds of structures" (Bod et al. 2003: 42)

⁷³ « Speech performances are here considered as statistical samples drawn from a probabilistic language competence » (Cedergren & Sankoff, 1972: 1, in Chevrot 2012: 57)

productions sur un continuum en tenant compte de toute la variation possible. La question que ces approches laissent pour l'instant en suspens dans mes travaux est celle de leurs applications possibles à l'enseignement des langues étrangères : j'y reviendrai brièvement dans la partie suivante (Synthèse, 1.3).

3.3 Applications à la didactique de la traduction (vers une langue seconde)

L'alternance du génitif nous ramène aussi à la question épineuse des équivalents de traduction. On constate en effet que l'anglais, langue germanique, met en œuvre des mécanismes de composition beaucoup plus étendus (cas des groupes nominaux complexes avec des suites telles que [Adj]+[Nom1]+[Nom2]+[Nom3]), là où le français, langue romane, privilégie une syntaxe plus contrainte, utilisant par exemple fréquemment la complémentation prépositionnelle (Maniez 2013; Scalise et Vogel 2010). La complémentation prépositionnelle existe cependant aussi en anglais, y compris dans le discours spécialisé, et c'est pour donner des repères d'usage aux étudiants que nous avons mené une étude sur un grand corpus d'anglais médical : le corpus d'écrits scientifiques en anglais de Scientext (Grossmann et Tutin 2014).

Le travail sur corpus a été fait en deux temps : il comprend une première série d'analyses réalisées sur un petit corpus d'écrits d'étudiants allophones non spécialistes de l'anglais, afin de repérer les difficultés qu'ils rencontrent, puis des analyses effectuées à plus grande échelle. Nous avons décrit les contextes gauches et droits de chacune des constructions analysées, ainsi que leur position dans le texte (en distinguant notamment les titres des différentes sections d'un article scientifique). Ces repérages sont conçus comme autant d'indices permettant d'aider les étudiants à faire fonctionner l'alternance dans leurs propositions de traduction en sélectionnant l'une ou l'autre construction en fonction du contexte, et l'article propose de consigner ces indices dans un outil existant, Dicorpus, qui s'apparente à un dictionnaire basé sur le corpus Scientext.

Les suggestions que nous avons faites dans cet article n'ont pas été mises en œuvre, mais on peut en retenir l'esprit sinon la lettre. L'intention était de sensibiliser les étudiants au fonctionnement de la langue, et les repérages comme leur mise à disposition dans un outil ont été pensés pour accompagner et encadrer les usages que pourraient faire nos étudiants d'un grand corpus d'anglais de spécialité. L'idée qu'une telle proposition ait des applications à la traduction ne doit pas être entendue comme un encouragement à adopter une vision purement linguistique de celle-ci, d'abord parce qu'il s'agit de traduction inverse (qui se pratique dans nos formations sans qu'une situation professionnelle soit visée, en tout cas pour les étudiants qui ne sont pas anglophones), ensuite parce que nous pensons que la démarche peut faciliter l'accès au sens, y compris en permettant une lecture plus juste des syntagmes nominaux complexes en anglais scientifique.

4. Le traducteur et ses outils

Les travaux récents en traductologie s'enrichissent des apports de la sociologie (Wolf et Fukari 2007 ; Vorderobermeier 2014), des sciences cognitives (Ehrensberger-Dow et al. 2015, Papavassiliou 2007) et de l'ergonomie (Lavault-Olléon 2011), pour mieux analyser les changements considérables auxquels les traducteurs doivent actuellement faire face. Avec l'évolution conjointe des pratiques commerciales (Moorkens 2017, Leblanc 2017) et des technologies numériques, ils sont de plus en plus souvent confrontés à des sorties de traduction automatique (TA), ainsi qu'à des environnements de traduction assistée par ordinateurs (TAO) intégrés à des plateformes virtuelles qui permettent aussi aux agences employant des traducteurs indépendants ou salariés de garder une trace de toutes leurs actions. De telles pratiques peuvent être perçues comme une forme de surveillance malvenue (Kenny 2016), et l'automatisation croissante, en dépit des gains de productivité qui lui sont souvent associés, peut aussi occasionner la fatigue, l'ennui et la démoralisation de celles et ceux à qui elles sont imposées (Moorkens and O'Brien 2017).

Dans ce contexte, réfléchir aux interactions entre le traducteur et ses outils ne revient pas uniquement à adopter le point de vue des interactions homme-machine, qui s'intéresse surtout à l'utilisateur mais vise d'abord et surtout l'optimisation des systèmes, et laisse trop souvent de côté les dimensions émotionnelles (Février et al. 2011) et sociales (Legris et al. 2003). Le point d'entrée de cette réflexion dans mes travaux est la traduction automatique, que j'étudie sous l'angle de ses usages et des perceptions auxquelles elle donne lieu. Deux grands ensembles de questionnements en découlent, qui sont abordés tour à tour dans ce qui suit.

4.1 La traduction automatique

L'avènement de la traduction automatique (TA) dite neuronale semble remettre en cause plus radicalement encore que les précédents modèles (TA statistique) le rôle du traducteur, et ce même si les progrès actuels ont aussi permis l'intégration de plus en plus fréquente de la TA aux outils de Traduction Assistée par Ordinateur (TAO) qui sont désormais d'un usage répandu et généralement bien acceptés par les traducteurs (Koskinen et Ruokonen 2017). Les pratiques évoluent, mais l'impact de la TA est encore mal connu et rarement décrit. Faut-il admettre que les changements actuels annoncent la fin d'une conception de la traduction comme activité essentiellement humaine ? Ou bien s'agit-il simplement d'un nouvel exemple de progrès technologique qui oblige le traducteur humain à renégocier sa place ? Et le cas échéant, les recommandations de Daniel Gouadec, qui préconisait il y a plus de dix ans déjà d'« abandonner à l'adversaire le terrain sur lequel on ne peut pas le battre » (2002 : 234) sont-elles toujours

d'actualité ? Dans des travaux parus ou en cours⁷⁴ j'ai essayé d'éclairer les débats actuels en m'appuyant sur des données de deux types.

Les premières sont un ensemble de trois « mini-corpus maison » (*DIY corpora*), que j'ai analysés en repérant les réseaux métaphoriques permettant de délimiter deux conceptions de la TA : ces analyses ont été présentées dans la première partie de ce document (2.1). En second lieu, afin d'évaluer l'impact des changements actuels, j'ai élaboré une série de questionnaires et recueilli des données sur les perceptions. Ces données concernent d'une part des étudiants de Master de traduction spécialisée [Rossi 2017a], et d'autre part, des traducteurs de la DGT [Rossi & Chevrot 2019]. L'enquête que j'ai menée auprès de mes étudiants à la fin de l'année 2016, lorsque j'ai mis en place un nouvel enseignement sur la TA au sein du Master TSM a constitué le point de départ de ce travail sur les perceptions. J'ai ensuite mis en œuvre une méthodologie d'inspiration ethnographique pour observer et interroger les traducteurs de la DGT lors de mon séjour à Bruxelles en janvier 2017 [Rossi 2017c-inv]. Ces deux étapes sont très liées, et j'ai d'emblée utilisé l'exemple de la traduction à la DGT et placé au centre de mon cours le moteur de TA de la Commission européenne (MT@EC, désormais intégré à la plate-forme eTranslation). J'ai ensuite appliqué les résultats de mes observations à la DGT à mon enseignement de deux manières : d'abord en travaillant à mieux reproduire l'environnement de travail des traducteurs institutionnels⁷⁵ pour renforcer la mise en situation des étudiants, et ensuite en engageant une réflexion très concrète sur l'enseignement de la TA dans le cadre d'un partenariat Hubert Curien avec l'université de Dublin (Dublin City University) en vue du montage d'un projet européen (voir ci-dessous, Synthèse, 3.1).

Pour synthétiser les résultats obtenus jusqu'ici, je partirai d'un point de méthode qui distingue ces travaux de l'ensemble de ceux que j'ai présentés jusqu'ici : le choix de combiner observations et analyses qualitatives à l'usage d'échelles et de modèles déjà établis afin d'obtenir des données quantitatives. Les observations effectuées en contexte professionnel et les entretiens individuels ou de groupe qui peuvent y être réalisés produisent des résultats très riches, comme le montrent par exemple les travaux de M. Leblanc (2014, 2017), mais ils ne permettent pas de mesurer et de comparer l'impact de certaines composantes ou catégories de perceptions. Or en interrogeant mes étudiants (au cours d'une séance de questions ouvertes de type *focus group* et par le biais de deux sondages en ligne successifs) j'ai rassemblé des observations remarquables

⁷⁴ Quatre articles rassemblés dans la troisième partie de mon recueil de publications, auxquels il faut ajouter un article en cours [Rossi et Dirand, 2019] et une nouvelle étude comparée de systèmes statistiques et neuronaux également en cours, que je présente brièvement dans ce qui suit.

⁷⁵ Les étudiants du Master TSM utilisent le logiciel Studio, qui fait partie de cet environnement. En y intégrant des mémoires de traduction ciblées, issues des bases européennes (DGT-TM), aux côtés de la sortie de TA, on se rapproche de l'environnement de travail des traducteurs de la DGT.

qui portaient sur leurs inquiétudes face aux usages et aux progrès actuels de la TA. C'est pour comprendre le rôle que pouvaient jouer de telles inquiétudes qu'un modèle d'acceptation de la technologie m'a semblé utile. Il s'agit en effet d'un modèle dont les versions successives (Venkatesh et Davis 2000, Venkatesh et Bala 2008) intègrent, notamment, une composante d'inquiétude liée à l'introduction de l'outil informatique (*computer anxiety*) et à ses nouveaux usages, telle qu'elle a été théorisée par Heinssen et ses collègues (Heinssen et al. 1987), et cherchent à en mesurer l'impact sur l'intention d'utiliser l'outil informatique et donc aussi sur les usages effectifs de celui-ci. Les questions se rapportant à une telle inquiétude cherchent à faire ressortir un malaise, ou toute réaction affective négative ressentie en situation et non dans l'attitude générale vis-à-vis des ordinateurs (*ibid.* p. 50). J'ai donc cherché à adapter ce concept à l'environnement informatique du traducteur et formulé une question relative aux efforts et aux difficultés que pouvaient ressentir les traducteurs de la DGT dans leur environnement actuel de TAO. Cette question a été difficile à interpréter : elle a suscité plusieurs commentaires des traducteurs et des réponses dont la répartition semble proche du hasard. En revanche, les autres questions se rapportant aux inquiétudes suscitées par la TA, que j'ai rassemblées pour calculer un indice mixte, ont donné lieu à des réponses beaucoup plus claires et tranchées. Dans l'évaluation des perceptions de mes étudiants, cet indice mixte est corrélé à la mesure de leur sentiment d'auto-efficacité [Rossi 2017a : 55], autrement dit, plus les étudiants se montraient inquiets face à la TA, moins ils se disaient capables de l'utiliser. Les situations observées à la DGT faisaient quant à elles ressortir une grande liberté d'utilisation de la TA et des variations importantes d'un traducteur à l'autre, et d'un projet à l'autre : je n'ai donc pas mesuré l'intention d'utiliser la TA. De plus, dans ce contexte institutionnel caractérisé par un environnement de travail bien encadré et où l'ergonomie fait l'objet d'une réflexion globale et régulière (Peters-Geiben 2016 : §69 et sq.), le sentiment d'auto-efficacité s'assimilait à la simplicité d'utilisation perçue et ne constituait pas un indice fort. Lors des entretiens individuels, plusieurs traducteurs ont cependant exprimé leur inquiétude face aux usages actuels et aux évolutions en cours. J'ai donc cherché à évaluer l'impact de telles appréhensions non pas sur l'intention d'utiliser la TA ou sur le sentiment d'auto-efficacité, mais sur les usages réels. Les résultats obtenus confirment l'existence d'un lien (corrélation) entre l'indice mixte calculé à partir des données recueillies auprès de 89 traducteurs institutionnels et l'usage réel de la TA : cet indice, que nous avons ajouté à notre modèle d'acceptation de la technologie [Rossi & Chevrot 2009 : 351], explique en fait une grande partie de la variabilité observée dans les usages (environ 37%, $p < 0,001$) alors même que l'utilité perçue (qui dépend fortement de la simplicité d'utilisation) rend compte de moins d'1% de celle-ci. La réaction émotionnelle des traducteurs est une composante à ne pas négliger : Cadwell et ses collègues en ont d'ailleurs également montré l'importance récemment (Cadwell et al. 2018 : 311) en analysant les résistances à l'adoption de la TA à partir de données de perception (*focus groups*)

recueillies à la DGT en 2016 et auprès de traducteurs salariés d'une entreprise britannique prestataire de services linguistiques (*language service provider*).

Le problème majeur que pose cependant l'analyse des perceptions est le manque de cohérence entre les résultats qu'elle permet d'obtenir et les mesures objectives qui portent sur les processus évalués. La mesure de l'effort cognitif requis lors de la post-édition de sorties de TA a déjà fourni de nombreux exemples de ce manque de cohérence (Koponen 2012 ; Teixeira 2014 ; Moorkens et al. 2015) : ce sont ces études qui constituent le point de départ d'un travail pluridisciplinaire qui comprend le recueil (toujours en cours) d'un complément de données de perception. On distingue depuis Krings (2001) des mesures temporelles (une durée plus longue correspond à un effort plus important) et techniques (nombre d'opérations effectuées sur la sortie de TA) de l'effort de post-édition. S'il est évident que l'effort cognitif ne correspondra pas à l'effort technique, puisque le choix de conserver une partie de la sortie de TA peut être tout aussi complexe que celui qui consistera à la modifier ou la déplacer, la corrélation entre durée et effort n'est pas non plus fiable, et des travaux récents suggèrent que le nombre de pauses soit plus révélateur que la durée de celles-ci (Nunes Vieira 2016 : 54). Les mesures subjectives sont également diverses, si bien que les travaux qui les utilisent sont assez peu comparables (ibid. p. 57).

Le travail pluridisciplinaire commencé il y a un peu plus d'un an ajoute à ces éléments la comparaison de systèmes neuronaux et statistiques, pour tenir compte des préoccupations actuelles des chercheurs qui développent et évaluent les nouveaux systèmes⁷⁶, et en faisant l'hypothèse selon laquelle les progrès des systèmes pourraient renforcer les perceptions négatives déjà décrites dans la littérature. Nous avons commencé par une évaluation simple et rapide, basée sur un corpus disponible au LIG, et constaté que les différences évaluées par un codage des erreurs et par les métriques automatiques traditionnellement utilisées pour évaluer la performance des systèmes étaient considérablement atténuées dans les jugements des étudiants interrogés [Rossi et al. 2017]. Cette étude pilote a fait ressortir plusieurs problèmes, nous nous contenterons de rapporter ici les deux principaux, pour faire état de la réflexion en cours. D'abord l'évaluation hors contexte, pratiquée sur des phrases isolées dans un tableau Excel, a créé des conditions défavorables à la production d'un jugement adéquat, car peu comparables à celles que rencontre le traducteur post-éditeur. Ensuite, et ce point est lié au

⁷⁶ Ce travail a été entrepris avec des chercheurs de l'équipe Getalp du Laboratoire d'Informatique de Grenoble, avec plusieurs chercheurs qui poursuivent une recherche de pointe sur la traduction automatique. Il s'agit d'ailleurs d'une spécificité grenobloise qui remonte à la fondation du Centre d'Etudes sur la Traduction Automatique par Jean Kuntzmann et Bernard Vauquois en 1960 <http://www.liglab.fr/fr/la-recherche/axes-et-equipes-de-recherche/getalp>

précédent, le corpus utilisé ne comprend pas un ensemble de textes mais une succession de phrases et d'expressions fréquemment utilisées en contexte touristique (il s'agit du *Basic Travel Expressions Corpus*, ou BTEC). Or si cette segmentation correspond bien aux besoins du traitement automatique des langues naturelles, elle ne permet pas à l'évaluateur humain de produire des jugements aussi fiables qu'il ne pourrait le faire dans une situation de traduction réelle ou simulée, en se fondant notamment sur la situation et sur les besoins du commanditaire. Nous travaillons actuellement⁷⁷ à la mise en place d'un dispositif expérimental susceptible de réduire l'écart ainsi décrit, pour mieux concilier l'évaluation automatique et l'évaluation humaine des systèmes.

4.2 Le corpus et les mémoires

Souvent rapprochés dans le cadre de l'intégration des corpus dans les outils de TAO (voir par exemple Bernardini & Castagnoli 2008 ; Désilets et al. 2009 ; Kübler 2011), parfois considérés comme identiques (les mémoires de traduction seraient des corpus "cachés" au sein des outils de TAO, ainsi que l'explique Loock 2016 : 34), mais finalement trop rarement comparés, les corpus et les mémoires de traduction constituent deux objets distincts. Issus de traditions épistémologiques sensiblement différentes, ces objets diffèrent à la fois du point de vue de leur constitution et par leurs usages. Comme je l'ai expliqué dans la première partie de cette synthèse, les corpus sont issus d'une tradition linguistique contextualiste et fonctionnaliste, née dans les années 1960 en réaction à l'analyse structurale. Construits pour apporter un ancrage empirique à l'analyse linguistique et permettre l'observation des mots en contexte, les corpus sont aussi et surtout les objets d'une linguistique appliquée, et dès l'origine ils doivent permettre d'améliorer l'enseignement des langues (Teubert 2009) mais leur application à la traduction professionnelle ne va pas de soi, ainsi qu'il m'a été donné de le remarquer lors de mon séjour à la DG-Traduction (voir [Rossi 2017c-inv] pour une présentation de mes activités à Bruxelles). La terminologie du département linguistique de langue française dans lequel j'étais accueillie m'a en effet sollicitée en me demandant de lui indiquer des applications possibles de l'outil Sketch Engine qui leur avait été proposé à l'essai. Suite au travail préparatoire que nous avons effectué en 2017, d'autres collègues sont intervenus pour accompagner la mise en place d'un travail sur corpus⁷⁸ qui à ma connaissance n'est toujours pas effective. Interrogé récemment, un traducteur du département linguistique de langue française à la DGT m'a confirmé que l'intégration n'y avait pas encore été

⁷⁷ Le stage de 5 mois de Nicolas Becker que j'ai co-encadré (voir Curriculum Vitae ci-dessous) a permis de mener à bien ce travail de réflexion, et le dispositif expérimental pourra être mis en place à l'automne 2018.

⁷⁸ Rudy Loock s'est rendu à la DGT en juin 2018 : il n'existe pas de compte-rendu de son intervention mais une émission de radio a été enregistrée à cette occasion (« Les corpus électroniques au service du traducteur professionnel », DGT Academy Radio Lingvistika, et l'émission peut être réécoutée : https://www.podomatic.com/podcasts/dgradiolingvistika/episodes/2018-06-25T05_01_04-07_00)

faite mais qu'une « nouvelle phase de test [avait] été organisée de manière informelle avec Sketch Engine » et était toujours en cours (B. Heyden, communication personnelle, août 2018).

Les acquis de la linguistique de corpus ne sont véritablement intégrés à l'analyse traductologique qu'à la fin des années 1990, dans des travaux qui ont principalement une visée descriptive (Baker 1995) et qui donnent à la discipline un plus fort ancrage empirique. Les mémoires de traduction sont issues d'une autre tradition (voir notamment Sommers 2003 et Garcia 2005, cités par Loock 2016 : 34) qui a très tôt fait usage de corpus mais avec une visée différente : le traitement automatique des langues. C'est aux recherches visant à améliorer le fonctionnement des systèmes de traduction automatique que l'on doit en effet les premiers travaux sur l'alignement de textes traduits (Véronis 2013), et ce sont les acquis de ce que Jacqueline Léon a nommé "l'automatisation des sciences du langage" (Léon 2015) qui ont permis le développement de la traduction assistée par ordinateur (TAO) et des outils à mémoire de traduction.

Si certains auteurs distinguent une branche théorique de la traductologie de corpus et ses applications pratiques, et intègrent à cette seconde branche les apports des corpus et des outils de TAO (voir par exemple McEnery et Xiao 2007 : 22), les considérations épistémologiques que nous venons d'esquisser plaident plutôt pour la distinction de deux objets : d'une part les corpus (et outils permettant leur exploitation par le linguiste et le traductologue) et de l'autre, les outils à mémoire de traduction. Constitués dans des contextes distincts, ces deux objets diffèrent peut-être surtout du point de vue de leurs usages.

Dans une étude en cours que nous présenterons lors des 11^{èmes} journées du réseau Lexicologie, Terminologie, Traduction [Rossi et al. 2018-com] nous revenons sur l'histoire des corpus et des mémoires de traduction afin d'en établir des définitions claires, puis nous proposons une comparaison systématique de ces deux objets lorsqu'ils sont utilisés comme outils par le traducteur spécialisé. Souvent critiquées pour la segmentation excessive qu'elles introduisent, les mémoires de traduction fonctionnent en général avec des segments phrastiques alors que l'unité la plus utilisée dans la constitution et la consultation de corpus est le texte. Ces différences permettent de comprendre que les corpus soient plus souvent utilisés en traduction culturelle, mais que les outils de TAO soient plus adaptés au travail du traducteur pragmatique (Bénel et Lacour 2014⁷⁹). Tout autant que le contexte linguistique droit et gauche que font apparaître les

⁷⁹ Le rapprochement récent effectué par l'outil TraduXio est une initiative remarquable qui se situe à l'articulation des deux traditions que nous avons distinguées.

concordances, c'est cependant aussi le contexte de situation qui intéresse le traducteur : "The focus is non-text that assists in interpreting a particular segment of text." (Melby et Foster 2010).

Il faut également remarquer que les mémoires de traduction ne font pas disparaître le contexte linguistique (puisque'il reste toujours possible d'accéder à une lecture de la mémoire), et surtout, que le traducteur peut garder (en l'enregistrant) la trace du contexte situationnel dans lequel la mémoire a été constituée. A l'inverse, les corpus « maison » qui peuvent être constitués pour accompagner le travail du traducteur et du terminologue ne conservent pas nécessairement d'indices concernant les auteurs et situations de production des textes qu'ils recensent ; c'est le cas notamment pour les corpus constitués grâce à l'outil désormais très répandu WebBootCaT (Baroni et Bernardini 2004).

Ces différences d'usage ne sont certainement pas irréconciliables. Des travaux ont, depuis plusieurs années, accompagné et encouragé la convergence en réfléchissant aux formes qu'elle pourrait prendre. A la suite des propositions faites par Bowker (2011), ils se sont notamment penchés sur l'utilité des Contextes Riches en Connaissances extraits de corpus comparables pour les traducteurs (Josselin-Leray et al. 2014, Planas et al. 2014) et ont proposé d'intégrer de tels contextes dans des outils de TAO. Ces contextes, qui présentent des informations de nature conceptuelle ou linguistique, permettent un gain de temps au traducteur en lui présentant des informations textuelles pré-triées, qui sont plus précises que la concordance brute. Dans l'ensemble, notre étude vise à encourager une convergence qui donne la priorité aux besoins du traducteur, et intègre pour ce faire une conception riche du contexte. Présenter les usages en proposant des descriptions riches du poste de travail du traducteur est d'autant plus important que, en dépit de tentatives relativement nombreuses, on n'observe toujours pas de convergence forte entre l'usage des corpus et celui des outils de TAO. La présence de *plug-ins* dans les outils de TAO (par exemple plug-ins Sketch Engine et Web Lookup! dans SDL Studio) suggère cependant que ces changements soient en cours, et il me semble que c'est en tenant compte des caractéristiques de chacun de ces deux objets que l'on pourra accompagner au mieux l'intégration.

4.3 Les applications à l'enseignement

L'ensemble des travaux que je viens de présenter est traversé par un questionnement sur la profession, qui peut s'énoncer ainsi : quelles sont les conditions de possibilité d'une conception durable et soutenable du métier de traducteur, face à la fois à une croissance rapide de la demande et à la place de plus en plus grande qu'occupent les nouvelles technologies ? En travaillant sur les applications de ma recherche à l'enseignement, je souhaite donc explorer des modèles de formation des traducteurs qui préparent les étudiants aux environnements de travail contemporains, en leur fournissant les compétences techniques et les connaissances dont ils

auront besoin, mais aussi les outils intellectuels qui leur permettront de faire face aux défis éthiques, juridiques et économiques à venir. Or pour tenir ensemble tous ces éléments, il faut non seulement intégrer à notre enseignement les tout derniers moteurs de traduction automatique neuronale, mais aussi développer une pédagogie adéquate, fondée sur une bonne connaissance des facteurs qui aident les traducteurs professionnels à accepter les technologies de traduction sans compromettre leur bien-être, et donc sans affecter l'ergonomie cognitive, organisationnelle et physique de leur poste et outils de travail.

Enseigner la TA neuronale constitue un défi auquel les formateurs commencent tout juste à s'attaquer (Moorkens 2018). Il existe, tout d'abord, des obstacles pratiques : l'entraînement d'un système neuronal demande plus de temps et de ressources, et il est susceptible de produire des résultats beaucoup plus étonnants et difficiles à expliquer qu'un système statistique (*ibid.* p. 2). Ces obstacles correspondent au défi épistémologique que représente la TA neuronale : alors que les modèles statistiques pouvaient être expliqués à partir de notions mathématiques relativement simples, les calculs vectoriels qui permettent aux modèles neuronaux de fonctionner sont moins accessibles (voir par exemple Arthur et al. 2016 qui produisent des explications détaillées du fonctionnement probabiliste de la TA neuronale). Il faut cependant remarquer que le même constat d'opacité avait été fait lors du passage de systèmes basés sur des règles à des systèmes que l'on appelait aussi « empiriques », parce qu'ils s'appuyaient sur des traductions authentiques et consacraient la quasi-disparition des règles. Ce sont les efforts de chercheurs comme A. Way (2009) et M. Hearne (Hearne & Way 2011) qui ont permis de lever le voile. Des chercheurs polyvalents ont d'ores et déjà entrepris d'en faire autant pour la TA neuronale et ils ont proposé des explications remarquablement accessibles (Forcada 2017, Poibeau 2017). Il est donc possible de présenter aux étudiants le fonctionnement des systèmes de TA les plus courants, sans présupposer une connaissance des modèles mathématiques ou des systèmes informatiques utilisés.

Les propositions pédagogiques que j'ai faites jusqu'ici ont consisté à considérer la TA comme outil heuristique [Rossi & Lavault-Olléon 2018-inv] et à en proposer des usages contextualisés (en me fondant sur les observations réalisées à la DGT), pour que la présence d'un cours dédié dans le cursus de Master ne suggère pas que la TA puisse représenter une fin en soi pour le traducteur. Considérer la TA comme outil heuristique, c'est parier que les sorties de TA fourniront des exercices qui permettront aux étudiants de renforcer les compétences acquises dans le cadre de leur formation. On trouve ainsi dans le référentiel de compétences établi par le réseau des Masters Européens de Traduction (EMT 2017) les aspects suivants, pour lesquels je propose une application à la TA lorsqu'elle n'est pas déjà présente dans le référentiel.

DOMAINE DE COMPETENCE	DESCRIPTION (numéro dans le référentiel)	APPLICATION A LA TRADUCTION AUTOMATIQUE
TRADUCTION	(10) Analyser et justifier leurs solutions et choix de traduction, en utilisant le métalangage approprié et en mettant en œuvre des approches théoriques adéquates	Analyser et justifier les choix de post-édition. Utiliser le métalangage approprié pour expliquer les erreurs et les solutions mises en œuvre pour y remédier
TRADUCTION	(14) post-éditer le produit de la TA en utilisant les niveaux et techniques de post-édition adéquats, en fonction des objectifs de qualité et de productivité, et en tenant compte des enjeux que représentent la propriété et la sécurité des données	Idem
TECHNOLOGIES	(18) Maîtriser les bases de la TA et son incidence sur le processus de traduction	Idem
TECHNOLOGIES	(19) Évaluer la pertinence des systèmes de TA dans un flux de traduction et mettre en œuvre le système de TA approprié si nécessaire	Idem

Tableau 5. Applications du référentiel de compétences EMT à l'enseignement de la TA

Si l'on considère tour à tour chacun des quatre points ci-dessus, on constate que les progrès de la TA, en particulier neuronale, ne bouleversent pas ces propositions. Au contraire, l'obtention de meilleurs résultats pourra permettre de travailler de façon plus ciblée, en se focalisant par exemple sur la production de niveaux de post-édition adéquats (14) ou sur une analyse plus fine des erreurs et des choix de post-édition (10). Ce sont ces choix qui définissent le rôle et la place du traducteur, tout comme l'ensemble des aptitudes à agir sur les systèmes que pourront développer les étudiants (Gaspari et al. 2015 : 350). Le rôle d'un enseignement qui fait de la TA un outil heuristique rejoint finalement les conclusions de D. Kenny et S. Doherty (2014 : 285) : c'est l'expérience d'usages critiques de la TA qui permet de redonner confiance aux traducteurs envers et contre le déterminisme technologique actuellement si répandu (Olohan 2017).

Synthèse : vers une conception critique de l'automatisation et de ses retombées en sciences du langage et en traductologie

Ce document de synthèse a suivi plusieurs pistes pour éclairer et expliquer le chemin qui va de l'analyse des faits de langue à la critique de l'automatisation : les premières partant d'une caractérisation méthodologique de mes travaux pour mener à la discussion de leurs orientations théoriques, et les secondes proposant un parcours thématique de mes recherches. L'objectif de cette dernière partie est de montrer comment ces éléments organisent la réflexion en cours et à venir. Il s'agit donc d'une synthèse prospective, qui commence par prolonger la réflexion méthodologique, pose ensuite les jalons épistémologiques qui constituent le socle d'un projet d'ouvrage inédit, et présente pour conclure les différents projets dans lesquels je suis actuellement impliquée et qui viendront nourrir une critique de l'automatisation.

1. Méthodes mixtes, objets variés

Dans la discussion méthodologique qui ouvre ce document, j'ai évoqué le choix d'une combinaison d'approches quantitatives et qualitatives (Première partie, 1.4) que j'avais présenté et expliqué dans l'une de mes premières publications [Rossi & Morgenstern 2008]. En travaillant sur le futur et son acquisition chez des enfants anglophones, j'ai mis en œuvre une méthodologie semblable à celle qui avait été décrite dans cet article de 2008 (Deuxième partie, 2.2), et on trouve dans mes travaux plusieurs illustrations de cette combinaison : elle sert notamment l'analyse des métaphores (Première partie, 2.1), mais elle s'illustre aussi, de façon sensiblement différente, dans la combinaison d'approches ethnographiques qualitatives et quantitatives visant à éclairer les usages et perceptions de la TA. De tels choix méthodologiques ont un coût à court terme en partie comparable à celui de l'interdisciplinarité (voir ci-dessus, Première partie, 2.5) : ils constituent une démarche plus complexe qui n'est pas toujours bien perçue par les évaluateurs, et requièrent souvent un travail collectif dont la mise en œuvre est également plus contraignante. Les éclairages multiples et la richesse des analyses qu'ils permettent sont cependant incomparables, et constituent une caractéristique forte de ces modèles, comme le montre la définition qu'en donnent Johnson et ses collègues :

Les méthodes mixtes définissent le type de recherche mené par un ou plusieurs chercheurs en empruntant à la fois aux approches qualitatives et aux approches quantitatives [...] l'objectif d'une telle combinaison d'approches étant d'atteindre une meilleure compréhension des phénomènes et

une confirmation plus forte des hypothèses formulées. (Johnson et al. 2007 : 123, notre traduction⁸⁰).

Même si elles sont désormais bien décrites, les méthodes mixtes correspondent à une tendance relativement récente, les travaux fondateurs se situant dans les années 1980 (Creswell 2014 : 48). En linguistique appliquée, leur utilisation est le plus souvent partielle, au sens où elle ne concerne que la collecte et l'analyse des données (Hashemi 2012 : 210), mais je montrerai dans ce qui suit l'apport des méthodes mixtes à la traductologie et à la psycholinguistique. Chacun des trois domaines abordés dans ce qui suit, à partir de mes travaux, permet d'illustrer un type de méthode mixte (Creswell 2014 : 50). J'aborde en premier lieu les apports de la méthode mixte de convergence parallèle par triangulation (*convergent parallel mixed methods*) à l'analyse du discours, puis ceux de l'explication séquentielle (*explanatory sequential mixed methods*) en traductologie, et je présente pour finir un cas d'exploration séquentielle (*exploratory sequential mixed methods*) en psycholinguistique.

1.1 Mettre en œuvre des synergies pluridisciplinaire pour analyser le discours sur la justice climatique

La méta-analyse de Hashemi fait ressortir un usage majoritaire de la triangulation en linguistique appliquée, et notamment en analyse du discours où elle décrit l'utilisation conjointe de données se situant à plusieurs niveaux (2012 : 210). Ce que Creswell appelle la convergence parallèle caractérise en outre le fait de rassembler dans un même mouvement et donc en même temps des données quantitatives et qualitatives, et d'intégrer les informations ainsi obtenues aux analyses. Si elle ne relève pas à proprement parler de méthodes mixtes, puisqu'elle s'appuie sur un seul recueil de données, on peut toutefois dire que la mise en regard d'analyses quantitatives basées sur corpus (comme les calculs de mots clés ou de catégories sémantiques discriminantes que permet le logiciel WMatrix) et d'une lecture « horizontale » des concordances en opérant un retour au texte (voir ci-dessous, figure 6) afin de repérer des phénomènes énonciatifs en contexte, s'en approche. Elle emprunte en tout cas à deux méthodologies distinctes, dont la convergence a été décrite sous le nom de « pragmatique de corpus » (Aijmer et Rühlemann 2015).

⁸⁰ Mixed methods research is the type of research in which a researcher or team of researchers combines elements of qualitative and quantitative research approaches (e.g., use of qualitative and quantitative viewpoints, data collection, analysis, inference techniques) for the broad purposes of breadth and depth of understanding and corroboration.

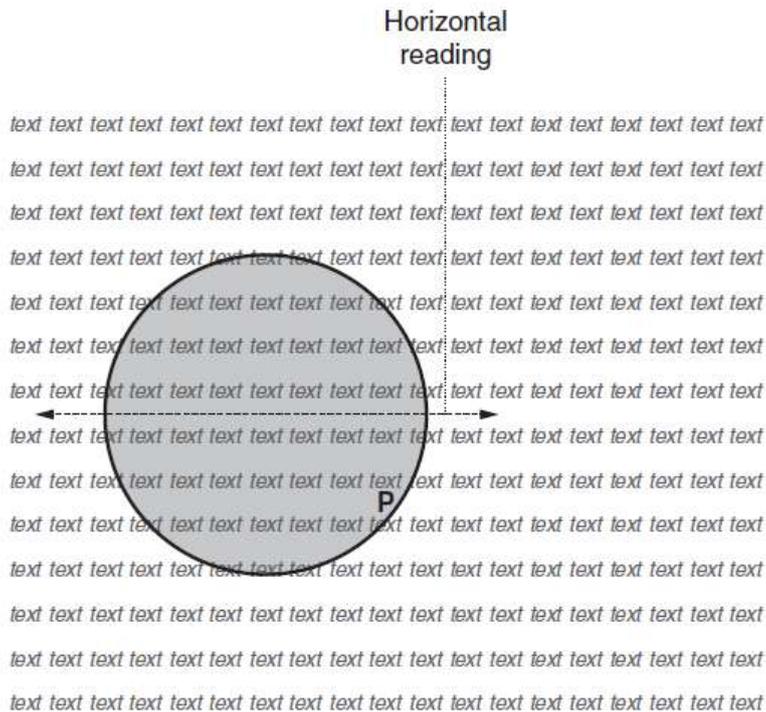


Figure 6. La lecture horizontale qui caractérise les approches pragmatiques (source : Aijmer et Rühlemann 2015 : 3)

Cette synergie correspond à la méthodologie que nous avons mise en œuvre pour constituer un corpus maison de discours sur la justice climatique. Ce projet nous est apparu après coup comme atypique : la première proposition de communication [Rossi et al. 2016-oral] puis d'article [Rossi et al. 2019b] que nous avons faite dans ce cadre a montré que la synergie que nous voulions mettre en œuvre avait un coût. En effet, pour cet article soumis à la fin de l'année 2017, nous avons dû répondre à des demandes de révisions multiples et procéder à la suppression partielle des contenus historiques, considérés comme autant de digressions dans une publication destinée à une revue de linguistique. Outre l'effort de clarification méthodologique qui nous a été demandé en plusieurs endroits, la dernière demande de modifications reçue précisait justement que la nature interdisciplinaire de l'étude la rendait plus difficile à aborder⁸¹. C'est donc avec une conscience d'autant plus aiguë des caractéristiques de notre démarche que nous avons envisagé la suite du travail sur ce projet. En travaillant sur le corpus, nous avons croisé et comparé deux types de recueils (la sélection manuelle des sources, informée par la lecture préalable d'un ensemble de textes situant les questions abordées, et le recueil automatique de contenus sur le web en utilisant des combinaisons de mots clés) : l'articulation de méthodes quantitatives et

⁸¹ Voici la version originale du commentaire qui nous a été transmis par l'un des éditeurs du volume : « Part of the difficulties probably lie in the interdisciplinary nature of the study ».

qualitatives est donc bien réelle dans l'article auquel nous avons travaillé ensuite⁸² [Rossi et al. 2018] mais les synergies pluridisciplinaires ont laissé place à une étude menée par trois linguistes.

La triangulation correspond à une étape plus tardive du travail sur la justice climatique, et c'est pour ajouter aux analyses monolingues une réflexion traductologique que nous l'avons mise en œuvre. L'étude présentée au colloque international Transius en juin 2018 comportait ainsi une analyse diachronique des traductions françaises des « Résumés à l'intention des décideurs » produits par le GIEC (*Summary for Policy Makers*). Pour interroger la cohérence des choix terminologiques qui s'y illustraient, nous avons consulté les bases de données terminologiques de référence qui étaient pertinentes dans ce contexte, en nous fondant en priorité sur la base UNTerm. Confrontées aux données chiffrées et contextualisées que nous fournissent les rapports et leurs traductions, les informations plus qualitatives que contiennent les fiches terminologiques (qui sont également datées et rapportées à un ou plusieurs contextes) nous ont permis d'analyser la stabilisation progressive des termes se rapportant aux degrés de certitude, et de mettre en évidence le rôle central que joue l'anglais dans la communication du GIEC.

Ces conclusions fournissent pour ainsi dire la première pierre d'une conception critique de l'automatisation. Elles ont en commun la mise en perspective de l'homogénéisation comme processus régi par un ensemble de conventions, qu'il s'agisse de la standardisation des méthodes de recherche qui autorise la publication dans une revue internationale, ou de celle qui caractérise désormais la communication des organisations internationales, en lien avec la priorité accordée à la langue anglaise. Ces conclusions ne sont certainement pas sans lien avec la traduction et son automatisation, car « [s]i l'anglais est devenu la langue translative mondiale, c'est primordialement parce qu'il se conçoit comme un pur système de signes permutables. » (Berman 1998 : 33), l'homogénéité étant une caractéristique des productions automatiques, bien davantage qu'elle ne caractérise les productions humaines.

⁸² Cet article a été écrit et soumis plus de six mois après [Rossi et al. 2019b] mais la publication a été plus rapide.

1.2 Revenir aux données qualitatives pour mieux expliquer : l'analyse des usages et perceptions de la TA à la DGT.

La possibilité de mettre en œuvre un autre type de méthode mixte (l'explication séquentielle) se situe dans le cadre de ce que j'ai appelé une traductologie empirique. J'ai présenté dans ce qui précède (Deuxième partie, 4.1) les différentes étapes de la méthodologie d'inspiration ethnographique utilisée pour étudier les usages et perceptions de la TA auprès des traducteurs de la Commission européenne. Les résultats obtenus sont principalement quantitatifs. Il me reste cependant à effectuer l'analyse de la partie qualitative des données recueillies, sous la forme de commentaires formulés en réponse à plusieurs questions ouvertes, facultatives mais utilisées par la plupart des traductrices et traducteurs interrogés, que j'énumère ci-dessous, en redonnant le numéro de la question correspondante [Rossi & Chevrot 2019 : 361 et sq.] :

9. Quant utilisez-vous la traduction automatique ? [Systématiquement et quel que soit le type de texte / Dans certains cas seulement / Jamais. Précisez si nécessaire.]
10. Quelle est la répartition qui correspond le mieux à votre conception du métier et à vos attentes ? [Utiliser la TA dans presque tous les cas / Utiliser la TA dans la moitié des cas / Ne jamais utiliser la TA. Autre : précisez].
11. Quelle(s) activité(s) ou tâches appréciez-vous d'accomplir; qui ne font pas appel à la traduction automatique ?
20. Qu'est-ce qui définit une traduction automatique utilisable, selon vous ?
21. Que signifie pour vous utiliser la TA avec discernement ?
23. Pensez-vous qu'il faille des compétences spécifiques pour utiliser la TA ? [Oui/Non. Précisez si nécessaire]

Cette première analyse qualitative ne sera pas séquentielle, au sens où le recueil n'aura pas été effectué après l'analyse quantitative, mais elle pourra être complétée. En effet, parmi les 89 traducteurs qui ont rempli le formulaire proposé, 15 ont également accepté d'être recontactés pour fournir des compléments d'information : ces entretiens pourront constituer une toute dernière étape dans la démarche d'explication, qui serait alors séquentielle.

Ce retour aux données qualitatives avait été envisagé d'emblée, et une première lecture des réponses aux questions ouvertes a été proposée sous la forme d'une catégorisation *ad hoc* permettant de saisir des tendances [Rossi & Chevrot 2019 : 356]. Même si elle intègre une analyse qualitative aux résultats quantitatifs, cette première lecture ne permet que l'élaboration d'hypothèses, qui imposent un retour aux données. Au vu des questions que l'analyse quantitative laisse en suspens, il me semble qu'une analyse fine des réponses (qui intègre par exemple l'analyse de régularités dans un logiciel comme NVivo) apportera un éclairage intéressant et justifiera probablement une nouvelle série d'entretiens. Ces questions concernent en effet le lien entre

connaissance de la TA et craintes exprimées, et l'identification des différents aspects de l'activité du traducteur que la TA modifie ou est susceptible de modifier. C'est en tout cas en mettant en regard les tendances que les données qualitatives pourraient faire apparaître et celles que nous avons décrites au plan quantitatif que l'on éclairera les liens suggérés par les indices de significativité lors du calcul statistique. On retrouve, là encore, une caractérisation en creux du traitement automatique (dont le calcul est une composante) comme première approximation, l'explication relevant nécessairement aussi de l'interprétation de données qualitatives.

1.3 Appliquer des modèles probabilistes à l'acquisition L2 : une exploration séquentielle complexe

J'ai décrit plus haut les différentes étapes qu'a suivies mon analyse de l'alternance du génitif et de son acquisition en anglais L2 (voir ci-dessus Deuxième partie, 3.2) en restituant simplement l'ordre chronologique de cette recherche et sans interroger la nécessité de ces étapes. Je souhaite donc revenir ici sur le caractère exploratoire de ce travail, qui se justifie en premier lieu par le choix d'aborder un phénomène encore peu décrit. L'alternance du génitif a certes fait l'objet de publications nombreuses et en grande partie récentes⁸³ (voir par ex. Kreyer 2003 ; Rosenbach 2008 ; Heller et al. 2017), mais son acquisition en langue seconde n'est à ma connaissance que très peu décrite dans la littérature (Romano 2016 s'intéresse davantage à la production qu'à l'acquisition). Certains manuels de grammaire fréquemment utilisés n'abordent pas cette question : c'est le cas de celui de Murphy (2004) ; d'autres proposent des descriptions très longues et rendues peu accessibles par la diversité des points abordés (voir par exemple Biber et al. 1999 : 292 et sq., ou celle de Huddleston et Pullum 2002 : 477). L'alternance serait donc considérée soit comme un faux problème, soit comme une acquisition tardive, que l'on peut rapprocher de celle des composés (González Alonso et al. 2016). La complexité et le caractère problématique de cette alternance pour l'apprenant francophone ne font à mon sens aucun doute : ils sont liés à l'existence d'une asymétrie importante entre le français, qui utilise principalement une construction (la modification par un syntagme prépositionnel en *de*) et l'anglais qui distingue trois ensembles de significations, véhiculées par trois constructions distinctes. Le petit livre de Huart et Larreya (2006) part de ce problème et consacre un chapitre à chacune des constructions, avant de faire des propositions orientant le choix de l'une ou l'autre (*ibid.* p. 5). Sa nature très descriptive montre combien il est difficile de trouver un ensemble de règles formelles. C'est cependant en résumant les conclusions de Huart et Larreya que j'ai élaboré l'ensemble suivant,

⁸³ Parmi les publications empiriques et quantitatives analysées par Rosenbach (2014 : 252 et sq.), on distingue trois articles datant d'avant 1940, neuf se situant entre 1980 et 2000, et vingt-huit entre 2001 et 2014.

qui a ensuite été utilisé pour évaluer l'impact d'une exposition aux règles ou à des exemplaires dans le premier dispositif empirique utilisé :

1. Le génitif ('s) est utilisé pour exprimer la possession (*Simon's cat*: le chat appartient à Simon), surtout lorsque le possesseur est un être animé.
2. Les autres utilisations du génitif comprennent certaines expressions temporelles (*yesterday's news*) ou des possesseurs inanimés (*on a knife's edge*).
3. Le génitif ne peut pas être utilisé avec des entités abstraites : on lui préférera alors les noms composés (~~*unemployment's rate*~~ > *unemployment rate*, ~~*information's channel*~~ > *information channel*)
4. Pour désigner une sous-classe ou une catégorie appartenant à un groupe plus large, on préfère généralement les syntagmes prépositionnels (*a new sort of event*, *the number of students in this group*)

Le premier dispositif expérimental nous a permis de constater que les règles n'étaient pas comprises par les étudiants et induisaient de nombreuses erreurs. Or ces règles sont particulières : à la différence des règles souvent utilisées dans les travaux sur l'acquisition des langues secondes, elles ne permettent pas une prise de décision simple et régulière, à l'exception peut-être de la première. Si l'on considère la règle 3, qui a donné lieu au plus grand nombre de difficultés, on constate que la caractéristique d'abstraction n'appartient pas à un ensemble d'entités préétablies mais se module en contexte. On peut ainsi comparer le génitif déterminatif *the hotel's lobby* au composé classifiant *the hotel lobby* (Rosenbach 2008 : 162), et trouver, assez facilement dans un grand corpus de référence, deux occurrences suffisamment contrastées pour illustrer cette différence. Les exemples (a) et (b) ci-dessous sont tirés du COCA, et on y voit bien la catégorisation que permet le composé *hotel lobby*, qui désigne un type de lieu et évoque ici le clinquant des résidences de stars de cinéma, alors que le génitif désigne un lieu précis (la scène se déroule dans un hôtel) :

- (a) Dark hair, dark suit, tie loosened, top button unbuttoned. He sat alone in a plush, red velvet booth with his arm casually sprawled out over the top. His fingers gently drummed against the fabric to the beat of the soft lounge music while his other hand methodically brought his nightcap to his mouth for another sip. I observed him, unseen, from the archway that led into the hotel's lobby.
- (b) He glanced at her nervously and tossed the keys onto the coffee table. " I don't know what Ted Metzler was thinking when he told me this place was in good taste. I wasn't expecting some goddamn Hollywood hotel lobby. Remember, this is just temporary, until we find a house.

On trouvera cependant tout aussi facilement, dans le même corpus, des occurrences du composé dans lesquelles la référence à une catégorie ou classe abstraite est nettement moins perceptible. Dans l'extrait (c) ci-dessous, également tiré du COCA, le déictique montre bien que l'on fait référence à un hôtel précis, celui-là même où se déroule la scène :

(c) It was really quiet a couple of minutes ago Dana, when a very subdued Senator Luther Strange and his wife Melissa stood at a podium that [sic] set up on a fireplace behind me in this hotel lobby to concede the election to Roy Moore.

Il n'est donc probablement pas étonnant que beaucoup d'étudiants se trompent, même si les passages du test proposé avaient été soigneusement sélectionnés. Cette première étape nous a donc permis de nous interroger sur les règles que nous souhaitions évaluer, et nous avons choisi de nous concentrer sur la seule règle qui était susceptible de contraindre les choix de façon simple et univoque : la première que nous avons énoncée ci-dessus.

Le débat qui oppose les modèles d'acquisition basés sur des règles à ceux qui se fondent sur la fréquence a été classiquement abordé à partir des verbes irréguliers, qui par contraste avec les verbes réguliers permettent d'évaluer facilement la prégnance de règles (abstraites, symboliques) ou d'exemplaires dont les productions ou choix préférentiels de formes seraient dérivés par analogie (Ambridge 2010). Des travaux récents apportent un nouveau regard sur ce débat : la régularité pourrait bien correspondre à un facteur auquel les apprenants sont plus ou moins sensibles, au même titre que les effets de fréquence (Domínguez et al. 2018). C'est donc sans exposer les étudiants à un ensemble de règles que nous leur avons proposé un questionnaire présentant des alternances, au sein desquelles il était possible de distinguer un choix fait en fonction de l'orientation fréquentielle ou de la régularité liée aux référents animés (cf. supra, Deuxième partie, 3.2). Le développement conjoint d'une sensibilité aux régularités et à la fréquence que nous observons rejoint les conclusions de Domínguez et al., qui montrent que la variation dépend (notamment) du niveau scolaire (2018 : 7).

Les premiers résultats de notre exploration de cette question montrent qu'on ne peut se contenter de prendre en compte la fréquence, et qu'il faut accompagner aussi la découverte de régularités. Ils plaident donc pour une introduction progressive de règles, qui tienne compte autant que possible de l'orientation fréquentielle des exemples présentés. Les implications didactiques d'une approche probabiliste de la grammaire sont encore peu décrites. En tant qu'elles permettent d'évaluer la pertinence des modèles probabilistes, qui sont actuellement au cœur de l'apprentissage machine (ou apprentissage profond) pour l'apprentissage humain, elles trouvent également leur place au sein d'une conception critique de l'automatisation.

2. Réflexion épistémologique

J'entends par conception critique l'expression d'un point de vue raisonné, qui repose sur une bonne compréhension des transformations actuelles. Ainsi, il ne s'agit pas de critiquer les modèles d'apprentissage profond pour leur caractère empirique, et le manque de fondement théorique de développements qui sont largement basés sur l'optimisation des systèmes (Poibeau 2017 : 189), mais bien plutôt d'en retrouver les fondements épistémologiques pour mieux situer et comprendre les enjeux auxquels nous sommes confrontés.

Je développe actuellement cette réflexion dans le cadre d'un projet d'ouvrage inédit intitulé *La traduction automatique, le linguiste et le traducteur*. Il abordera et situera l'émergence d'un nouveau paradigme, inspiré par l'utilisation de grandes bases de données, fouillées par des méthodes statistiques de plus en plus complexes. Ce tournant des *Big Data*, qui s'appuie sur notre capacité nouvelle à recueillir, traiter et modéliser des masses de données issues de la communication numérique, est en cours et il est encore difficile d'en mesurer toutes les conséquences. Il est toutefois probable qu'il modifie notre conception même de la cognition humaine. En effet, l'efficacité de systèmes de traduction (ou de reconnaissance visuelle ou de la parole) fondés sur l'analyse statistique de données massives suggère que des secteurs importants de la connaissance humaine puissent fonctionner de la même façon, par le traitement analogique et de plus en plus profond des régularités rencontrées dans les flux d'information qui constituent l'environnement de l'homme (Dupoux 2018 ; Becerra-Bonache & Jimenez Lopez 2016).

L'objet de cet ouvrage, qui s'appuie sur l'histoire de la TA et sur le succès actuel de systèmes de TA statistiques puis neuronaux, est de s'interroger sur les implications épistémologiques et sociétales de ce succès. Je reprends ici les principaux points de la première partie (réflexion épistémologique sur les nouveaux modèles) sans en restituer le fonctionnement : chacun des développements proposés y est en effet illustré par un texte de référence que je traduis et commente. La seconde partie présente et analyse plusieurs types de données documentant les usages actuels de la traduction automatique par le traducteur professionnel et l'apprenti traducteur (voir ci-dessus, Deuxième partie, 4.1 et Synthèse, 1.2).

2.1 Des règles aux statistiques : destins croisés de la linguistique et de la traduction automatique

De la grammaire universelle aux grammaires de construction, en passant par le cognitivisme et le connexionnisme, les conceptions cognitives du langage ont opéré plusieurs changements de paradigme et modifié les outils de description et d'analyse linguistique, en même temps que notre compréhension des phénomènes langagiers. La première série de changements

a été abordée dans la partie théorique de ce document : il s'est agi principalement d'interroger les représentations qu'utilisent les modèles linguistiques, psycholinguistiques et traductologiques, en mettant l'accent sur les évolutions parallèles et les points de rencontre. Dans l'ensemble, mes analyses font ressortir une relative autonomie de la traductologie, qui en sortant du « tout linguistique » a surtout montré que les formalismes linguistiques, quels qu'ils soient, n'étaient pas de nature à éclairer notre compréhension des processus et compétences à l'œuvre dans la traduction. La traductologie se situe donc plutôt du côté de ce que Rastier nomme le paradigme « rhétorique-herméneutique » (1998 : 100) aux antipodes du réductionnisme des modèles computationnels.

Les évolutions actuelles de ces modèles opèrent cependant un rapprochement inédit entre le traitement statistique de grandes bases de données et l'activité du traducteur. En particulier, deux types de modélisations non linguistiques actuellement en plein essor permettront peut-être de renouveler notre compréhension des processus de traduction : l'apprentissage statistique, et les modélisations connexionnistes.

L'apprentissage statistique désigne une hypothèse de recherche en psychologie développementale et cognitive, à laquelle les travaux pionniers de Saffran et ses collègues (Saffran, Aslin et Newport 1996) ont apporté une validation empirique forte. A partir d'expériences basées sur des suites de pseudo-mots de trois syllabes, ces chercheurs ont montré que des nourrissons de 8 mois étaient capables de segmenter un flux (signal acoustique) continu en se basant uniquement sur les probabilités de transition d'une syllabe à l'autre. Par exemple pour « rebidakupadotigolabubidaku.... », la probabilité de transition de « bida » était de 1 alors que celle de « kupa » (qui se trouve à cheval sur deux pseudo-mots) était de 0,33. Ils ont ainsi mis en évidence une aptitude fondamentale et jusque-là méconnue :

la capacité d'utiliser les informations distributionnelles présentes dans l'input, qui joue un rôle moteur dans l'acquisition du langage, et qui implique une série de calculs à partir d'unités ou de patrons (par ex. des sons, des syllabes, ou des catégories syntaxiques). (Rebuschat 2015 : xv, notre traduction⁸⁴)

En tant qu'il définit une capacité humaine et éclaire l'acquisition du langage, l'apprentissage statistique rejoint l'apprentissage implicite et les dispositifs expérimentaux utilisés sont bien souvent similaires, si bien que plusieurs auteurs ont proposé de parler d'apprentissage statistique implicite (voir par ex. Conway et al. 2010). En statistique, l'apprentissage a commencé à intéresser

⁸⁴ "Statistical learning, i.e. our ability to make use of distributional information in the input to bootstrap language acquisition, involves computations based on units or patterns (e.g. sounds, syllables, syntactic categories)."

les scientifiques dès les années 1960 et sa théorisation a marqué un tournant (Vapnik 2000 : x) : l'apprentissage statistique désigne donc aussi un champ de recherche très porteur⁸⁵, désormais souvent rattaché au paradigme de l'intelligence artificielle. Il existe ainsi deux communautés, bien distinctes, qui partagent le même intérêt pour l'apprentissage non supervisé, c'est-à-dire pour la découverte de catégories et de structures dans un input qui peut être linguistique. Ces deux communautés ont des objectifs également dissemblables : alors que l'apprentissage automatique concerne surtout le développement d'applications (technologies linguistiques, si l'on considère le travail qui est fait à partir d'un input linguistique) les chercheurs en sciences cognitives cherchent à mieux comprendre le fonctionnement de l'esprit humain. La rencontre de ces deux communautés n'en reste pas moins remarquable, et la convergence est bien réelle, tant du point de vue des questions de recherche que des données utilisées⁸⁶. Suite aux interactions les plus récentes entre ces deux communautés, on peut espérer que des cercles vertueux se mettent en place, qui permettent un questionnement pluridisciplinaire sur les modèles et les méthodes, et une réflexion sur le rôle de l'humain en début comme en bout de chaîne.

Les apports de ces synergies à notre compréhension des processus de traitement du langage sont d'ores et déjà considérables : le passage des modélisations basées sur des règles à des modèles statistiques, qui caractérisent par exemple les grammaires de constructions, a permis de mettre en évidence des régularités qui n'étaient pas apparues jusqu'ici, et les modèles d'apprentissage statistique ont renouvelé notre compréhension des débuts de l'acquisition du langage. Les évolutions les plus récentes de l'apprentissage automatique, devenu apprentissage profond (*deep learning*) sont à ce titre porteuses d'espoirs parfois démesurés. Ici encore, il me semble qu'on ne peut se dispenser d'une réflexion épistémologique, pour se défaire des discours sur le miracle technologique et envisager les intérêts et les limites des progrès actuels.

2.2 Des modèles connexionnistes aux réseaux de neurones

Les réseaux connexionnistes ont été longtemps vus comme une modélisation alternative, plus sophistiquée mais moins puissante que celles que mettaient en avant les cognitivistes. Ce sont pourtant des modèles informatiques basés sur des réseaux connexionnistes qui ont permis

⁸⁵ En témoignent notamment les conférences annuelles StatLearn organisées par la Société française de statistique. L'édition 2017 a eu lieu à Lyon et j'ai eu la chance de pouvoir y assister pour partie.

⁸⁶ On peut citer par exemple l'atelier intitulé « Zero Resource Speech Technologies and Models of Early Language Acquisition » qui a eu lieu aux États-Unis en 2013 (cf. <https://www.clsp.jhu.edu/workshops/12-workshop/zero-resource-speech-technologies-and-models-of-early-language-acquisition/>) ou encore la conférence sur l'« Apprentissage naturel et artificiel » qui a eu lieu à Paris les 5 et 6 juillet 2018.

d'obtenir depuis la fin de l'année 2016 des résultats inégalés en traduction automatique, grâce au développement conjoint de processeurs de plus en plus puissants et de grandes bases de données.

Le succès de la TA neuronale est issu du passage de l'apprentissage implicite et statistique à sa systématisation dans des réseaux connexionnistes. Pour expliquer le fonctionnement de ces nouveaux systèmes, je rappellerai donc brièvement ce qu'est le connexionnisme, et quels en ont été les principaux développements, avant de définir trois notions clés de ce nouveau paradigme : les réseaux neuromimétiques, l'apprentissage profond, et les représentations avec lesquelles les systèmes opèrent, souvent appelées plongement (*embedding*).

Même si elles ne représentent qu'une approche connexionniste parmi d'autre (celle qui repose sur le traitement parallèle distribué : *parallel distributed processing*) les propositions de Rumelhart et McClelland (1986) sur l'apprentissage du passé des verbes anglais ont fait date ; elles sont d'ailleurs publiées dans un ouvrage dont la publication a conféré au connexionnisme une « notoriété considérable » (Andler 1990 : 95, note 1). Ils y présentent un système qui, à la suite d'un apprentissage supervisé, est capable de produire le passé à partir de l'infinitif d'un verbe donné⁸⁷. Daniel Andler donne une description très claire du fonctionnement d'un tel système :

La tâche consiste à compléter une configuration dont l'environnement ne fournit qu'une partie. Au contact d'un grand nombre d'exemples de configurations complètes [...] le système s'adapte aux régularités de l'environnement en ajustant ses poids synaptiques, ce qui lui permet d'une part de réagir sans aucune erreur aux exemples qui lui ont été présentés au cours de l'apprentissage, et d'autre part de réagir « intelligemment » à des configurations incomplètes - soit en les assimilant à des parties de configurations connues, soit en discernant un mélange de configurations connues, et en les complétant en conséquence. Bref, le système se comporte en détecteur de régularités statistiques multidimensionnelles (*ibid.* p. 103).

Les réseaux neuromimétiques qui sont capables de tels apprentissages peuvent être de plusieurs types, mais tous sont faits d'unités interconnectées dont l'activation est fonction de celle des unités auxquelles elles sont liées (les mesures d'activation étant multipliées par le poids des connexions). C'est ainsi que « dans la plupart des architectures de réseaux neuromimétiques, les activations de neurones individuels ne sont pas pertinentes en elles-mêmes, mais en lien avec

⁸⁷ Supervisé veut dire que les réponses du système sont corrigées pour établir à chaque fois la forme correspondant à l'infinitif proposé, jusqu'à obtention de résultats satisfaisants.

celles des autres neurones (Forcada 2017 : 293, notre traduction⁸⁸). Même si les poids sont calculés, ce que l'on appelle l'apprentissage profond n'est pas un apprentissage statistique, au sens où il n'est plus basé sur des structures ou des connaissances préétablies. Celles-ci émergent lors de l'apprentissage, qui permet un « assemblage hiérarchique » de l'information (Poibeau 2017 : 183 et sq.). Ce sont ces structures hiérarchiques ajustables qui permettent aux systèmes de TA neuronale de tenir compte de contextes à dimension variable, à la différence de la TA statistique, dont les modèles les plus courants se fondent sur des groupes de mots contigus (*phrase-based statistical machine translation*). Les représentations obtenues sont multidimensionnelles, et on parle de plongement lexical (*word embedding*) pour désigner « les mots qui apparaissent dans des contextes traductionnels semblables » (*ibid.* p. 186). Le plongement permet la prise en compte de segments bien plus longs que ne pouvaient le faire les modèles statistiques, puisque chaque unité reçoit en fait une représentation vectorielle qui permet la prise en compte d'informations beaucoup plus riches que les probabilités d'occurrences de mots ou groupes de mots (Forcada 2017 : 293 et sq.). Ce fonctionnement explique cependant aussi la production d'erreurs inattendues, en particulier dans le traitement des phrases longues et complexes. Dans l'ensemble, les évaluations humaines qui comparent la TA neuronale à des systèmes statistiques montrent une amélioration de la fluidité (*fluency*) mais pas de gain significatif des mesures d'adéquation (*adequacy*) si bien que ces progrès ne constituent pas nécessairement un gain pour le traducteur, au sens où ils ne permettront pas toujours d'atténuer les efforts du traducteur (*ibid.* p. 306).

Il faut cependant rappeler que ces mesures subjectives, qui sont de plus en plus fréquemment utilisées aux côtés de l'évaluation automatique, sont encore peu satisfaisantes parce que trop souvent réalisées sur des phrases hors contexte. Tributaires d'outils qui permettent notamment d'obtenir des mesures chronométriques comparables d'une étude à l'autre, elles ne restituent pas la complexité du travail du traducteur. Il manque donc, face aux progrès importants qui ont été accomplis ces dernières années, une implication des traducteurs permettant d'effectuer un travail d'évaluation plus riche et plus qualitatif. Il importe par conséquent que les études sur le traducteur (celles que Chesterman a appelées *translator studies*⁸⁹) parviennent à promouvoir l'intégration du regard et du point de vue du traducteur sur ces données. Or, intégrer aux travaux sur la TA des analyses situées, ou socio-cognitives, suppose aussi que l'on repense le rôle et la place du traducteur dans des flux d'activité complexes.

⁸⁸ In most neural network architectures, the activations of individual neurons do not make sense by themselves, but rather when grouped with the activations of the other neurons,

⁸⁹ Chesterman 2009.

3. Construire une critique de l'automatisation

Pour mettre ensemble toutes les pièces du puzzle ainsi esquissé, je me suis engagée dans des projets qui se situent le plus souvent à l'intersection de plusieurs champs disciplinaires, et au sein desquels s'organise une réflexion théorique appuyée sur des dispositifs empiriques et/ou sur l'analyse des pratiques.

Les orientations et les questionnements qui caractérisent chacun de ces projets sont ici présentés comme autant d'étapes dans la construction d'une critique de l'automatisation. Je me suis attachée, dans ce document de synthèse, à montrer comment la nécessité d'une telle critique avait émergé dans mes travaux, et à retisser les liens qui la rattachent à mes publications antérieures. Je propose donc ci-dessous une conclusion en quatre temps, qui annonce la suite de mes travaux telle que je l'envisage actuellement.

3.1 Repenser le rôle de l'utilisateur : le projet M2TLabs

Les traducteurs spécialisés sont confrontés à de profonds défis liés aux pratiques commerciales contemporaines et aux technologies numériques. On leur demande de plus en plus de corriger (ou «post-éditer») des sorties de traduction automatique plutôt que de traduire de toutes pièces. La collecte automatisée de traductions humaines sur ces plateformes peut conduire à déposséder le traducteur humain de ces données, malgré les apparents droits de propriété intellectuelle (Commission européenne 2014), et les configurations technologiques particulières avec lesquelles travaillent les traducteurs peuvent entraîner des problèmes ergonomiques (Ehrensberger-Dow et al. 2016). Malgré tous ces défis, la traduction est toujours perçue comme une profession qui offre de «brillantes perspectives». Elle est moins sujette au chômage technologique que d'autres secteurs et devrait connaître une croissance non négligeable au cours de la prochaine décennie (Frey et Osborne 2016, Bureau of Labor Statistics 2016).

Face à ce paradoxe, nous avons proposé dans le cadre de ce projet de placer les formations de traducteurs de l'espace européen au cœur d'une dynamique innovante qui permette de valoriser l'expertise du traducteur et de repenser les interactions homme-machine. En particulier, et dans la continuité des premières applications à l'enseignement que j'ai décrites ci-dessus (Deuxième partie, 4.3), nous souhaitons proposer et évaluer de nouveaux modèles de formation de traducteurs qui préparent au mieux les étudiants à ces défis. Les travaux pionniers de l'équipe de Dublin City University ont permis de concevoir et de mettre en œuvre un cours de traduction automatique statistique (SMT) à destination des apprentis traducteurs (Doherty et Kenny 2014). L'initiative a eu des répercussions importantes, car elle a permis de sensibiliser la communauté sur la nécessité de mieux comprendre les bouleversements en cours, pour mieux les enseigner.

Des enquêtes à grande échelle sur les compétences en traduction automatique (Gaspari et al. 2015) et des propositions de nouveaux cours ont ainsi été faites (Koponen 2015, Rico 2017, [Rossi-2017a]). Il existe aussi désormais des outils dédiés à la formation des étudiants à l'utilisation de la TA (Bouillon et al. 2017, mais aussi des solutions développées grâce aux partenariats universitaires que proposent plusieurs fabricants de moteurs). À notre connaissance, cependant, les propositions de cours sur la traduction automatique neuronale (NMT) pour les traducteurs sont encore embryonnaires. Il y a là un défi d'autant plus important que la NMT est une technologie difficile, qui diffère notamment de la SMT en ce sens que de nombreux aspects du traitement de l'information restent opaques (Kenny 2017).

M2T Labs cherche à relever ce défi de deux façons. Il s'agit principalement de savoir comment intégrer les dernières technologies de NMT aux flux de traduction (une préoccupation très actuelle au sein de la communauté de la traduction automatique) et de développer une pédagogie adaptée à celles-ci. Ce faisant, nous espérons renouveler la compréhension des facteurs qui aident les traducteurs professionnels à accepter les technologies de traduction et à travailler avec ceux-ci de manière à augmenter leurs capacités sans mettre en péril leur bien-être.

Les laboratoires proposés sont pensés sur le modèle des séances de traduction intensives (« Tradutech ») telles qu'elles ont été conçues et déployées à grande échelle dans le projet européen OCTC (<http://www.otct-project.eu/fr/about/abstracts/>). Il s'agit donc de proposer aux étudiants un travail sur projet (simulé) qui serve de cadre à l'évaluation de différentes solutions de traduction automatique, puis au choix et à la mise en œuvre de l'une d'elles. Des ateliers de créativité doivent ensuite permettre aux étudiants d'analyser leurs choix et de faire des propositions pour améliorer les flux qu'ils auront évalués. Cette dernière proposition correspond à une possibilité que j'ai identifiée, avec l'aide du programme Promising de l'Université Grenoble Alpes, comme susceptible de sensibiliser les étudiants à la possibilité d'une participation plus active du traducteur aux développements qui sont en cours. Pensant que ces ateliers pourraient jouer un rôle facilitateur (Cruickshank et Evans 2012) et constituer une expérience intéressante, je me suis formée aux méthodes de créativité (voir ci-dessous, Annexe, 5.2) et serai à même d'animer un premier atelier au printemps 2019. La mise en œuvre partielle de la méthodologie que nous avons proposée dans la réponse à l'appel à projet permettra d'évaluer le dispositif pour faire une nouvelle proposition plus convaincante, en même temps que de recueillir des données de perception.

3.2 Mieux articuler cognition individuelle et comportements collectifs : le projet Co3i

Les interactions homme-machine sont aussi des interactions sociales : prendre en main un moteur de traduction automatique dans le cadre d'un projet de traduction implique donc aussi que l'on s'interroge sur l'usage qui est fait des données du traducteur. De même, les progrès de l'apprentissage profond au sein de systèmes entraînés sur des grandes bases de données ne constituent pas uniquement une prouesse technique ; elles s'appuient aussi sur un travail qu'elles effacent en l'intégrant : celui du traducteur. Enfin, les textes ne sont pas des échantillons massifs et aisément dissous dans une masse plus grande ; ils se lisent à plusieurs niveaux qui correspondent d'abord à ceux de l'énonciation, et qu'il faudrait pouvoir conserver dans des corpus structurés, ou distinguables :

Un texte ne saurait être assimilé à une masse de connaissances directement exploitable par la machine. Il faut dans un premier temps prévoir des traitements complexes pour identifier l'information pertinente, la normaliser, la catégoriser et éventuellement la mettre en contexte. (Poibeau, 2014: 27).

Ma participation au projet Co3i m'a permis de reprendre la question du sujet parlant et de l'intégrer à un questionnement plus large sur les échelles et les niveaux de granularité : le langage et la cognition se situent à l'articulation de l'individuel et du collectif, mais cette distinction est-elle tenable ? Que l'on adopte le point de vue saussurien de la langue comme système de signes et comme savoir partagé, ou celui, chomskyen, qui situe la compétence dans l'individu tout en ne retenant de celui-ci que des caractéristiques supposées universelles, la vision du collectif que l'on convoque est extrêmement homogène, et elle intègre une multitude indéfinie d'individus idéalisés.

Avec l'essor de la sociolinguistique, les limites de cette idéalisation ont été mises en exergue (Baley et al. 2013) en même temps que la psycholinguistique battait en brèche les hypothèses modulaires qui fondaient de telles conceptions. Les linguistiques basées sur l'usage, quant à elles, ont permis de penser l'émergence du langage sans postuler de structures universelles (Bybee et McClelland 2005), mais en s'appuyant sur la montée en puissance de la linguistique de corpus, indissociable de la multiplication de grandes bases de données documentant les usages linguistiques. Certaines différencient les locuteurs enregistrés ou les différentes variétés d'une même langue et d'autres les amalgament (Gries 2015 : 99; Laks 2013) : elles s'inspirent alors d'avantage de l'analyse du discours et des méthodes interactionnistes. Au total, des synergies importantes se font jour entre les différents champs disciplinaires dont l'objet d'étude est la communication (de l'anthropologie aux neurosciences cognitives, et à la traductologie) et bouleversent l'étude du langage. Ainsi que

je l'ai proposé dans la première partie de ce document, ces changements sont pour moi autant d'incitations à réfléchir à nouveau à la définition du sujet parlant.

La réflexion peut s'organiser selon trois axes principaux. Elle consistera d'abord à se demander comment sont pris en compte le locuteur individuel et la communauté linguistique dans la linguistique de corpus, et dans les applications d'intelligence artificielle basées sur l'apprentissage machine à partir de grands corpus. Le deuxième axe interrogera ensuite la portée des changements en cours, en examinant la question suivante : comment sont conçus le social et le cognitif dans les nouveaux champs à l'intersection de la psycholinguistique et de la sociolinguistique (Chevrot et Nardy, à paraître) ? Pour finir, on analysera le cas des grammaires de construction, en cherchant à définir la conception du cognitif et du social, du collectif et de l'individuel qui les sous-tend. Ces grammaires modélisent-elles le fonctionnement individuel ou le savoir de la communauté ? Et dans quelle mesure tiennent-elles la promesse qu'annonce leur « programme socio-pragmatique » (Desagulier 2011), celle d'intégrer les aspects sociaux du langage à la description linguistique ? Nourries par la réflexion pluridisciplinaire qu'organise ce projet, ces questions seront abordées tour à tour dans la préparation d'un chapitre co-écrit qui s'intégrera à un ouvrage collectif intitulé : *Cognition individuelle et connaissances collectives : changements conceptuels et apports interdisciplinaires*.

3.3 Développer une théorie du contexte : le projet CL3i

Ce projet pluridisciplinaire, porté par un philosophe de l'Université Grenoble Alpes (Denis Perrin) s'ancre dans ce que D. Andler a appelé une pratique réaliste de la philosophie des sciences (souvent aussi qualifiée de naturaliste) pour interroger le contexte linguistique du point de vue des modèles qui l'ont théorisé, mais aussi de ses fondements cognitifs et de son opérativité informatique :

Cette pratique, qu'on nomme aussi naturaliste en un sens particulier mais parfaitement régulier du terme, consiste à examiner la science et les diverses sciences en se déprenant autant que possible de tout projet normatif. (Andler 2016 : 16).

En repartant de la dépendance contextuelle telle qu'elle a été théorisée en pragmatique du langage, et telle qu'elle est utilisée par plusieurs courants de la linguistique contemporaine, le projet fait sien le constat de Finkbeiner, qui propose que nous ne manquions pas seulement d'une définition du contexte mais d'une théorie unifiée et susceptible de faire consensus :

La dépendance contextuelle est une notion centrale pour toutes les linguistiques empiriques, ce qui suggère que la communauté des linguistes dispose au moins d'une compréhension fondamentale de ce qu'est un contexte, sur laquelle tous s'accordent. Or même si les tentatives d'expliquer la notion de contexte en se focalisant sur un aspect de celle-ci ne manquent pas, il s'agit de tentatives

variées qu'il n'est pas évident de faire converger pour aboutir à une théorie unifiée du contexte. (Finkbeiner 2012: 1, notre traduction⁹⁰)

Le projet propose de mettre en œuvre une démarche que l'on pourrait nommer « pragmatique naturalisée », qui consistera à croiser les méthodes expérimentales de la psycholinguistique et les contraintes d'opérativité que font ressortir les réalisations informatiques pour bâtir plusieurs modèles du contexte linguistique. C'est à partir de ces modèles qu'une définition intégrée de la notion de contexte pourra être proposée, qui éclairera la nature et la composition du contexte en même temps que ses opérations, c'est-à-dire la détermination du sens qu'il opère. Les phénomènes qu'une telle définition est susceptible de nous aider à aborder sont variés : ils comprennent notamment la présupposition, la deixis, l'anaphore, ou encore le contenu et le format de la signification lexicale des expressions linguistiques.

Suite à l'obtention d'un financement de l'IDEX grenoblois au titre des « Initiatives de Recherche Stratégiques », une post-doctorante très polyvalente travaillera dès l'automne 2018 en lien avec les différentes équipes qu'il fédère. Ma participation s'inscrit dans le prolongement du projet « Evaluer la traduction automatique » [P6], et me permettra de poursuivre avec la même équipe pluridisciplinaire la réflexion que nous avons commencée. Nous interrogerons l'intégration du contexte dans les nouveaux systèmes de TA neuronale, à partir de deux dispositifs expérimentaux. Le premier concerne l'évaluation de la performance des systèmes avec et sans adaptation au domaine (il consistera à comparer évaluation automatique et évaluation humaine dans quatre domaines distincts) et le second se placera du point de vue de l'utilisateur des sorties, en comparant des évaluations produites sur des phrases isolées à celles qui peuvent être obtenues suite à une mise en situation (explication de la commande et de la fonction du texte). En combinant ces deux dispositifs, nous espérons montrer les limites de l'évaluation des systèmes telle qu'elle est actuellement pratiquée, et faire ressortir la spécificité de l'évaluation humaine.

Voici, pour finir, un exemple du matériau linguistique sur lequel nous travaillerons, choisi à dessein pour montrer que ce travail sera également lié à la réflexion sur l'anaphore que j'ai entamée avec L. Gardelle et L. Vincent-Durox [Rossi et al. 2019c] :

[Extrait source.] Wash a bunch of fresh spinach and then shred it finely. Sauté it in a little butter until it is wilted, drain, then put a little into each ramekin.

⁹⁰ "Since context dependency is a core notion of linguistics as an empirical science, one would suspect that there is at least a shared, basic understanding in the linguistic community of what a context is. Yet while there are numerous attempts at explaining single aspects of the notion of context, these attempts are quite diverse and do not easily converge to a unified theory of context."

[Sortie Google Translate⁹¹.] Laver un bouquet d'épinards frais et déchiqueter finement. Faites-le sauter dans un peu de beurre jusqu'à ce qu'il soit fané, égouttez-le, puis mettez-le un peu dans chaque ramequin.

[Traduction de référence proposée] Lavez un bouquet d'épinards frais et ciselez les branches. Mettez une noix de beurre dans une poêle et faites-y sauter les épinards jusqu'à ce qu'ils soient bien tendres, égouttez puis répartissez en petite quantité dans vos ramequins.

Cet extrait d'une recette de cuisine nous montre un problème classique du fonctionnement anaphorique : celui des référents évolutifs. L'anaphorique *it* n'opère pas une reprise stricte du bouquet mais fait plutôt référence à l'ingrédient. Cette évolution est très claire à la fin de l'extrait lorsque le prélèvement de petites quantités est évoqué, puisque le quantifieur *a little* se rapporte nécessairement à un indénombrable. J'ai proposé une traduction de référence qui met en évidence le référent évolutif et la cohésion textuelle que garantit le choix homogène du mode (infinitif ou impératif) dans ce type de texte. L'écart avec la sortie de TA est alors assez franc, mais on pourra par exemple se demander dans quelle mesure ces éléments seraient pris en compte par l'évaluation automatique (qui repose souvent sur plusieurs traductions de référence) et comment ils seraient intégrés à une mesure d'adéquation.

3.4 Interroger nos pratiques : l'éthique de la recherche

Les données constituent une source de questionnement éthique autant que scientifique, et les évolutions actuelles, qui permettent de développer des bases de données de très grande taille, s'accompagnent d'un encadrement accru, marqué notamment par de nouvelles réglementations (RGPD). La distinction de plus en plus franche des recherches interventionnelles et non-interventionnelles, conjuguée à l'encadrement de ces deux types de recherches par des comités d'éthique et des comités de protection des personnes, marque un tournant qui rend moins évidente notre affirmation de 2008 : « De nos jours, les chercheurs qui travaillent sur des suivis longitudinaux d'enfants en milieu naturel peuvent filmer les enfants chez eux. » [Rossi & Morgenstern 2008 : 1036]. Il existe donc un impact de cet encadrement sur les pratiques des chercheurs : il est probable que, pour être sûrs de rester dans le cadre d'une recherche non interventionnelle (puisque c'est bien ainsi que s'entendent les suivis longitudinaux visant à documenter l'acquisition du langage), on préfère désormais laisser aux familles le soin de filmer leurs enfants. Si tel était le cas, il faudrait tenir compte de l'effet de telles variations du dispositif sur les données recueillies. Cet exemple issu d'une conversation récente avec une collègue montre

⁹¹ Cet exemple ne correspond pas au dispositif que nous envisageons, car nous avons prévu d'utiliser en priorité le système neuronal développé au LIG (Seq2seq) et de choisir des textes dans les domaines pour lesquels une adaptation a été préparée.

que le choix de contrôler et d'encadrer n'est pas neutre, au même titre d'ailleurs qu'une recherche fondamentale, qui pourrait se sentir peu concernée par une réflexion sur ses impacts, pourra occasionner des effets qu'elle n'avait pas escomptés.

Mes interrogations actuelles sur nos pratiques de chercheurs se développent sur deux fronts : le premier correspond à mon implication au sein du Pôle Grenoble Cognition, et le second à la direction éditoriale que j'assume.

Les rencontres annuelles du Pôle Grenoble Cognition se sont cette année déroulées sur le thème des impacts divers que peuvent avoir les recherches en sciences cognitives⁹² et nous avons travaillé à en faire un dialogue sur des questions éthiques, avec le soutien de philosophes grenoblois et la participation d'une invitée lyonnaise très active dans ce domaine (Sarah Carvalho, voir par ex. Carvalho et Roux 2007). Ces journées se distinguaient des précédentes, en proposant un thème auquel la plupart des chercheurs de la communauté n'étaient pas habitués. Nous avons choisi d'aborder cette nouveauté avec prudence, en distinguant les présentations du matin du dialogue organisé en petits groupes l'après-midi, et en sollicitant une participation des chercheurs qui présentaient leurs travaux le matin sur la base du seul volontariat. Ces journées ont rencontré un franc succès et la discussion se poursuivra désormais lors de plusieurs rendez-vous annuels, dont le format sera semblable.

En acceptant de reprendre la direction éditoriale de la revue de linguistique cognitive *CogniTextes*, j'ai également été confronté à des questions parfois difficiles, qui ont révélé la nécessité de mieux encadrer les pratiques. J'ai fait récemment deux propositions qui ont été acceptées par le comité éditorial et seront mises en œuvre rapidement. La première concerne la mise en place d'une charte éthique pour cette revue, c'est-à-dire d'un document de cadrage, publié sur le site de la revue, auquel nous puissions nous référer lorsque nous sommes confrontés à des cas comme la soumission d'une proposition à plusieurs revues en même temps, ou encore le dépôt d'un article déjà publié ailleurs sous un autre titre. J'ai fait cette proposition en espérant qu'elle nous épargnerait une partie au moins de ces problèmes, tout en sachant bien qu'une vigilance accrue ne suffirait pas à enrayer les effets néfastes des pressions que subit l'ensemble de la communauté universitaire pour augmenter la quantité des travaux qui sont publiés. L'autre type d'action que nous souhaitons mener doit permettre de mieux encadrer la mise en œuvre de dispositifs expérimentaux, auxquels les linguistes ne sont pas toujours habitués. Elle consiste à encourager les auteurs à publier également les données qu'ils ont utilisées pour produire l'article qu'ils

⁹² Il s'agissait des 9^{èmes} journées du Pôle, qui ont eu lieu le 5 juin 2018 (<http://www.grenoblecognition.fr/index.php/animation/journees-du-pole/59-les-journees/463-9e-journee-du-pole-grenoble-cognition-5-juin-2018>).

soumettent à la revue. Plusieurs collègues ont déjà travaillé à la mise en place d'une archive internationale librement accessible (Conzett et al. 2014), où chaque dépôt est consigné avec des métadonnées qui permettent aux chercheurs de parcourir facilement la base : il s'agit de TROLLing (The Tromsø Repository of Language and Linguistics). Créée en 2014, l'archive rassemble déjà 60 bases de données et 560 fichiers, qui ont fait l'objet de plus de 3000 téléchargements. Outre le fait qu'il facilite le dialogue entre chercheurs et encourage la publication de travaux répliquables, ce choix montre que tout en gardant une grande ouverture sur l'ensemble des linguistiques cognitives telles qu'elles sont actuellement utilisées en France et à l'étranger, notre revue a su prendre acte du tournant quantitatif récent.

Bibliographie

- AIJMER K & RÜHLEMANN C (Éd.). (2015). *Corpus Pragmatics: A Handbook*. Cambridge University Press.
- ALONSO E & CALVO E. (2015). Developing a Blueprint for a Technology-mediated Approach to Translation Studies. *Meta : Journal Des Traducteurs / Meta : Translators' Journal*, Vol. 60, No. 1, p. 135-157.
- AMBRIDGE B. (2010). Children's judgments of regular and irregular novel past-tense forms: New data on the English past-tense debate. *Developmental Psychology*, Vol. 46, No. 6, p. 1497-1504.
- AMBRIDGE B. (2017). Syntactic Categories in Child Language Acquisition: Innate, Induced, or Illusory? In Cohen et Lefebvre (Eds.) *Handbook of Categorization in Cognitive Science*, p. 567-580.
- ANDLER D. (2016). *La silhouette de l'humain. Quelle place pour le naturalisme dans le monde d'aujourd'hui ?* (Gallimard).
- ANDLER D. (1990). Connexionnisme et cognition: À la recherche des bonnes questions. *Revue de synthèse*, Vol. 111, No. 1-2, p. 95-127.
- ANDLER D, COLLINS, THÉRÈSE, & TALLON-BAUDRY, CATHERINE. (2018). *La cognition. Du neurone à la société*.
- ARTHUR P, NEUBIG G, & NAKAMURA S. (2016). Incorporating Discrete Translation Lexicons into Neural Machine Translation. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing* (p. 1557-1567). Austin, Texas, Association for Computational Linguistics.
- AUROUX S. (1998). Les enjeux de la linguistique de terrain. *Langages*, Vol. 32, No. 129, p. 89-96.
- BAAYEN R. HARALD. (2010). Demythologizing the word frequency effect: A discriminative learning perspective. *The Mental Lexicon*, Vol. 5, No. 3, p. 436-461.
- BAKER M. (1993) Corpus linguistics and translation studies: Implications and applications. In M. Baker et al. (eds.), *Text and Technology*, p. 233-250. John Benjamins.

- BAKER, M. (1995). Corpora in translation studies: An overview and some suggestions for future research. *Target. International Journal of Translation Studies*, Vol. 7, No. 2, p. 223-243.
- BAKER P, GABRIELATOS C, KHOSRAVINIK M, KRZYŻANOWSKI M, MCENERY T, & WODAK R. (2008). A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press. *Discourse & Society*, Vol. 19, No. 3, p. 273-306, mai.
- BARONI M. & KILGARRIFF, A. (2006). Large linguistically-processed web corpora for multiple languages. In *Proceedings of the Eleventh Conference of the European Chapter of the Association for Computational Linguistics: Posters & Demonstrations* p. 87-90. Association for Computational Linguistics.
- BARONI M. & BERNARDINI S. (2004). BootCaT: Bootstrapping corpora and terms from the web. *Proceedings of LREC 2004*.
- BAYLEY R, CAMERON R, & LUCAS C (Éd.). (2013). *Oxford Handbook of Sociolinguistics*. Oxford University Press.
- BECERRA-BONACHE, L. & JIMÉNEZ-LÓPEZ, M.D. (2016). Could Machine Learning Shed Light on Natural Language Complexity? *Proceedings of the Workshop on Computational Linguistics for Linguistic Complexity*, p. 1-11.
- BENAZZO S. (2009). Introduction : l'intérêt de croiser différents types d'acquisition. *Acquisition et interaction en langue étrangère*, No. Aile/ Lia 1, p. 7-15.
- BENEL A. & LACOUR P. (2014). Towards a participative platform for cultural texts translators. *Cross-Cultural Interaction: Concepts, Methodologies, Tools, and Applications*, p. 1695-1705. IGI Global.
- BERETTA, ILARIA. (2012). Some Highlights on the Concept of Environmental Justice and its Use. *e-cadernos ces*, Vol. 17.
- BERMAN A. (1988). De la translation à la traduction. *TTR : traduction, terminologie, rédaction*, Vol. 1, No. 1, p. 23.
- BERNARDINI S. & CASTAGNOLI S. (2008). Corpora for translator education and translation

- practice. *Topics in language resources for translation and localisation*, p.39-55.
- BIBER D. & ET AL. (1999). *Longman Grammar of Spoken and Written English*. Longman.
- BIBER D. (1993). Representativeness in Corpus Design. *Literary and Linguistic Computing*, Vol. 8, No. 4, p. 241-257.
- BIROS C. (2014). Les couleurs du discours environnemental. *Mots. Les langages du politique*, Vol. 105, No. 2, p. 45-66.
- BLANCHET P. (2012). *La linguistique de terrain*. Presses Universitaires de Rennes.
- BLOOM L, LIGHTBOWN P, HOOD L, BOWERMAN M, MARATSOS M, & MARATSOS M P. (1975). Structure and Variation in Child Language. *Monographs of the Society for Research in Child Development*, Vol. 40, No. 2, p. 1, mai.
- BOD R, HAY J, & JANNEDY S (Éd.). (2003). *Probabilistic linguistics*. Cambridge, MA, MIT Press.
- BOISSEAU M. (2016). Lire et relire Jacqueline Guillemin-Flescher. In M. Boisseau, C. Chauvin, C. Delesse, & Y. Kerommes (Éd.), *Linguistique et traductologie. Les enjeux d'une relation complexe*. Artois Presses Université.
- BOISSEAU M, CHAUVIN C, DELESSE C, & KEROMNES Y. (2016). *Linguistique et traductologie. Les enjeux d'une relation complexe*. Artois Presses Université.
- BORGWALDT S. & BENCZES R. (2011). Word-formation patterns in a cross-linguistic perspective. In D. Schönefeld (Éd.), *Converging Evidence: Methodological and Theoretical Issues in Linguistic Research*. John Benjamins.
- BOTTINEAU D. (2004). Traductologie, linguistique et cognition : les procédés de traduction comme correction des écarts typologiques entre l'anglais et le français. In Ballard M. & Hewson L. (Eds.) *Correct, incorrect* (p. 109-122). Artois Presses Université,
- BOUILLON P, GIRLETTI S, & ESTRELLA P. (2017). MTTT–Machine Translation Training Tool: A tool to teach MT, Evaluation and Post-editing. Présenté à The 20th Annual Conference of the European Association for Machine Translation.
- BOWERMAN M. (2011). Linguistic typology and first language acquisition. In J. J. Song (Éd.), *The Oxford handbook of linguistic typology* (p. 591–617). Oxford University Press.

- BOWKER L, (2011). Off the record and on the fly: Examining the impact of corpora on terminographic practice in the context of translation. In A. Kruger, K. Wallmach & J. Munday (Eds.): *Corpus-based Translation Studies: Research and Application*, p. 211-236. Continuum.
- BULLARD, R. D. (1990). *Dumping in Dixie: Race, class, and environmental quality*. Westview.
- BUREAU OF LABOR STATISTICS (2016) U.S. Department of Labor, Occupational Outlook Handbook, 2016-17 Edition, 'Interpreters and Translators', <https://www.bls.gov/ooh/media-and-communication/interpreters-and-translators.htm>
- BYBEE J, & MCCLELLAND J L. (2005). Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition. *The Linguistic Review*, Vol. 22, No. 2-4, p. 381-410.
- CADWELL P, O'BRIEN S, & TEIXEIRA C S C. (2018). Resistance and accommodation: factors for the (non-) adoption of machine translation among professional translators. *Perspectives*, Vol. 26, No. 3, p. 301-321.
- CARDINI F. (2008). Manner of motion saliency: An inquiry into Italian. *Cognitive Linguistics*, Vol. 19, No. 4, p. 533-569.
- CARRERES Á, MUÑOZ-CALVO M, & NORIEGA-SÁNCHEZ M. (2017). Translation in Spanish language teaching: the fifth skill/La traducción en la enseñanza del español: la quinta destreza. *Journal of Spanish Language Teaching*, Vol. 4.
- CARTON F, NARCY-COMBES J.-P. NARCY-COMBES M.-F. & TOFFOLI D. (2015). *Cultures de recherche en linguistique appliquée*. Riveneuve éditions.
- CARVALLO S, & ROUX S. (2007). *Du nouveau dans les sciences*. Vrin.
- CHABERT E. (2011). Le génitif anglais et la métaphore adamczewskienne de soudure. Exemples et contre-exemples. *Anglophonia. French Journal of English Linguistics*, No. 15 (30), p. 209-224, novembre.
- CHENU F. & JISA H. (2009a). Reviewing some similarities and differences in L1 and L2 lexical development. *Acquisition et Interaction En Langue Étrangère*, No. 1 (Aile... Lia), p. 17-38.

- CHEU F. & JISA H. (2009b). Les rapports entre méthodologie et théorie : le cas des corpus en acquisition. *Cahiers de linguistique. Revue de sociolinguistique et de sociologie de la langue française*, Vol. 33, No. 2.
- CHESTERMAN A. (2009). The name and nature of translator studies. *Hermes*, Vol. 42.
- CHESTERMAN A. (2015). Models of what processes? In Ehrensberger-Dow et al. (Eds.) *Describing Cognitive Processes in Translation*, p. 7-20.
- CHEVROT, J.-P. (2012). Que reste-t-il de la distinction compétence-performance dans les théories fondées sur l'usage ? In *Mémoires de la société de linguistique de Paris*.
- CHEVROT J.-P. (2016a). « La Sociolinguistique au risque de l'interdisciplinarité : le cas des sciences cognitives et des STIC ». Lyon, ENS de LYON/DGESCO. Consulté à l'adresse <http://cle.ens-lyon.fr/plurilingues/a-trier/la-sociolinguistique-au-risque-de-l-interdisciplinarite-le-cas-des-sciences-cognitives-et-des-stic>
- CHEVROT J.-P. (2016b). « What do we mean by interdisciplinary research? » Conférence invitée présenté à Workshop. Bridging gaps: Conceptual and epistemological approaches, Center for Multilingualism in Society across the Lifespan, Université d'Oslo, Norvège.
- CHEVROT J.-P. & NARDY A. (A paraître). Sociolinguistique et sciences cognitives : promesses d'une rencontre inattendue. In M.-H. Côté, J. Durand, C. Lyche, & J. Peuvergne (Éd.), *Dynamiques linguistiques : variation, changement et cognition*.
- CHOI S, MCDONOUGH L, BOWERMAN M, & MANDLER J M. (1999). Early Sensitivity to Language-Specific Spatial Categories in English and Korean. *Cognitive Development*, Vol. 14, No. 2, p. 241-268.
- CHUQUET, HÉLÈNE, & PAILLARD, MICHEL. (1987). *Approche linguistique des problèmes de traduction. Anglais-français*. Ophrys.
- CHUQUET HÉLÈNE, & PAILLARD MICHEL. (1992). Enonciation et traduction chez les linguistes francophones. *French Language Studies*, Vol. 2, p. 237-259.
- CLARK, E. V. (1973). What's in a word: on the child's acquisition of semantics in his first language, in T.E. Moore (Ed.) *Cognitive development and the acquisition of language*, New York and

- London, Academic Press, pp. 65-110.
- CLARK, E. V. (2001). Some emergent categories in English acquisition, in M. Bowerman & S. C. Levinson (Eds.) *Language Acquisition and Conceptual Development*, Cambridge, Cambridge University Press, pp. 379-405.
- CLARK, E. V. & CARPENTER K. L. (1989). The notion of source in language acquisition., *Language*, Vol. 64, p. 1-30.
- COL G, & VICTORRI B. (2007). Comment formaliser en linguistique cognitive ? Opération de fenêtrage et calcul du sens temporel. *Corela. Cognition, représentation, langage*, No. HS-6, novembre.
- COLLOMBAT I. (2003). La Stylistique comparée du français et de l'anglais : la théorie au service de la pratique. *Meta: Journal des traducteurs*, Vol. 48, No. 3, p. 421.
- COLTIER D, DENDALE P, & DE BRABANTER P. (2009). La notion de prise en charge : mise en perspective. *Langue française*, Vol. 162, No. 2, p. 3.
- COMMISSION EUROPÉENNE (2014). Translation and Intellectual Property Rights. *Studies on Multilingualism and Translation*. Publications Office of the European Union.
- CONWAY C M, BAUERNSCHMIDT A, HUANG S S, & PISONI D B. (2010). Implicit statistical learning in language processing: Word predictability is the key. *Cognition*, Vol. 114, No. 3, p. 356-371.
- CONZETT P, ANDREASSEN H N, HØYDALSVIK S, JANDA L A, LONGVA L, NESSET T, & ODU O. (2014, 11-27.11). « TROLLing - The Tromsø Repository of Language and Linguistics ». Poster présenté à 9th Munin Conference on Scholarly Publishing, Tromsø.
- CRESWELL J W. (2014). *Research Design. Qualitative, Quantitative and Mixed Methods Approaches*. (4ème édition). Sage.
- CROFT W. (2010). Relativity, linguistic variation and language universals. *CogniTextes. Revue de l'Association Française de Linguistique Cognitive*, No. Volume 4, février. doi:10.4000/cognitextes.303
- CRUICKSHANK L, & EVANS M. (2012). Designing creative frameworks: design thinking as an

- engine for new facilitation approaches. *International Journal of Arts and Technology*, Vol. 5, No. 1, p. 73.
- CULIOLI A. (1971). À propos d'opérations intervenant dans le traitement formel des langues naturelles. *Palimpsestes*, Vol. 34, p. 7-15.
- CULIOLI A. (1973). Sur quelques contradictions en linguistique *Communications*, Vol. 20, Le sociologique et le linguistique. p. 83-91.
- CULIOLI, A. (1990). *Pour une linguistique de l'énonciation. Opérations et représentations* (Ophrys).
- DANKS J H, SHREVE G M, FOUNTAIN S B, & MCBEATH M K. (1997). *Cognitive Processes in Translation and Interpreting* (Sage).
- DELISLE J. (1980). *L'analyse du discours comme méthode de traduction : initiation à la traduction française de textes pragmatiques anglais : théorie et pratique* /. Ottawa, Canada ; Éditions de l'Université d'Ottawa.
- DESILETS A., MELANÇON C., PATENAUDE G. & BRUNETTE L. (2009). How translators use tools and resources to resolve translation problems: An ethnographic study. In *MT Summit XII-Workshop: Beyond Translation Memories: New Tools for Translators MT*.
- DELMAS C. (1993). De l'extralinguistique au métalinguistique. In J.-R. Lapaire & W. Rotgé (Éd.) *Séminaire pratique de linguistique anglaise*, p. 195-212.
- DEMUTH, K., CULBERTSON, J. & ALTER, J. (2006), Word-minimality, epenthesis, and coda licensing in the acquisition of English, *Language & Speech*, Vol. 49, p. 137-174.
- DEMUTH, K. & TREMBLAY A. (2008). Prosodically-conditioned variability in children's production of French determiners. *Journal of Child Language*, Vol. 35, No. 1, p. 99-127.
- DESAGULIER G. (2011). Le programme socio-pragmatique des grammaires de constructions : bilan et perspectives. *Intellectica*, Vol. 56, No. 2, p. 99-123.
- DESAGULIER G. (2015). Le statut de la fréquence dans les grammaires de constructions. Simple comme bonjour ? *Langages*, Vol. 197, No. 1, p. 99-128.
- D'HULST L. (1992). Sur le rôle des métaphores en traductologie contemporaine. *Target*, Vol. 4, No. 1, p. 33-51, janvier.

- DIESSEL, H. (2004). *The Acquisition of Complex Sentences*. [Cambridge Studies in Linguistics 105]. Cambridge University Press.
- DIESSEL, H. (2014). Usage-based linguistics. In Mark Aronoff (Eds.), *Oxford Bibliographies in Linguistics*. Oxford University Press.
- DIVJAK D, & GRIES S T (Éd.). (2012). *Frequency effects in language representation*. Berlin ; Boston, De Gruyter Mouton.
- DOHERTY S. & KENNY D. (2014). The Design and Evaluation of a Statistical Machine Translation Syllabus for Translation Students. *The Interpreter and Translator Trainer*, Vol. 8, No. 2, p. 295-315.
- DOMÍNGUEZ A, IZURA C, & MEDINA A C. (2018). Developmental and Gender Differences in the Production of the Past Tense in Spanish Children. *The Spanish Journal of Psychology*, Vol. 21.
- DUARTE J F, ASSIS ROSA A, & SERUYA T (Éd.). (2006). *Translation studies at the interface of disciplines*. Amsterdam ; Philadelphia, J. Benjamins Pub. Co.
- DUFAYE L. (2012). Analyse des emplois non locatifs de BY. *Anglophonia. French Journal of English Linguistics*, No. 16 (32), p. 85-97.
- DUGUA C, SPINELLI E, CHEVROT J-P, & FAYOL M. (2009). Usage-based account of the acquisition of liaison: Evidence from sensitivity to the singular/plural orientation of nouns. *Journal of Experimental Child Psychology*, Vol. 102, No. 3, p. 342-350.
- DUPOUX, E. (2018). Cognitive Science in the era of Artificial Intelligence: A roadmap for reverse-engineering the infant language-learner. *Cognition*, Vol. 173, p. 34-59.
- DURY P. (2013). Que montre l'étude de la variation d'une terminologie dans le temps. Quelques pistes de réflexion appliquées au domaine médical. *Debate Terminológico*. No. 9, p. 2-10.
- EHRENSBERGER-DOW M, & ENGLUND DIMITROVA B. (2018). *Exploring the Situational Interface of Translation and Cognition*. John Benjamins.
- EHRENSBERGER-DOW M, GÖPFERICH S, & O'BRIEN S. (2015). *Interdisciplinarity in Translation and Interpreting Process Research*. John Benjamins.

- EHRENSBERGER-DOW M, HUNZIKER HEEB A, MASSEY G, & ET AL. (2016). An international survey of the ergonomics of professional translation. *ILCEA. Revue de l'Institut des langues et cultures d'Europe, Amérique, Afrique, Asie et Australie*, Vol. 27.
- EMT. (2017). Référentiel de compétences de l'EMT. Consulté à l'adresse https://ec.europa.eu/info/resources-partners/european-masters-translation-ent/european-masters-translation-ent-explained_fr
- ELLIS, N. C. (2006). Current issues in the teaching of grammar: An SLA perspective. *TESOL Quarterly*, Vol. 40, No. 1, p. 83-107.
- ELLIS, N. C. (2013). Second language acquisition. In G. Trousdale & T. Hoffmann (Eds.) *Oxford Handbook of Construction Grammar* (pp. 365-378), Oxford University Press.
- FALAISE A., TUTIN A. & KRAIF O. (2011). Une interface pour l'exploitation de corpus arborés par des non informaticiens : la plate-forme ScienQuest du projet Scientext, *Revue TAL*, Vol. 52, No. 3, p. 103-128.
- FENSON, L., MARCHMAN, V.A., THAL, D., DALE, P., REZNICK, J.S., & BATES, E. (2007). *MacArthur-Bates Communicative Development Inventories: User's Guide and Technical Manual*. Baltimore, Brookes Publishing Co.
- FERNALD A, MARCHMAN V A, & WEISLEDER A. (2013). SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental Science*, Vol. 16, No. 2, p. 234-248, mars.
- FÉVRIER F, GAUDUCHEAU N, JAMET É, ROUXEL G, & SALEMBIER P. (2011). La prise en compte des affects dans le domaine des interactions homme-machine : quels modèles, quelles méthodes, quels bénéfices ?, The study of affects in human-computer interactions: theories, methods and benefits. *Le travail humain*, Vol. 74, No. 2, p. 183-201.
- FILLMORE C. J. (1992). "Corpus linguistics" or "Computer-aided armchair linguistics" , *Directions in Corpus Linguistics*, Proceedings of Nobel Symposium 82 Stockholm, 4-8 August 1991, J. Svartvik (eds), Mouton de Gruyter, p. 35-60.
- FINKBEINER R, MEIBAUER J, & SCHUMACHER P B (Éd.). (2012). *What is a context? linguistic*

- approaches and challenges*. Amsterdam ; Philadelphia, John Benjamins Pub. Co.
- FORCADA M L. (2017). Making sense of neural machine translation. *Translation Spaces*, Vol. 6, No. 2, p. 291-309.
- FORTIS J-M. (2012). La linguistique cognitive : histoire et épistémologie. *Histoire Épistémologie Langage*, Vol. 34, No. 1, p. 5-17.
- FRANÇOIS J, & CORDIER F. (2006). Psycholinguistique vs psychologie cognitive du langage : une simple variante terminologique ?, Abstract. *Syntaxe et sémantique*, Vol. N° 7, No. 1, p. 57-78.
- FRANK M C, BRAGINSKY M, YUROVSKY D, & MARCHMAN V A. (2017). Wordbank: an open repository for developmental vocabulary data. *Journal of Child Language*, Vol. 44, No. 03, p. 677-694.
- FREY, C. B. & M. A. OSBORNE (2016). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting & Social Change*.
<http://dx.doi.org/10.1016/j.techfore.2016.08.019>
- FROELIGER N. (2013). « En plein milieu des confins -- Eléments pour la construction d'une réflexion en traduction pragmatique. » (Synthèse de travaux (HDR)). Univ. Grenoble Alpes. Consulté à l'adresse <https://tel.archives-ouvertes.fr/tel-00971732>
- FROELIGER N. (2017). Traduction et trahison -- Tout est dans le contexte. In M. Grégoire & B. Mathios (Éd.), *Traductions et contextes, contextes de la traduction* (p. 33-52). L'Harmattan.
- FUCHS C. (Éd.) (2004). *La linguistique cognitive*. Editions de la Maison des sciences de l'homme.
- GAMBIER Y. (2000). Traduction et analyses de discours: typologie croisée. *Studia Romanica Posnaniensia*, Vol. 25/26, p. 97-108.
- GAMBIER Y, & VAN DOORSLAER L. (2016). *Border Crossings: Translation Studies and other disciplines*. John Benjamins.
- GASPARI F, ALMAGHOUT H, & DOHERTY S. (2015). A survey of machine translation competences: Insights for translation technology educators and practitioners. *Perspectives*, Vol. 23, No. 3, p. 333-358.

- GEERAERTS D, GRONDELAERS S, & BAKEMA P. (1994). *The Structure of Lexical Variation: Meaning, Naming, and Context*. De Gruyter Mouton.
- GONZÁLEZ ALONSO J, VILLEGAS J, & GARCÍA MAYO M del P. (2016). English compound and non-compound processing in bilingual and multilingual speakers: Effects of dominance and sequential multilingualism. *Second Language Research*, Vol. 32, No. 4, p. 503-535.
- GOUADEC D. (2002). Professionnaliser - *Actes de la huitième université d'été en traduction*, p. 217-246.
- GRIES S. T. (s. d.). Some proposals towards more rigorous corpus linguistics. *Zeitschrift für Anglistik und Amerikanistik*, Vol. 54, No. 2, p. 191-202.
- GRIES S. T. (2010). Useful statistics for corpus linguistics. In A. Sanchez & M. Almela (Éd.), *A Mosaic of Corpus Linguistics: Selected Approaches* (p. 23).
- GRIES S.T. (2015). The most under-used statistical method in corpus linguistics: multi-level (and mixed-effects) models. *Corpora*, Vol. 10, No. 1, p. 95-125, avril.
- GRIES S. T. & WULFF S. (2012). Regression analysis in translation studies. In M. P. Oakes & M. Ji (Éd.), *Studies in Corpus Linguistics* Vol. 51, p. 35-52. John Benjamins.
- GROSSMANN F, & TUTIN A. (2014). *L'écrit scientifique : du lexique au discours. Autour de Scientext*. Presses universitaires de Rennes.
- GROUSSIÉ M.-L. (1997). Prépositions et primarité du spatial : de l'expression de relations dans l'espace à l'expression de relations non-spatiales. *Faits de langues*, Vol. 14, p. 221-234.
- GUILLEMIN-FLESCHER J. (1981). *Syntaxe comparée du français et de l'anglais : problèmes de traduction* (Ophrys).
- GUILLEMIN-FLESCHER J. (1986). Le linguiste devant la traduction. *Fabula*, Vol. 7, p. 59-68.
- HABERT B. (2000). Des corpus représentatifs : de quoi, pour quoi, comment. In *Linguistique sur corpus. Etudes et réflexions*. Presses Universitaires de Perpignan.
- HARDIE A. (2014). Statistical identification of keywords, lockwords and collocations as a twostep procedure. *ICAME 35 conference*, University of Nottingham.
- HARVEY, M. (1996). A translation course for French-speaking students. *Teaching Translation in*

- Universities. Present and Future Perspectives*. Association for French Language Studies & Centre for Information on Language Teaching and Research, p. 45-65.
- HASHEMI, M. R. (2012). Reflections on mixing methods in applied linguistics research. *Applied linguistics*, Vol. 33, no. 2, p. 206-212.
- HEARNE M, & WAY A. (2011). Statistical Machine Translation: A Guide for Linguists and Translators. *Language and Linguistics Compass*, Vol. 5, No. 5, p. 205-226.
- HEFFRON. R.J. & D. MACCAULEY. (2017). The concept of energy justice across the disciplines, *Energy Policy* Vol. 105, p. 658-667.
- HEIDEN, S., MAGUÉ, J-P. ET PINCEMIN, B. (2010). TXM : Une plateforme logicielle open-source pour la textométrie – conception et développement. *Proc. of JADT'2010* (10^{èmes} Journées internationales d'Analyse Statistique des données Textuelles), Rome.
- HEINSSEN R K, GLASS C R, & KNIGHT L A. (1987). Assessing computer anxiety: Development and validation of the Computer Anxiety Rating Scale. *Computers in Human Behavior*, Vol. 3, No. 1, p. 49-59.
- HELLER B, SZMRECSANYI B, & GRAFMILLER J. (2017). Stability and Fluidity in Syntactic Variation World-Wide: The Genitive Alternation Across Varieties of English. *Journal of English Linguistics*.
- HEWSON L. (2016). Les incertitudes du traduire. *Meta: Journal des traducteurs / Meta: Translators' Journal*, Vol. 61, No. 1, p. 12-28.
- HICKMANN, M. (2012). Diversité des langues et acquisition du langage : espace et temporalité chez l'enfant. *Langages*, Vol. 188, No. 4, p. 25-39.
- HOUSE J. (2006). Text and context in translation. *Journal of Pragmatics*, Vol. 38, p. 338-358.
- HUART R, & LARREYA P. (2007). *Pour en finir avecLes constructions Nom + Nom*. Ophrys.
- HUDDLESTON R, & PULLUM G K. (2002). *Cambridge Grammar of the English Language*. Cambridge University Press.
- JACKENDOFF, R. (1983). *Semantics and Cognition*. MIT Press.
- JANDA L. A. (Éd.). (2013). *Cognitive linguistics: the quantitative turn: the essential reader*. Berlin,

De Gruyter Mouton.

JOHNSON R B, ONWUEGBUZIE A J, & TURNER L A. (2007). Toward a Definition of Mixed Methods Research. *Journal of Mixed Methods Research*, Vol. 1, No. 2, p. 112-133.

JOSSELIN-LERAY, A., FABRE, C., REBEYROLLE, J., PICTON, A., PLANAS, E. (2014). Good Contexts for Translators-A First Account of the Cristal Project. *XVI EURALEX International Congress: The User in Focus*. Bolzano/Bolzen.

KAIL M. (2002). *Variabilités en psycholinguistique développementale*. Éditions de la Maison des sciences de l'homme.

KELLER-COHEN D. (1978). Context in Child Language. *Annual Review of Anthropology*. Vol. 7, p. 453-482.

KENNY D. (2001). *Lexis and Creativity in Translation. A corpus-based approach*. Routledge.

KENNY D. (2016). Translator vs. Machine. *The Linguist*, Vol. 55, No. 6.

KENNY D. (2017). The Translation Studies guide to disruption. Présenté à Translation and Disruption: Global and local perspectives, Portsmouth.

KENNY D, & DOHERTY S. (2014). Statistical machine translation in the translation curriculum: overcoming obstacles and empowering translators. *The Interpreter and Translator Trainer*, Vol. 8, No. 2, p. 276-294.

KERBRAT-ORECCHIONI C. (2012). Le contexte revisité. *Corela*, No. HS-11, avril. doi:10.4000/corela.2627

KERN S, & GAYRAUD F. (2010). *IFDC*. Les éditions la cigale, Grenoble, France.

KILGARRIFF A. (2005). Language is never, ever, ever, random. *Corpus Linguistics and Linguistic Theory*, Vol. 1, No. 2, janvier. doi:10.1515/cllt.2005.1.2.263

KLEIBER G. (2009). D'un contexte à l'autre : aspects et dimensions du contexte. *L'Information Grammaticale*, Vol. 123, No. 1, p. 17-32.

KOPONEN M. (2012). Comparing human perceptions of post-editing effort with post-editing operations. In *Proceedings of the 7th workshop on statistical machine translation* (p. 181-190).

- KOPONEN M. (2015). How to teach machine translation post-editing? Experiences from a post-editing course. (p. 2). Présenté à 4th Workshop on Post-Editing Technology and Practice (WPTP4).
- KOSKINEN, K. & M. RUOKONEN (2017). Love letters or hate mail? Translators' technology acceptance in the light of their emotional narratives. In D. Kenny (Ed.), *Human issues in translation technology*, Routledge, p. 8-24.
- KREYER R. (2003). Genitive and of-construction in modern written English. Processability and human involvement. *International Journal of Corpus Linguistics*, Vol. 8, No. 2, p. 169-207.
- KRINGS H P. (2001). *Repairing Texts: Empirical Investigations of Machine Translation Post-Editing Processes*. (G. S. Koby, Éd.). The Kent State University Press.
- KRÜGER R. (2016). The textual degree of technicality as a potential factor influencing the occurrence of explicitation in scientific and technical translation. *JoSTrans*, Vol. 26. Consulté à l'adresse http://www.jostrans.org/issue26/art_kruger.php
- KÜBLER, N. (Ed.) (2011) *Corpora, language, teaching, and resources: From theory to practice*. Peter Lang.
- KÜBLER N. (2014). Traduction pragmatique, linguistique de corpus, traducteur : un ménage à trois explosif ? *Tralogy*. [en ligne] <http://lodel.irevues.inist.fr/tralogy/index.php?id=288>
- L'HOTE E. & M. LEMMENS (2009). Reframing treason: metaphors of change and progress in new Labour discourse, *CogniTextes* Vol.3.
- LABOV, W. (1972). *Language in the Inner City*. University of Pennsylvania Press.
- LADMIRAL J-R. (1970). *Traduire : théorèmes pour la traduction*. Payot.
- LADMIRAL J-R. (2004). Dichotomies traductologiques. *La linguistique*, Vol. Vol. 40, No. 1, p. 25-50.
- LAKOFF G. (1987). *Women, Fire and Dangerous Things. What categories reveal about the mind*. University of Chicago Press.
- LAKOFF G, & JOHNSON M. (1980a). *Metaphors we live by*. University of Chicago Press.
- LAKOFF G, & JOHNSON M. (1980b). The Metaphorical Structure of the Human Conceptual System. *Cognitive Science*, Vol. 4.

- LAKOFF, G. (1990). The Invariance Hypothesis. *Cognitive Linguistics*, Vol. 1, No. 1.
- LAKS B. (2013). Why is there variation rather than nothing? *Language Sciences*, Vol. 39, p. 31-53.
- LANSARI, L. (2009). *Les périphrases verbales aller + infinitif et be going to*. Ophrys.
- LAPAIRE, J.R, & ROTGÉ, W. (2002). *Linguistique et grammaire de l'anglais*. Presses Universitaires du Mirail.
- LAVAUULT É. (1985). *Fonctions de la traduction en didactique des langues : Apprendre une langue en apprenant à traduire*. Didier.
- LAVAUULT-OLLÉON É. (1994). Former des étudiants LEA à la traduction technique et scientifique : un défi didactique ? *ASp. la revue du GERAS*, No. 3, p. 65-82, février.
- LAVAUULT-OLLÉON É. (2006). Le skopos comme stratégie de déblocage : dialecte et scotticité dans Sunset Song de Lewis Grassie Gibbon. *Meta: Journal des traducteurs*, Vol. 51, No. 3, p. 504.
- LAVAUULT-OLLÉON É. (2011). L'ergonomie, nouveau paradigme pour la traductologie. *ILCEA. Revue de l'Institut des langues et cultures d'Europe, Amérique, Afrique, Asie et Australie*, No. 14.
- LAVAUULT-OLLÉON É. (2018). Regard diachronique et prospectif sur un demi-siècle de traductologie pragmatique. *ASp. la revue du GERAS*, Vol. 74.
- LEBLANC M. (2014). Traduction, bilinguisme et langue de travail : une étude de cas au sein de la fonction publique fédérale canadienne. *Meta: Journal des traducteurs*, Vol. 59, No. 3, p. 537.
- LEBLANC M. (2017). "I can't get no satisfaction!" Should we blame translation technologies or shifting business practices? In D. Kenny (Éd.), *Human Issues in Translation Technology*. Routledge.
- LEGALLOIS, D. & TUTIN, A. (2013). Présentation : Vers une extension du domaine de la phraséologie. *Langages*, Vol. 189, No. 1, p. 3-25.
- LEGRIS P, INGHAM J, & COLLERETTE P. (2003). Why do people use information technology? A critical review of the technology acceptance model. *Information and Management*, Vol. 40, No. 3.
- LEJEUNE, P. (2016.) Structures linguistiques et problèmes de traduction : schémas nominaux

- renvoyant à la transformation dans le discours de spécialité. In Boisseau, Maryvonne, Chauvin, Catherine, Delesse Catherine & Kerommes Yvon, (Eds.). *Linguistique et traductologie : les enjeux d'une relation complexe*. Collection Traductologie. Artois Presses Université
- LÉON J. (2008). Aux sources de la « Corpus Linguistics » : Firth et la London School. *Langages*, Vol. 171, No. 3.
- LÉON, J. (2015) *Histoire de l'automatisation des sciences du langage*, Lyon, ENS Éditions, coll. « Langages », 218 p.
- LIEVEN E. (2014). First language development: a usage-based perspective on past and current research. *Journal of Child Language*, Vol. 41, No. S1, p. 48-63.
- LINN A. (2011). Impact: Linguistics in the Real World. *Histoire Épistémologie Langage*, Vol. 33, No. 1, p. 15-27.
- LINN A., CANDEL D. & LEON J. (2011) Présentation: Linguistique appliquée et disciplinarisation. *Histoire Épistémologie Langage*, Vol. 33, No. 1. p. 7-14
- LOOCK R. (2016). *La traductologie de corpus*. Presses Universitaires du Septentrion.
- LOWREY B, & TOUPIN F. (2010). L'invariant à l'épreuve de la diachronie. *Corela. Cognition, représentation, langage*, Vol. 8, No. 2.
- MACWHINNEY B. (2000). *The CHILDES project: Tools for analyzing talk*. 3^{ème} Edition. Lawrence Erlbaum Associates.
- MACWHINNEY B. (2014). What we have learned. *Journal of Child Language*, Vol. 41, No. S1, p. 124-131.
- MACWHINNEY B. (2018). CLAN Manual. Doi:10.21415/t5g10r
- MACWHINNEY, BRIAN, & ANDERSON J. (1986). The acquisition of grammar. In I. Gopnik & M. Gopnik (Éd.), *From Models to Modules*. Ablex Publishing Corporation.
- MALMKJAER K. (2010). *The Routledge Linguistics Encyclopedia* (Troisième édition). Routledge.
- MANEGRE M. (2001). « The combination of lexical morphemes: Forming co-ordinate compounds and blends in child language » (Mémoire de Master). Université d'Alberta.

- MANIEZ F. (2013). The translation into French of adjectives formed with a noun and a past participle in English-language medical articles. *Panace@*, Vol. XIV, No. 38.
- MCENERY, T. & R. XIAO (2007). Parallel and comparable corpora: What is happening? *Incorporating Corpora. The Linguist and the Translator*, p. 18-31.
- MELBY, ALAN & FOSTER C. (2010). Context in translation: Definition, access and teamwork, *Translation and Interpreting*, vol.2, No. 2, p. 1-15.
- MININNI G. (2007). *La psycholinguistique à l'œuvre dans la traduction*. (C. Cortesi, Trad.). L'Harmattan Italia.
- MOELLER, A. K. & O. KETSMAN (2010). Can We Learn a Language Without Rules? *Faculty Publications: Department of Teaching, Learning and Teacher Education*. Vol. 170. <http://digitalcommons.unl.edu/teachlearnfacpub/170>
- MOLINE E, & STOSIC, DEJAN. (2016). *L'expression de la manière en français*. Ophrys.
- MONDADA, L. (2000). Les effets théoriques des pratiques de transcription. *Linx - Revue des linguistes de l'université Paris Nanterre*, Vol. 42, p. 131-150.
- MONDADA L. (2008). L'analyse de « collections » de phénomènes multimodaux en linguistique interactionnelle : À propos de l'organisation systématique des ressources gestuelles en début de tour *Cahiers de praxématique*, Vol. 50, p. 21-66.
- MOORKENS J. (2017). Under pressure: translation in times of austerity. *Perspectives: Studies in Translation Theory and Practice*, Vol. 25, No. 3.
- MOORKENS J. (2018). What to expect from Neural Machine Translation: a practical in-class translation evaluation exercise. *The Interpreter and Translator Trainer*, p. 1-13.
- MOORKENS J, & O'BRIEN S. (2017). Assessing User Interface Needs of Post-Editors of Machine Translation. In D. Kenny (Éd.), *Human Issues in Translation Technology*. Routledge.
- MOORKENS J, O'BRIEN S, DA SILVA I A L, DE LIMA FONSECA N B, & ALVES F. (2015). Correlations of perceived post-editing effort with measurements of actual effort. *Machine Translation*, Vol. 29, No. 3-4.
- MORGENSTERN A, & PARISSÉ C. (2010). Transcrire et analyser les corpus d'interactions adulte-

- enfant. In E. Veneziano & A. Salazar Orvig (Éd.), *Acquisition du langage et interaction* (p. 201-222). L'Harmattan.
- MORGENSTERN A, PARISSÉ C, & SEKALI M. (2009). À la source du futur : premières formes verbales dans les productions spontanées de deux enfants français de 18 mois à 3 ans. *Faits de langues*, Vol. 33, p. 163-175.
- MORGENSTERN, A. & PARISSÉ C. (2007). Codage et interprétation du langage spontané d'enfants de 1 à 3 ans. *Corpus*, Vol.6, p. 55-78.
- MURPHY R. (2004). *English Grammar in Use*. Cambridge University Press.
- NELSON K. (1973). Structure and strategy in learning to talk. *Monographs of the Society for Research in Child Development* (149.38), p. 1-135.
- NORRIS, J., & ORTEGA, L. (2002). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, Vol. 50, No. 3, p. 417-528.
- NORMAN D A. (1980). Twelve issues for cognitive science. *Cognitive Science*, Vol. 4, p. 1-32.
- NUNES VIEIRA L. (2016). How do measures of cognitive effort relate to each other? A multivariate analysis of post-editing process data. *Machine Translation*, Vol. 30, p. 41-62.
- OCHS, E. (1979). Transcription as Theory. In Ochs E. and Schieffelin B. (Ed.), *Developmental Pragmatics*. Academic Press.
- OLOHAN M. (2017). Technology, translation and society: A constructivist, critical theory approach. *Target*, Vol. 29, No. 2, p. 264-283.
- ORERO, P., DOHERTY, S., KRUGER, J. L., MATAMALA, A., PEDERSEN, J., PEREGO, E. et al. (2018). Conducting experimental research in audiovisual translation (AVT): A position paper. *Journal of Specialised Translation*, Vol. 30, p. 105-126.
- OSGOOD C. & SEBEOK T. (1965). *Psycholinguistics*. Indiana University Press.
- PAPAVASSILIOU P. (2007). Traductologie et sciences cognitives : une dialectique prometteuse. *Meta: Journal des traducteurs*, Vol. 52, No. 1, p. 29.
- PARISSÉ C. (2002). Oral language, written language and language awareness. *Journal of Child Language*, Vol. 29, No. 2, p. 478-481.

- PERGNIER M. (1980). *Les fondements sociolinguistiques de la traduction*. Honoré Champion.
- PETERS-GEIBEN L. (2016). La prévention comportementale et contextuelle : intégrer une approche ergonomique dans la formation des traducteurs. *ILCEA. Revue de l'Institut des langues et cultures d'Europe, Amérique, Afrique, Asie et Australie*, Vol. 27.
- PLANAS, E., PICTON, A., JOSSELIN-LERAY, A. (2014) Exploring the Use and Usefulness of KRCs in Translation: Towards a Protocol. *Terminology and Knowledge Engineering*, Berlin.
- POIBEAU T. (2014). La linguistique est-elle soluble dans la statistique ? *Revue Sciences/Lettres*, Vol. 2, février.
- POIBEAU T. (2017). *Machine Translation*. MIT Press.
- PULLUM G. K. (1989). Formal linguistics meets the boojum. *Natural Language and Linguistic Theory*, Vol. 7, p. 137-143.
- RANGER G. (2014). « Fire away » Suggestions pour une caractérisation énonciative de la particule adverbiale anglaise AWAY. *Faits de Langues*, Vol. 2, No. 44, p. 147-158.
- RASTIER F. (1998). Le problème épistémologique du contexte et le statut de l'interprétation dans les sciences du langage. *Langages*, Vol. 129, p. 97-111.
- RAYSON P. (2008). From key words to key semantic domains. *International Journal of Corpus Linguistics*, Vol. 13, No. 4, p. 519-549.
- RAYSON P. & GARSIDE R. (2000). Comparing corpora using frequency profiling. In *WCC '00 Proceedings of the workshop on Comparing corpora* (Vol. 9, p. 1-6).
- REBUSCHAT P. (2015) *Implicit and explicit learning of languages*. John Benjamins.
- REDDY M. (1979). The Conduit Metaphor: a Case of Frame Conflict in our Language about Language. In A. Ortony (Éd.), *Metaphor and Thought*. Cambridge University Press.
- REFFAY C, CHANIER T, & NORAS M. (2008). Contribution à la structuration de corpus d'apprentissage pour un meilleur partage en recherche, Vol. 15, p. 25.
- RENNER V. (2015). Panorama rétro-prospectif des études amalgamatives. *Neologica : revue internationale de la néologie*, Vol. 9, p. 97-112.
- REY, O. (2016). *Quand le monde s'est fait nombre*. Stock.

- RICO C. (2017). The ePortfolio: constructing learning in translation technology. *The Interpreter and Translator Trainer*, Vol. 11, No. 1, p. 79-95.
- RISKU H, & WINDHAGER F. (2013). Extended Translation: A Sociocognitive Research Agenda. *Target*, Vol. 25, No. 1, p. 34-46.
- ROMANO F. (2016). Syntactic planning of English genitives in L1 and L2 production: The Animacy Rule Model. *Lingua*, Vol. 184, p. 104-121.
- ROSCH E. (s. d.). Cognitive reference points. *Cognitive Psychology*, Vol. 7, p. 532-547.
- ROSENBACH A. (2008). Animacy and grammatical variation—Findings from English genitive variation. *Lingua*, Vol. 118, No. 2, p. 151-171.
- ROSENBACH A. (2014). English genitive variation – the state of the art. *English Language and Linguistics*, Vol. 18, No. 02, p. 215-262.
- ROTGÉ, W. (2002). Model/Models in linguistics, *ASp*, Vol. 35-36, p. 5-22.
- ROTH P. (1996). *Sabbath's Theater*. Vintage.
- ROY B. C. (2014). Précis of The birth of a word (summary of a TED talk). http://web.media.mit.edu/~bcroy/papers/bcroy-phd_precis.pdf.
- ROY, B. C., M.C. FRANK, P.DECAMP, M. MILLER & D. ROY (2015) Predicting the birth of a spoken word. *Proceedings of the National Academy of Sciences*, Vol. 112, p. 12663–12668.
- ROY D. (2009). New horizons in the study of child language acquisition. *Proceedings of Interspeech 2009*, [en ligne] <http://dspace.mit.edu/handle/1721.1/65900>
- RUMELHART D E, & MCCLELLAND J L. (1986). Past Tenses of English Verbs. In *Parallel distributed processing: explorations in the microstructure of cognition* (Vol. 2, p. 56).
- SAFFRAN J R, ASLIN R N, & NEWPORT E N. (1996). Statistical learning by 8-month-old infants. *Science*, Vol. 274. Consulté à l'adresse <https://www.ncbi.nlm.nih.gov/pubmed/8943209>
- SCALISE I, & VOGEL S (Éd.). (2010). *Cross-disciplinary issues in compounding*. John Benjamins.
- SCHNEIDER R M, YUROVSKY D, & FRANK M C. (2015). Large-scale investigations of variability in children's first words. In *Proceedings of the 37th Annual Conference of the Cognitive Science Society*. (p. 6).

- SEKALI M. (2012). The emergence of complex sentences in a French child's language from 0;10 to 4;01: causal adverbial clauses and the concertina effect. *Journal of French Language Studies*, Vol. 22, No. 1.
- SHREVE G M. (2010). Bilingualism and translation. In D. Rocher (Trad.), *Handbook of Translation Studies* (p. 1-6). John Benjamins.
- SINCLAIR J., JONES S., DALEY R. (2004). *The OSTI Report (1970)*. Krishnamurthy R. (Éd). Continuum.
- SINHA C, & KUTEVA T. (1995). Distributed Spatial Semantics. *Nordic Journal of Linguistics*, Vol. 18, No. 2.
- SNELL-HORNBY M. (1988). *Translation Studies. An Integrated Approach*. John Benjamins.
- SNELL-HORNBY M. (2006). *The turns of translation studies: new paradigms or shifting viewpoints?* John Benjamins.
- SOLER L. (2000). *Introduction à l'épistémologie*. Ellipses.
- STEGU M. (2011). La linguistique appliquée : discipline ou groupement de disciplines indépendantes ? *Histoire Épistémologie Langage*, Vol. 33, No. 1, p. 129-139
- STEINER, G. (2010). *Langage et silence*. (P.-E. Dauzat, Trad.). Les Belles Lettres.
- STUBBS M. (1996). *Text and corpus analysis: Computer-assisted studies of language and culture*. Blackwell.
- SWALES J. (1990). The Concept of Discourse Community. *Genre Analysis: English in Academic and Research Settings*, p. 21-32. Cambridge University Press.
- SWALES J M. (2016). Reflections on the concept of discourse community. *ASp*, Vol. 69, p. 7-19, mars.
- SZMRECSANYI B. (2017). Variationist sociolinguistics and corpus-based variationist linguistics: overlap and cross-pollination potential. *Canadian Journal of Linguistics/Revue Canadienne de Linguistique*, Vol. 62, No. 04, p. 685-701, décembre.
- TALMY L. (2000). *Toward a Cognitive Semantics* (Vol. Vol. 1 Concept Structuring Systems). MIT Press.

- TEIXEIRA C S C. (2014). Perceived vs. measured performance in the post-editing of suggestions from machine translation and translation memories. In S. O'Brien, M. Simard, & L. Specia (Éd.), *Proceedings of the 11th Conference of the Association for Machine Translation in the Americas: Workshop on Post-Editing Technology and Practice (WPTP3)* (p. 450-459).
- TEUBERT W. (2005). My version of corpus linguistics. *International Journal of Corpus Linguistics*, Vol. 10, No. 1, p. 1-13.
- TEUBERT, W. (2009). « La linguistique de corpus : une alternative [version abrégée] », *Semen* [En ligne], Vol. 27, <http://journals.openedition.org/semen/8914>.
- THAL, D., MARCHMAN, V.A., & TOMBLIN, J.B. (2013). Late talking toddlers: Characterization and prediction of continued delay. In Rescorla, L. & Dale, P. (Éd.), *Late Talkers: Language Development, Interventions, and Outcomes*. Baltimore, Brookes Publishing Co.
- THEAKSTON A. & LIEVEN E. (2017). Multiunit Sequences in First Language Acquisition. *Topics in Cognitive Science*, Vol. 9, No. 3.
- TOMASELLO M (Éd.). (1998). *The new psychology of language: cognitive and functional approaches to language structure*. Mahwah, N.J, L. Erlbaum.
- TOMASELLO M. (2001). First steps toward a usage-based theory of language acquisition. *Cognitive Linguistics*, Vol. 11, No. 1-2.
- TOMASELLO M. (2007). Cognitive Linguistics and First Language Acquisition. In D. Geeraerts & H. Cuyckens (Éd.), *The Oxford Handbook of Cognitive Linguistics*.
- TRAVERSO, V. (2011). Analyser un corpus de langue parlée en interaction : questions méthodologiques. *Verbum, Akadémiai Kiadó*, Vol. 4, p. 313-329
- TREVISE A. (1996). Réflexion, réflexivité et acquisition des langues. *Acquisition et interaction en langue étrangère, 8 / 2011*, Vol. 8, p. 5-39.
- TUMMERS J, HEYLEN K, & GEERAERTS D. (2005). Usage-based approaches in Cognitive Linguistics: A technical state of the art. In *Corpus Linguistics and Linguistic Theory* (Vol. 1, p. 225-261).
- VAN NOORDEN R. (2015). Interdisciplinary research by the numbers. *Nature News*, Vol. 525, No.

- 7569, p. 306.
- VANDAELE S. (2007). Quelques repères épistémologiques pour une approche cognitive de la traduction. Application à la traduction spécialisée en biomédecine. *Meta: Journal des traducteurs*, Vol. 52, No. 1, p. 129.
- VANDAELE S, & LUBIN L. (2005). Approche cognitive de la traduction dans les langues de spécialité : vers une systématisation de la description de la conceptualisation métaphorique. *Meta: Journal des traducteurs*, Vol. 50, No. 2, p. 415.
- VAPNIK V N. (2000). *The Nature of Statistical Learning Theory* (Deuxième édition). Springer.
- VENKATESH, V. & F. D. DAVIS (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science* Vol. 46, No. 2, p.186–204.
- VENKATESH, V. & H. BALA (2008) Technology Acceptance Model 3 and a Research Agenda on Interventions, *Decision sciences*, Vol. 39, No. 2, p. 273–315.
- VÉRONIS, J. (2013). *Parallel Text Processing: Alignment and use of translation corpora* (Vol. 13). Springer Science & Business Media.
- VINAY J-P, & DARBELNET J. (1958). *Stylistique comparée du français et de l'anglais. Méthode de traduction*. Didier.
- VON HOLZEN K, NISHIBAYASHI L-L, & NAZZI T. (2018). Consonant and Vowel Processing in Word Form Segmentation: An Infant ERP Study. *Brain Sciences*, Vol. 8, No. 2, p. 24, janvier.
- VORDEROBERMEIER G M. (2014). *Remapping Habitus in Translation Studies*. Brill | Rodopi.
- WAY A. (2009). A Critique of Statistical Machine Translation. *Linguistica Antverpiensia*, Vol. 8.
- WIERZBICKA A. (1987). The semantics of modality. *Folia Linguistica*, Vol. 21, No. 1. doi:10.1515/flin.1987.21.1.25
- WOLF M, & FUKARI A. (2007). *Constructing a Sociology of Translation*. John Benjamins.
- WU, Y., SCHUSTER, M., CHEN, Z., et al. (2016). Google's neural machine translation system: Bridging the gap between human and machine translation. *ArXiv preprint* (arXiv:1609.08144).

Index

A

alternance du génitif, 4, 23, 73, 76, 80, 95
amalgamation (amalgames), 49-50, 73-74
apprenant, 8, 50, 60, 63, 69, 73, 95, 97, 125,
140, 146
apprentissage profond, 31, 36, 53, 97, 98,
100, 101, 102, 105
apprentissage statistique, 99, 102
automatisation, 7, 9, 87, 93, 98, 100, 105,
112

C

catégorie(s), 4, 8, 14, 31, 36, 42, 44-45, 60-
61, 63-66, 68, 70, 72, 82, 91, 96-97, 99-
100
catégorisation, 4, 8, 19, 31, 44, 49, 60, 61,
62, 63, 64, 66, 68, 94, 96
codage, 14, 65, 68, 70, 71, 84
conceptualisation, 42, 45, 49, 61, 69, 133,
145
contexte de situation, 25-28, 87
contexte situationnel, 26, 28-30, 49, 55, 87

E

ergonomie, 9, 56, 81, 83, 88, 125
environnement, 9, 15-16, 22, 27, 35, 48, 60-
61, 87-89, 94, 106, 109

G

grammaires de construction, 33, 45, 98, 106

L

langue maternelle, 8, 11, 25, 30, 46, 49, 50,
60, 64, 66, 73

M

méthodes mixtes, 5, 90
machine (homme-machine), 7, 51, 87, 11,
114-115

O

observation, 11, 14, 20, 24, 38, 50, 65, 67,
68, 85, 149

P

perception, 11, 24, 31, 44, 61, 64, 69, 81, 82,
83, 84, 90, 94, 104, 141
pluridisciplinaire, 5, 27, 37, 56, 57, 58, 59,
84, 91, 93, 100, 106, 107, 140, 141, 144

Q

quantitatif (méthodes, analyses), 22, 30,
45, 79, 91, 94-95, 110, 122, 128
qualitatif (méthodes, analyses), 39, 47,
56; 76, 89, 97-98, 100-102, 110

R

règles, 12, 19, 23, 48, 51, 53, 58, 60, 76, 77,
78, 88, 96, 97, 98, 100, 152

S

sciences cognitives, 37, 53, 57, 81, 100, 109, 115, 128, 138

suivis longitudinaux (suivi longitudinal), 11, 13, 19, 20, 21, 48, 60, 65, 69, 70, 108, 141

T

traducteur (professionnel, humain), 4, 8, 9, 43, 44, 45, 47, 53, 55, 56, 58, 59, 61, 71, 81, 83, 84, 85, 86, 87, 88, 89, 95, 98, 99, 102, 103, 104, 105, 124, 142-143, 147, 150

traduction automatique (TA), 10-11, 24-25, 39, 46-47, 57, 65, 77, 87, 92, 94-95, 101, 106, 109, 112-114

traductologie de corpus, 53, 54, 86, 126

traitement automatique, 12, 85, 86, 95, 139, 142, 144

transcription(s), transcrire, 12-14, 65, 127, 140

typologie(s), 7, 20, 60, 63, 65, 66, 67, 69, 72, 120

U

usage (basé sur ~, UB), 7, 12, 14, 18-19, 23, 26, 29-32, 34, 36, 43, 45, 46, 50-51, 53-54, 60, 63, 68-69, 78-82, 86-87, 91, 105, 115, 126, 132, 152

utilisateur, 87, 112, 116

V

variation, 17, 18, 19, 20, 21, 22, 24, 35, 41, 49, 62, 67, 75, 80, 97, 115, 116, 118, 125, 130, 146, 149, 150

variationniste (approche), 22-24, 85

Annexe : Curriculum Vitae

1. Trajectoire professionnelle

Maître de conférences à l'Université Grenoble Alpes et membre de l'ILCEA4 (depuis 2013)

Maître de conférences à l'Université Lumière Lyon 2 et membre du CRTT (2011-2013)

Docteure en linguistique anglaise et psycholinguistique (2010, Université Lumière Lyon 2)

Agrégée d'anglais (2004, rang 35, option linguistique)

Ancienne élève de l'Ecole Normale Supérieure de Lyon (2001)

2 Expérience pédagogique

2.1 Enseignements

Grammaire et traduction appliquées (L1, CM et TD)

Grammaire et traduction appliquées (L2, TD, traduction anglais – français)

Grammaire et traduction appliquées (L3, TD, traduction français – anglais)

Découverte des métiers de la traduction (L2)

Communication appliquée au monde professionnel (L3)

Communication professionnelle – anglais (M1, Master de Traduction Spécialisée Multilingue)

Introduction à la traduction automatique (M2, Master de Traduction Spécialisée Multilingue)

Traduction automatique et post-édition (M2, Master de Traduction Spécialisée Multilingue)

Introduction à la linguistique et à la terminologie (L3-M1, Master de Traduction Spécialisée Multilingue)

Traduction spécialisée : traduction technique (M1 français – anglais, Master de Traduction Spécialisée Multilingue)

Traduction à vue (L2 LLCER)

2.2 Implication dans des activités liées au numérique

2018. Responsable de l'anglais pour le projet IDEX formation HELD (hybridation des enseignements en langues débutées) porté par l'UFR de Langues étrangères de l'Université Grenoble Alpes.

Depuis 2015. Développement d'un parcours hybride en anglais pour les étudiants de L1 débutant une langue B (année préparatoire) dans la filière LEA. Ce travail a été soutenu par le Service universitaire de pédagogie de l'ex. Université Stendhal. Le parcours est intégré à la plateforme de cours de l'Université, les principaux outils utilisés sont donc ceux qui sont disponibles dans Moodle, certaines activités ont également été créées avec HotPotatoes. Ces développements ont fait l'objet de plusieurs présentations, notamment lors d'une rencontre sur le numérique en langues (décembre 2017) : <https://prezi.com/view/PvykohpJlqPGgEw5CXT5/>

Depuis 2014. Utilisation d'outils numériques pour animer un cours magistral de grammaire anglaise en L1 LEA : boîtiers de vote, Plickers, et outil QCM papier (pour la création et la gestion de QCM auto-correctifs) dans Moodle. L'hybridation de ce cours se met en place peu à peu, et un fonctionnement en classe inversé a été expérimenté de façon ponctuelle : j'ai témoigné sur cet essai lors d'une rencontre de la communauté de pratiques le 4 juin 2018.

2.3 Responsabilités administratives

2014-2016 : Responsable adjointe de la sous-section d'anglais (LEA).

2013-2015 : Responsable du site web du laboratoire ILCEA4

2011-2013 : Correspondante du Service commun universitaire d'information et d'orientation (SCUIO) de Lyon 2

3 Activités d'animation et d'encadrement scientifique

3.1 Participation à la recherche collective à l'Université Grenoble Alpes

- Membre du Groupe de Recherche Multilingue en Traduction Spécialisée (GREMUTS) au sein de l'Institut des Langues et des Cultures d'Europe, Amérique, Afrique, Asie et Australie (ILCEA4), je co-anime avec Camille Biros les travaux du groupe sur la justice climatique et énergétique (constitution de corpus et veille lexicologique)

Notre objectif est de contribuer à la caractérisation de la lexicologie et des genres de discours spécifiques au domaine de la justice climatique et énergétique dans une perspective multilingue. Nous nous appuyons sur la constitution d'un corpus regroupant des documents

spécialisés issus de différents types d'acteurs travaillant sur ces notions : les Nations-Unies, organisations non-Gouvernementales environnementales, entreprises du secteur énergétique, auteurs d'articles scientifiques, journalistes. Un financement du consortium CORLI, au sein du TGIR Huma-Num, a permis d'embaucher quatre stagiaires qui effectuent actuellement les dernières tâches de constitution du corpus sous ma direction, et un ingénieur d'études qui préparera les données avant publication sur ORTOLANG.

- Membre du bureau du Pôle Grenoble Cognition (depuis novembre 2017)

Le Pôle Grenoble Cognition a été créé en 2010 sous forme d'une Structure Fédérative de Recherche (FR 3381) qui est soutenue par le CNRS (INS2i), Grenoble INP et l'Université Grenoble Alpes. Il rassemble aujourd'hui près de 200 scientifiques de plusieurs disciplines, qui sont issus de 15 laboratoires grenoblois. Sa mission est d'impulser une dynamique de site autour des sciences cognitives, de favoriser les collaborations et de faciliter la progression des connaissances dans un domaine qui demeure complexe, fortement interdisciplinaire et aux enjeux sociétaux importants pour l'éducation, la santé, le handicap, les télécommunications, etc. Avec le développement de travaux dans le domaine de l'Intelligence Artificielle, de nouvelles questions émergent, d'autres plus anciennes se posent avec une acuité nouvelle.

3.2 Travail éditorial

- Depuis 2017 : direction éditoriale de la revue en ligne *CogniTextes*, revue de linguistique cognitive à comité de lecture.

Membre du comité éditorial de la revue depuis 2012, j'ai notamment travaillé (avec Florent Perek et Maarten Lemmens) à l'obtention du référencement ERIH plus en 2015. Depuis 2017, j'ai fait des propositions pour améliorer le fonctionnement de la revue, étendu le comité scientifique et nommé deux responsables des recensions. Je souhaite également que la revue puisse proposer la traduction de textes fondateurs de langue anglaise, mais face à l'impossibilité d'obtenir les droits pour publier en libre accès nous nous orientons vers la publication de courts textes inédits et d'entretiens que nous traduirons ou ferons traduire.

- Evaluations ponctuelles pour les revues LIDIL (*Revue de linguistique et de didactique des langues*), LIA (*Language, Interaction and Acquisition*) (<https://benjamins.com/#catalog/journals/lia/main>), CoReLa et *Cahiers de Praxématique*
- Relectrice pour *Language Science Press* (<http://langsci-press.org/hallOfFame>)

3.3 Colloques et conférences (organisation et évaluation)

- GLAD2018 : Gérer l’anaphore en discours, vers une approche interdisciplinaire (avril 2018)

De nombreux travaux sur l’anaphore ont été consacrés à la définition du concept et à ses limites (deixis, anaphores sans antécédent), ainsi qu’au choix du type d’expression anaphorique en discours (statut cognitif, pragmatique). Plus récemment, des recherches se sont intéressées au traitement des anaphoriques par le récepteur humain (expériences chronométriques ou oculométriques par exemple) ou à sa modélisation computationnelle (traitement automatique des co-références notamment). Mais la question de la gestion de l’anaphore est encore en partie méconnue, notamment du fait de l’absence d’échanges systématiques entre les champs disciplinaires qui abordent le phénomène. Ce colloque visait à favoriser les échanges interdisciplinaires sur cette question, avec deux objectifs :

1) construire une meilleure connaissance des processus de compréhension et de production des anaphores, en prenant en compte des contextes de communication et des locuteurs très variés, notamment dans des situations de communication perturbées ou contraintes.

2) élaborer une réflexion sur la place de l’anaphore dans les théories du traitement de l’information. Cette question implique également une réflexion métalinguistique sur la pertinence de l’anaphore comme outil d’étude, face notamment au concept plus récent de chaîne de référence.

- AFLiCo6 : Langage, Cognition, Société (mai 2015)

L’objectif principal de ce colloque a été de confirmer et de renforcer le réseau de collaboration et de discussion des chercheurs en linguistique cognitive, en France dans un cadre international, comme avaient pu le faire les colloques AFLiCo précédents à Bordeaux (2005), Lille (2007), Nanterre (2009), Lyon (2011) et Lille (2013), et les journées JET à Bordeaux (2010) et Paris (2012 et 2014). Les thèmes du colloque s’inscrivaient dans l’orientation des colloques AFLiCo antérieurs, tout en prenant acte d’évolutions récentes du domaine : l’émergence d’une sociolinguistique cognitive, ainsi qu’un rapprochement entre sociolinguistique et travaux sur l’acquisition du langage. Une partie des contributions présentées ont été publiées dans la revue *CogniTextes*, réparties en trois numéros thématiques.

- Fabriques de la langue (octobre 2010)

Ce colloque a rassemblé des intervenants des domaines de la linguistique, de la psychanalyse et de la création littéraire ainsi qu’un public très divers autour d’un intérêt commun

pour le langage, ses mutations et (re)fabrications constantes. Les actes ont pris la forme d'un ouvrage collectif publié aux Presses Universitaire de France.

- **AcquisiLyon** (décembre 2009)

Ce colloque avait pour objectif de rassembler les doctorants et les jeunes chercheurs en acquisition du langage, en leur permettant de présenter et de confronter leurs travaux. Des chercheurs confirmés en acquisition du langage ont été invités à communiquer sur les thématiques en lien avec ce domaine de la linguistique. Les actes ont été publiés en ligne sur le site du colloque.

- Membre du comité scientifique de CEDIL'14, CEDIL'18, AFLiCo7 et pour le colloque « L'apprenant en langues et dans les métiers de la traduction » qui se tiendra à Rennes en janvier-février 2019 (<https://apprenant.sciencesconf.org/>)

3.4 Bases de données, dispositifs empiriques, documentation

[BD1] Corpus COP21 (bientôt disponible sur Ortolang)

[BD2] "Using corpora in translation teaching - a test for students", instrument accessible sur IRIS <https://www.iris-database.org/iris/app/home/detail?id=york%3a931044&ref=search>

[BD3] Guide Dynamique du logiciel CLAN, créé dans le cadre du projet ANR CoLaJe

3.5 Participation à des projets de recherche financés

[P1] Projets ANR Léonard : acquisition du langage et grammaticalisation (porté par A. Morgenstern, de 2006 à 2011)

Ce projet a pour objectif de départ de constituer une base de données de vidéos d'enfants francophones et bilingues, alignées avec leurs transcriptions. Il s'agit ensuite d'étudier l'apparition et le fonctionnement des marqueurs grammaticaux chez l'enfant de douze mois à trois ans. Le projet interroge l'acquisition du langage en lien avec la mise en place des marqueurs grammaticaux : nous avons en particulier cherché à établir l'ordre d'apparition des outils grammaticaux et leur lien avec les conduites linguistiques de l'enfant (explication, demande, récit...). <http://anr-leonard.ens-lsh.fr/>

[P2] Projet ANR ColaJe : Communication Langagière chez le Jeune enfant (porté par A. Morgenstern, de 2010 à 2013)

L'objectif du projet est de reconstituer l'émergence et le développement de la communication langagière chez le jeune enfant, avec une approche pluridisciplinaire et

multimodale. Le projet a permis de continuer les suivis longitudinaux commencés dans le cadre du projet LEONARD. L'analyse simultanée de la phonologie, de la prosodie, de la morpho-syntaxe, du dialogue et du mimo-gestuel nous offre une perspective enrichie du développement linguistique de l'enfant. <http://colaje.scicog.fr/>

[P3] Projet ANR Prens (porté par S. Kern, de 2011 à 2015)

Ce projet porte sur le développement lexical chez le jeune enfant. Il a pour ambition de parvenir à une approche holistique qui prend en compte l'interface entre la production et la perception. Dans ce cadre, l'acquisition lexicale est vue comme un processus émergeant de l'interaction entre les capacités motrices et l'apprentissage à partir de l'input de l'environnement linguistique. Ainsi, le principal objectif de ce projet est de prendre en compte à la fois l'influence des contraintes articulatoires et les propriétés de l'input dans le choix des mots et dans leur réalisation chez les jeunes enfants avant le stade des 50 mots et après celui-ci, et ceci, dans une perspective translinguistique. Le projet utilise des données existantes en français et en anglais américain, mais comprend aussi la collectes de nouvelles données, en arabe tunisien et en tachelhit : deux langues dans lesquelles l'acquisition des premiers mots n'a jamais été étudiées. (<http://www.ddl.cnrs.fr/projets/prens/>)

[P4] Projet « Connaissances collectives et cognition individuelle » : Co3i (porté par J-P. Chevrot)

Il s'agit d'un projet pluridisciplinaire dont l'objectif est de mieux articuler la double opposition entre collectif et individuel, et entre social et cognitif. La distinction entre collectif et individuel est en effet souvent superposée à l'opposition entre social et cognitif. Les phénomènes sociaux sont perçus comme collectifs alors que le cerveau de l'individu serait le siège de la cognition. Dans le cadre de ce projet, une première série de rencontres ont eu lieu, elles ont pris la forme d'ateliers confrontant quatre points de vue : les approches sociales de la cognition, les approches cognitives du social, l'approche systèmes complexes, la philosophie des sciences.

Le projet a été financé par la Mission pour l'Interdisciplinarité du CNRS, dans le cadre du défi Sciences Sociales et Cognitives des Comportements Collectifs, et porté par l'IXXI (Institut rhônalpin des systèmes complexes) et par le laboratoire ICAR (ENS Lyon) en 2017-2018. Il sera financé jusqu'en 2020 par l'IXXI et la MSH de Lyon et la rédaction d'un ouvrage collectif est en cours.

[P5] Projet « Le contexte linguistique : modèles théoriques, fondements cognitifs, opérativité informatique » : CL3i (porté par D. Perrin)

Ce projet part d'un constat : nous manquons d'une définition unifiée et satisfaisante de la notion de contexte, et a fortiori d'une théorie susceptible de faire consensus, ne serait-ce que minimalement, au sein de la communauté scientifique. Ce manque est dû notamment à la dispersion prononcée que manifestent les sciences du langage sur la notion de contexte. En effet, chaque discipline (informatique, linguistique, philosophie, psychologie, sociologie) emploie la notion, mais sans connaissance réelle de ce que ses partenaires potentielles seraient susceptibles de lui apporter. Pour cette raison, les apports respectifs des disciplines qui travaillent à la constitution de modèles théoriques généraux et de celles qui se consacrent principalement à la recherche expérimentale ne sont jamais ni confrontés ni alimentés les uns par les autres. Notre projet vise à combler ce manque : il cherche à mettre en œuvre une démarche théorique qui permette de construire des modèles du contexte linguistique en recourant aux méthodes expérimentales de la psycholinguistique et aux contraintes d'opérativité que met en évidence la modélisation informatique des modèles en question.

Ce projet a obtenu un financement « Initiative de recherche stratégique » IDEX, qui permettra de recruter un post-doctorant pour deux ans à l'automne 2018.

[P6] Projet « Evaluer la traduction automatique » (porté par C. Rossi)

Cette étude collaborative que je porte rassemble des chercheurs de trois équipes grenobloises (ILCEA4, LIG et PPL) pour proposer, dans un contexte où les approches statistiques et neuronales ont permis une amélioration extrêmement rapide de la qualité de la traduction automatique (TA), une approche interdisciplinaire liant traductologie, traitement automatique des langues et philosophie de la connaissance. L'étude s'est donné trois objectifs : dégager les usages et les perceptions des systèmes de TA statistique/neuronale chez des traducteurs professionnels et des apprentis traducteurs ; comparer ces usages et perceptions à l'architecture, au fonctionnement et aux potentialités effectives des systèmes ; et mettre en perspective ces comparaisons avec les conceptions de l'action et de la cognition sous-tendant la TA statistique/neuronale.

Le projet a reçu un financement du Pôle Grenoble Cognition, puis a fait l'objet d'un partenariat Hubert-Curien également financé (PHC Ulysses n° 41088TC, Rossi & Kenny)

3.6 Blogs (carnets)

J'ai créé deux carnets Hypothèses que j'alimente régulièrement, car ils correspondent à des projets de recherche en cours :

<https://evaluerlata.hypotheses.org/>

<https://co3i.hypotheses.org/>

3.7 Réponse à des appels à projet et participation à la constitution de réseaux internationaux

- **EcoTrad** : Montage d'un réseau scientifique international de recherche sur la traduction de textes sur l'environnement et leur réception, et préparation d'une réponse à l'appel MRSEI de l'ANR à l'automne 2018.
- **M2TLabs: Machine Translation Training Labs** (porté par Olga Torres Hostench, Université Autonome de Barcelone), partenariat stratégique Erasmus+ déposé en mars 2018.

Ce projet, qui a été classé sur liste d'attente en juillet 2018 et sera représenté en mars 2019, propose la mise en place de laboratoires de formation à la traduction automatique, qui placent traducteurs et étudiants en traduction au cœur de dynamiques innovantes. Des ateliers de créativité accompagnent le dispositif, pour permettre de valoriser l'expertise du traducteur et de repenser les interactions homme-machine. Les dispositifs pilote seront évalués et consolidés, puis diffusés dans le réseau des Masters Européens de Traduction (EMT). En particulier, nous souhaitons explorer des modèles de formation de traducteurs qui préparent les étudiants à des flux de travail qui font la part belle aux nouvelles technologies, en leur fournissant des compétences techniques et des connaissances appropriées, mais aussi des outils intellectuels qui leur permettront de faire face aux défis éthiques, juridiques et économiques.

Outre l'Université Grenoble Alpes, les partenaires universitaires sont : Dublin City University, l'université Rennes 2, l'université d'Alicante et les partenaires industriels sont KantanMT (située sur le campus de Dublin City Université) et Prompsit (qui travaille de près avec les chercheurs de l'université d'Alicante).

- **PEPS TRAMES : Traduction vers l'anglais médical d'expressions complexes.** Ce projet que j'ai soumis en 2015, en réponse au « Programme Exploratoire Premier Soutien », et qui n'a pas été retenu, proposait de regrouper des chercheurs de 3 équipes du site Grenoble-Alpes (ILCEA4, LIDILEM, LIG) spécialistes de

linguistique française et anglaise, de traductologie, de traitement automatique des langues et de traduction automatique. En tant que projet exploratoire premier soutien, il avait pour principal objectif d'évaluer la complémentarité des approches linguistiques et automatiques de la traduction sur une problématique linguistique, les groupes nominaux complexes et dans un domaine spécialisé.

- **ANR LIP- AV: Language Input and Production: Adaptability and Variation** (porté par C. Parisse). Ce projet soumis en réponse à l'appel à projets générique 2018 de l'ANR propose de travailler sur des événements langagiers rares, à partir de corpus denses d'acquisition de langage. Une nouvelle proposition sera envoyée en réponse à l'appel de 2019.
- **ANR CONTTRAD : Automatisation de Méthodologies d'Usages de Contextes en Traduction** (porté par E. Planas). Ce projet soumis en réponse à l'appel à projets générique 2018 de l'ANR propose de mettre en place un accès automatisé à des Contextes Riches en Connaissances (CRC) sélectionnés pour leur richesse linguistique et conceptuelle à partir de corpus comparables et issus de la mise en œuvre d'une méthodologie de repérage systématique. Ce projet ne sera à ma connaissance pas représenté en 2019.

3.8 Encadrement de la recherche

- Encadrement régulier de mémoires de L3 et de M1

Plusieurs mémoires de projet tuteuré (L3) : les étudiants travaillent par groupe de trois ou quatre sur un projet de recherche pluridisciplinaire, ils bénéficient d'un accompagnement tout au long de l'année et le mémoire est réalisé collectivement.

Encadrement de mémoires de traduction longue commanditée (M1, Traduction spécialisée multilingue). Les textes source comportent 5000 mots, la traduction est présentée et commentée, et les étudiants livrent également une liste terminologique.

- Encadrement de mémoires de traductologie (M2 Traduction Spécialisée Multilingue)

2016-2017 : co-encadrement du mémoire de Baptiste Dirand (avec Elisabeth Lavault-Olléon) : *Le néologisme biotraducteur : symbole de l'opposition homme-machine en traduction.*

2017-2018 : encadrement de quatre mémoires :

Elise Domingues : *Concilier l'inconciliable : TAO et littérature*

Virginie Libert : *Évaluation comparative de deux systèmes de traduction automatique pour la paire de langues allemand>français : MT@EC vs DeepL*

Julie Martin : *Usages de la traduction automatique pour la paire de langues japonais>français : réflexion sur les bonnes pratiques.*

Angelika Pinisch : *Mémoire sélective, conceptualisation bilingue et compétence linguistique d'un point de vue traductologique.*

- Co-encadrement d'un stage de recherche de 5 mois, sur les usages et perceptions de la traduction automatique

Ce stage a été effectué par un jeune diplômé du Master de Traduction spécialisée multilingue (Nicolas Becker), et financé dans le cadre du projet « Evaluer la traduction automatique » [P5].

Trois types de tâches ont été confiées à Nicolas :

1. L'annotation d'erreurs de sorties de TA issues des deux moteurs développés par l'équipe Getalp du laboratoire d'informatique où il était accueilli (LIG Moses et Seq2seq). Ce travail a été encadré par Emmanuelle Espérança-Rodier (LIG).

2. La réalisation d'une revue de littérature sur les principaux points abordés dans le cadre du projet. Ce travail a été encadré par Caroline Rossi (ILCEA4).

3. Sur la base des compétences acquises en accomplissant les deux premiers ensembles de tâches, Nicolas a ensuite été sollicité pour nous accompagner dans la mise en place du dispositif empirique qui constituera l'aboutissement du projet. La réflexion a porté sur le choix des outils, des moteurs et des textes utilisés pour cette tâche.

- Participation à des jurys de thèse

Fazia Khaled. *La communication des émotions chez l'enfant et l'adulte (colère, joie, tristesse) ; études de cas et confrontation des théories linguistiques.* Thèse dirigée par Aliyah Morgenstern (Université Paris 3 Sorbonne Nouvelle) et soutenue le 3 décembre 2016.

Olena Blagodarna. *Enhancement of Post-Editing Performance: Introducing Machine Translation Post-Editing in Translator Training.* Thèse dirigée par Pilar Sánchez Gijón (Université Autonome de Barcelone, Espagne) et soutenue le 15 octobre 2018.

- Co-encadrement de thèse

En collaboration avec Jean-Pierre Chevrot (LIDILEM), j'encadrerai la thèse de Wenjie Hong, étudiante diplômée en traduction de l'Université de Pékin. La thèse portera sur la traduction des métaphores dans le discours sur l'environnement, et intégrera un recueil de données auprès d'apprenants de langue seconde et de traducteurs (étudiants en M2 et professionnels).

3.9 Membre de sociétés savantes

- Membre du bureau de l'Association française de linguistique cognitive (Aflico, <http://www.aflico.fr/organisation.html>), je suis très impliquée dans ses activités courantes et j'ai participé à l'organisation de la sixième conférence internationale de l'association à Grenoble en 2015 (cf. supra).
- Membre de l'Association Française de Linguistique Appliquée (AFLA), de l'Association pour la Recherche en Didactique de l'Anglais et en Acquisition (ARDAA), de l'International Cognitive Linguistics Association (ICLA) et de la Société des Anglicistes de l'Enseignement Supérieur (SAES).

4 Publications et travaux

4.1 Articles dans des revues internationales à comité de lecture

[Rossi et al. 2019a] Glas, L., Rossi, C., Hamdi, R. et Batailler, C. (sous presse). Activity types and child-directed speech: a comparison between French, Tunisian Arabic & English. *Revue Canadienne de Linguistique*, 63(4).

[Rossi et al. 2019b] Rossi, C., Biroş, C. et Schmutz, H. (accepté). The case for environmental justice: tracking variation in do-it-yourself corpora, *Studia Neophilologica*.

[Rossi & Chevrot 2019] Rossi, C. et Chevrot J.-P. (2019) Uses and perceptions of Machine Translation at the European Commission. *Jostrans*, 31.

[Rossi 2017a] Rossi, C. (2017). Introducing statistical machine translation in translator training: from uses and perceptions to course design, and back again. *Tradumàtica*, 15.

[Rossi & Parisse 2012] Rossi, C. & Parisse, C. (2012) Categories in the making: Assessing the role of semantics in the acquisition of noun and verb categories, *Journal of French Language Studies*, 22.

4.2 Articles dans des revues nationales à comité de lecture

[Biros et Rossi 2019] Biros, C. et Rossi, C. (soumis)⁹³ Nommer les victimes du changement climatique : approche variationniste sur un corpus multilingue. *Mots. Les langages du politique*.

[Rossi 2019] Rossi, C. (accepté). L'apprenti traducteur et la machine : des connaissances aux perceptions de la traduction automatique. *Des mots aux actes*, 8 (numéro spécial, « Traduction et technologie : regards croisés sur de nouvelles pratiques » dir. E. Lavault-Olléon).

[Rossi & Dirand 2019] Rossi, C. et Dirand, B. (soumis)⁹⁴ Traducteur et biotraducteur. *Des mots aux actes*, 8 (numéro spécial, « Traduction et technologie : regards croisés sur de nouvelles pratiques » dir. E. Lavault-Olléon).

[Rossi et al. 2018] Biros, C., Rossi, C. et Sahakyan, I. (2018). Discourse on climate and energy justice: a comparative study of Do It Yourself and Bootstrapped corpora. *Corpus* (numéro spécial : petits corpus, sous la direction de C. Danino).

[Rossi 2013a] Rossi, C. (2013) L'expression de la manière de mouvement et son acquisition en français et en anglais. *Faits de langues* vol. 42.2 (Sémantique des relations spatiales, numéro spécial sous la direction de C. Chauvin) p. 199-216.

[Rossi et al. 2013] Rossi, C., Caët, S. & Morgenstern, (2013) Le futur en construction : premiers usages de *will* et *be going to* chez l'enfant anglophone, *Cercles*, 29. <http://www.cercles.com/n29/rossietal.pdf>

4.3 Directions de numéros de revues ou d'ouvrages

[Rossi et al. 2019c] Gardelle, L., Rossi, C. et Vincent-Durroux, L. (accepté). « La gestion de l'anaphore en discours : nouvelles perspectives théoriques et pratiques », *Cahiers de Praxématique*.

[Rossi 2017b] Rossi, C. (2017). Récits d'expérience : numéro spécial Aflico6 de la revue *Cognitextes*, publié suite à la sixième conférence de l'Association française de linguistique cognitive.

[Rossi et al. 2012] Nassikas, Prak-Derrington & Rossi (2012). *Fabriques de la langue*, Paris : PUF.

4.4 Chapitres d'ouvrages (publiés après évaluation par un comité de lecture)

⁹³ Les articles soumis ne figurent pas dans le recueil de publications et travaux.

⁹⁴ Les articles soumis ne figurent pas dans le recueil de publications et travaux.

[Rossi & Nassikas 2019] Rossi, C. et Nassikas, K. (accepté). « Langages perceptifs » du bébé et de l'adolescent. *Cures de langage* (Dir. F. Moncomble) collection Etudes Linguistiques, Artois Presses Université.

[Rossi 2017c] Rossi, C. (2017) Mouvement et développement du langage. In A. Morgenstern et C. Parisse (eds.) *Le langage de l'enfant : de l'éclosion à l'explosion*. Presses Universitaires de la Sorbonne Nouvelle.

[Parisse & Rossi 2017] Parisse, C. & Rossi, C. (2017) Semantic discrimination of noun-verb categories in French children aged 1;6 to 2;11. In E. Veneziano & V. Vapnarsky (eds.) *Lexical Polycategoriality Cross-linguistic, Cross-Theoretical and Language Acquisition Approaches*, John Benjamins.

[Rossi et al. 2016] Rossi, C., Frérot, C. & Falaise, A. (2016) Integrating controlled corpus data in the classroom: a case-study of English NPs for French students in specialised translation. In: Alonso Almeida, F., Cruz García, L., González Ruiz, V. (eds) *Corpus-based studies on language varieties*. Bern: Peter Lang, coll. Linguistic Insights, pp 167-190.

[Rossi 2015] Rossi, C. (2015). Les mots complexes en français contemporain : pour qui n'a pas grandi avec le Dinotrain. In M. Abecassis & G. Ledegen (eds.) *De la genèse de la langue à Internet*. Bern: Peter Lang.

[Rossi 2013b] Rossi, C. (2013) Relativité linguistique et innéité : ennemis d'un jour ? *L'innéité aujourd'hui*, éditions Matériologiques, pp 101-116.

[Rossi 2012] Rossi, C. (2012), « Des gestes qui font signe : fabriques mimétiques de la langue », in Nassikas, Prak-Derrington et Rossi (Eds.) pp 211-234 *Fabriques de la langue*, PUF.

4.5 Communications avec actes édités (après évaluation par un comité de lecture)

[Rossi et al. 2017] Esperança-Rodier, E., Rossi, C., Besacier, L. et Bérard, A. (2017) "Evaluation of NMT and SMT systems: A study on uses and perceptions". *Translating and the Computer 39* (Londres, 16-17 novembre 2017). Les actes sont disponibles en ligne : <https://www.asling.org/tc39/> (notre contribution se trouve p. 11-24)

[Rossi & Morgenstern 2008] Rossi, C. et Morgenstern, A. (2008). "Outils et méthodes de recherche en acquisition du langage : de la complémentarité entre statistiques et analyse linguistique" *9e Journées internationales d'Analyse statistique des Données Textuelles (JADT)*, Lyon, 12-14 mars, Heiden, S. & Pincemin, B. (eds), Lyon : Presses universitaires de Lyon /2, p. 1198 & sq.

[Rossi et al. 2007] Kochan, A., Morgenstern, A., Rossi, C. & Sekali, M. (2007). "Children's early prepositions in English and French: a social-interactional device" *LingO*, Oxford, 21-22 septembre. Accessible en ligne: <http://www.ling-phil.ox.ac.uk/events/lingo/papers/rossi.kochan.pdf>.

4.6 Diffusion de la recherche (publications à visée de vulgarisation)

[Rossi 2017-v] Rossi, C. (2017). De quoi la traduction automatique est-elle le nom ?, *En attendant Nadeau, journal de la littérature, des idées et des arts* ; dossier sur la traduction. <https://www.en-attendant-nadeau.fr/2017/07/25/traduction-automatique/>

[Rossi 2015-v] Deux contributions au volume de M. Abecassis et M. Block (2015). *French Cinema in Close-Up: La Vie d'un acteur pour moi*. « Brigitte Bardot, The Body Artist » et « Jamel Debbouze, Jack of All Trades ».

[Rossi 2012-v] Rossi, C. (2012) « Thinking in Tongues. An interview with John A. Lucy », *La Vie des idées*, 8 octobre 2012. <http://www.booksandideas.net/Langua...>

La version française également publiée par *La vie des idées* correspond à une traduction que j'ai réalisée : « Ce que le langage nous fait penser. Entretien avec John A. Lucy », <http://www.laviedesidees.fr/Ce-que-...>,

[Rossi 2009-v] Rossi, C. (2009) "Le langage en construction(s) : l'acquisition du langage, de l'observation aux approches théoriques.", *La Clé des Langues*, dossier "Acquisition du langage". <http://cle.ens-lyon.fr/plurilingues/langue/le-changement-linguistique/le-langage-en-construction-s>

4.7 Principales communications orales

[Rossi et al. 2018a-oral] Josselin-Leray, A., Rossi, C. & Sauron, V. (2018) Corpus et mémoires de traduction : deux approches du lexique et de la phraséologie en traduction spécialisée. *LTT*, Grenoble, 27-28 septembre 2018.

[Rossi et al. 2018b-oral] Biros, C., Talbot, A. & Rossi, C. (2018) "Translating the International Panel on Climate Change reports - Standardization of terminology in summaries for policy makers from 1990 to 2014". *Transius*, Genève, 18-20 juin 2018.

[Rossi et al. 2018c-oral] Rossi, C., Biros, C. & Talbot, A. (2018) « La variation terminologique en langue de spécialité : pour une analyse à plusieurs niveaux ». JET Aflico, Nanterre, 3-4 mai 2018.

[Rossi et al. 2017-oral] Rossi, C., Besacier, L. & Esperança-Rodier, E. (2017) « Uses and perceptions of Machine Translation at the European Commission ». Translation and Disruption: Global and local perspectives (Portsmouth, le 4 novembre 2017)

[Rossi & Renner 2017-oral] Rossi, C. et Renner, V. (2017). « L'acquisition du procédé morphologique d'amalgamation lexicale ». La néologie lexicale à travers les âges. Lyon, 8-10 juin 2017. Enregistrement de la conférence disponible en ligne : <https://webtv.univ-lyon3.fr/channels/NEOLEX/media/MEDIA170612190504315>

[Rossi 2017a-oral] Rossi, C. (2017). « Zero-shot translation: le traducteur humain et la traduction neuronale ». Vers une anthropologie de la traduction ? Divergences et convergences dans les cultures et les patrimoines. Avignon, 17-18 mai 2017.

[Rossi, Chevrot, Sahakyan et Batailler 2017-oral] Rossi, C., J.-P. Chevrot, I. Sahakyan et C. Batailler (2017). Teaching complex grammar: Do less advanced learners prefer examples? Thinking, Doing, Learning. Munich, Allemagne, 19-21 avril 2017.

[Rossi 2017b-oral] Rossi, C. (2017). L'apprenti traducteur et la machine : des connaissances aux perceptions de la traduction automatique. Congrès mondial de traductologie, Paris ouest Nanterre, 10-14 avril 2017.

[Rossi 2017c-oral] Rossi, C. (2017). Introducing SMT in translator training: From course design to professional issues and back again. 8^e Congreso Internacional de la Asociación Ibérica de Estudios de Traducción e Interpretación (AIETI8), Alcalá de Henares, Espagne, 8-10 mars 2017.

[Rossi et al. 2016-oral] Rossi, C., Biros, C. Roncato, C. et Schmutz, H. (2016). The case for environmental justice: tracking variation in do-it-yourself vs. web corpora. Nouvelles approches du corpus en linguistique anglaise, Avignon 9-10 juin 2016.

[Rossi & Nassikas 2015-oral] Rossi, C. & Nassikas, K. (2015) 'Langages perceptifs' du bébé et de l'adolescent. Colloque *Cures de langage*, Université d'Artois, 10-11 décembre 2015.

[Rossi & Frérot 2014a-oral] Rossi, C. & Frérot, C. (2014) "From Learner to Specialised Corpora: Integrating a Corpus-based Analysis of English NPs in French-English Translation Training", 4th Using Corpora in Contrastive and Translation Studies Conference (UCCTS 4), Lancaster, Royaume-Uni, 24-26 juillet 2014.

[Rossi & Frérot 2014b-oral] Rossi, C. & Frérot, C. (2014) Integrating controlled corpus data in the classroom: a case-study of English NPs for French students in specialized translation, CILC

2014: 6th International Conference on Corpus Linguistics, Las Palmas de Gran Canaria, Espagne, 22-24 mai 2014.

[Glas & Rossi 2014-oral] Glas, L. & Rossi, C. (2014) Activity contexts and early vocabulary development : a cross linguistic study, Georgetown University Round Table on Languages and Linguistics (GURT), Washington D.C. 13-16 mars 2014.

[Parisse & Rossi 2012-oral] Parisse, C. & Rossi, C. (2012) Statistical discrimination of semantic noun-verb categories in children aged 1;6 to 2;11. Fourth UK Cognitive Linguistics Conference, King's College, Londres, 10-12 juillet 2012.

[Rossi 2011-oral] Rossi, C. (2011) "Manner of motion in English and French child language: beyond lexicalization patterns", AFliCo IV, Lyon, du 24 au 27 mai 2011.

[Rossi 2010a-oral] Rossi, C. (2010) "L'acquisition des prédicats complexes en anglais : une étude des constructions avec UP." CEDIL 2010 (Colloque international des Etudiants chercheurs en Didactique des Langues et en Linguistique), Grenoble, du 29 juin au 2 juillet.

[Rossi 2010b-oral] Rossi, C. (2010) "UP in motion: Motion Events and the acquisition of complex predicates in English", Child Language Seminar, Londres, 24-25 juin.

[Rossi et al. 2010-oral] Morgenstern, A., Caët, S. et Rossi, C. (2010) "Un horizon en construction : l'expression de l'avenir dans les productions d'enfants anglophones de 1 ;06 à 3 ;06 ans", 50ème congrès SAES, Lille, du 21 au 23 mai.

[Rossi 2010c-oral] Rossi, C. (2010) "You finished that juice up, huh? Etude des premières constructions avec UP." 50ème congrès SAES, Lille, du 21 au 23 mai.

[Rossi & Morgenstern 2008-oral] Rossi, C. et Morgenstern, A. (2008) "Outils et méthodes de recherche en acquisition du langage : de la complémentarité entre statistiques et analyse linguistique" 9e Journées internationales d'Analyse statistique des Données Textuelles, Lyon, 12-14 mars, Heiden, S. & Pincemin, B. (eds), Lyon : Presses universitaires de Lyon /2, p. 1198 & sq.

[Rossi et al. 2008-oral] Turkay, F., Kern, S. et Rossi, C. (2008) "Pragmatic and contextual differences in French and Turkish Caregivers' input", XI International Congress For The Study Of Child Language (IASCL), Edimbourg, 28 juillet au 1er août.

[Rossi & Shimanek 2007-oral] Rossi, C. et Shimanek, C. (2007), "Contradiction et émergence du langage", Journée d'étude de l'Ecole doctorale HSH, Université Lyon 2, 1er juin 2007.

4.8 Principales conférences à titre d'invitée

[Rossi & Chabert 2019-inv] Rossi, C. et Chabert, E. (2019) « Réflexions sur l'enseignement de l'alternance entre génitif et syntagme prépositionnel en OF » Journée du CRELINGUA, Université Paris 3, Sorbonne Nouvelle.

[Rossi & Lavault-Olléon 2018-inv] Rossi, C. & Lavault-Olléon E. (2018). "Introducing MT as a heuristic tool in translator training". NMT Workshop for Teaching Staff, ZHAW, Zurich.

[Rossi 2017a-inv] Rossi, C. (2017). « Enseigner la grammaire anglaise en Licence LEA : pour une introduction progressive des règles ». Journée d'études sur l'enseignement de la grammaire anglaise, Université Paris 3, Sorbonne Nouvelle, 24 novembre 2017.

[Rossi 2017b-inv] Rossi, C. (2017). "Uses and perceptions of MT at the European Commission". *SALIS Seminar*, 18 oct. 2017, Dublin City University.

[Rossi 2017c-inv] Rossi, C. (2017). « Trois semaines à la DG-Traduction ». Séminaire interne du GREMUTS, 03 février 2017. <https://prezi.com/fccmpdu-o5s1/trois-semaines-a-la-dgt-bruxelles/>

[Rossi 2017d-inv] Rossi, C. (2017). « Traductologie appliquée : à la recherche du sens perdu ». Emission de *Radio Lingvistika*, Bruxelles, DGT, 17 janvier 2017. <https://webcast.ec.europa.eu/radio-lingvistika-170117>

[Rossi & Lavault-Olléon 2015-inv] Rossi, C. et Lavault-Olléon, E. (2015) "Intérêts du multilinguisme". Conférence débat du 4 novembre 2015 : *Une encyclopédie de l'énergie : pourquoi ? Comment ?*

[Rossi 2015-inv] Rossi, C. (2015). "A usage-based approach to translation: implications for translation teaching". *Approaches to Translation: An Interdisciplinary Workshop*, Warwick University, 27 février 2015.

[Rossi 2012-inv] Rossi, C. (2012). Relativité linguistique et innéité : ennemis d'un jour. *L'innéité aujourd'hui : questions scientifiques et perspectives philosophiques*, vendredi 30 mars 2012 à l'université Paris Ouest Nanterre.

5. Formations suivies

5.1 Tests statistiques dans R Studio

Cette formation de 18 heures, financée dans le cadre du projet Co3i, m'a permis de me familiariser avec l'outil libre R et l'environnement R Studio, en même temps que de développer une meilleure compréhension des principaux tests statistiques. L'accent était mis, en fin de parcours, sur les modèles de régression et sur l'apport des modèles mixtes.

5.2 Initiation aux démarches de créativité

Dispensée par Promising au sein de l'UGA, cette formation de 14 heures présentait la méthode CPS (*Creative Problem Solving*) et permettait d'en faire l'expérience. Elle rassemblait des enseignants de l'UGA venus de composantes variées et avait pour objectif de nous préparer à mettre en œuvre une démarche de créativité dans un de nos enseignements. J'ai suivi cette formation pour pouvoir appliquer la démarche dans mon cours sur la traduction automatique avec les étudiants de M2 (parcours TSM).

<https://www.promising.fr/>