

**RAPPORT DE SYNTHÈSE
DE L'ACTIVITÉ SCIENTIFIQUE**

présenté par :

Eva SCHAEFFER-LACROIX

pour obtenir l'**habilitation à diriger des recherches de l'université de Strasbourg**

Soutenance le **11 septembre 2020**

Spécialité : Études germaniques et sciences du langage

**RECHERCHES ET PRATIQUES OUTILLÉES
EN ALLEMAND ET FRANÇAIS**

Didactique, linguistique de corpus et traduction

GARANT :

Thierry GRASS

Professeur des universités, Université de Strasbourg

RAPPORTEURS :

Anne BAILLOT

Professeure des universités, Le Mans Université

Günter SCHMALE

Professeur des universités, Université Lyon III Jean Moulin

Denyze TOFFOLI

Professeure des universités, Université Paul Sabatier Toulouse III

AUTRES MEMBRES DU JURY :

Alex BOULTON

Professeur des universités, CNRS & Université de Lorraine

Laurent GAUTIER

Professeur des universités, Université de Bourgogne

Rapport de synthèse de l'activité scientifique
présenté par Eva Schaeffer-Lacroix
Maître de conférences en sciences du langage
à l'Inspé de l'académie de Paris (école interne de Sorbonne Université)
CeLiSo (Centre de linguistique en Sorbonne) - EA 7332, Paris
Membre associé de l'ILC (Institut du langage et de la communication), Strasbourg

Table des matières

Introduction	7
Positionnement	7
Présentation des chapitres	8
1. L’outillage comme point de convergence entre didactique et domaines voisins	11
1.1 Didactique et linguistique.....	13
1.2 Didactique et TAL.....	15
1.3 Didactique et traductologie	21
2. La recherche dans les formations universitaires	25
2.1 Enseignement et didactique des langues	25
2.1.1 Masters de type « Enseignement des langues »	25
2.1.2 Masters de type « Recherches en didactique des langues »	28
2.2 Outils numériques et linguistique outillée.....	30
2.2.1 Formations continues suivies	30
2.2.2 Formations continues dispensées	32
3. Un parcours de recherche outillé.....	35
3.1 Communications.....	35
3.2 Publications	38
3.3 Expertise scientifique	40
3.4 Animation et co-organisation d’événements scientifiques.....	41
3.5 Édition numérique	41
3.6 Veille scientifique	43
3.7 Insertion dans des réseaux scientifiques.....	45
3.7.1 Échange et partage	46
3.7.2 Entraide	47
4. La mise en œuvre de la recherche outillée	51

4.1 Réflexions méthodologiques et pratiques	51
4.2 Défis à relever	54
4.3 Récolte de données	57
4.3.1 Opinions	57
4.3.2 Actions	59
4.3.3 Productions	61
4.4 Transformation des données	62
4.4.1 Anonymisation	62
4.4.2 Transcription	62
4.2.3 Annotation	63
4.5 Analyse des données	64
4.5.1 Analyse qualitative	64
4.5.2 Analyse quantitative	66
4.6 Corpus d'experts et corpus d'apprenants	68
5. Un choix de projets outillés	73
5.1 Corpus et didactique des langues étrangères	73
5.1.1 <i>Weissensee</i>	73
5.1.2 <i>Weissensee Transfer</i>	74
5.1.3 <i>Kommas</i>	77
5.1.4 <i>Audiodeskription</i>	78
5.1.5 <i>Cahier des clauses techniques particulières</i>	81
5.1.6 <i>Kleinanzeigen</i>	84
5.2 Corpus et linguistique	88
<i>Noms communs de personnes dans le langage administratif suisse</i>	88
5.2.1 Constitution de l'équipe de recherche	91
5.2.2 Conception de la recherche	91

5.2.3 Valorisation de la recherche	92
5.2.4 Perspectives ouvertes par la participation au projet FNS	93
5.3 Corpus et traduction	94
<i>Les corpus numériques au service de la traduction d'audiodescriptions</i>	94
5.3.1 Mise en place d'une collaboration scientifique franco-allemande.....	94
5.3.2 Description du projet.....	95
5.3.3 Constitution de l'équipe de recherche	100
5.3.4 Gestion du projet	101
5.3.5 Constitution du corpus	101
5.3.6 Récolte de fonds	102
5.3.7 Valorisation de la recherche	102
6. Les évolutions visées.....	103
Conclusion.....	105
Références	107
Liste des principaux acronymes et sigles	125
Liste des figures	126
Liste des graphiques	127
Liste des tableaux	127
Annexes	129
Annexe 1 – Formations continues suivies.....	129
Annexe 2 – Expertise scientifique.....	130
Annexe 3 – Événements scientifiques animés ou (co-)organisés	131
Annexe 4 – Communications	132
Conférences invitées	132
Colloques à portée internationale	133
Colloques et journées d'étude à portée nationale.....	137

Annexe 5 – Engagements pour des sociétés savantes	138
Annexe 6 – Corpus mentionnés dans le rapport de synthèse	138
Annexe 7 – Traductions	139
Annexe 8 – Direction de mémoires de master 2	139

Introduction

Positionnement

Peu de spécialistes des langues étrangères travaillant dans le milieu universitaire en France réclament pour elles-mêmes l'étiquette de « didacticienne »¹ ; les pays voisins comme l'Allemagne et la Suisse semblent lui attribuer davantage de prestige. Ce fait m'amène à interroger le statut de la didactique des langues étrangères en France. Les didacticiennes travaillant dans ce pays peuvent parfois avoir l'impression que leur métier y est classé, de façon injustifiée, dans la même catégorie professionnelle que celui des formatrices en langue étrangère et de voir ainsi leur travail réduit aux aspects liés à la pratique qui concernent le terrain. Ces derniers, tout en étant fondamentaux, ne représentent qu'une des deux faces de la médaille de leur profession. L'équilibre entre théorie et pratique, qui est l'essence même de l'approche didactique, est rappelé dans le descriptif de l'ouvrage de Keck et Kim (2014) portant sur la grammaire pédagogique.

The book balances research and theory with practical discussions of the decisions that teachers must make on a daily basis, offering guidance in such areas as materials development, data-driven learning, task design, and classroom assessment.

[Ce livre établit un équilibre entre, d'une part, la recherche et la théorie et, d'autre part, les discussions pratiques sur les décisions que les enseignants doivent prendre au quotidien, offrant des conseils dans des domaines tels que l'élaboration de matériel, l'apprentissage porté par les données, la conception des tâches et l'évaluation en classe.]²

La recherche d'un tel équilibre entre théorie et pratique complexifie le travail de recherche en didactique : il n'est pas toujours facile de faire le tri parmi les données multiples du terrain d'enseignement-apprentissage, les unes pouvant être considérées comme toutes aussi précieuses que les autres. Dans ce domaine, il n'est pas d'usage de sélectionner, parmi les membres d'une communauté, ceux qui seraient « représentatifs » et d'enlever les « individus

¹ Je renonce, à regret, à adopter une écriture intégralement inclusive dans ce rapport de synthèse. La langue française ne semble pas faciliter sa mise en pratique (Elmiger et al., 2019). Toutefois, comme la didactique et l'enseignement des langues étrangères sont fortement marqués par des femmes, je leur fais honneur dans ce texte en appliquant le féminin générique, que je considère comme le pendant du masculin générique, pour des termes comme « didacticienne », « enseignante » et « formatrice ».

² Dans ce rapport de synthèse, toutes les traductions, sauf mention contraire, ont été soutenues par DeepL (DeepL GmbH, 2017) (avec post-édition).

en marge » du profil du groupe ; la recherche en didactique des langues est inclusive de ce point de vue. Toutefois, les didacticiennes ne renoncent pas systématiquement à la création d'échantillons représentatifs pour mener leurs recherches. Dans le cadre de certains projets et en lien avec certains questionnements, elles peuvent trouver opportun d'avoir recours à des méthodes d'analyse statistique (voir paragraphe [4.5.2](#)). Selon mon positionnement, les outils numériques, dont les outils de corpus, ont un rôle particulier à jouer dans la mise en synergie de la recherche et du terrain. Comme exemples, on peut citer la mise en perspective de l'enseignement-apprentissage de la grammaire par les corpus numériques proposée dans Schaeffer-Lacroix (2018b) et la création d'activités d'apprentissage faisant appel à un corpus, développées dans le cadre d'une formation d'élèves-ingénieurs à un écrit professionnel répandu (Salengros Iguenane, Schaeffer-Lacroix, 2020). Ce positionnement est compatible avec les pistes pour une méthode intégrative offertes dans Schmale (2015).

Présentation des chapitres

Le chapitre 1 de ce rapport de synthèse situe la didactique des langues par rapport à trois domaines voisins.

Dans la section 1.1, la didactique des langues (à l'exemple de la didactique de l'allemand) est considérée comme le dénominateur commun entre la linguistique et le domaine pratique représenté par l'enseignement d'une langue.

La section 1.2 s'intéresse au TAL (traitement automatique des langues), compris comme un domaine de recherche qui a comme pendant pratique la linguistique outillée. La linguistique de corpus joue le rôle de pivot entre les deux domaines, comparable au rôle attribué à la didactique, qui se situe, selon mon positionnement, à l'intersection entre la linguistique et la pratique de l'enseignement.

La section 1.3 présente la traductologie comme le versant scientifique d'un domaine dont les pendants pratiques sont le secteur de l'enseignement de la traduction et la pratique professionnelle de la traduction. On notera une évolution inégale dans le secteur de l'enseignement secondaire, où l'activité de traduction vient de traverser une période de délaissement presque total, et dans les cursus de l'enseignement universitaire, portés par les besoins professionnels importants ainsi que par les évolutions dans le domaine de la traduction outillée (semi-)automatique et, de façon plus générale, dans le domaine de l'apprentissage par la machine (Carstensen, 2017, p. 184).

Le chapitre 2 est dédié au rôle que jouent la recherche et l’outillage dans le secteur de la formation. Le premier volet s’intéresse à la formation initiale des enseignantes et des didacticiennes. Dans le deuxième volet, je précise d’abord les formations continues en linguistique outillée que j’ai suivies et ensuite celles que j’ai dispensées dans les murs de l’Inspé (Institut national supérieur du professorat et de l’éducation), de Sorbonne Université et de l’EHESS (École des hautes études en sciences sociales).

Le chapitre 3 retrace mon parcours de chercheuse, qui est mis en perspective à l’aide d’un aperçu d’organes français de la recherche en didactique des langues. J’y présente les revues et colloques qui me paraissent centraux pour ce secteur, et je mentionne mon implication dans ce domaine. De plus, on trouve dans ce chapitre un bilan des activités de recherche que j’ai menées pendant ma carrière universitaire. J’y mentionne les événements que j’ai co-animés ou (co-)organisés ainsi que les activités d’édition, de veille et d’insertion dans des réseaux scientifiques.

Le chapitre 4 se concentre sur l’usage d’outils numériques lors de projets de recherche. J’y décris la façon dont les outils sont mis au service de l’observation du terrain, de la récolte et de l’analyse de données. Il y est également question d’interventions possibles sur les données. La dernière section de ce chapitre précise le rôle que peuvent jouer les corpus d’experts et les corpus d’apprenants pour faire de la recherche.

Le chapitre 5 présente les projets outillés que j’ai conçus et/ou auxquels j’ai participé dans le passé. De plus, on y trouve le descriptif de deux projets en cours d’exécution et un projet qui est en train d’être développé. Ce chapitre est divisé en trois sections, faisant toutes appel aux corpus numériques. La première est adossée à la didactique des langues, le deuxième à la linguistique et le troisième à la traduction.

Le chapitre 6 évoque des évolutions possibles que j’entrevois pour mes activités dans le domaine de la didactique, de la linguistique de corpus et de la traduction, tous les trois mis au service de l’inclusion sociale.

1. L’outillage comme point de convergence entre didactique et domaines voisins

Un des thèmes qui ont particulièrement façonné mon parcours d’enseignante et de chercheuse

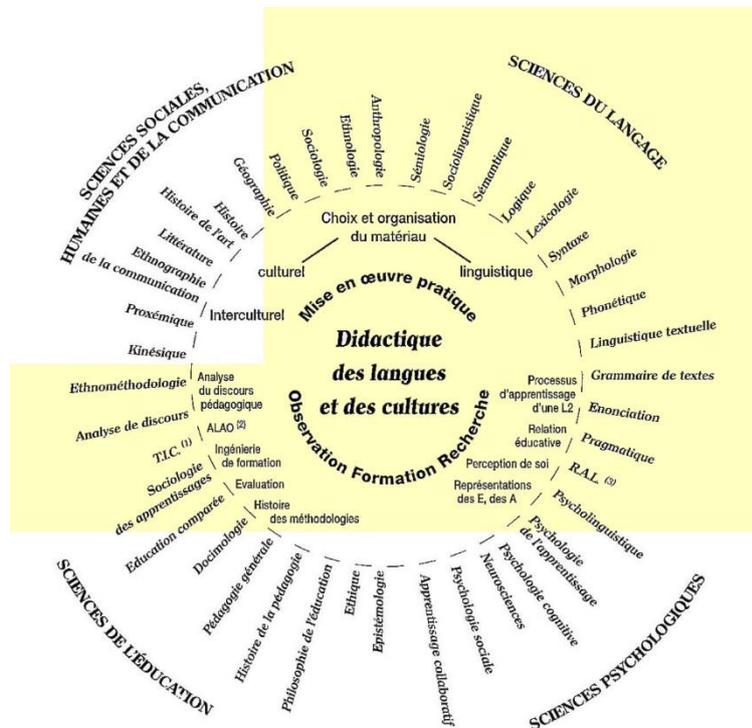


Figure 1 – La marguerite de la didactique.

est celui de l’exploration des « zones de collaboration » (Véronique, 2010) entre plusieurs domaines de recherche. De sa part, Pothier (2001) présente la didactique comme « lieu de rencontre de divers spécialistes, chacun travaillant dans son optique particulière pour le plus grand bien de tous ». Sa « marguerite de la didactique » présente toute une palette de disciplines pouvant être envisagées pour la collaboration entre ce qu’elle appelle « didactique des langues et des

cultures » et d’autres domaines appartenant à un des quatre groupes de disciplines suivants : les sciences sociales, humaines et de la communication, les sciences du langage, les sciences psychologiques et les sciences de l’éducation.

Deux des secteurs (surlignés en jaune en Figure 1) ont été retenus pour présenter mes travaux de recherche, celui des sciences du langage, élargi à gauche vers les sciences sociales, humaines et de la communication, et un bout du secteur se situant entre les sciences de l’éducation et les sciences sociales, humaines et de la communication, incluant l’analyse de discours et les « T.I.C. » (Technologies de l’information et de la communication), que l’on appellerait aujourd’hui « technologies éducatives ». Ces dernières s’intéressent au rôle des outils numériques pour soutenir l’enseignement-apprentissage. On notera mon évitement de deux autres zones, celle des sciences de l’éducation et celle des sciences psychologiques. Je ne me suis pas non plus beaucoup aventurée dans la partie de la zone « sciences sociales, humaines et de la communication » incluant les mots-clés « culturel » et « interculturel ». On peut s’en

étonner : depuis quelques années, ces termes ont beaucoup marqué la didactique des langues, comme en témoigne l'évolution des publications affichées par le moteur de recherche Ifindr (Ifindr, 2019) pour les mots-clés anglais « intercultural, language learning ». Pour l'année 2018, 399 publications sont recensées.



Figure 2 – Mots-clés « intercultural, language learning » (Ifindr).

Mon absence de positionnement explicite par rapport au domaine du culturel et interculturel ne veut pas dire que les contenus concernés par ces entrées ne m'intéressent pas. Toutefois, si je devais analyser les phénomènes qu'ils recouvrent, j'aurais tendance à les aborder par l'angle de la pragmatique, autrement dit la science du contexte (Kerbrat-Orecchioni, 2012), de la linguistique contrastive (Guillemin-Flescher, 1981 ; Janitza, Travers, 2003 ; Gautier, Bach, 2017) ou du plurilinguisme (Mochet, Coste, 2005 ; Castellotti, 2010 ; Elmiger, Matthey, 2017). Dans Schaeffer-Lacroix (2015b), j'utilise le terme d'« extralinguistique », pouvant être associé à ce que d'autres appelleraient « culturel » ou « interculturel », comme en témoigne la citation suivante.

Pourquoi avoir fait le choix d'intégrer les discussions métalinguistiques et le travail sur corpus dans un scénario d'enseignement-apprentissage ? Le sens se construit de façon complexe. Quand on cherche à comprendre un fait de langue, il ne suffit pas toujours de regarder une ligne de concordance isolée et figée : il convient de tenir également compte du cotexte qui doit parfois être élargi pour permettre une interprétation adéquate (...). Les contextes discursif et extralinguistique peuvent également fournir des éléments d'interprétation des marqueurs linguistiques. Les indications d'action et d'agencement de personnes ou choses fournies dans les didascalies sont mises en scène dans la version filmée de Weissensee. Le visionnement du DVD [pendant la formation] donne accès à des informations susceptibles de soutenir la définition du sens d'une particule ou préposition.

D'ailleurs, je rajouterais volontiers une entrée à la marguerite représentée en figure 1, à savoir la traduction (peut-être à proximité de l'entrée « interculturel »), qui, à la période à laquelle ce schéma a été publié, n'était pas souvent proposée comme activité d'enseignement-apprentissage d'une langue : pour l'année 2001, seulement 146 publications sont recensées par Ifindr pour les mots-clés anglais « translation, language learning ». En 2018, on peut observer un pic de 747 publications.



Figure 3 – Mots-clés « translation, language learning » (Ifindr).

Ce rapport de synthèse vise à questionner les échanges possibles entre la didactique des langues étrangères (exemplifiée principalement à l'aide de l'allemand, mais aussi à l'aide du français) et trois des domaines de recherche ou des praxéologies situés dans des champs voisins, en l'occurrence la linguistique, le TAL (en particulier son volet « linguistique de corpus ») et la traductologie, dont le pendant pratique est l'activité de traduction. L'outillage structure de façon récurrente mes activités de recherche et d'enseignement. Il est l'élément pivot de l'approche interdisciplinaire évoquée ci-dessous : je considère l'emploi de ressources et d'outils numériques comme dénominateur commun entre linguistique, didactique, TAL et traductologie, domaines que j'invite à entrer en dialogue dans ce texte.

1.1 Didactique et linguistique

La séparation entre pratique et recherche, en l'occurrence entre l'enseignement de l'allemand et la linguistique allemande, est souvent déplorée en France. Dalmas (2019) plaide pour un rapprochement entre les deux, qui pourrait être effectué à l'aide de la didactique des langues.

Die bis heute starre Trennung zwischen Lehre an den Schulen und Forschung an den Universitäten, die sich bis in die Lehrerausbildung und -Weiterbildung auswirkt, führt oft zu Unkenntnis oder gar Unverständnis auf beiden Seiten. Dies ist einer der Gründe, warum es nur wenig Forschungsarbeiten gibt, die sich mit der Didaktik des Deutschen befassen, Arbeiten, die endlich eine Brücke zwischen Forschung und Lehre schlagen würden und dem gesamten Fach nur zugute kommen könnten.

[La séparation, à ce jour encore stricte, entre l'enseignement dans les établissements scolaires et la recherche à l'université, qui a un impact sur la formation initiale et continue des enseignants, conduit souvent à l'ignorance, voire à l'incompréhension des deux côtés. C'est l'une des raisons pour lesquelles il n'y a que peu de travaux de recherche sur la didactique de l'allemand, des travaux qui permettraient enfin d'établir une passerelle entre recherche et enseignement et qui profiteraient grandement à toute la discipline.]

Cette demande de rapprochement entre recherche et pratique rappelle celle prononcée par Johns (1997) dans le contexte du *kibbitzing*, une activité consistant à explorer des corpus numériques dans le but de réfléchir sur la langue et de réviser un texte écrit en langue étrangère. Je m'y

réfère dans Schaeffer-Lacroix (2016c).

Johns (1997) reports that these kibbitzing sessions were felt to be very helpful by the participants, but he regrets that they did not have a strong impact on foreign language teaching.

[Selon Johns (1997), ces sessions de kibbitzing ont été ressenties comme très aidantes par les participants, mais il regrette qu'elles n'aient pas eu de retombées sensibles sur l'enseignement des langues étrangères.]

Demaizière et Berthoud en 1998 et Chini (2009), de leur côté, argumentent en faveur d'un partenariat égalitaire entre didactique et linguistique anglaises. Le manuel de collègue *Behind the words*, conçu à la suite de recherches en linguistique et d'expériences sur le terrain par le collectif Charlirelle auquel appartenaient Chini et Demaizière, en est un exemple parlant (Charlirelle, 1975). Est-ce qu'un rapprochement comparable existe ou pourrait être envisagé entre linguistique germanistique et DaF³ (allemand langue étrangère) ? Et si oui, quels types de relation sont envisageables ?

Pieklarz-Thien (2018), chercheuse germaniste en Pologne, présente trois conceptions selon lesquelles la didactique de l'allemand langue étrangère...

(i) dépendrait de la linguistique germanistique ;

(ii) serait une discipline extérieure à la linguistique germanistique ;

(iii) créerait des savoirs et descriptions de la langue spécifiques, adaptés aux besoins et réalités de l'enseignement-apprentissage des langues.

La vision (iii), illustrée en Figure 4, permet de placer le DaF au centre des questionnements et actions de recherche s'intéressant à l'enseignement-apprentissage de l'allemand. Elle illustre le fait qu'en didactique des langues, il est d'usage de sélectionner, parmi les méthodes, ressources et résultats des champs voisins disponibles (p. ex. les sciences de l'éducation, la psychologie ou la linguistique), les éléments susceptibles de contribuer à fournir des réponses pertinentes aux questions des chercheurs et à celles des praticiens.

³ J'utilise le sigle "DaF" (Deutsch als Fremdsprache) en raison de l'absence d'une abréviation établie pour "allemand langue étrangère".

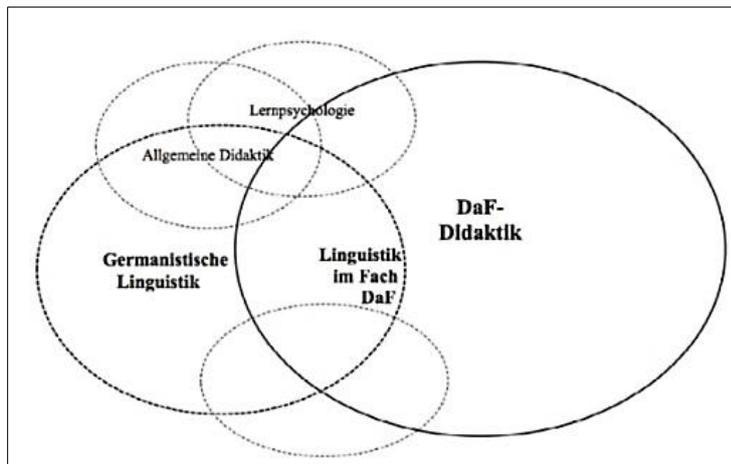


Figure 4 – Le DaF comme point de départ des questionnements scientifiques.

Cette idée rappelle le point de vue suivant formulé par Pothier (2001).

Pour préciser les rapports de la linguistique et de la didactique, je dirais qu’il n’est pas question de prendre la linguistique comme modèle pour penser la didactique, mais de se servir des analyses et des descriptions de la linguistique pour fonder son travail sur la langue.

Mon article sur le développement de compétences de trois apprenantes dans le domaine de la réflexion sur la langue rapproche les méthodes d’analyse de linguistes, traducteurs et enseignants de langue.

‘Expert-like language observation’ means using language analysis methods which can be compared to those applied by academic linguists, professional translators and—such is my claim—even by language teachers. (Schaeffer-Lacroix, 2016c)

[L’observation experte de la langue veut dire que l’on se sert de méthodes d’analyse comparables à celles utilisées par des linguistes universitaires, des traducteurs professionnels et – selon mon positionnement – même par des enseignants de langue (Schaeffer-Lacroix, 2016c).]

Ce positionnement dénote une vision intégrative, voire interdisciplinaire. Mais il me semble nécessaire d’aller plus loin encore et de s’ouvrir à l’international, aussi bien dans le domaine de la didactique que de la linguistique allemande, afin de poursuivre la « tentative de médiation entre la tradition française et la tradition allemande de l’étude de la langue allemande » recommandée par Rühl (2000) il y a 20 ans déjà.

1.2 Didactique et TAL

Selon Habert (2004), le traitement automatique des langues est un domaine susceptible de mettre à disposition de la linguistique (et selon moi, aussi de la didactique et de la traduction) des outils qui, bien compris et utilisés à bon escient, peuvent compléter de façon avantageuse

l'appareil méthodologique existant.

Une bonne partie de la recherche linguistique ne suppose pas d'instrument particulier. Les travaux en linguistique ont cependant à gagner à intégrer les outils fournis par le TAL (traitement automatique des langues) et par l'informatique de manière plus large. L'apport sera d'autant plus fructueux si l'on ne se cantonne pas à des emprunts de ressources ou de techniques et si s'approfondit le dialogue entre les savoirs de chaque discipline.

Tanguy et Fabre (2014) considèrent le TAL comme une technologie et la linguistique outillée comme une pratique qui se sert de méthodes issues du TAL. Ils citent la linguistique de corpus comme exemple illustrant la relation entre les deux. Dans la figure 5, ces chercheurs retracent l'évolution du TAL en la juxtaposant à celle de la linguistique de corpus. Au fil du temps, la taille des corpus s'agrandit, de concert avec l'évolution d'outils TAL permettant d'annoter et d'interroger les données rassemblées dans les corpus.

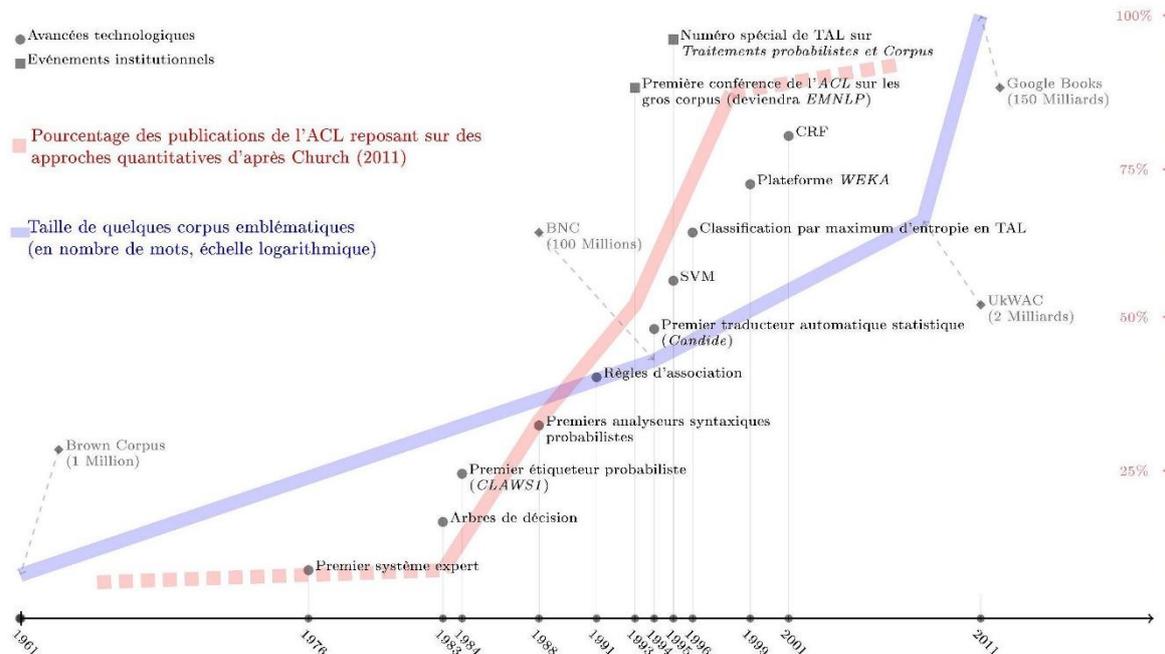


Figure 5 – Quelques repères chronologiques de l'évolution du TAL.

Les outils de TAL facilitent l'accès à une grande quantité de données. L'application de méthodes d'investigation quantitatives permet de découvrir des régularités et des variantes d'usage que l'on ne peut pas déceler autrement ou pas de façon aussi immédiate et confortable (Tanguy, Fabre, 2014). Selon Abney (2011), la scientificité de la linguistique en serait accrue : « la linguistique a désormais la capacité de tester ses hypothèses en opérant un traitement rigoureux des données ». L'outillage de la linguistique a toutefois des limites : tout ne se laisse pas concevoir, décrire et expliquer à l'aide de procédés automatisés. Si la machine est experte

dans le domaine du repérage d'éléments à haute fréquence et/ou de surface (p. ex., l'identification de suites de mots ou de combinaisons de mots) et d'éléments standardisés, elle peine à analyser les structures profondes, et elle ne détecte pas toujours de façon satisfaisante la variété liée à des usages créatifs ou régionaux de la langue (Fuchs, 2014). Certains des résultats obtenus lors du projet *Politique linguistique et usage de la langue dans la suisse plurilingue: noms communs de personne dans le langage administratif* (v. section 5.2), auquel j'ai participé en tant que linguiste de corpus, reflètent les limites d'une analyse de discours menée à l'aide de données mises en corpus. Ce constat concerne principalement les données en langue française, dont l'étiquetage automatique à l'aide de TreeTagger (Schmid, 1994) pose de réels problèmes conceptuels et techniques (Elmiger et al., 2019). Le projet *Les corpus numériques au service de la traduction d'audiodescriptions* (v. section 5.3) vise à distinguer, dans un corpus de scripts d'audiodescriptions de séries télévisées, les éléments standardisés, traitables par des outils de traduction, des éléments nécessitant une intervention humaine.

Les difficultés de la machine reflètent d'ailleurs souvent celles des apprenants d'une langue étrangère et parfois même celles des linguistes. Elles peuvent concerner le domaine des unités pluricatégorielles (Kravchenko-Biberson, 2012), dont les caractéristiques ouvertes ont pour conséquence qu'il n'est pas toujours aisé de distinguer entre les différentes fonctions et les sens variés d'une même forme. Pour donner un exemple, « ohne » [sans] peut être une préposition (« ohne Behinderung » [sans handicap]), introduire un groupe infinitif avec « zu » (« ohne zu... machen » [sans faire...]) ou s'associer à une conjonction de subordination (« ohne dass » [sans que]). Ces différents traits de « ohne » ont fait l'objet d'une discussion métalinguistique lors du projet *Kommas* (v. paragraphe 5.1.3), exécuté avec de futures enseignantes d'allemand. L'extrait ci-dessous provient de Schaeffer-Lacroix (2018d).

(...) certains des éléments observés sont pluricatégoriels (...). Ce fait est également l'objet d'une discussion métalinguistique menée entre Germiné et Sarah. Germiné s'étonne de découvrir *ohne* [sans] dans la liste des occurrences des conjonctions de subordination.

Germiné : *ohne* c'est pas une conjonction de subordination, non ? . Ah, mais c'est parce qu'il y a *zu* ? Ah non.

Sarah : Mais SI. à, dans, par, pour, envers, de, sans, sous . ça, c'est les conjonctions de subordination.

Germiné : =[Ah non,]

Sarah : c'est des conjonctions [de coordination]

Germiné : *ohne*, °°c'est (pas)°° une conjonction de (léger rire) subordination.

Sarah : xxx

Germine : On va le: garder en tête.

Sarah : Oui.

Germine : On va (le noter.)

Sarah : Oui, tu peux noter que: que: hein, on se posé la question du *ohne*.

Germine : Oui. Hein .

Dans cette interaction, un flottement catégoriel peut être constaté pour Sarah. En effet, la préposition *ohne* (sans) peut apparaître au sein de la locution *ohne dass* (sans que) introduisant un groupe conjonctionnel et dans *ohne zu* (sans) introduisant un groupe infinitif. Un questionnement comparable se trouve dans les interactions verbales de Germine et Sarah et de Chantal et Fanny quand elles cherchent le métaterme correspondant à *um zu (machen)* [pour (faire)].

Sarah et Germine ont observé la forme « ohne » dans le corpus *Aufsätze* (v. section 4.6), de très petite taille (4907 mots), à l'aide de Sketch Engine (Kilgarriff et al., [sans date]). L'étiqueteur TreeTagger y analyse cette forme comme préposition située à gauche de l'élément qu'elle spécifie (« APPR – Präposition: Zirkumposition links ») (Schiller et al., 1995, p. 7) ou comme conjonction de subordination avec « zu » et infinitif (« KOUI – Unterordnende Konjunktion mit 'zu' und Infinitiv ») (Schiller et al., 1995, p. 58). Si la forme « ohne » précède « dass », elle est étiquetée comme préposition (ibid.).

POS =	Beschreibung	Beispiele
KOUI	unterordnende Konjunktion mit Infinitiv	{ <i>um [zu], ohne [zu], anstatt [zu], statt [zu]</i> }
Aber:		
APPR	Präposition	ohne /APPR <i>daß er es weiß</i>

Figure 6 – Étiquetage de « ohne »[sans] par TreeTagger.

Comme l'illustrent la première et la troisième ligne de concordance de la figure 7, TreeTagger a tendance à attribuer, de façon non pertinente, l'étiquette « APPR » à des occurrences relevant de la catégorie « KOUI » (conjonctions de subordination avec « zu » ou un infinitif).

word="ohne" & depos="APPR"]			
	Contexte gauche	Pivot	Contexte droit
	dann muss vorsichtig versucht werden, zurückzulenken,	ohne	die andere (auch richtige / mögliche Antwort) zu übergehen.
	, dass die ganze Sache zuende ist,	ohne	dass er " sein " Gesicht " verlieren musste ". In
	Bildung zu einer Summe abfragbarer Fakten zu verringern	ohne	den entscheidenden Bildungsgehalt zu bewahren. 2. oder durch

Figure 7 – Étiquettes erronées de TreeTagger.

Afin d'évaluer s'il s'agit d'une erreur liée au traitement des données du corpus *Aufsätze*, j'ai vérifié les étiquettes de « ohne » dans le corpus *Instruktionen* d'une taille de 575 068 mots (v. Annexe 6). Cette forme y apparaît 100 fois. L'étiquette « KOUI » est quatre fois attribuée

de façon erronée dans un ensemble de 29 occurrences. Concernant l'étiquette « APPR », j'ai constaté 24 erreurs pour 71 occurrences. En tout, le score d'erreur s'élève à 28% ce qui dépasse de beaucoup les 5% que l'on considère comme seuil acceptable dans le domaine de l'étiquetage automatique. Le problème est donc structurel et ne peut être résolu sans post-édition des étiquettes annotées de façon automatique.

Les liens entre TAL et didactique semblent moins établis que ceux entre TAL et linguistique. Pour les mettre en synergie, le détour par la linguistique, partenaire privilégié des deux domaines d'étude, peut s'avérer utile. André Salem, spécialiste en TAL et en lexicométrie, par un concours de circonstances devenu mon directeur de thèse vers la phase finale de rédaction, a démontré par son ouverture d'esprit que l'intercompréhension entre les deux domaines peut être cultivée. De mon côté, ce rapprochement a été facilité par deux formations à l'université de Constance (Allemagne), une en lexicologie française, dispensée par le romaniste Christophe Schwarze, et une en TAL auprès du romaniste Achim Stein, connu pour avoir créé le modèle de l'ancien français pour l'étiqueteur automatique TreeTagger. À l'occasion de la conférence ARANEA organisée par la Chaire UNESCO de l'Université Comenius de Bratislava (v. [Annexe 4](#)), j'ai noué des liens avec Serge Sharoff, créateur de l'étiqueteur TreeTagger pour le russe et de l'étiqueteur et *tokenizer* pour le chinois⁴.

Mon tout premier exposé scientifique a eu lieu lors d'une Journée d'ATALA (Association pour le Traitement Automatique des Langues). Il a porté sur les outils comme support pour la production écrite. Lors de mon inscription à cette journée, son organisateur, Aurélien Max, m'a encouragée à y présenter mon embryon de thèse. Cette première excursion dans le monde des « talistes » m'a fait connaître trois personnalités du domaine qui m'ont marquée, à savoir Agnès Tutin (pour le domaine des corpus), Jean Véronis (pour des outils numériques, dont l'humoristique « pourrisseur de textes ») et Michael Zock, un chercheur qui travaille au CNRS (Limsi Orsay, puis LFI Marseille) et qui s'intéresse aux conditions favorisant l'apprentissage humain. Zock établit un lien entre le développement d'outils et la prise en compte des besoins et contraintes des usagers (il évoque « *the human factor* », donc le facteur humain), et il plaide pour une démarche interdisciplinaire pour faire vivre cette approche. Dans Schaeffer-Lacroix (2011), j'ai défendu un point de vue comparable en disant que l'interface d'outils de corpus

⁴<https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>

devrait être adaptée aux usagers à l'aide d'une collaboration entre linguistes, didacticiens et spécialistes de TAL.

Le choix de considérer les corpus comme espaces discursifs fournit des arguments forts en faveur du recours aux corpus réfléchis pour l'enseignement-apprentissage des langues étrangères. Il est parfois nécessaire de constituer de telles ressources soi-même. Ce procédé est coûteux en temps, et il demande certaines compétences techniques. Cela peut justifier la collaboration entre linguistes, didacticiens et spécialistes en TAL, au moins pour des projets d'une certaine ampleur. De tels projets devraient viser à proposer des corpus offrant deux types d'interfaces : une interface pour les professionnels en linguistique de corpus et une interface gérable par des non-spécialistes en informatique, en proposant un nombre suffisant de types de requête.

Quelques années plus tard, pendant une recherche que j'ai menée avec de futures enseignantes d'allemand, Heike, une des participantes, a réclamé la création d'un « concordancier pour les jeunes » (Schaeffer-Lacroix, 2019a, p. 61).

(...) both teachers and learners are willing to learn to use quite complex tools (e.g. Glogster) if they perceive that they offer exciting pedagogical features (...). Heike suggested further development of the interface of Sketch Engine was needed to tailor it to a young audience: "Perhaps is it the interface. It is too complicated, in fact. Perhaps one should develop a concordancer for the young". Heike's claim for user-friendly corpus tools must be taken seriously (...); encouragingly, the new Sketch Engine interface (designed in 2018) is evolving in this direction, and SkELL (Sketch Engine for Language Learning) (...) is finally available for German. What is more, this corpus management system can now easily be accessed on tablets and smartphones, corresponding to the digital devices most often used by young people.

[(...) aussi bien les enseignants que les apprenants sont prêts à apprendre à utiliser des outils assez complexes (par exemple Glogster) s'ils estiment qu'ils offrent des caractéristiques pédagogiques intéressantes (...). Heike a suggéré de développer davantage l'interface de Sketch Engine pour l'adapter à un public jeune : « Peut-être est-ce l'interface. En fait, c'est trop compliqué. Peut-être faudrait-il développer un concordancier pour les jeunes ». La revendication de Heike pour des outils de corpus conviviaux doit être prise au sérieux (...); la nouvelle interface Sketch Engine (conçue en 2018) évolue dans ce sens, et SkELL (Sketch Engine for Language Learning) (...) est enfin disponible en allemand. De plus, ce système de gestion de corpus est désormais accessible sur des tablettes et des smartphones, préférés par les jeunes à d'autres machines.]

Mon premier article scientifique a été publié dans les actes de colloque de la Journée ATALA (Schaeffer-Lacroix, 2006). Il témoigne de mon appétence pour les apports potentiels de la machine pour soutenir l'apprentissage du DaF écrit. Les possibilités des ressources et outils de corpus y sont également évoquées. La pédagogie reste toutefois le point de repère principal.

Pendant ma carrière d'enseignante et de chercheuse, j'ai exploré de façon récurrente l'apport des outils de corpus, principalement – mais pas exclusivement – pour l'enseignement-apprentissage des langues, et j'ai fait mes premiers pas dans le domaine des outils de statistique textuelle.

Je suis sensible à la question de l'usage d'outils libres (c'est-à-dire, gratuits, non intrusifs et,

pour certains, ouverts à des développements futurs par la communauté des usagers). Toutefois, pour des raisons d'ergonomie et d'accès à des fonctionnalités avancées, d'autres outils seront également mentionnés dans ce qui suit. Il ne faut pas ignorer le fait que la durée de vie des outils peut être courte. Il convient alors de penser en termes de type d'outil et d'en présenter un exemplaire représentatif qui, à l'heure de la rédaction de ce rapport de synthèse, est connu comme suffisamment stable et qui a de bonnes chances d'être pérenne.

1.3 Didactique et traductologie

Une des expériences qui m'ont particulièrement marquée pendant les années passées comme professeure agrégée d'allemand dans le milieu de l'enseignement secondaire est l'usage du traducteur électronique fait par des élèves pour qui c'était apparemment la seule option envisageable pour produire de l'écrit en allemand langue étrangère. Il est aisé de critiquer cette pratique et même de se moquer des résultats obtenus, comme, par exemple, de l'expression « *der heilige Bohrer* » [la sainte perforeuse], traduction automatique du nom de la ville de « Saint leu la forêt ». Toutefois, les outils de traduction sont devenus très performants ces dernières années. Des professionnels de la traduction s'en emparent pour gagner du temps en se dégageant du travail portant sur des passages standardisés et déjà traduits auparavant par des experts humains, passages que la machine arrive en général à gérer de façon très satisfaisante. L'outil DeepL (DeepL GmbH, 2017) en est un exemple parlant. Si les professionnels utilisent des traducteurs dits « automatiques », pourquoi les apprenants d'une langue étrangère ne devraient-ils pas s'en servir ? Ne serait-il pas pertinent et même nécessaire de les encourager à les utiliser et de les former à la relecture critique des résultats obtenus et à la post-édition ?

Pendant de longues années – en fait, depuis la publication du Cadre européen commun de référence pour les langues (CECRL) (Conseil de l'Europe, 2001) jusqu'à l'édition de son volume complémentaire (Conseil de l'Europe, 2018), la relation entre la didactique et la traduction a été difficile, voire absente, dans l'enseignement secondaire en France ; on peut presque parler d'une interdiction de traduire en cours de langue étrangère pendant cette période. La situation n'était pas la même dans l'enseignement universitaire, où, paradoxalement, les épreuves de traduction des concours (CAPES et agrégation) ont été maintenues, toutefois non sans subir des variations par rapport à leur nature, leur importance et le traitement parfois inégal des deux sens de traduction. Le volet « médiation », qui apparaît dans le volume complémentaire du CECRL, ainsi que l'intérêt croissant porté au plurilinguisme (Elmiger,

Matthey, 2017 ; Castellotti, 2010 ; Weissmann, 2012) ces dernières années contribuent au rétablissement d'une relation plus étroite entre didactique des langues et traduction. Cette évolution est renforcée par la prise en compte croissante de la situation de personnes à besoins spécifiques – qu'elles soient en situation de migration ou porteuses d'un handicap –, qui mène à la production massive de sous-titres et d'audiodescriptions multilingues (v. section 5.3). Depuis 2017, je m'intéresse à l'audiodescription. J'en ai fait un objet d'enseignement-apprentissage en l'insérant dans une formation de création de scénarios d'enseignement de l'allemand soutenus par les technologies (master MEEF), ainsi que dans un cours d'allemand Lansad (Langues pour spécialistes d'autres disciplines) à l'Inspé de l'académie de Paris (Schaeffer-Lacroix, soumis).

De plus, le souci d'aller vers une société inclusive motive le développement d'outils soutenant la traduction instantanée textuelle ou de type *text to speech* [le texte vers la parole] et *speech to text* [la parole vers le texte]. D'autres outils développés pour des personnes ayant des besoins dits « spécifiques » sont présentés dans le chapitre 7 d'Amadiou et Tricot (2014). L'Inspé de l'académie de Paris a complété son offre de formation par un master « Mention pratiques et ingénierie de la formation – parcours métiers de l'adaptation et de la scolarisation », abrégé en ASH (Adaptation scolaire et scolarisation des élèves handicapés).

Transcrire et définir les codes temporels : Allemand Supprimer le brouillon Enregistrer les modifications

Actions Raccourcis clavier | Aide

Saisissez le sous-titre ici, puis appuyez sur Entrée +

0:00.0	in münchen wird das spiel zur bitteren
0:05.2	realität die wagner straße im stadteil
0:05.2	schwabing
0:08.2	auch hier müssen die mitarbeiter fehlt
0:08.2	weiter ziehen und ihr zuhause
0:16.2	zurücklassen
0:16.2	ja ich wohne jetzt glaube ich schon seit
0:20.3	bald fünf jahren hier
0:20.3	das glaub ich meine vierte fünfte
0:24.0	wohnung in münchen aber die erste die

München wird zur Schlossallee

Figure 8 – Transcription automatique avec YoutubeStudio.

En printemps 2019, lors d'une des séries de formation de formateurs organisée par des collègues impliqués dans le parcours ASH, j'ai proposé une formation de formateurs à l'usage de

YouTube Studio (ANON., 2019e), un outil de transcription automatique avec possibilité de post-édition (v. Figure 8), pouvant être détourné pour la création de sous-titres, répondant à des besoins de personnes malentendantes.

L'évolution actuelle du secteur de la traduction se caractérise par un outillage de plus en plus sophistiqué et par la création de liens avec le secteur porteur de l'intelligence artificielle. En été 2019, j'ai choisi de combiner les domaines « personnes à besoins spécifiques », « traduction », « intelligence artificielle » et « corpus » en initiant un projet, qui est en train d'être monté avec une équipe de spécialistes en traduction audiovisuelle et en audiodescription, travaillant à l'université de Hildesheim en Allemagne (v. section [5.3](#)).

2. La recherche dans les formations universitaires

Dans ce chapitre, l'attention sera portée sur la façon dont la recherche est prise en compte dans des formations initiales et des formations continues en France portant sur les thèmes « enseigner une langue étrangère », « didactique des langues » et « linguistique outillée ».

2.1 Enseignement et didactique des langues

2.1.1 Masters de type « Enseignement des langues »

Qu'en est-il en matière de recherches en didactique dans le domaine de la formation universitaire en France ? Dans le contexte du master MEEF (Métiers de l'enseignement, de l'éducation et de la formation), le terme de « recherche » est clivant. Selon certains, la recherche serait, de façon exclusive, du ressort des partenaires universitaires, tandis que la pratique serait du ressort des partenaires de l'Inspé. Dans quelle mesure ces derniers peuvent-ils ou même doivent-ils offrir des enseignements menant à des activités de recherche ? Leurs domaines d'intervention peuvent être l'observation du terrain et la production du mémoire de master 2. L'observation du terrain peut être abordée au sein de formations transversales et généralistes (tronc commun) et lors de formations de type tutorat ou accompagnement de stage. En ce qui concerne la préparation à la production du mémoire, il paraît utile de rappeler le cadrage officiel.

Dans le cadre du master “ MEEF ”, chaque étudiant réalise un mémoire de master qui doit avoir un contenu disciplinaire et de recherche en relation avec la finalité pédagogique, les pratiques et l'environnement professionnels. Pour les fonctionnaires stagiaires, le mémoire prend appui sur le stage en alternance comme sur d'autres enseignements au sein de la formation.

Selon l'article 18 de l'Arrêté du 27 août 2013 fixant le cadre national des formations dispensées au sein des masters « métiers de l'enseignement, de l'éducation et de la formation » (2013), dont un extrait est cité ci-dessus, le mémoire doit conjuguer les observations du terrain avec les apports de la recherche. Il n'est pas construit sur le mode de deux volets indépendants, l'un étant dédié à la recherche (souvent réduit aux aspects disciplinaires) et l'autre à la pratique. Il convient de le penser comme un objet hybride : les deux types de réflexion et d'observation doivent se côtoyer, se croiser et s'alimenter mutuellement.

Le plan d'un mémoire en didactique des langues avec enquête du terrain contiendra des éléments se référant à la littérature scientifique et aux documents de cadrage institutionnel

rassemblés sur le site Eduscol⁵ : CECR, référentiel des compétences professionnelles des métiers du professorat et de l'éducation, socle commun de connaissances, de compétences et de culture, programmes officiels. La figure 9 rappelle le plan du mémoire de Gabriel Rauschert, inscrit en master 2 enseignement à l'IUFM de Paris en 2013/2014. L'enquête qu'il a menée sur la pratique d'enseignement DNL (Discipline non linguistique) en cours d'histoire y est motivée par un obstacle observé sur le terrain. Elle s'alimente par des lectures scientifiques et institutionnelles, et elle mène à des propositions destinées à modifier certains éléments de cette pratique afin de l'optimiser.

Les limites de l'enseignement en immersion dans le cadre de la DNL

1. Introduction
2. Le principe d'immersion en débat : potentiel et limites de l'enseignement de contenus disciplinaires en langue étrangère
3. Enjeux et limites de la DNL dans l'institution
 - 3.1. La diversité des cadres de l'enseignement en DNL
 - 3.2 Enjeux et limites de la coordination pédagogique des cours de DNL
 - 3.3. Les Programmes et l'épreuve du Baccalauréat en DNL
4. Les limites de l'enseignement en DNL sur un terrain - Le lycée Faidherbe de Lille
 - 4.1. Présentation de l'enquête et du terrain
 - 4.2. « Bains linguistiques » et « renforcements » - les cours de DNL histoire-géographie au lycée Faidherbe
 - 4.3. Qu'est-ce que l'immersion au lycée Faidherbe ? De la diversité des situations.
 - 4.4. Quelques idées pour tenter d'améliorer l'enseignement de la DNL sur le terrain
5. Conclusion

Figure 9 – Plan d'un mémoire de master 2 Enseignement en didactique de l'allemand.

Les outils numériques ont un rôle particulier à jouer lors de la production du mémoire. Ils peuvent aider à recenser des opinions, à récolter des données de terrain (voir section 4.3) et à mettre en forme le mémoire, en particulier en ce qui concerne la gestion des références. Ces dernières années, les stagiaires d'allemand de l'Inspé de l'académie de Paris se sont vus offrir une formation à Zotero (Corporation for Digital Scholarship, [sans date]) dispensée par Ingrid Momein, une collègue documentaliste, et ils ont été encouragés à appliquer cette formation lors

⁵ <https://eduscol.education.fr/cid58018/textes-de-reference.html>

de la rédaction de leur mémoire de master 2 en gérant leurs références à l'aide de cet outil.

The image shows a Google Docs document on the left and a chat window on the right. The document contains a questionnaire titled "Questionnaire 1 : idées de questions pour le questionnaire des enfants en situation de handicap." with 10 numbered questions. The chat window shows three messages from Eva Lacroix, dated 06:34, 06:35, and 06:36. The messages discuss the questionnaire's content and suggest improvements.

Questionnaire 1 : idées de questions pour le questionnaire des enfants en situation de handicap.

1. Êtes-vous élève dans un établissement spécialisé ou dans une école publique ?
2. Quel type de handicap avez-vous ? (dyslexie, trouble de l'apprentissage, syndrome, ect).
3. Etes-vous dans une classe adaptée ou dans un classe mixte ?
4. Quel âge avez-vous et en quelle classe êtes-vous ?
5. Vous sentez-vous intégré-e dans la classe ? Merci d'expliquer et/ ou de donner des exemples.
6. Avez-vous eu des supports adaptés à votre type de handicap ? donner des exemples.
7. Avez-vous remarqué une différence d'évaluation ? expliquer et/ou donner des exemples
8. Avez-vous remarqué si le professeur adapte son contenu de cours et la manière de vous l'apprendre ? Expliquer et/ou donner des exemples
9. Bénéficiez-vous d'un AVS ? (auxiliaire de vie scolaire) Si oui, avez-vous ressenti des progrès ? Si non, en ressentez vous le besoin.
10. As-tu des difficultés en langue vivante ? Si oui, lesquelles et penses-tu qu'elles soient liées à ta situation de handicap ?

Chat messages:

- 06:34 Avant-hier: "Si oui : Quel...?" (ou bien, redirection à l'aide de l'outil de création de questionnaire).
- 06:35 Avant-hier: Quelles en sont les définitions ? Est-ce clair pour les enfants (parents), à votre avis ?
- 06:36 Avant-hier: Je remonterais cette question tout en haut du questionnaire (elle concerne tous les élèves). De plus, je proposerais ici un qcm (plus facile à traiter pour la machine que les réponses libres).

Figure 10 – Suivi du mémoire de Kim Gehere à l'aide de Google documents.

Certains des étudiants se servent de Google drive (Google, 2020) comme espace d'écriture en ligne. Comme en témoigne l'extrait du mémoire de Kim Gehere (inscrite en Master 2 à l'Inspé en 2019/2020) présenté en Figure 10, la version provisoire du mémoire peut ainsi être régulièrement commentée et annotée par la directrice de mémoire.

The image shows a Trello board with three columns: "Informationen und Anregungen", "Kim", and "Solène". The "Kim" column contains several task cards related to a thesis on "Quelle intégration en classe d'allemand des élèves en situation de handicap?". The "Solène" column contains a card about "Appports du jeu Classcraft".

Trello Board Content:

- Informationen und Anregungen:**
 - Mémoire de master MEEF et FSTG
 - Inspirieren lassen vs. plagieren
 - INITIATION A LA METHODOLOGIE DE RECHERCHE EN SHS - Réussir son mémoire ou sa thèse
- Kim:**
 - Vorläufiges Thema der Hausarbeit: Welche intégration en classe d'allemand des élèves en situation de handicap ?
 - TD: Schulräte und Schülerrätinnen kontaktieren, um Klassen für eine Umfrage zu gewinnen.
 - TD: Fragebogen fertigstellen
 - Begriffe über Behinderung von A bis Z
 - Albarellero : Apprendre à chercher (2015), De Boeck
- Solène:**
 - Vorläufiges Thema der Hausarbeit: "Appports du jeu Classcraft pour l'apprentissage institutionnel de l'allemand"

Figure 11 – Suivi de mémoires à l'aide de Trello.

La direction du mémoire peut également être soutenue à l'aide de l'outil de gestion de projets

Trello (Atlassian, 2010), qui permet de diffuser des conseils et d'échanger en ligne autour de questions qui concernent soit le groupe entier de stagiaires soit des personnes suivies individuellement.

La liste des mémoires de master que j'ai (co-)dirigés est disponible en Annexe 8.

2.1.2 Masters de type « Recherches en didactique des langues »

Le master MEEF ne permet pas de s'inscrire directement en thèse, mais d'autres masters, créés en France à partir des années 80, s'adressent à des personnes désirant faire de la recherche avancée en didactique. Ces formations répondent majoritairement aux besoins d'accueillir des spécialistes en FLE (Français langue étrangère). Elles sont en général intégrées dans des parcours de type lettres modernes ou sciences du langage. Le master mention « Didactique des langues », proposé par l'université Rennes 2, offre une spécialisation en humanités numériques : « La bi-diplomation avec la mention 'Humanités numériques' encourage des programmes de recherche innovants associant analyse de données langagières expertes et humanités numériques »⁶.

À côté des masters à dominante FLE, bien établis, il existe plusieurs formations plus récentes qui mettent l'accent sur la recherche en didactique d'une autre langue que le français et qui sont en concurrence avec les masters de recherche faisant suite à des formations de type LLCE (Langues, littératures et civilisations étrangères) ou LEA (Langues étrangères appliquées). J'ai repéré deux formations affichant la possibilité de poursuivre en thèse à l'issue du master 2.

L'Inspé de Bretagne propose un master, parcours « Recherches en didactique »⁷.

Le parcours « Recherches en Didactique » doit conduire les étudiants à mettre à l'épreuve un travail de recherche sur un terrain particulier. Les objectifs du parcours sont :

- Construire des outils conceptuels et méthodologiques qui permettent de comprendre et d'expliquer les pratiques d'enseignement, de formation et d'apprentissage, ainsi que les contextes dans lesquels ils s'inscrivent ;
- Former aux méthodes de recueil et d'analyse de données sur les pratiques d'enseignement et d'apprentissage ;
- Faire connaître et utiliser les apports des recherches en didactique dans toutes ses

⁶ <https://formations.univ-rennes2.fr/fr/formations/master-37/master-mention-didactique-des-langues-JFC6ZDIE.html>

⁷ <http://www.inspe-bretagne.fr/devenir-enseignant/parcours-recherches-en-didactique>

dimensions (disciplinaires, comparées...) sans négliger les apports de disciplines contributives (psychologie, sociologie...).

Le choix de cette perspective a pour objectif de conduire à l'acquisition d'un positionnement, de connaissances et de méthodes d'analyse pertinents aussi bien pour le futur chercheur, expert, consultant que pour un professionnel de l'éducation amené à s'inscrire dans une dynamique évolutive tout au long de sa vie professionnelle.

Descriptif des semestres S07-S08 (M1) et S09-S10 (M2)				
Intitulés des unités d'enseignement (UE)	Volumes horaires des enseignements			
	S07	S08	S09	S10
UE1 : Techniques de recueil et d'analyse de données	24h	24h	6h	6h
UE2 : Recherche et didactique(s)	36h	24h	24h	-
UE3 : Langues et langages	12h	12h	12h	12h
UE4 : Ateliers de recherche (Travail d'étude et de recherche en M1 - Mémoire de recherche en M2)	6h	12h	12h	24h
Total années : 150h M1 et 96h en M2	78h	72h	54h	42h

Figure 12 – Contenus du master « Recherches en didactique » de l'Inspé de Bretagne.

60 heures de cette formation d'une durée de deux ans sont dédiées aux techniques de recueil et d'analyse de données⁸ (v. Figure 12).

L'Inspé de Versailles évoque également la possibilité de poursuivre en thèse à l'issue d'un master appelé REDEF (Recherches en éducation, didactique et formation).

Ce cursus s'adresse essentiellement à des professionnels de l'éducation de l'espace francophone ; il propose une formation scientifique permettant de mener à bien une thèse en 3 ans à l'issue de ce master⁹.

Dans la maquette, l'observation de terrain n'est pas mentionnée. Par contre, l'offre comprend 36 heures en recherche quantitative et 24 heures en recherche qualitative (v. section 4.5).

L'université Paris 3 Sorbonne Nouvelle offre une formation qui combine les études germaniques et la linguistique outillée : les deux parcours actuellement accessibles au sein du

⁸ <http://www.inspe-bretagne.fr/sites/default/files/documents/plaq-red-2017.pdf>

⁹ <http://www.inspe-versailles.fr/Recherches-en-education-didactique-et-formation-REDEF>

master « Humanités numériques » s'appellent « Numérisation, gestion et exploration de sources textuelles (NGEST) » et « Études germaniques – humanités numériques (EGHN) »¹⁰. Les autres cursus universitaires combinant l'étude d'une langue étrangère et du TAL et/ou linguistique outillée semblent majoritairement privilégier l'anglais.

2.2 Outils numériques et linguistique outillée

À côté des formations universitaires diplômantes, il est possible de se former lors d'actions ponctuelles, dont certaines sont adossées à des événements scientifiques, p. ex. des ateliers pré- ou post-colloque. D'autres sont organisées par des unités de recherche d'établissements universitaires. Suivre une formation continue est une des possibilités de continuer à se spécialiser et à consolider ses connaissances et compétences en didactique et/ou en linguistique outillée. Cette action permet aussi de développer un réseau scientifique (voir section 3.7). Dans ce qui suit, je présenterai d'abord les formations continues auxquelles j'ai participé pour enchaîner avec celles que j'ai dispensées, toujours en mettant l'accent sur leur lien avec la recherche.

2.2.1 Formations continues suivies

Les formations que j'ai eu l'occasion de suivre depuis 2007 (voir la liste complète en [Annexe 1](#)) ont toutes visé mon perfectionnement dans le domaine de la linguistique outillée, en particulier dans le domaine des outils de corpus numériques (notamment Sketch Engine, Cosmas, Corpus workbench, CQPweb, DWDS, TXM). Pour certaines des formations, j'ai fait un dossier de candidature afin de profiter d'une bourse (formation CLARA Summer School, 2010). Pour d'autres, j'ai dû demander un droit d'accès à la formation (École thématique CNRS *MISAT*, 2016). Cette dernière a été subventionnée par le laboratoire CeLiSo (Centre de linguistique en Sorbonne). La formation « Initiation à l'édition électronique et la pratique de l'encodage XML/TEI », que j'ai suivie en 2015, a été financée par le dispositif DIF (Droit individuel à la formation).

Plusieurs des personnes que j'ai rencontrées à l'occasion de ces formations et certains des contenus présentés ont posé les jalons pour mes recherches futures. En 2012, j'ai invité Heike

¹⁰ <http://www.univ-paris3.fr/master-humanites-numeriques-538585.kjsp>

Zinsmeister (maintenant professeure à l'université de Hambourg) ainsi que Thomas Schmidt (maintenant à l'*Institut für deutsche Sprache* Mannheim) à une journée d'étude sur la réflexion sur la langue et les corpus, que j'ai co-organisée en 2012 ensemble avec Martine Dalmas, professeure à Sorbonne Université. J'ai recroisé Thomas Schmidt lors d'une des rencontres scientifiques organisées par Corli (Consortium corpus, langues et interactions), lors de laquelle il a fait un exposé. Par la suite, j'ai exploré les possibilités d'un outil dont Thomas Schmidt est un des développeurs, à savoir EXMARaLDA, un outil pour l'annotation de données multimodales (voir un exemple en paragraphe 4.3.1). Une rencontre très marquante, aussi bien du point de vue intellectuel qu'humain, était celle que j'ai faite en 2014 avec Bénédicte Pincemin et Serge Heiden, membres de l'équipe de l'ENS (École normale supérieure) de Lyon, deux des développeurs de l'outil d'analyse textométrique TXM (Heiden et al., 2010). En paragraphe 3.7.2, je présente de façon plus détaillée les échanges réguliers que nous avons poursuivis depuis. TXM est l'outil phare du projet le plus récent dans lequel je suis impliquée, intitulé *Les corpus numériques au service de la traduction d'audiodescriptions* (voir section 5.3). Là aussi, le support de l'équipe de l'ENS de Lyon a été précieux. De plus, Bénédicte Pincemin est pour moi un modèle en matière de formation aux outils de corpus. Elle a fortement inspiré mes propres méthodes de formation dans ce domaine (v. paragraphe 2.2.2). Ses tutoriels filmés concernant la prise en main de TXM sont accessibles en ligne (Pincemin, 2012).

En lien avec mon engagement éditorial pour la revue *Alsic*, j'avais besoin d'apprendre à gérer l'étape finale de publication en ligne des articles. Cela m'a amenée à suivre deux formations au logiciel d'édition électronique Lodel, organisées par Cléo (Centre pour l'édition électronique ouverte)¹¹. Ces dernières années, l'initiative OpenEdition¹² en faveur d'une science ouverte a pris de l'ampleur, ce qui a mené à la mise en disponibilité en libre accès d'articles, de revues, de livres ou de blogues scientifiques. Ces derniers, nommés « carnets de recherche », sont publiés sur la plateforme Hypotheses¹³ d'OpenEdition (v. 2.2.2 et 3.7.1).

¹¹ https://fr.wikipedia.org/wiki/Centre_pour_l%27%C3%A9dition_%C3%A9lectronique_ouverte

¹² <https://www.openedition.org/>

¹³ <https://hypotheses.org/>

2.2.2 Formations continues dispensées

Pendant ma carrière académique, j'ai régulièrement dispensé des formations continues, portant sur l'usage d'outils numériques pour l'enseignement des langues et/ou pour la recherche en linguistique ou, de façon plus générale, en sciences humaines.

Formation de formateurs

Au sein de mon lieu de travail (appelé IUFM, puis Espé, puis Inspé de Paris), j'ai proposé plusieurs formations de type « formation de formateurs ». Voici les titres de ces formations.

- Utiliser un espace numérique de travail pour former en langue vivante
- Utiliser des corpus numériques pour apprendre une langue étrangère
- Atelier Sketch Engine

Formation de doctorants

Ensemble avec Kim Oger, professeure agrégée de l'enseignement supérieur (PRAG) et docteure en linguistique anglaise à Sorbonne Université, j'ai dispensé en 2017/18 une formation à l'usage de l'outil TXM organisée par Sorbonne Université pour un public de doctorants. Ci-dessous le descriptif officiel de ce cours.

« Les outils de corpus au service de la thèse » par Mme Eva Lacroix (EA 7332) en quatre séances

Lieu : Maison de la recherche, 28 Rue Serpente, 75006 Paris ; horaire : 18-20h.

Les corpus numériques sont des collections de textes qui peuvent être abordés à l'aide d'outils de textométrie offrant des services de comptage et de visualisation graphique des mots. Il est possible d'y faire appel pour se rendre compte des caractéristiques générales d'un genre textuel ou d'un discours spécialisé, pour comparer un texte en gestation à des textes comparables rédigés par des experts du domaine et pour bien d'autres actions encore. Lors de la formation, les doctorant·e·s auront l'occasion de créer et explorer un corpus représentant le domaine principal qu'ils abordent dans leur recherche et d'y comparer la version actuelle de leur fichier de thèse, que leur travail soit en stade précoce ou déjà bien avancé.

Espace numérique de travail du cours : <https://initxm.hypotheses.org/>

Séance du 9 octobre 2017 : découverte des fonctionnalités de base de l'outil de gestion et d'exploration de corpus TXM (Heiden, Magué & Pincemin, 2010) ; recensement des projets et besoins des participants.

Séance du 6 novembre 2017 : rappel de quelques caractéristiques de base de TXM ; phase pratique : déposer un corpus jetable (pressepapier) et tester les fonctionnalités de TXM déjà rencontrées ; extraction d'un corpus d'experts (Boch, Frier, 2015, p. 91) à partir du *Corpus Chambers-Le Baron* (Chambers, Le Baron, 2006) et constitution d'un fichier de métadonnées ; découverte de fonctionnalités supplémentaires de TXM.

Séance du 20 novembre 2017 : intégration de la version actuelle de sa thèse dans le corpus d'experts et rajout d'un choix d'articles cités dans la thèse ; intégration des nouvelles entrées dans le fichier de métadonnées. Explorations comparatives, par exemple, l'étude de spécificités lexicales ou l'observation de trigrammes (Charles, 2016).

Séance du 5 février 2018 : questions techniques et éthiques autour de la récupération de données en ligne pour leur mise en corpus ; explorations libres ; échange autour de questions communes et bilan de la formation.

Pour accompagner le processus d'installation de TXM lors de ce cours, j'ai rédigé un tutoriel validé par l'équipe de TXM et publié sur HAL (Schaeffer-Lacroix, 2018c).

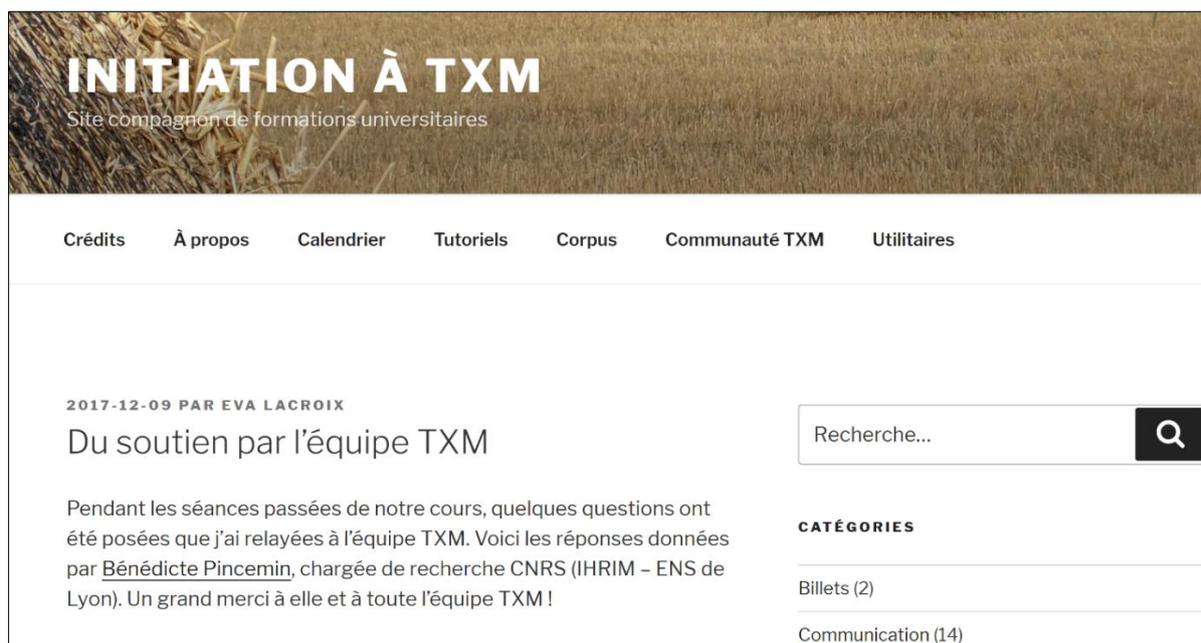


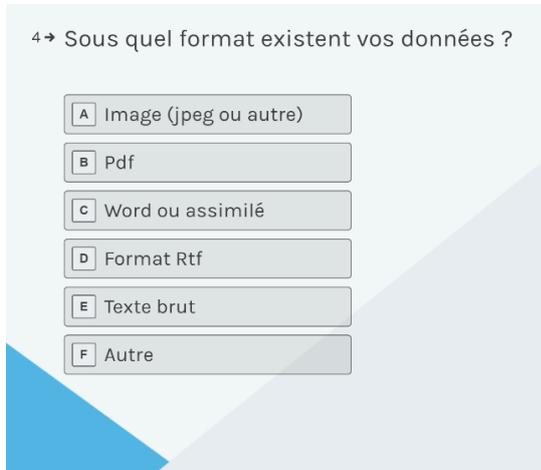
Figure 13 – *Initxm*, site compagnon d'une formation pour doctorants.

De plus, j'ai créé un carnet de recherche Hypothèses intitulé *Initxm* (Schaeffer-Lacroix, 2018a) (v. Figure 13). Ce carnet est le site compagnon de la formation dispensée. Il donne l'occasion aux participants ou à tout internaute intéressé de réagir en ligne aux articles qui y sont publiés en y apposant un commentaire ou en posant une question. On y trouve aussi des réflexions et fruits de discussions dépassant le cadre de cette formation. Cet espace numérique reflète notre volonté de mise en réseau du public formé avec les développeurs et développeuses de TXM qui ont répondu aux questions survenues pendant la formation.

Formation d'un public académique varié

En 2018, j'ai répondu à un appel à formation à l'outil TXM de l'EHESS (École des hautes études en sciences sociales) pour un public académique varié (chercheurs, doctorants, mastérants). Afin d'optimiser cette formation d'une durée de 6 heures, dispensée ensemble avec

Kim Oger, les participants préinscrits ont été invités à renseigner un questionnaire en ligne¹⁴ créé à l'aide de Typeform ([sans date]).



4 → Sous quel format existent vos données ?

- A Image (jpeg ou autre)
- B Pdf
- C Word ou assimilé
- D Format Rtf
- E Texte brut
- F Autre

Figure 14 – Questionnaire avant la formation à TXM.

Nous avons opté pour cet outil européen en raison de la réticence du public par rapport à l'outil américain de création de questionnaires Google forms ([sans date]), qui, jusqu'à présent, a eu la réputation d'être peu respectueux des données publiées en ligne (mais les évolutions récentes aident à remédier aux problèmes de confidentialité). Ce questionnaire a recensé les champs de recherche des participants, p. ex. anthropologie, économie, arts et langages, ainsi

que leurs questions de recherche. Les personnes ont aussi été priées de décrire la nature et la langue des données qu'elles souhaitaient explorer à l'aide de TXM. Une des questions s'est intéressée au format dans lequel ces données étaient disponibles (voir Figure 14) et une autre à leur structuration. L'analyse des réponses nous a permis de préparer les données et de les utiliser lors de nos démonstrations le jour de la formation.

¹⁴ <https://evalacroix.typeform.com/to/ie0kxR>

3. Un parcours de recherche outillé

Que veut dire « faire de la recherche en didactique des langues » ? En quoi consiste ce travail ? Les activités les plus connues sont la mise en place d'expériences, la présentation de conférences, la publication d'articles et l'expertise scientifique. Quatre occupations supplémentaires peuvent être mentionnées, pouvant toutes être optimisées à l'aide d'outils numériques : l'organisation d'événements scientifiques, des activités éditoriales, la pratique d'une veille scientifique et l'insertion dans un réseau scientifique. Dans ce qui suit, tous les volets que je viens de citer seront traités sauf celui concernant la mise en place d'expériences : les projets que j'ai conçus ou auxquels j'ai participé seront présentés en chapitre 5.

3.1 Communications

De nombreux colloques sont organisés en France pour soutenir la diffusion des travaux en didactique des langues¹⁵ : le colloque de l'ACEDLE (Association des chercheurs et enseignants didacticiens des langues étrangères), CEDIL (Colloque international des jeunes chercheurs en didactique des langues et en linguistique), EPAL (Échanger pour apprendre en ligne), le colloque du PERL (Pôle d'élaboration de ressources linguistiques) et le congrès RANACLÈS (Rassemblement national des centres de langues de l'enseignement supérieur). Comme on peut le constater, à part le colloque de l'ACEDLE, ces événements font explicitement le lien entre la didactique et un autre domaine, voire plusieurs, dont la linguistique, les sciences de l'information et de la communication et l'informatique. À l'occasion de l'édition de 2020 d'EPAL, les contours pluridisciplinaires de ce colloque sont affichés de façon explicite¹⁶.

Comme le montre la composition du comité scientifique, la perspective proposée est résolument pluridisciplinaire, conjuguant les sciences du langage et de la didactique avec les sciences de l'éducation, les sciences de l'information et de la communication, l'informatique (notamment dans ses composantes EIAH et IHM) afin de continuer à étudier les échanges en ligne dans les dispositifs pédagogiques. Cette édition se situera, toujours d'un point de vue pluridisciplinaire, à la croisée de différentes thématiques liées aux échanges en ligne : concepts d'interaction, littératie digitale, conception et usages des outils et dispositifs d'apprentissage médiatisés, stratégies en ingénierie de formation, ainsi que les démarches et outils d'analyse exploités par les différentes disciplines.

Le tableau 1 liste mes communications et mes engagements collectifs dans le domaine de la

¹⁵ Je ne cite ici que les colloques proches de mes domaines de recherche principaux.

¹⁶ <https://www.sciencesconf.org/browse/conference/?confid=8030>

didactique des langues en France.

Tableau 1 – Mes activités en lien avec des colloques de didactique en France.

Colloque	Communications	Engagements collectifs
ACEDLE	<p>« Pour un apprenant réfléchissant », présentation avec Elsa Chachkine et Françoise Demaizière. <i>Colloque Acedle 2012 - Recherches en didactique des langues - Apprendre les langues autrement</i>. Organisé par le Centre International des Langues, Université de Nantes, 6-9 juin 2012.</p> <p>« L'intertextualité comme procédé d'écriture en L2 ». Journée d'études organisée par l'ACEDLE (Association des Chercheurs et Enseignants Didacticiens des Langues Étrangères) : <i>Recherches émergentes en didactique des langues</i>. IUFM de Paris, 13 juin 2008.</p> <p>« Corpus et concordances au service de nouvelles pratiques de formulation en L2 ». Colloque ACEDLE (Association des Chercheurs et Enseignants Didacticiens des Langues Étrangères) 2008 - <i>Recherches en didactique des langues</i>. Strasbourg, 17-19 janvier 2008.</p>	<p>Co-organisation de la journée d'étude de l'ACEDLE (Association des Chercheurs et Enseignants Didacticiens des Langues) du 4 juin 2010 à IUFM de l'académie de Paris, en coopération avec Muriel Grosbois.</p> <p>Co-organisation de la journée d'étude de l'ACEDLE <i>Recherches émergentes en didactique des langues</i> du 12 juin 2009 à IUFM de l'académie de Paris, en coopération avec Muriel Grosbois.</p>
Alsic	<p>« Utiliser les corpus avec des apprenants Lansad ». Séminaire « Le numérique pour l'enseignement et l'apprentissage des langues pour non-spécialistes », Montpellier, 2013.</p>	<p>Co-organisation du séminaire <i>Alsic Le numérique pour l'enseignement et l'apprentissage des langues pour non-spécialistes</i> à l'université Paul Valéry Montpellier 3 sous la responsabilité de Cécile Poussard, avec le soutien de Claude Bertrand (Mines) et des équipes de l'université Paul Valéry, 12 et 13 décembre 2013.</p>
CEDIL	<p>« Création de corpus réfléchis pour l'enseignement de la production écrite. L'exemple de l'allemand, langue étrangère ». <i>CEDIL 2010</i> (Colloque international des jeunes chercheurs en didactique des langues et en linguistique). Laboratoire LIDILEM, Université Stendhal, Grenoble, 29 juin - 2 juillet 2010.</p>	<p>Expertises (2014,2018)</p>
PERL	<p>Avec Isabelle Salengros Iguenane. « Structurer un module de FOS (Français sur objectifs spécifiques) dédié à un écrit professionnel à l'aide d'analyses du corpus CCTP (Cahier des clauses techniques particulières) ». 2ème colloque du PERL : Entre présence et distance. Enseigner et apprendre les langues à l'université à l'ère numérique. Organisé par le PERL, Université de Paris & USPC. Paris, 12-13-14 décembre 2019.</p>	<p>Expertises (2018, 2019)</p>

Ranacles	<p>« Les futurs enseignants de langues face aux corpus numériques ». XXIV^e congrès du Rassemblement National des Centres de Langues de l'Enseignement Supérieur (RANACLES), 24-26 novembre 2016, <u>Espé de l'académie de Paris</u> – École interne de l'Université Paris-Sorbonne.</p> <p>« Faire pratiquer les corpus au collège lors d'un projet d'écriture en allemand ». 16^{ème} congrès de RANACLÈS (Rassemblement national des centres de langues de L'enseignement supérieur). Université de Nice Sophia-Antipolis. <i>Didactique et pratique des langues, quelles priorités ?</i>. Nice, 20-22 novembre 2008.</p>	<p>Membre du comité d'organisation du XXIV^e congrès du Rassemblement National des Centres de Langues de l'Enseignement Supérieur (RANACLES), 24-26 novembre 2016, <u>Espé de l'académie de Paris</u> – École interne de l'Université Paris-Sorbonne.</p>
----------	--	---

Le contenu du tableau 1 témoigne d'une activité équilibrée en termes de communication et d'expertise en France. On notera ma collaboration avec des spécialistes de l'anglais (Françoise Demaizière, Muriel Grosbois) et du français langue étrangère (Elsa Chachkine, Isabelle Salengros Iguenane). Mon positionnement ne se restreint donc pas à la didactique de l'allemand. Comme l'illustre la figure 15, j'interviens également à l'étranger.

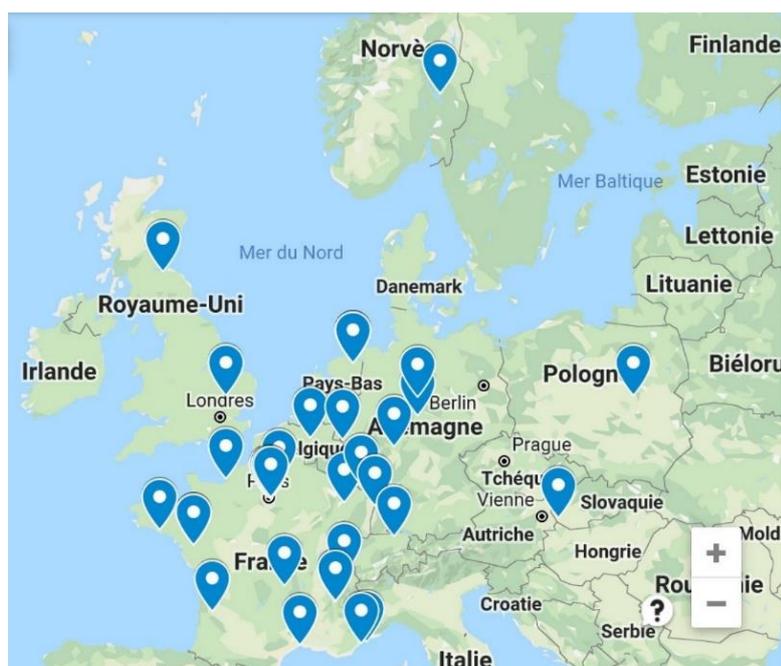


Figure 15 – Lieux de communication.

Dans le passé, mes langues de présentation ont été l'allemand (quatre communications), l'anglais (neuf communications) et le français (23 communications). La liste complète de mes communications peut être consultée en [Annexe 4](#).

3.2 Publications

Il existe des revues dédiées à la didactique des langues, dont les suivantes, classées ici par ordre alphabétique : *Alsic* (Apprentissage des langues et systèmes d'information et de communication), *ÉLA* (Études de linguistique appliquée – Revue de didactologie et de lexiculurologie des langues-cultures), *LIA* (Language, Interaction, and Acquisition) faisant suite à la revue *AILE* (Acquisition et interaction en langue étrangère), *Lidil* (Revue de linguistique et de didactique des langues), *Mélanges Crapel* (Revue en didactique des langues et sociolinguistique), *LINX* (Revue des linguistes de l'Université Paris Ouest Nanterre la Défense), *RDLC* (Recherches en didactique des langues et des cultures) et *RPPLSP* (Recherche et pratiques pédagogiques en langues de spécialité – Cahiers de l'Apliut). Ces revues annoncent toutes un deuxième domaine de référence ou de spécialisation, p. ex. la culture, les technologies éducatives, la linguistique ou la langue de spécialité.

Dans son descriptif, *Lidil* met en avant son ancrage fort en linguistique, mais d'autres domaines sont également évoqués : « La revue *Lidil* publie des travaux en linguistique, sociolinguistique, didactique des langues, traitement automatique des langues (TAL) et en apprentissage et acquisition du langage »¹⁷. La priorité donnée à la linguistique est encore plus marquée en ce qui concerne la revue *LINX*. Son descriptif inclut la didactique des langues et il affiche un positionnement interdisciplinaire.

Elle [la revue *LINX*] aborde les thèmes les plus variés des Sciences du langage : syntaxe, sémantique, pragmatique, analyse du discours, lexicologie, histoire et épistémologie de la linguistique, langage et cognition, linguistique de l'oral, linguistique de l'écrit, acquisition des langues, didactique des langues, traitement automatique des langues et linguistique informatique, etc. La revue *LINX* se caractérise par son ouverture et sa volonté de dialogue entre tous les courants et toutes les écoles¹⁸.

Dans son titre de revue, *RPPLSP* (Recherche et pratiques pédagogiques en langues de spécialité – Cahiers de l'Apliut) ne mentionne pas le terme de didactique, mais celui de pédagogie, ce qui accentue la prise en compte du terrain et des relations entre acteurs impliqués dans l'enseignement-apprentissage. La revue *LIA* occupe une position particulière dans le paysage des revues citées : bilingue anglais-français, elle est publiée par le prestigieux éditeur John Benjamins, ce qui accroît considérablement la portée de ses publications. D'autres solutions

¹⁷ <https://journals.openedition.org/lidil/170>

¹⁸ <https://journals.openedition.org/linx/>

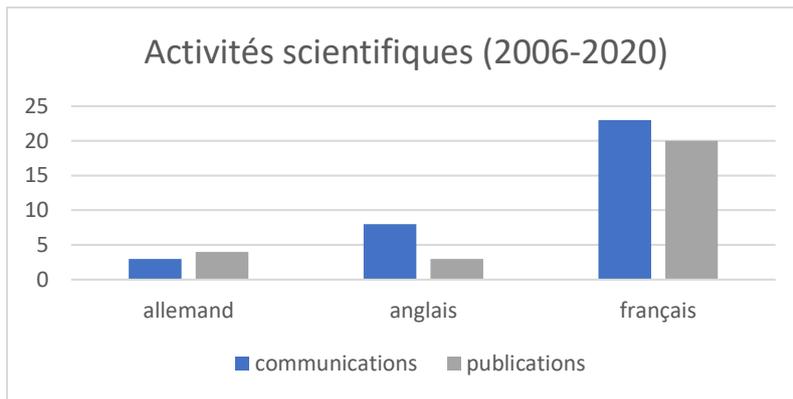
d'ouverture à l'international existent, par exemple, le choix de proposer un titre, un résumé et des mots-clés en une autre langue que celle de la publication, ce qui est courant pour les revues hébergées par OpenEdition Journals. Certaines revues, comme *Lidil* et *RPPLSP*, accueillent, en plus du français, des articles écrits en anglais. *Alsic* va plus loin encore en signalant sur son site que la revue accepte des articles « en français et en anglais et dans d'autres langues, notamment en allemand »¹⁹.

Tableau 2 – Mes activités en lien avec des revues de didactique en France.

Revue	Publications	Engagements collectifs
Alsic	Schaeffer-Lacroix, 2010 ; 2014 ; 2015 ; 2016a	Membre de l'équipe éditoriale et membre du comité scientifique
Lidil	Schaeffer-Lacroix, 2018d ; 2017	
LINX	Schaeffer-Lacroix, 2015b	
RDLC	Schaeffer-Lacroix, 2010a	
RPPLSP		Membre du comité scientifique

J'ai publié dans quatre des huit revues mentionnées au début de cette section et je m'engage activement pour deux d'entre elles. Mon activité de publication relativement restreinte en didactique des langues en France s'explique par mon positionnement élargi : je suis didacticienne, certes, mais en même temps, je suis fortement intéressée par la linguistique générale et/ou outillée. Certaines de mes publications s'insèrent dans ce deuxième domaine (Schaeffer-Lacroix et al., 2020 ; Alghisi et al., 2017 ; Elmiger et al., 2017b ; 2017a ; Schaeffer-Lacroix, 2016b). De plus, je cherche activement à agir et à m'engager au-delà des frontières françaises, ce qui se reflète dans mes publications dans les revues publiées à l'étranger, comme *Language Awareness Journal* (Schaeffer-Lacroix, 2016c), *Linguistik Online* (Chachkine et al., 2013) et le *Bulletin de Vals-Asla* (Alghisi et al., 2017) ainsi que dans mon chapitre dans un volume publié par Routledge (Schaeffer-Lacroix, 2019a). Les langues représentées dans mes publications majeures sont l'allemand (quatre publications), l'anglais (trois publications et une quatrième en cours d'évaluation) et le français (20 publications). Les chiffres pour mes communications sont comparables.

¹⁹ <https://journals.openedition.org/alsic/960>



Graphique 1 – Mes activités scientifiques en didactique et linguistique outillée.

3.3 Expertise scientifique

L'expertise scientifique est une activité qui permet aux chercheurs et chercheuses de s'engager pour la communauté et de se tenir au courant de l'actualité et des évolutions dans les secteurs qui relèvent de leurs domaines de compétence. Il s'agit de rendre compte, avec précision et bienveillance, du potentiel et de la pertinence d'un résumé pour une conférence ou une publication. Quand il s'agit de propositions faites par de jeunes chercheurs, p. ex. des articles soumis à la revue *ÉLIS* (Échanges Linguistiques en Sorbonne), il est particulièrement important de peser ses mots, d'encourager les jeunes collègues et de leur fournir des pistes d'amélioration. Ma pratique d'expertise est soutenue par Antidote (Druide informatique inc., 2019), un correcteur dont je me sers pour identifier les erreurs de langue anglaise ou française dans un texte avant de l'aborder du point de vue de son contenu. Basé sur l'exploitation de banques de données lexicales, Antidote peut être intégré comme macro dans un traitement de texte.

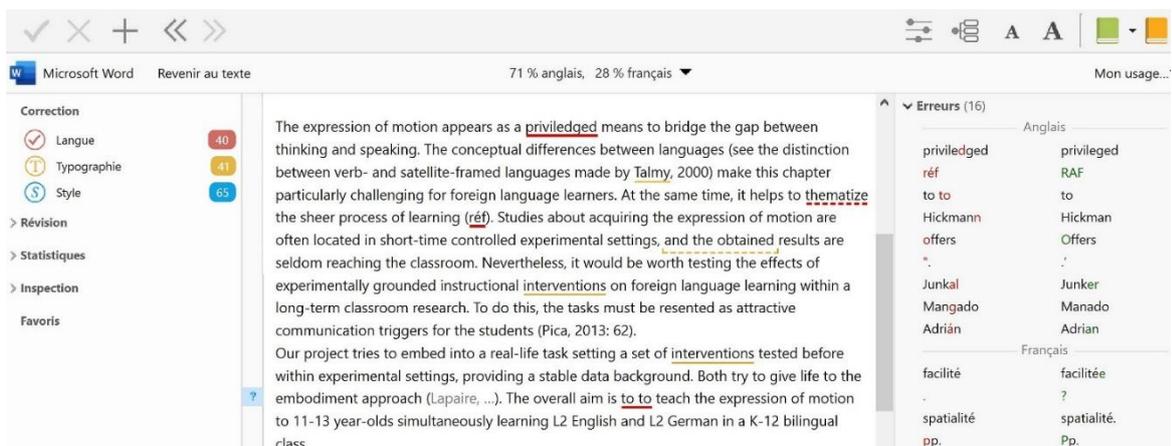


Figure 16 – Application du correcteur Antidote à une proposition de communication.

La figure 16 présente l'interface de cet outil, appliqué au brouillon d'une proposition de

communication rédigée en 2015 ensemble avec l'acquisitionniste Tatiana Iakovleva (maintenant maître de conférences en linguistique à l'université Paris Diderot-Paris 7).

Je suis membre permanent du comité scientifique de cinq revues, dont trois publiées en France, une en Italie et une en Suisse. De plus, j'ai expertisé de façon ponctuelle des propositions de communication pour dix colloques ou journées d'étude, organisés en Chypre, France, Italie ou en Slovaquie (voir la liste complète de mes activités d'expertise en [Annexe 2](#)).

En 2013, j'ai expertisé un projet ANR (Agence nationale des recherches) visant le développement d'un logiciel d'aide à la lecture en langue étrangère.

3.4 Animation et co-organisation d'événements scientifiques

Animer, organiser ou co-organiser un événement scientifique est une activité à caractère non seulement administratif, mais aussi social : elle permet de cultiver son réseau scientifique (v. aussi section [3.7](#)). La liste fournie en [Annexe 3](#) présente un choix de mes activités de ce type au niveau national ou international.

En étroite collaboration avec Cornelia Steinmann, directrice du *Selbstlernzentrum* [Centre d'auto-apprentissage] adossé au *Sprachenzentrum der UZH und der ETH Zürich*, j'ai dirigé une des sections proposées à l'IDT 2013 (Internationaler Deutschlehrrtag). Les outils numériques y ont joué un rôle important, en particulier pour coordonner à distance nos activités d'expertise, gérées à l'aide de Google documents. Après le colloque, Cornelia Steinmann et moi-même avons ouvert un blogue, de type Wordpress, sur lequel les personnes ayant communiqué ont eu l'occasion de déposer elles-mêmes leurs présentations.

En ce qui concerne le milieu des germanistes supérieur en France, il m'a été offert la possibilité de modérer, le 7 juin 2018, une table ronde lors du 51^e Congrès de l'AGES (Association des Germanistes de l'Enseignement Supérieur) « Voyages, migrations, mobilités » à l'Institut Goethe à Paris.

Je suis impliquée dans l'organisation de l'édition 2021 d'[Eurocall](#), une des manifestations internationales majeures pour la didactique outillée des langues, qui aura lieu au CNAM (Conservatoire national des arts et métiers) à Paris.

3.5 Édition numérique

Parmi les revues françaises spécialisées en didactique des langues étrangères, *Alsic*

(Apprentissage des langues et systèmes d'information et de communication) se distingue par les trois caractéristiques suivantes : créée en 1998, elle fait partie des revues les plus anciennes du domaine. Elle fait un lien explicite entre didactique et technologies éducatives, et elle est, depuis son existence, disponible sur Internet, en accès libre. Voici le descriptif qui en est fait sur la page d'accueil²⁰ :

Son objectif est d'offrir aux chercheurs, confirmés ou débutants, ainsi qu'aux praticiens un lieu d'expression, d'information et d'échange accessible à tous. Depuis sa création en 1998, *Alsic* est, en effet, publiée uniquement en ligne, s'inscrivant dès le départ dans le mouvement pour l'accès libre aux publications scientifiques.

Depuis 2010, je m'engage activement pour *Alsic*. Depuis 2013, j'occupe la fonction de trésorière de l'association *Adalsic* qui gère cette revue. De concert avec Elke Nissen, professeure des universités à l'université Grenoble-Alpes, je m'occupe de la rubrique « Pratique et recherche ». De plus, je fais partie du comité scientifique d'*Alsic*. Je relis régulièrement des articles prêts à être publiés.

The screenshot shows the Alsic journal website interface. On the left is a navigation sidebar with links for various volumes (Vol. 3, 2, 1) and sections like 'Nouvelles', 'Revue Alsic', and 'Informations'. The main content area is titled '2. Enregistrement de la conférence' and features a video player with a play button. Below the video is a 'Table des illustrations' section containing a small thumbnail image of a person, a URL, and file information.

▲ Table des illustrations	
	URL http://journals.openedition.org/alsic/docannexe/image/3179/img-1.jpg
	Fichier image/jpeg, 14k

Figure 17 – Insertion d'un document multimédia dans un article publié en ligne.

Ensemble avec Christian Ollivier, professeur des universités à l'université de la Réunion, je suis responsable de la qualité des images insérées dans les articles. Comme l'illustre la figure

²⁰ <https://journals.openedition.org/alsic/960>

17, je me suis également spécialisée dans le domaine de la publication de documents multimédia au sein des articles. Cette implication forte, récompensée par des échanges humains et scientifiques durables et très satisfaisants, m'a donné l'occasion de regarder derrière les coulisses du travail de rédaction menant à la publication de numéros de revues scientifiques.

3.6 Veille scientifique

La veille scientifique numérique est un des moyens pour se tenir au courant de l'actualité en didactique des langues, linguistique, TAL et traduction. Les moteurs de recherche spécialisés, comme 1findr (1findr, 2019), Isidore²¹ ou Google scholar²², peuvent soutenir cette activité.

Apprentissage phonologique pluriel dans une perspective didactique plurilingue

par Yvon Rolland

Recherches en *Didactique des Langues et Cultures - les Cahiers de l'Acedle* · 2013 · [Voie verte](#)

Le contexte est celui d'une classe de 6e bilangue L3 anglais/ L4 allemand à la Réunion, où le créole est la L1 et le français L2. ... Après une analyse des paramètres didactiques d'un apprentissage phonologique dans un premier temps, puis des freins à la construction de la compétence plurilingue dans un deuxième temps, une expérience est lancée, afin de proposer

 [Texte intégral \(PDF\)](#)

 Export de données

La médiation linguistique à l'Université : proposition pour un changement d'approche de la traduction dans les filières LLCE

par Dirk Weissmann

Etudes de Linguistique Appliquée · 2012 · [Voie verte](#)

Prenant appui sur les descripteurs développés pour l'enseignement de la langue allemande (Profile deutsch), nous illustrerons les possibilités d'application du CECR dans le cadre universitaire. ... Au final, nos propositions vont dans le sens d'une généralisation des entraînements à la médiation linguistique sous forme de leur intégration dans d'autres types de

 [Texte intégral \(PDF\)](#)

 Export de données

Langues allemandes de spécialité: implications pédagogiques de la recherche au niveau du lexique et de la syntaxe

par Jean-Marc Delagneau

Les Cahiers de l'Aplut · 2003 · [Voie dorée](#) · DOI

Une didactique ciblée des langues de spécialité doit s'appuyer sur les avancées de la recherche en linguistique appliquée pour intégrer les différents aspects, aussi bien lexicaux que syntaxiques.

Figure 18 – Moteur de recherche 1findr.

Le moteur de recherche 1findr (1findr, 2019) se présente comme « une plateforme de recherche inclusive visant à inclure les articles de toutes les revues savantes examinées par les pairs, dans tous les domaines, dans toutes les langues et dans l'ensemble des pays »²³. Il permet de recenser

²¹ <https://isidore.science/>

²² <https://scholar.google.com/>

²³ <https://www.1science.com/fr/1findr/>

des publications scientifiques librement accessibles en ligne, classables par domaines et/ou années de publication. La figure 18 fournit l'exemple de requête à l'aide des mots-clés « didactique, allemand, linguistique », avec restriction aux publications en accès libre. Le résultat affiche des publications appartenant à trois secteurs variés, à savoir la phonologie au collège, la traduction en filière LLCE et le lexique et la syntaxe en allemand de spécialité.

La veille scientifique profite particulièrement de l'usage d'outils de curation qui alertent de façon automatique sur l'actualité dans le monde de la recherche, qui est diffusée via *Linguist list* ou par d'autres canaux numériques. Certains de ces outils fonctionnent à l'aide de flux RSS (*Really Simple Syndication* [syndication vraiment simple]), aussi appelés « fils RSS ». L'outil de veille Feedly (ANON., [sans date]) récupère les fils RSS à l'aide de l'adresse Internet des blogs ou sites auxquels on s'intéresse. Les éléments retenus peuvent ensuite être classés dans des catégories définies par les usagers.

Figure 19 – Outil de curation Feedly.

Dans mon espace Feedly (v. Figure 19), j'ai choisi de créer, entre autres, une catégorie appelée « Didactique et linguistique ». Des publications relevant de la recherche dans un de ces deux domaines, p. ex. le blogue *Hochschuldidaktik* [didactique de l'enseignement supérieur] (Reinmann, 2019), y côtoient des informations d'ordre pédagogique ou institutionnel orientées terrain, comme *Éduveille* (Collectif, 2019) ou *Bob Blume / Unterricht digital* (Blume, 2019). D'autres s'intéressent plus particulièrement aux technologies éducatives, p. ex. *Les Outils Tice* de Navamuel (2019) ou le site suisse *Web2 Unterricht* (Lechner et al., 2019).

Des outils de curation comme Feedly informent les chercheurs et chercheuses de façon quasiment automatique ; l'information vient vers eux. D'autres outils permettent de garder une trace de l'information que l'on côtoie au gré de ses recherches sur Internet, en la classant dans des catégories que l'on aura établies auparavant et en lui ajoutant des étiquettes. Diigo (ANON., [sans date]) est un tel outil. Il accueille des liens vers des pages Internet. Ces dernières peuvent être brutes ou surlignées et/ou annotées par les usagers. La figure 20 illustre la façon dont cet espace représente des documents que j'ai identifiés et dont j'ai surligné et étiqueté certains. La colonne des « *Tags* » renvoie aux étiquettes que j'ai choisies afin de retrouver l'information plus tard dans un autre contexte. Il peut, par exemple, m'intéresser de regrouper, pour les besoins d'une recherche, tous les documents ayant l'étiquette « corpus ». Les passages surlignés figurent également dans cet espace, comme on peut le voir dans la première des références grâce au nombre indiqué à la fin du titre et grâce aux traits verticaux.

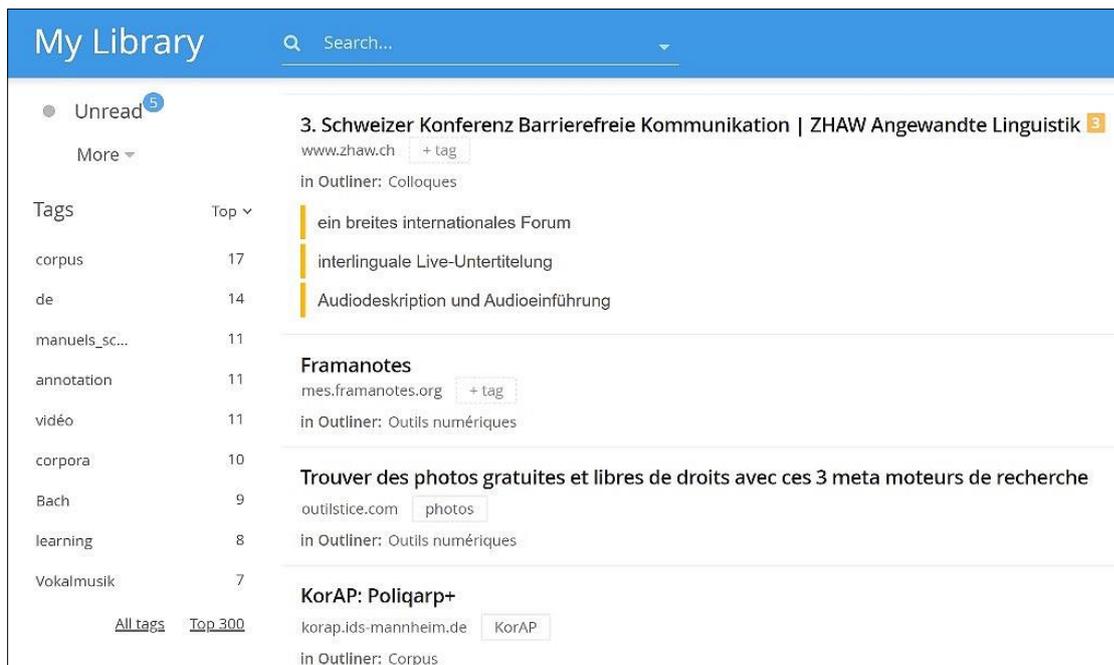


Figure 20 – Outil de curation Diigo.

La section 3.7 illustre le fait que la veille scientifique peut également prendre la forme d'une insertion dans des réseaux sociaux scientifiques.

3.7 Insertion dans des réseaux scientifiques

L'insertion dans des réseaux scientifiques peut bénéficier de rencontres faites pendant des colloques, des journées d'étude ou des formations. S'y ajoutent des échanges en ligne organisés

au niveau local (p. ex. via la mailing list *parislinguists*), national (p. ex. via la mailing liste *DH* (*Digital Humanities*) ou au niveau international (p. ex. via la *Linguist list*, les groupes Facebook *CorpusCaLL* et *TXM* ou grâce à l'abonnement à l'hashtag « *corpuslinguistics* » sur Twitter).

3.7.1 Échange et partage

Mon insertion dans des réseaux numériques scientifiques a été favorisée par mes activités de blogging d'abord pédagogique, puis scientifique. Depuis 2005, je tiens divers blogues, le tout premier diffusant les activités développées pour l'enseignement de l'allemand dans le secondaire, destiné à donner aux apprenants l'occasion de participer à la publication en ligne en déposant des commentaires aux articles publiés par moi (v. son interface en Figure 23, section 4.1). Le blogue *Aufgabenbörse*, que j'ai ouvert en 2009, pendant la période où j'ai

encore travaillé dans le secondaire, a eu pour but de partager avec des collègues d'autres établissements des activités pour enseigner l'allemand. J'ai abandonné cette entreprise en 2015 en raison du manque de publications régulières par d'autres personnes que moi-même.

En 2010, j'ai ouvert un blogue professionnel académique qui expose mes activités d'enseignement universitaire et de recherche scientifique (Schaeffer-Lacroix, 2019c) (v. Figure 21).



Figure 21 - Blogue professionnel académique.

Depuis 2016, j'alimente, ensemble avec d'autres chercheurs et chercheuses, le carnet de recherche du Club Corpus du laboratoire CeLiSo (ANON., 2019a), un groupe de personnes intéressées par les corpus numériques, qui se rencontre régulièrement pour se former mutuellement (v. le paragraphe 3.7.2 pour plus de détails). Ce carnet fait partie du catalogue des carnets publiés dans la section Hypothèses d'OpenEdition (<https://hypotheses.org/>). Il permet de faire le lien entre les séances lors desquelles se

rencontrent les membres du Club Corpus. De plus, nous y publions des informations par rapport à des événements pouvant intéresser les membres du groupe, et nous y stockons des données servant de matériel de démonstration lors des rencontres.

Malgré le fait que je sois souvent une des rares personnes à publier dans les blogues ou carnets dits « collectifs » que je viens de citer, l'idée de mutualisation, échange et partage a pris racine en moi pour ne plus me quitter tout au long de mon chemin d'enseignante et de chercheuse. À la rentrée universitaire 2019, l'adoption de l'outil de gestion de projets Trello (Atlassian, 2010) à des fins d'enseignement par deux de mes collègues du département d'allemand de l'Inspé de Paris a comblé mes attentes par rapport à la mutualisation et l'échange en ligne (voir les paragraphes [2.1.1](#) et [5.1.6](#)).

3.7.2 Entraide

Le premier réseau d'entraide scientifique que j'ai côtoyé de façon intense et durable est le cercle RIDL (Recherches informelles en didactique des langues), composé d'un groupe de chercheuses qui se rassemblent presque tous les mois depuis 2005 chez Françoise Demaizière, leur ancienne directrice de DEA (Diplôme d'études approfondies) et de thèse. Lors des rencontres, nous nous présentons mutuellement nos projets, nos brouillons d'exposés et d'articles, et nous nous entraînons pour passer des auditions. Des collaborations scientifiques se sont mises en place et ont donné lieu à des communications, publications et expériences communes (Chachkine et al., 2013 ; Salengros Iguenane, Schaeffer-Lacroix, 2020) (v. aussi paragraphe [5.1.5](#)). L'ouvrage de Grosbois (2012) présente des recherches de plusieurs membres de ce groupe, dont les miennes (p. 102-108).

Le deuxième réseau scientifique dans lequel je me suis insérée de façon durable est la communauté autour de l'outil d'analyse textométrique TXM. Je suis inscrite sur la liste de diffusion courriel de TXM, qui donne aux usagers l'occasion de s'entraider et de profiter de l'expertise des développeurs et développeuses de l'outil. Dans un article co-écrit avec Thierry Grass et Gilbert Magnus (Schaeffer-Lacroix et al., à paraître 2020), je mentionne les différents volets d'aide de cette communauté active et agréable.

L'aide fournie aux usagers est très bien pensée : ils peuvent compter sur le soutien d'une communauté réactive composée de développeurs et d'usagers de *TXM*, que ce soit par le biais d'une liste de courriel, d'un wiki ou d'un groupe *Facebook* dédié à la prise en main de cet outil.

Il ne s'agit pas seulement de profiter de l'aide fournie par les experts, mais aussi de faire profiter

la communauté de ses propres expériences d’usager en rapportant la façon dont on a résolu des problèmes. Dans Schaeffer-Lacroix (2015a), je cite le contenu de trois courriels que j’ai échangés à ce sujet avec Serge Heiden et Bénédicte Pincemin, membres de l’équipe qui développe TXM. De plus, je participe régulièrement aux échanges dans des groupes Facebook spécialisés dans les domaines qui m’intéressent, comme la linguistique de corpus ou la traduction automatique. La figure 22 retrace un échange que j’ai mené au sein du groupe Facebook TXM avec une chercheuse qui avait une question concernant la formulation d’une requête CQL (*corpus query language*) appliquée à des données de corpus gérées par l’outil TXM.

Bonjour! Est-ce que quelqu'un peut me dire s'il y a une requête qui permet de repérer les mots en majuscule?

3 commentaires Vu par 11 personnes

 J'aime

 Commenter

 Partager



Eva Lacroix Bonjour [redacted], pour trouver tous les mots qui commencent par une majuscule dans un corpus en langue française, la requête est la suivante : `[frlemma="[A-Z].*"]` . Pour info : de telles questions peuvent être résolues à l'aide de documents comme celui-ci : <http://cwb.sourceforge.net/temp/CQPTutorial.pdf>

J'aime · Répondre · 2 ans · Modifié



[redacted] Merci Ève! Je vais l'essayer!

J'aime · Répondre · 2 ans



Eva Lacroix De rien, [redacted] (Eva 😊)

J'aime · Répondre · 2 ans

Figure 22 – Entraide dans le groupe Facebook TXM.

Le troisième réseau qui m’a portée et réconfortée pendant mes années d’insertion dans le monde de la recherche est le Club Corpus du laboratoire CeLiSo de Sorbonne Université. Ce groupe a été créé à l’initiative de doctorants et doctorantes en vue de mutualiser les compétences disponibles. Des rencontres informelles sont ouvertes à toute personne intéressée par la linguistique assistée par les outils de corpus. Les formations proposées pendant les premières années ont porté sur l’usage de divers outils de corpus (AntConc, TXM, Sketch Engine, #LancsBox) et sur des méthodes d’analyse statistique. Elles ont été dispensées par les membres du Club Corpus, en particulier par Daniel Henkel (maintenant maître de conférences spécialisé en traduction assistée par ordinateur à l’université Paris 8 Vincennes Saint Denis), Pierre Labrosse (maître de conférence en linguistique anglaise de Sorbonne Université, maintenant à

la retraite), Kim Oger (PRAG et docteur en linguistique anglaise de Sorbonne Université), Naomi Truan (maintenant « Wissenschaftliche Mitarbeiterin » [Assistante de recherche] à l'Institut für Germanistik de l'université de Leipzig en Allemagne) et moi-même. Les membres de ce groupe ont eu l'occasion de présenter leurs travaux de recherche et de profiter d'échanges constructifs destinés à les aider à trouver des réponses à leurs questions techniques ou scientifiques. Il a également été possible d'inviter des spécialistes, comme Sylviane Granger, qui nous a fait connaître le domaine des corpus d'apprenants, Alexei Lavrentev de l'ENS de Lyon pour une formation approfondie TXM et XML et de la spécialiste en phraséologie Elena Krotova de l'université de Moscou, qui a présenté les possibilités d'interrogation des corpus écrits en langue allemande accessibles via le portail Cosmas II (Das Portal für die Korpusrecherche in den Textkorpora des Instituts für Deutsche Sprache, [sans date]). Comme déjà évoqué en paragraphe [3.7.1](#), un carnet Hypothèses a été ouvert afin d'offrir un lieu d'information, de stockage de données à partager et d'échange entre les séances.

4. La mise en œuvre de la recherche outillée

Le chapitre 4 est dédié aux méthodes de recherche outillées et aux défis à relever lors de recherches impliquant le terrain. Il y est question de récolte, de transformation et d'analyse de données. La recherche en didactique des langues se nourrit, pour la plupart des questionnements, du va-et-vient entre empirie et théorie (Euler, 2017). Elle part d'obstacles à l'apprentissage identifiés dans les fruits de la recherche (articles, ouvrages, rapports de projets) et/ou dans les données du terrain. Le procédé de récolte de données gagne à viser une triangulation (Berger et al., 2010 ; Demaizière, Nancy-Combes, 2007) : les opinions récoltées ou, de façon plus générale, les discours, sont confrontés aux actions observables et aux traces des productions (écrites, orales ou matérielles) réalisées par les participants. Les traces peuvent être rassemblées dans des corpus d'apprenants (voir section 4.6). Ces trois entrées font, à des degrés divers, appel à l'emploi d'outils numériques.

La veille scientifique ayant déjà été abordée en section 3.6., je vais me concentrer, dans ce qui suit, sur les autres aspects de la recherche terrain. Ils seront présentés après une section dédiée à des réflexions méthodologiques et pratiques portant sur la nature des données que l'on peut récolter avec un public d'apprenants d'une langue étrangère.

4.1 Réflexions méthodologiques et pratiques



Figure 23 – Blogue pédagogique.

Récolter de données sur le terrain est pour moi une évidence et une envie que j'ai cultivée déjà à une étape de ma carrière à laquelle la recherche ne faisait pas encore partie de mes tâches professionnelles. Voilà en quoi peut consister une telle activité : prendre des photos, faire des enregistrements audio, filmer des gestes physiques ou des écrans et récupérer des traces numériques. La capture de telles traces peut être motivée par des raisons pédagogiques : on peut avoir l'intention de s'en servir pour créer des activités ou des supports. Il peut aussi y avoir des raisons scientifiques, comme l'intention d'analyser les données pour la recherche. Parfois, les raisons sont d'ordre relationnel : les traces récoltées peuvent être utilisées pour témoigner de moments passés ensemble, un peu comme les albums photo produits dans le cadre familial ou amical. Les photos publiées sur le blogue *Didaktikel*,

que j'ai tenu pendant la période où j'ai travaillé comme professeure agrégée dans l'enseignement secondaire, en sont une illustration parlante (v. aussi Schaeffer-Lacroix, 2009b). Elles retracent, entre autres, des voyages scolaires effectués avec la classe. L'album « Die Schlösser Ludwigs II. » [Les châteaux de Ludwig II] a été créé par un des élèves. Il contient des photos qui ont été faites pendant un voyage dans le cadre familial et que cet élève a souhaité partager avec la classe via le blogue *Didaktikel*.

D'autres des albums publiés sur *Didaktikel* présentent les fruits de travail des élèves lors de projets ayant une composante visuelle, comme, par exemple, le projet *Konkrete Poesie* [Poésie concrète] exécuté avec des classes de troisième. Il s'agissait de créer un dessin en se servant de mots comme moyens d'expression sémantique et graphique (Gomringer, 1997). Le dessin représenté en Figure 24 illustre la séparation des deux Allemagnes entre 1949 et 1990. Une

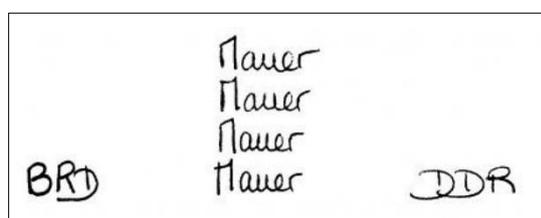


Figure 24 – *Konkrete Poesie*.

colonne comportant quatre fois le mot « Mauer » [mur] se dresse entre les mots « BRD » (Bundesrepublik Deutschland) et « DDR » (Deutsche Demokratische Republik), correspondant aux noms des deux pays de l'époque, à savoir la RFA (République fédérale

d'Allemagne) et la RDA (République démocratique d'Allemagne).

Deux des projets figurant sur la capture d'écran représentée en Figure 23, proposés à des classes de sixième, font le lien entre l'enseignement-apprentissage de faits linguistiques identifiés et leur représentation graphique. Ce travail s'inscrit dans le mouvement *language awareness* [prise de conscience linguistique] (Hawkins, 1987 ; Schaeffer-Lacroix, 2019b, p. 85-90 ; voir aussi le site *Internet Grammatik in Bildern* [Grammaire en images] (Dautel, [sans date]). L'un des deux projets, appelé *Kein oder nicht?*, a été dédié à la conceptualisation de la négation en allemand. Le deuxième projet, appelé *Spezialisierte Hunde* [Chiens spécialisés], a invité les élèves à trouver un mot composé se terminant par « Hund » [chien] et de faire un dessin qui le représente. Ce travail a eu comme but de soutenir la conceptualisation de la structure des énoncés en allemand. Les membres de la classe concernée ont trouvé des mots comme « Automatenhund » [chien automate], « Fahrerhund » [chien chauffeur] et « Jazzhund » [chien jazzy].

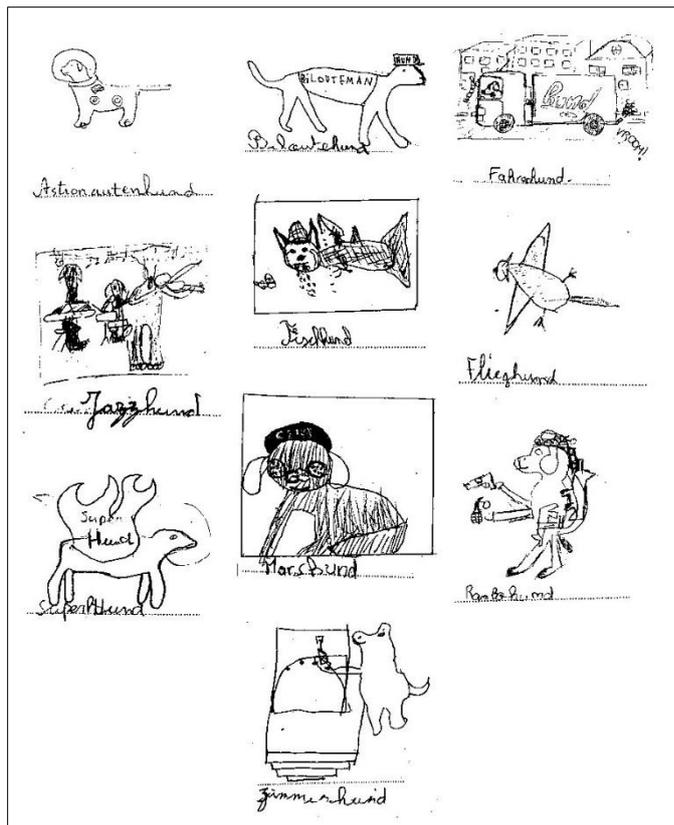


Figure 25 – Dessiner des chiens spécialisés.

J'ai fait des enregistrements de productions orales dans le secondaire, tout d'abord dans le but de pouvoir les évaluer de façon précise et pour donner aux élèves l'occasion de se réécouter et de s'écouter les uns les autres. J'ai aussi produit quelques films et j'ai fait des photos, surtout lors de sorties et pendant des échanges scolaires, afin de documenter ces événements et afin d'y inclure les parents. L'action de filmer a également fait partie de certains projets pédagogiques : je pense aux autoportraits filmés dans le cadre du projet *Qui est qui ?*, basé sur l'échange de courriels, effectué avec une classe en Allemagne. Ces réalisations étaient destinées à faire repérer aux élèves de l'autre pays dans le lot des films qui avait été leur partenaire dont l'identité avait été cachée pendant le projet.

Ces dernières années, la récolte de traces dans le cadre d'activités pédagogiques a commencé à être cadrée par l'institution de façon de plus en plus précise : sur le site Eduscol, une boîte à outils est mise à disposition, permettant au personnel enseignant ou encadrant de disposer de formulaires à proposer pour signature aux parents d'enfants mineurs et les enfants (Ministère de l'Éducation nationale et de la Jeunesse - Direction générale de l'enseignement scolaire [sans date]). En ce qui concerne le droit à l'image et le droit d'auteur (que je propose d'appliquer aussi aux productions des apprenants), la législation évolue constamment, et il est nécessaire

de se tenir au courant des lois en vigueur au moment d'une expérimentation.

En même temps, on observe une évolution vers une surveillance des usagers de plus en plus étroite par des moyens numériques, qui ne sont pas tous transparents : géolocalisation, stockage en ligne ou sur des serveurs de productions écrites ou orales, films ou photos produits dans le cadre d'événements publics, etc. (v. aussi les caractéristiques du Web 4 évoquées dans (Schaeffer-Lacroix, 2019b, p. 11).

Tracer les processus d'apprentissage est en train de devenir un mode (et peut-être même une mode ?) dont on met en avant les avantages : mesurer et quantifier à l'aide de statistiques faites à partir de données mises à disposition par les personnes apprenantes aiderait à mieux les servir, et on serait en mesure de renseigner sur leur évolution précise dans un contexte de formation et de leur proposer des parcours d'apprentissage individualisés. L'évaluation en serait facilitée et objectivée. Cette tendance est annoncée dans la section *Growing Focus on Measuring Learning* [Quantifier l'apprentissage de plus en plus] du document *Horizon Report, Higher Education Edition* (Alexander et al., 2019, p. 10). Il paraît important d'avoir conscience des limites à ne pas dépasser dans ce domaine, par exemple le stockage de données personnelles sans raison apparente et une confiance aveugle en les informations fournies par la machine par rapport à ce que font, savent et savent faire les usagers du numérique.

4.2 Défis à relever

L'observation de la façon dont des personnes apprennent une langue étrangère dans des situations formelles, organisées dans des institutions scolaires ou universitaires, est une entreprise délicate.

Les défis à relever sont premièrement d'ordre relationnel. Dans le cadre de projets menant à des communications et publications, il convient d'informer toute la communauté concernée (apprenants, parents d'apprenants mineurs, responsables de la formation et/ou de l'institution) sur ce qui est prévu de faire en matière de documentation de l'expérience et pourquoi on compte le faire. Minimale, il faut proposer aux participants (et aux parents des élèves mineurs) de remplir et signer un document pour donner leur accord explicite par rapport à l'utilisation des données récoltées. Un exemple pour un tel formulaire, proposé lors du projet *Audiodeskription* (v. 5.1.4), est fourni en figure 26.

Autorisation d'utilisation des données collectées lors du cours "UE Langue (allemand)" 2018/19

Je, soussigné·e (nom et prénom),
étudiant·e en M1 MEEF à l'Espé de l'académie de Paris, accepte que les données collectées en présentiel ou en ligne (plateformes *Edmodo*, *VideoAnt*, *YouDescribe*) sur la période du 1^{er} octobre 2018 au 28 février 2019 soient utilisées à des fins de recherche (publications, présentations orales lors de colloques, dépôt sur des plateformes de partage de données comme Clarin, Ortolang, etc.).

J'ai bien noté que mon anonymat sera toujours respecté.

Fait à Paris, le (date)

Signature :

Figure 26 – Formulaire d'autorisation d'utilisation des données.

Si l'on a affaire à des apprenants mineurs, les activités générant des traces numériques peuvent être rendues visibles pour les parents. Il peut s'agir de publications sur un blogue ou sur une plateforme d'apprentissage, rendus accessibles

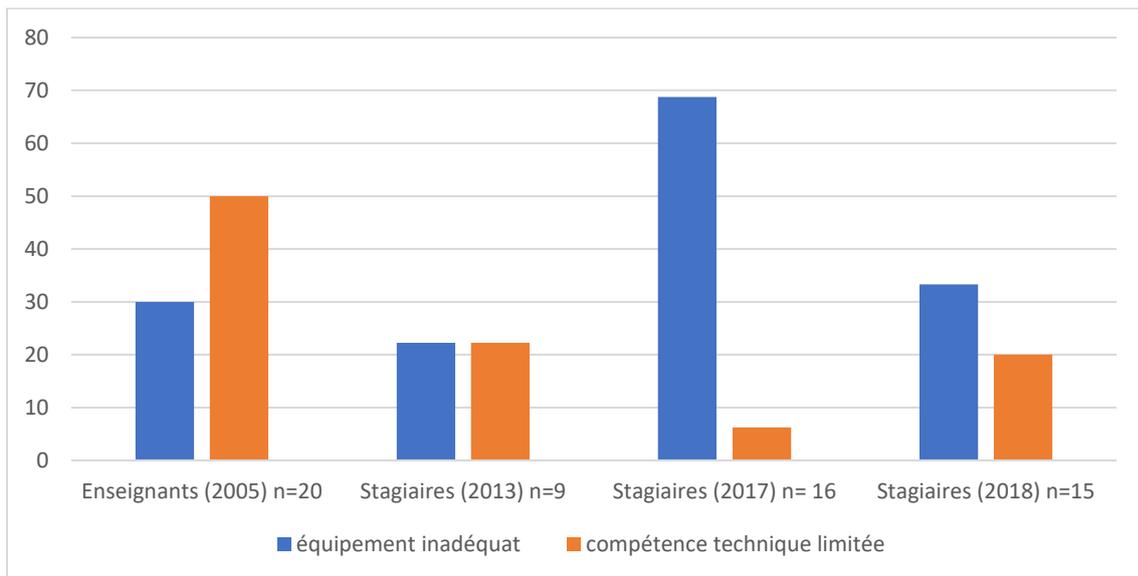


Figure 27 – Page d'accueil d'Edmodo.

grâce à des mots de passe ou des accès privés. Edmodo (ANON., 2008), une plateforme conçue pour de jeunes apprenants, prévoit des comptes réservés aux parents.

Un deuxième défi à relever lors d'une recherche impliquant le terrain (aussi appelée « recherche-action » ; pour une définition, v. Schaeffer-Lacroix, 2009a, p. 150-151) est la présence régulière des participants, moins facile à gérer que pendant une expérimentation en laboratoire, qui, par définition, est restreinte et contrôlée du point de vue temps, lieu et modalités d'exécution. Lors d'une recherche ayant lieu dans un établissement d'enseignement, il peut arriver que certains des apprenants viennent de façon irrégulière ou quittent définitivement le groupe, voire l'établissement, pendant la période concernée par la recherche. Un troisième défi est la disponibilité et/ou la fiabilité du matériel numérique dont on a besoin

pour exécuter et/ou documenter le projet. Les sept années passées, afin d'évaluer la faisabilité d'activités incluant le numérique dans l'enseignement secondaire, j'ai cherché à évaluer dans quelle mesure les stagiaires inscrits à l'IUFM/Espé/Inspé en master 2 MEEF ou FSTG, parcours allemand, pouvaient s'appuyer sur du matériel numérique pour implémenter, dans leur établissement d'exercice, un projet alimentant leur mémoire de fin d'études. Pour ce faire, j'ai créé un questionnaire en ligne, incitant les futures enseignantes à se renseigner sur la possibilité de faire, sur le lieu de stage, des cours impliquant du matériel informatique (ordinateurs, tablettes graphiques, téléphones portables, caméras). J'ai également cherché à savoir dans quelle mesure elles pouvaient compter sur un accès stable à Internet. J'ai complété ces sondages par une enquête menée en 2005 auprès d'enseignantes d'allemand en poste pour les besoins de mon mémoire de DEA (Diplôme d'études approfondi) en didactologie des langues et des cultures (Lacroix, 2005, p. 75) ayant pour titre *L'allemand, les TICE et les tâches*.



Graphique 2 – Freins perçus par rapport à l'usage du numérique en cours d'allemand.

Dans les enquêtes représentées en Graphique 2²⁴, la plupart des personnes répondantes ont cité comme freins (potentiels, perçus ou réels) à l'usage du numérique en cours d'allemand l'équipement non adéquat et leur propre compétence considérée comme limitée. Certaines des réponses contiennent des réserves par rapport au fonctionnement des équipements, en

²⁴ Le graphique présente un échantillon des données intégrales. Afin de rendre les données comparables, le nombre des réponses dans le graphique 2 a été ramené à des groupes de 100 enquêtés.

particulier par rapport à l'accès à Internet, comme le précise une des stagiaires inscrites à l'Espé de l'académie de Paris en 2018/19 : « Il devrait y avoir Internet dans les salles, mais en réalité ça ne marche jamais ». Il faudrait compléter cette liste de desiderata par la nécessité de mettre à disposition des enseignantes une personne les aidant à résoudre de façon rapide et efficace les problèmes d'ordre technique, mais aussi administratif qu'elles peuvent rencontrer sur le terrain lors de l'utilisation d'outils numériques et/ou de machines (Schaeffer-Lacroix, 2012, p. 221).

4.3 Récolte de données

4.3.1 Opinions

La récolte de données peut inclure l'identification d'opinions des personnes impliquées dans une recherche. Pour y avoir accès, on peut soumettre aux participants un pré- et/ou post-questionnaire en ligne, contenant des questions fermées et/ou ouvertes (Framaforms, Google formulaire, Typeform). Un exemple de questionnaire réalisé à l'aide de Typeform a été présenté en paragraphe [2.2.2](#). Des entretiens semi-directifs filmés, menés par une personne (de préférence experte du domaine) que les participants ne connaissent pas, sont proposés à des groupes constitués de trois ou quatre apprenants. Ce type d'entretien, dit « compréhensif » (Kaufmann, Singly, 2014), vise à poser des questions permettant de recueillir les opinions et le vécu des participants. Les questions sont préparées avec soin, mais suffisamment ouvertes pour éviter de suggérer des réponses ou de diriger la parole dans un sens ou un autre. Ci-dessous, à titre d'exemple, le questionnaire proposé pour le projet *Kommas* (voir paragraphe [5.1.3](#)).

1. Pouvez-vous me dire en quoi a consisté l'atelier auquel vous venez de participer ?
2. Mme Lacroix vous a proposé de travailler avec Sketch Engine. Qu'est-ce qu'on peut faire avec cet outil ? Est-ce que cela vous a servi de l'utiliser ? Si oui, à quoi ?
3. Pouvez-vous me parler des activités autour des virgules que vous avez faites ? Est-ce que ces activités vous ont apporté quelque chose ? Si oui, quoi exactement ?
4. Cet après-midi, vous avez réécrit vos textes. Comment avez-vous fait pour les améliorer ? Auriez-vous préféré réviser votre texte d'une autre façon, avec d'autres outils ?
5. Qu'est-ce que vous avez envie de proposer si un projet comparable devait être refait pour d'autres – des changements, des idées à retenir, des idées nouvelles... ?

La transcription des entretiens est ensuite effectuée à l'aide d'un outil numérique, par exemple oTranscribe (Bentley, [sans date]) ou EXMARaLda (Schmidt, Wörner, 2013). Ce deuxième a l'avantage d'être très interopérable, mais il nécessite plus de temps de prise en main que le premier. La figure 28 présente la transcription multimédia d'un bout de l'entretien post-recherche du projet *Weissensee*, qui est présenté plus en détail en paragraphe [5.1.1](#). Cette

transcription permet la lecture et l'écoute simultanée de l'entretien. De plus, le texte de la transcription peut être parcouru à l'aide du concordancier EXAKT (EXMARaLDA Analysis- and Concordance Tool) (Schmidt, 2019).

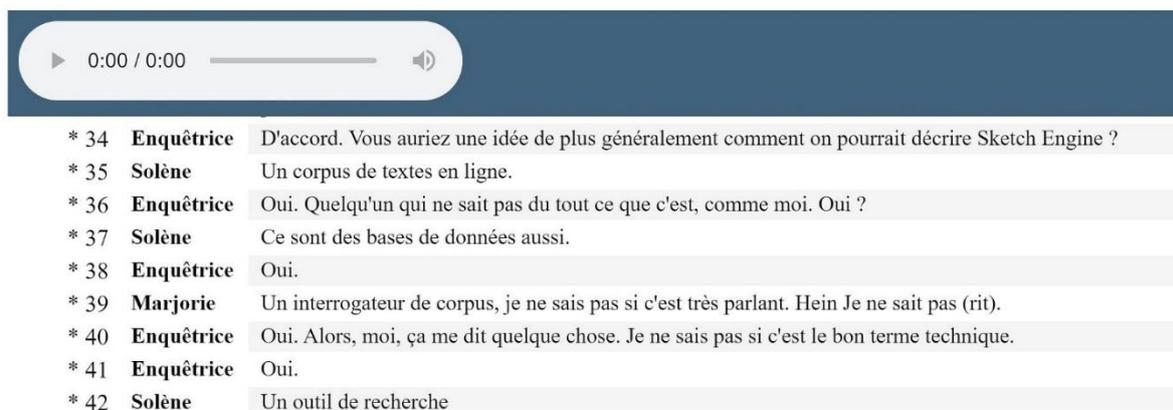


Figure 28 – EXMARaLDA pour transcrire l'entretien du projet Weissensee.

Entre 2009 et 2019, j'ai mis en place cinq entretiens filmés avec un public de germanistes, par groupes de trois ou personnes, et deux autres sont prévus pour 2020.

Tableau 3 – Entretiens menés entre 2005 et 2020.

Année	Public	Nom du projet	Groupes	Enquêtrice
2009	Enseignement secondaire (collège)	<i>Projekt Prospekt</i>	5	Isabelle Salengros Iguenane (maître de conférences de l'École des Ponts ParisTech)
2011	Enseignement secondaire (lycée)	<i>Filmlokal</i>	4	Isabelle Salengros Iguenane
2012	Master 1 documentation	<i>Weissensee</i>	1	Françoise Demaizière (Ingénieure de recherches HDR, université Paris 7)
2013	Master 1 Enseignement	<i>Weissensee Transfer</i>	3	Françoise Demaizière
2016	Master 1 Enseignement	<i>Kommas</i>	2	Ulrike Petzold (ingénieure pédagogique, Inspé Paris)
2020	Élèves-ingénieurs de l'École des Ponts ParisTech	<i>Cahiers des clauses techniques particulières</i>	5	Françoise Demaizière
2020	Enseignement secondaire (collège)	<i>Kleinanzeigen</i>	5	Claire Laurin (PRAG allemand, Inspé Paris)

4.3.2 Actions

Les actions des personnes participant à un projet de recherche peuvent consister en des gestes physiques, des interactions verbales et des activités laissant une trace numérique (p. ex. l'usage d'un forum en ligne). Ces dernières actions ne nécessitent pas d'intervention technique supplémentaire (sauf pour l'extraction des données). Les autres actions doivent être capturées à l'aide d'enregistrements audio et/ou vidéo et à l'aide d'outils permettant de filmer les écrans numériques. Jusqu'à présent, les gestes physiques n'ont pas été retenus dans mes recherches, et le volet « interactions verbales » sera traité dans le paragraphe suivant, s'intéressant aux productions des apprenants.

Les activités numériques des personnes participant à une recherche impliquant le terrain laissent des traces dont la nature et la granularité dépendent de l'outil. L'espace numérique de travail Moodle (ANON., 2019d) fournit beaucoup de détails sur les activités (heure et durée de la connexion, nombre d'essais pour les tests, dépôts fournis, participation à des interactions sur forum, etc.) tandis que d'autres espaces comme le portfolio Mahara (ANON., 2019c) en fournissent très peu. Edmodo (ANON., 2008) représente un compromis entre un traçage pouvant être ressenti comme intrusif et une absence de traces empêchant le repérage dans l'espace et la récolte d'informations nécessaires pour une recherche. Les outils de type wiki ou traitement de texte en ligne sont également une source précieuse pour la récolte de données numériques. Je me suis plusieurs fois servi de TitanPad, un outil qui n'est plus disponible depuis 2017, mais qui peut être remplacé par son pendant gratuit Framapad (ANON., [sans date]), un éditeur de texte collaboratif. Des exemples d'utilisation et d'exploitation de TitanPad pour la recherche en didactique de l'allemand sont offerts dans Schaeffer-Lacroix (2016c) et aussi dans Schaeffer-Lacroix (2015b), d'où provient l'extrait ci-dessous.

Enseignante (EL) / Cornélia (C) / Marjorie (M) / Solène (S)

(EL) **Was können wir Solène, V2, hierzu sagen?** [Que pourrions-nous dire à Solène concernant cela ?]

*Ich habe die Kassette **zugehört**...* [J'ai écouté la cassette.] (S) **anhören** [écouter]

(C) Peut être revoir la particule verbale, en fonction de ce qu'on a dit en cours. Ou bien utiliser Pons pour les différences entre écouter/entendre (S) ce qui est bizarre c'est que je trouve bien le sens écouter pour *zuhören*?

Dans cet exemple, les personnes sont identifiées par un moyen visuel : l'outil attribue à chaque

auteur une couleur en laquelle son texte sera surligné²⁵. Le processus de rédaction et d'interaction peut être retracé à l'aide d'une fonctionnalité de type « historique des versions », apparaissant sous forme d'une capture d'écran filmée.

Si le projet s'intéresse à la façon dont les participants s'emparent d'un outil numérique, il peut être utile de filmer les écrans à l'aide d'un outil qui enregistre, en parallèle, ce que disent les usagers (Camtasia (ANON., 2020) ; Screencast-O-Matic (ANON., [sans date])). Les projets que j'ai conçus et exécutés en didactique des langues ont été principalement soutenus par ce dernier outil (voir section 5.1). Un des exemples d'exploitation des données récupérées à l'aide d'écrans filmés est fourni dans Schaeffer-Lacroix (2018d).

Le tableau 4 donne un aperçu des outils ayant servi à récolter des données dans le cadre des cours que j'ai dispensés dans le secondaire et le postsecondaire.

Tableau 4 – Données récoltées.

Nature des données	Objet/public	Cadre/raison de la récolte de données	Techniques et exemples	Outils
Photos prises par l'enseignante, par d'autres personnes encadrantes ou par les apprenants	- Salle de classe - Lieux visités avec les apprenants - apprenants	Documenter des sorties et des échanges scolaires	Visite de la <i>Brotbackhaus</i> [maison pour cuire du pain] à Wendlingen am Neckar (Allemagne).	Blogue
		Projets pédagogiques alliant photos et production écrite; avec publication sur un blogue de classe.	Les élèves prennent des photos d'une situation qu'ils trouvent surprenante en Allemagne ; preuves de gestes écologiques.	Blogue
Photos			Les élèves font cinq photos qui deviennent la base d'une histoire publiée sur un blogue.	Blogue
Films			Autoportraits filmés pour le projet <i>Qui est qui ?</i>	
Films		Documentation d'un projet de classe	Film produit par l'enseignante de la pièce de théâtre <i>Hänsel und Gretel</i> (école primaire).	

²⁵ Afin de rendre le texte compréhensible pour d'autres personnes, j'ai inséré la première lettre des prénoms des participantes et j'ai fourni des traductions vers le français.

Films		Mise en scène, contextualisation, <i>learning by doing</i>	- Films produits avec les élèves : la météo, les infos. - Situations prototypiques pour la culture allemande, p. ex. attendre au feu rouge.	
Enregistrements audio	- Oral en continu. - Interactions orales entre apprenants ou entre apprenants et enseignante.	- Faciliter l'évaluation orale et la rendre plus objective. - Faire entendre aux apprenants comment ils parlent ou comment parlent les autres (prise de conscience).	- Parler dans les blancs laissés à des fins d'entraînement. - Enregistrements en cours CLES B2 et en cours Lansad B2)/ - Descriptions orales d'une image en vue de la faire dessiner (projet <i>Audiodeskription</i>).	Audacity avec commentaires insérés
Clips numériques	- Fournir des modèles. - Créer des supports.		- Enregistrements avec la voix des apprenants ou de l'enseignante (personnages et fonds customisables).	Voki (avatar parlant)
Traces numériques	- Partager des ressources. - Interagir en ligne (<i>chat</i>). - Outils d'évaluation.		- Accompagnement de mémoires de master 2 - TD M1 MEEF "Création de scénarios". - CM M1 MEEF "Connaître les approches pédagogiques"	Moodle Mahara Google drive Edmodo Trello

4.3.3 Productions

Les productions écrites numériques laissent des traces qui sont facilement exploitables pour la recherche. Le confort par rapport à la récolte de telles données est encore augmenté quand on fait appel à un outil qui retrace le processus d'écriture, soit en affichant les différentes étapes et les changements effectués (Google document), soit en mettant à disposition l'historique dynamique, c'est-à-dire un film ou un diaporama qui donne l'occasion de suivre « en direct » la genèse du texte. Cette fonctionnalité a été offerte par TitanPad mentionné en section 4.3.2 et elle existe aussi pour Framapad. L'enregistreur d'actions utilisateur fourni par Windows 10²⁶ est également une source d'information précieuse.

²⁶ <https://support.microsoft.com/fr-fr/help/22878/windows-10-record-steps>

Enregistrer les interactions orales de groupes de deux apprenants permet d'étayer les opinions récoltées à l'aide de traces matérielles. Le fait de conjuguer les actions à l'écran et les interactions entre pairs (deux personnes qui commentent leurs actions en les effectuant) permet de se faire une idée de leurs représentations (sur la langue, sur les qualités des outils, etc.) qui, par ailleurs, sont difficiles d'accès. Le choix de ce procédé s'inspire du protocole *Think aloud* (penser à haute voix) (Ericsson, Simon, 1987), décrit en section 1.4 de Schaeffer-Lacroix (2019b, p. 25-26). Il a été utilisé dans plusieurs des recherches en didactique de l'allemand que j'ai eu l'occasion d'exécuter. Des exemples de son exploitation peuvent être trouvés dans Schaeffer-Lacroix (2015b ; 2016c), s'intéressant au sens et à la fonction d'éléments polysémiques et pluricatégoriels, et dans Schaeffer-Lacroix (2018d), présentant une recherche portant sur la conceptualisation de la virgule (v. les paragraphes [5.1.1](#) et [5.1.3](#)).

4.4 Transformation des données

4.4.1 Anonymisation

Plusieurs raisons peuvent justifier une intervention technique sur les données récoltées. Premièrement, il faut systématiquement anonymiser les éléments liés à l'identité des participants, au moins leurs noms et prénoms. Si une personne n'a pas signé l'autorisation d'utiliser les données pour la recherche, il faut supprimer ses données, et il faut, dans le cas de données multimodales, couper les passages la concernant dans des enregistrements et/ou flouter des parties d'une image. Pour retravailler les données audio, on peut se servir d'outils comme l'enregistreur éditeur audio Audacity (2007). Pour modifier les images, Photofiltre (ANON., 2004) ou Redacted ([sans date]) rendent de bons services.

4.4.2 Transcription

Il existe des outils sophistiqués pour transcrire les données récoltées à l'aide d'enregistrements, par exemple EXMARaLDA (données audio ou vidéo ; voir l'exemple fourni en paragraphe [4.3.1](#)) ou Folker (données audio) (Schmidt, 2012). Ces deux outils ont l'avantage de maintenir le lien numérique entre le fichier d'origine et sa transcription, et ils permettent de créer plusieurs pistes d'annotation, ce qui peut être intéressant si l'on s'intéresse aux actions non verbales ou si l'on souhaite ajouter une traduction. Ces outils nécessitent une certaine expertise. Si l'on souhaite transcrire assez vite et si l'on pense uniquement avoir besoin du texte (sans maintenir le lien avec la vidéo), l'outil oTranscribe (Bentley, [sans date]) est à recommander.

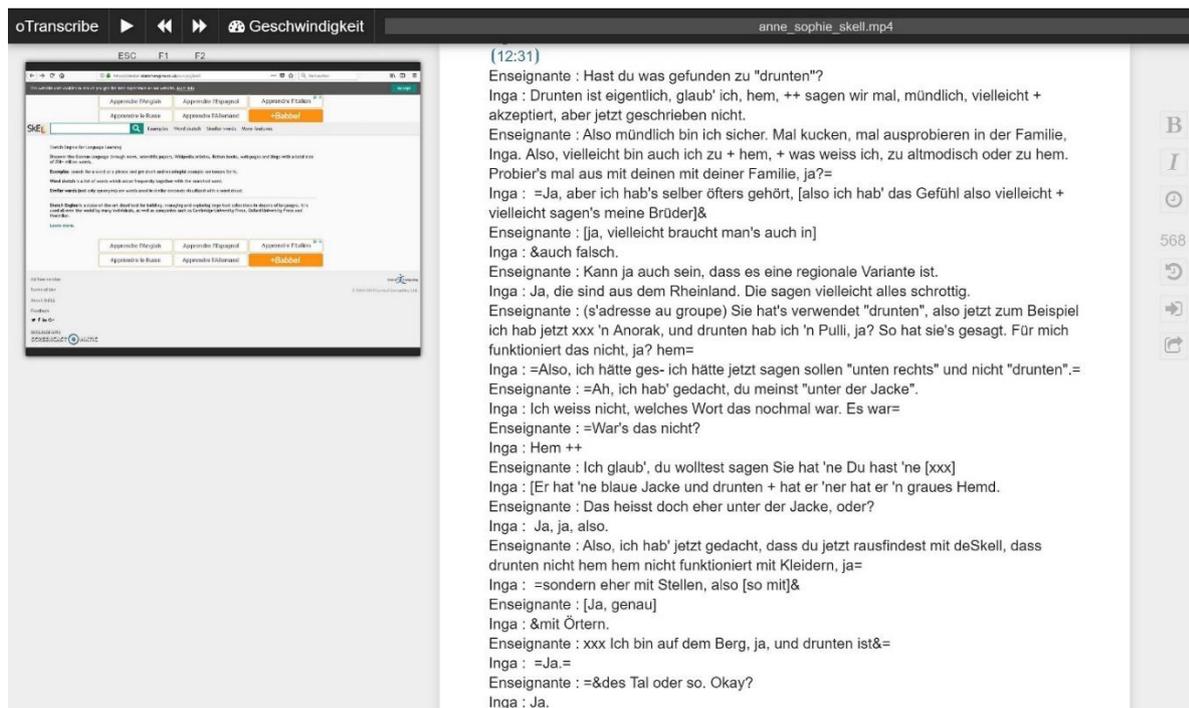


Figure 29 – Transcription d'un fichier audiovisuel avec oTranscribe.

La figure 29 fournit un exemple de l'interface d'oTranscribe pour des données provenant du projet *Audiodeskription* (voir paragraphe 5.1.4). Il est possible d'insérer des repères de temps et d'exporter la transcription finale sous plusieurs formats : texte brut (.txt), Markdown (.md), qui est lisible par un outil de XML, et un format propre à oTranscribe (.otr).

Notons aussi que l'outil VLC (ANON., [sans date]) donne l'occasion d'extraire une partie d'une vidéo ou de ralentir son débit, ce qui peut être utile pour faire des transcriptions de l'oral.

4.2.3 Annotation

L'annotation se situe au carrefour entre recherche et pratique. Elle est centrale dans le domaine des corpus numériques, où elle renseigne sur les caractéristiques linguistiques et autres des données. Il est également courant de fournir les métadonnées qui précisent les conditions de production du document (auteur, lieu, date, etc.). La figure 30 présente un exemple d'annotation (en langue anglaise) que j'ai faite pour le script d'audiodescription du premier épisode de la série *Neues aus Buettengewarden* (Eberlein, 1997). L'étiquette « stage » veut dire « didascalies ». Les passages entourés de l'étiquette « record » doivent être enregistrés par l'audiodescripteur, et le terme de « mode » informe sur la façon dont il faut parler (p. ex. « ss » veut dire « parler très vite »).

```

10:05:06
<stage>(Türöffnen)</stage>
<record>In der Scheune stehen viele Käfige. Brakelmann stellt den Korb ab und sieht zur Katze hinein. </record>
10:06:05 "die Fütterungszeiten eingestellt."
<mode>ss</mode>
<record>Draußen stehen Wecker ... </record>
("SO")
<record>... auf einem Sims. </record>

```

Figure 30 – Annotation d'un script d'audiodescription à l'aide de XML.

Appliquée au domaine de l'enseignement-apprentissage, l'annotation des données peut être offerte dans le but d'offrir une aide à la compréhension d'un document. Je trouve particulièrement formateur la mise à disposition de fenêtres surgissantes renseignant sur les étiquettes disponibles pour les corpus déposés sur TXM. La figure 31 illustre cette possibilité, accessible à partir de la fonctionnalité « Édition », à l'exemple de la forme allemande « ab » [à partir de, être détaché, être cassé]. Cette forme peut avoir la fonction de préposition ou de préverbe. L'étiquette qui lui est attribuée dans l'exemple ci-dessous, issu du corpus Buettenwarder (Eberlein, 1997), renseigne sur le fait que TreeTagger a identifié cette forme comme préverbe détachable (« PTKVZ: abgetrennter Verbzusatz », v. Schiller et al., 1995, p. 66), ce qui correspond au contexte fourni : « Brakelmann stellt den Korb ab und sieht zur Katze hinein » [Brakelmann dépose le panier et va voir le chat.].

```

10: 05: 06
(Türöffnen)
In der Scheune stehen viele Käfige. Brakelmann stellt
den Korb ab und sieht zur Katze hinein.
depos:PTKVZ delemma:ab

```

Figure 31 – Fenêtre surgissante pour une étiquette TreeTagger (TXM).

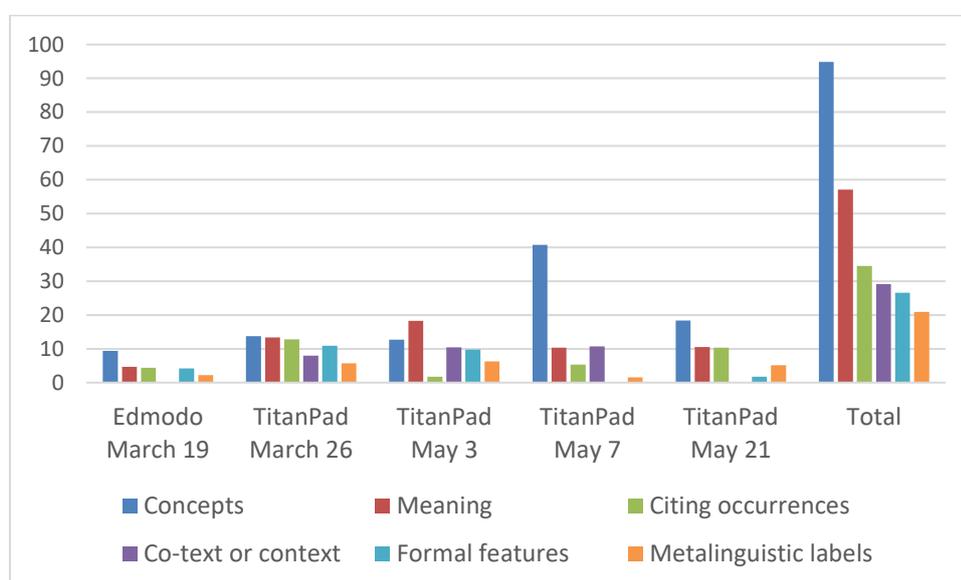
Les annotations peuvent également être ajoutées par les usagers. S'il s'agit d'apprenants, ils peuvent créer des étiquettes dans le but d'exprimer ce qu'ils ont compris (ou pas), ou en vue d'enrichir un objet de façon créative. Il peut être intéressant de faire utiliser un outil d'annotation par des apprenants pour leur donner l'occasion d'écrire et de faire commenter leurs premières versions de texte (v. les annotations d'un générique de film, faites à l'aide de l'outil VideoAnt (ANON., [sans date]) pendant le projet *Audiodeskription*, paragraphe 5.1.4).

4.5 Analyse des données

4.5.1 Analyse qualitative

L'analyse qualitative (Silverman, 2017) porte sur des éléments récoltés sur le terrain (opinions, discours, comportements, productions). Elle peut aller de pair avec une lecture linéaire des

données, qui mène à l'identification de caractéristiques remarquables et à la création de catégories ou de grilles de lecture. Les résultats peuvent fournir des points d'entrée pour des analyses supplémentaires, faites, par exemple, à l'aide d'outils de statistique textuelle ou de corpus. Pour les besoins d'une analyse qualitative, on fait couramment appel à des catégories existantes, déjà mentionnées dans des recherches précédentes, démarche adoptée par Stranovská et al. (2019). Parfois, il semble nécessaire de créer des catégories afin de pouvoir répondre aux questions liées à un contexte particulier. J'ai opté pour cette deuxième solution à l'occasion de l'analyse des données du projet *Weissensee* (Schaeffer-Lacroix, 2016c), proposé à trois étudiantes en master 1 documentation. La recherche s'est intéressée à la façon dont ces étudiantes comprennent le sens et la fonction des préverbes et prépositions dans un script de film allemand (voir paragraphe 5.1.1).

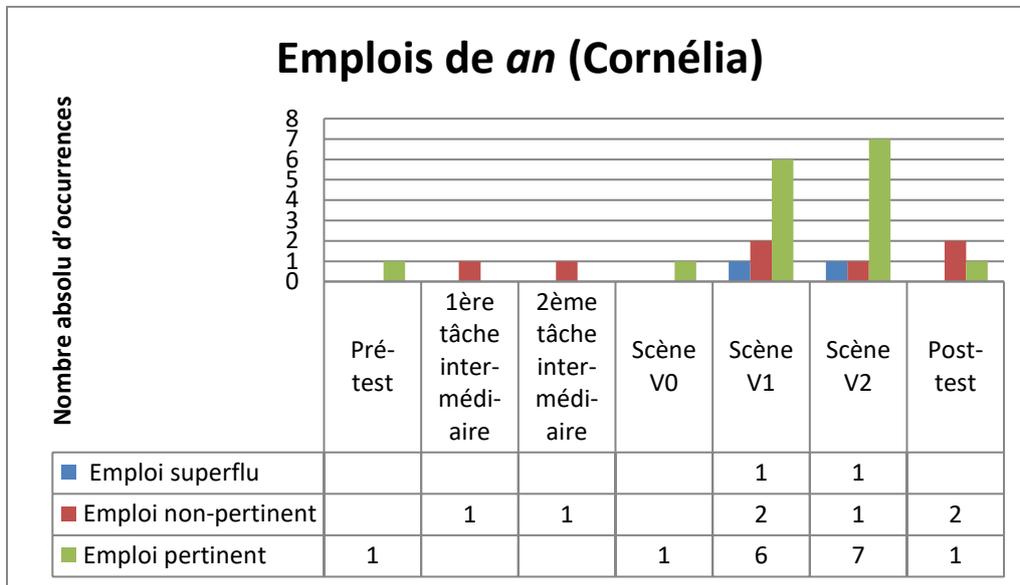


Graphique 3 – Thèmes des interactions verbales du projet *Weissensee*.

J'ai choisi les catégories suivantes (v. Graphique 3), détectées dans le discours des participantes et susceptibles de renseigner sur (l'évolution de) leur expertise métalinguistique : mentionner un concept, parler du sens d'un terme ou énoncé, citer une occurrence dans le texte, mentionner le cotexte (défini comme environnement immédiat d'un terme) ou le contexte (défini comme environnement plus large d'un terme), mentionner des caractéristiques formelles d'un terme, citer une étiquette métalinguistique. Les catégories établies ont, par la suite, servi à évaluer dans quelle mesure les activités proposées ont permis aux participantes de développer un discours métalinguistique plus expert, donc plus proche de ce que diraient des spécialistes de la langue, qu'au début du projet.

4.5.2 Analyse quantitative

L'analyse quantitative (Glynn, Robinson, 2014) cherche à interpréter des données de recherche en les transformant en unités quantifiables. Elle est soutenue par des outils d'analyse, par exemple des tableurs, des outils de corpus et des outils d'analyse statistique.



Graphique 4 – Évolution de l'emploi de « an » par Cornélia (projet *Weissensee*).

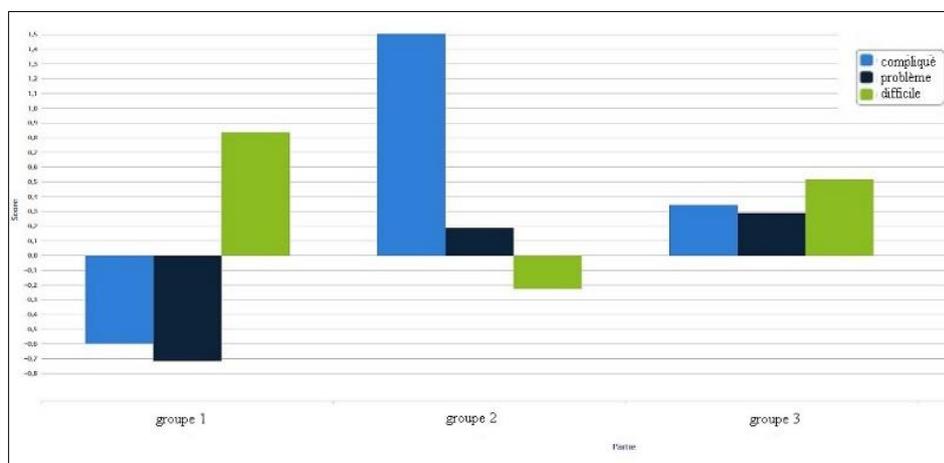
Pour analyser mes données, j'utilise en général un tableur pour créer des graphiques, comme celui présenté ci-dessus. Il retrace l'usage que fait une étudiante de la forme polycatégorielle *an* [à, près de] au fil d'un projet d'écriture soutenu par la consultation d'un corpus réflexif (Schaeffer-Lacroix, 2015b), c'est-à-dire, d'un corpus contenant des textes représentant tous un même genre textuel ou discursif (Mayaffre, 2002).

Pour procéder à une analyse quantitative des données des recherches terrain auxquelles j'ai participé (voir section 5.1), je me suis servie de plusieurs outils de corpus : Lexico3 (Salem, 2009) pour *Projekt Prospekt* (Schaeffer-Lacroix, 2009a), Sketch Engine (Kilgarriff et al., [sans date]) pour *Weissensee* (Schaeffer-Lacroix, 2015b ; 2016c), ainsi que de Sketch Engine, TXM (Heiden et al., 2010) et R (ANON., [sans date]) pour *Kommas* (Schaeffer-Lacroix, 2018d). Ces outils ont principalement permis de calculer des fréquences absolues ou relatives ou d'identifier des formules (semi-)figées. Je me suis servie de R pour mettre en regard plusieurs types de données du projet *Kommas*.

nombre_requêtes_	type_requête	nombre.occurrences.obtenues	formation.préalable
bertille_agnès:17	CQL	:24 0	:11 non:11
chantal_fanny : 7	CQL_context	: 1 100<	: 8 oui:27
germine_sarah :10	phrase	: 1 entre 1 et 5	: 2
habiba_laura : 4	simple	: 4 entre 10 et 49:	1
	simple_context:	8 entre 50 et 99:	3
		entre 6 et 49 :	5
		NA's	: 8

Figure 32 – Analyse quantitative des types de requêtes (projet Kommas).

Pour les besoins de la publication d'un article faisant partie d'un volume portant sur l'usage des outils de corpus par et pour de jeunes apprenants (Schaeffer-Lacroix, 2019a), j'ai analysé, à l'aide de TXM, certaines des données du projet *Weissensee Transfer* (v. paragraphe 5.1.2), proposant à de futures enseignantes de créer un scénario pour enseigner l'allemand au collège ou lycée. L'éditeur du volume n'a pas souhaité retenir les graphiques obtenus, trouvant qu'ils nécessiteraient un appareil explicatif trop lourd. J'avais, par exemple, prévu l'insertion de graphiques permettant d'évaluer des freins à l'usage de Sketch Engine par les trois groupes ayant participé au projet. Leur scénario devait inclure au moins une activité faisant appel à l'usage de Sketch Engine. Certains des participants ont accepté cette contrainte de bonne grâce ou même avec enthousiasme, tandis que d'autres ont émis des réserves.



Graphique 5 – Termes « compliqué, problème, difficile » (*Weissensee Transfer*).

Afin de mesurer la résistance à l'usage de Sketch Engine (en vue de son utilisation avec un public du secondaire), j'ai choisi comme indices les termes « compliqué, problème, difficile ». J'ai analysé leurs fréquences pondérées à l'aide de la fonctionnalité « Spécificités » de TXM²⁷.

²⁷ Le graphique d'origine contient un groupe test supplémentaire, nécessaire pour effectuer des calculs de type "spécificités".

Le groupe 2, ayant obtenu un score plus élevé que les deux autres groupes pour le terme de « compliqué », a, en effet, résisté le plus à cette idée d'utiliser Sketch Engine avec une classe de jeunes élèves, tandis que le groupe 3, modéré dans son usage des trois termes, lui a réservé un accueil favorable, à en juger par les autres données recueillies (écrans filmés, tâches produites et entretiens).

Une approche combinant lecture linéaire et analyse textométrique est mise en pratique dans (Salengros Iguenane, Schaeffer-Lacroix, 2020) et mentionnée dans Schaeffer-Lacroix et al. (2020).

Les outils de textométrie donnent l'occasion de confronter des hypothèses linguistiques à une grande quantité d'occurrences et d'en créer à partir des données observées. Leur usage permet d'aborder le texte non seulement par le biais d'une lecture linéaire, mais aussi en observant de façon ciblée certains de ses éléments et caractéristiques sélectionnés en fonction des questions que l'on se pose.

Pour résumer, on peut constater qu'en matière d'analyse quantitative, j'ai fait des essais, dont certains plus concluants que d'autres. Ce que l'on peut regretter le plus est le nombre faible et instable des participants dont les données ont été explorées. Il s'agit premièrement d'un problème institutionnel : le nombre des personnes souhaitant apprendre ou enseigner l'allemand dans les établissements d'exercice dans lesquels j'ai eu l'occasion de travailler est en général faible. C'est d'ailleurs une des raisons pour lesquelles j'ai cherché à collaborer avec des collègues travaillant à des endroits où les groupes d'apprenants sont plus grands, p. ex. avec Caroline Rakotondravao, enseignante au collège, et/ou avec des collègues enseignant d'autres langues que l'allemand, notamment avec Isabelle Salengros Iguenane, spécialiste du FLE à l'école des Ponts ParisTech. Un deuxième obstacle à surmonter est lié à mes compétences encore à parfaire en matière de statistique générale et textuelle. Il est possible d'y parvenir grâce à des formations pré- ou post-colloque ou des formations en ligne, p. ex. des MOOCs (Massive open online course) portant sur les outils de corpus, la statistique et l'apprentissage machine. La participation à des projets impliquant des partenaires spécialistes du TAL est également un des facteurs pouvant favoriser le développement de compétences dans le domaine des outils d'analyse quantitative. C'est une des raisons qui ont motivé la conception du *TVseriesAD Translation Project* (v. section [5.3](#)), impliquant une équipe aux compétences plurielles.

4.6 Corpus d'experts et corpus d'apprenants

Un corpus d'experts contient des textes produits par des personnes considérées comme spécialistes de la langue et du domaine traité (Boch, Frier, 2015, p. 91). Lors d'une recherche

en didactique des langues, il peut servir de corpus comparable permettant de se rendre compte de différences entre ce que des apprenants produisent et ce qui est attendu dans le contexte d'un genre textuel ou d'un discours déterminé. Dans le cadre de la conscientisation linguistique (*language awareness*), les apprenants sont encouragés à aborder la langue comme le feraient des linguistes. Ils sont alors considérés comme « apprenants-chercheurs » (Cheng et al., 2003), et leurs observations de données d'un corpus d'experts peuvent avoir une valeur heuristique (v. le projet *Weissensee* en paragraphe 5.1.1 et le projet *Kommas* en paragraphe 5.1.3).

Les corpus d'apprenants contiennent les productions récoltées lors d'une recherche impliquant le terrain. La mise en corpus des données produites par les participants permet d'interroger toutes les productions ou de cibler les productions d'un échantillon déterminé. Pour donner un exemple, dans le cas d'un cours Lansad (Langues pour spécialistes d'autres disciplines), regroupant des spécialistes de plusieurs disciplines universitaires autres que la langue étudiée, on peut consulter uniquement les données produites par les personnes spécialistes d'une discipline en particulier.

Le public ayant, dans le passé, assisté aux cours de didactique et de technologie éducative que j'ai dispensés à l'Espé (maintenant Inspé) de l'académie de Paris s'est composé de personnes dont la langue maternelle est le français ou une autre langue, dont l'allemand. En 2013/2014, j'ai récolté des résumés de textes du domaine de la pédagogie, faisant partie des lectures obligées d'un cours magistral en didactique de l'allemand. De plus, j'ai récolté les commentaires que les étudiants ont rédigés par rapport aux résumés accessibles sur la plateforme éducative Edmodo (ANON., 2008). Dans le corpus obtenu, appelé *Aufsätze* [compositions], j'ai distingué entre un sous-corpus d'experts (Boch, Frier, 2015, p. 91), contenant les données des personnes ayant l'allemand comme langue maternelle, et un sous-corpus d'apprenants, contenant les données des personnes ayant une autre langue maternelle que l'allemand.

Name	Tokens	Words	%
apprenants	10,233	~8,413	68.9
experts	4,628	~3,805	31.1

Figure 33 – Corpus *Aufsätze* sur Sketch Engine.

Cette distinction permet d'observer des différences entre des traits des écrits des apprenants et de ceux des personnes considérées comme expertes de la langue. On peut, par exemple,

constater, à l'aide de la requête [word=", "[tag="CONJ.Coord"], recensant les virgules suivies de conjonctions de coordination, que les personnes ayant une autre langue maternelle que l'allemand utilisent plus de virgules (6449,72 par million) placées avant les conjonctions de coordination que les personnes germanophones du groupe (4753,67 par million). Il est possible de faire explorer par des apprenants soit l'un, soit l'autre des deux sous-corpus (v. projet *Kommas*, paragraphe 5.1.3).

En 2016/17, les participantes du projet *Kommas* ont également rédigé des résumés d'articles pédagogiques, appelés *Lesefrüchte* [fruits de lecture], ainsi que des commentaires portant sur les textes produits par les pairs. En vue d'une exploration des données à l'aide de TXM, j'ai structuré les données obtenues à l'aide d'un tableur. Les métadonnées retenues concernent les genres textuels (résumé, commentaire), les prénoms anonymisés des participantes, les identités des binômes (p. ex. « agnesmagda » est le binôme composé d'Agnès et de Magda), l'université dans laquelle les participantes se sont inscrites (P3 = Université Paris Sorbonne nouvelle, P4 = Sorbonne Université), ainsi que leur participation (ou pas) à l'atelier de prise en main de Sketch Engine, signalée par « oui » ou « non ». Voici la représentation des métadonnées dans le menu « Propriétés des structures » de TXM du corpus *Aufsaetze*.

text

genre (2) = résumé, commentaire.

id (33) = agnes, agnesmagda, agnessarah, bertille, bertillegermine, bertillelaura, bertillemaya, chantal, chantalbertille, chantalfiona, fanny, fiona, fionaagnes, fionabertille, fionachantal, germine, germinesarah, germinesarahbertille, ghislaine, habiba, laura, lindalaura, maya, mayaagnes, mayamadga, sarah, sarahgermine, sarahmaya, setraagnes, vahide, vahidebertille, vahidelaura, vahidemagda.

name (15) = "", Agnes, Bertille, Chantal, Fanny, Fiona, Germinesarah, Ghislaine, Habiba, Laura, Linda, Maya, Sarah, Setra, Vahide.

université (3) = "", p3, p4.

atelier (3) = "", non, oui.

La conjonction de subordination « dass » [que] doit être précédée d'une virgule sauf si elle est précédée d'une conjonction de coordination. Sa recherche dans le corpus a permis d'observer les contextes dans lesquelles les apprenants non germanophones se sont trompés. La figure 34 permet d'observer une source d'erreur fréquente : l'emploi de la structure de mise en relief « ... ist, dass... » (exemple : « Was mir auffällt, ist, dass... » = « Ce que je remarque, c'est que... ») est nativisé, donc probablement liée à l'existence d'une structure représentée dans une langue dont les apprenants sont familiers, en l'occurrence en français.

Requête	dass	Contexte gauche	Pivot	Contexte droit
text_id, text_genre				
laura, abstract		verlassen. Der Unterschied zu Übungen ist xxx	dass	die Aufgaben selten Fehlerfrei bleiben. Im Englischen
fionachantal, commentary		← [mir] auffällt, ist xxx	dass	eine ganz innovative und revolutionäre Lernmethode
laura, abstract		erklären. In seinem Dokument wird gezeigt xxx	dass	es sich um einen Fortschritt handelt xxx den der Lerne
laura, abstract		beim Fremdsprachenlehren und Lernen) wird gezeigt xxx	dass	Handlungsorientierung und Aufgaben sehr eng zusan
laura, abstract		sich etwas Einprägt xxx wovon man denkt xxx	dass	es wichtig und richtig ist, man übt solange xxx bis ma
laura, abstract		Man charakterisiert oft diese neuen Aufgaben dadurch xxx	dass	sie dem Lernenden die Möglichkeit geben xxx den Lei
laura, abstract		Aufgabenorientierung in einer Fremd- oder Zweitsprache, und	dass	nicht nur im Deutschen Bereich xxx sondern auch in ε
vahidemagda, commentary		Voraussetzung, dass sie begleitet sind, und	dass	sie weiter mit ihrer Verwandten in Verbindung bleiben.
germine, abstract		, um den Schülern zu helfen, und	dass	, sie nicht als demotivierend betrachtet soll. Deshalb k
bertille, abstract		ein Kind keine angeborene Intelligenz hat, und	dass	es nicht alle Aufgaben in bestimmten Alter ausführen l
chantal, abstract		Sie haben nämlich die Lernmethoden geprägt, ohne	dass	es mir bewusst war. Meiner Ansichten nach erweisen :
chantal, abstract		auf die Nachahmung seiner Umwelt angewiesen, ohne	dass	Instinkt und eingeborene mentale Eigenschaften zu b
fionabertille, commentary		, was mich auch besonders auffällt xxx ist	dass	Piagets und Bruners Theorien sich nicht nur ganz erg
bertillelaura, commentary		Was auch wichtig zu betonen wäre, ist	dass	Ingo Thonhauser den Begriff der Autonomie erwähnt:
germine, abstract		Korrektur notwendig. Was fällt auf, ist	dass	, die Studenten eine Korrektur nach ihrem Redebeitra
fanny, abstract		William Crozier, der ihn davon überzeugte.	dass	das Tier als Mittelpunkt der Forschung xxx sei und xx
germine, abstract		eine Katze geseht ", was zeigt,	dass	der Schüler die Bildung und die Regeln des Perfektes
habiba, abstract		anzuregen. Allerdings bin ich überzeugt xxx,	dass	alle Menschen in der Welt nicht über die gleichen Bilc
fanny, abstract		Hund jedes mal, wenn er wusste,	dass	es Futter gibt, anfang zu sabbern. Dieser Lernprozess
chantalfiona, commentary		von Erwachsenen anzugebracht wird. Ich wundere,	dass	ein Forscher so fern von der Realität stehen kann.
fiona, abstract		der Sprache erwerben, wenn berücksichtigt wird,	dass	sie unter unterschiedenen Bedingungen aufwachsen

Figure 34 – Comparaison de l'usage de la virgule en LE et LM (TXM).

5. Un choix de projets outillés

Les projets que j'ai développés ou dans lesquels j'ai pu m'insérer intègrent de façon exclusive des ressources et outils de corpus. L'arrière-plan théorique qui sous-tend l'intégration des corpus dans des projets d'enseignement-apprentissage de l'allemand ainsi que la méthodologie appliquée et les procédés sont décrits en détail dans Schaeffer-Lacroix (2019b). Dans ce chapitre, je reviendrai sur certains des projets qui y sont mentionnés et j'en rajouterai d'autres qui ont été développés depuis. On observera un élargissement des domaines de recherche abordés : les champs couverts par les projets cités sont toujours la didactique de l'allemand, mais aussi celle du français, la linguistique et la traduction. Une attention particulière est portée au rôle des outils numériques ainsi qu'aux leçons que je tire de ces expériences par rapport à ma façon de faire de la recherche.

5.1 Corpus et didactique des langues étrangères

La section 5.1 présente, par ordre chronologique, les projets que j'ai développés sans partenaire universitaire, mises à part les personnes sollicitées pour les entretiens post-recherche. Ces projets s'inscrivent dans le domaine de la didactique de l'allemand ou du français, et ils contiennent, à des proportions variables, les composantes « production écrite en langue étrangère », « réflexion sur la langue », « exploration de corpus ». Les projets implémentés ont été exécutés avec des étudiants inscrits en première ou deuxième année de master MEEF, qui suivaient le parcours allemand (paragraphe 5.1.1, 5.1.2, 5.1.3) ou un autre parcours d'enseignement (paragraphe 5.1.4). Les deux derniers projets, en cours d'implémentation, concernent un public d'élèves-ingénieurs allophones (FLE) d'une grande école (paragraphe 5.1.5) et un public d'élèves germanistes au collège (paragraphe 5.1.6). La méthodologie adoptée contient les éléments suivants ou un choix parmi ces éléments : pré-test, post-test, questionnaire en ligne avant la formation, enregistrements audio des séances de cours, écrans d'ordinateur filmés, entretiens post-recherche.

5.1.1 *Weissensee*

Description du projet dans Schaeffer-Lacroix (2015b) :

Le projet *Weissensee* a été développé pour un groupe de trois étudiantes, inscrites en Master 1 documentation à l'IUFM de Paris (Institut universitaire de formation des maîtres). Nous avons profité de ces conditions de travail particulières pour observer finement les interactions verbales. Les 24 heures de formation ont été distribuées sur neuf séances de deux heures en

présentiel et six séances d'une heure, proposées à distance. Les séances en présentiel ont été consacrées à la compréhension et à l'expression (écrit et oral), en s'appuyant sur le DVD et la version papier du premier épisode du script de la série *Weissensee* (Hess, 2010). Tout au long de la formation, la tâche principale, la production d'un texte pouvant remplacer ou prolonger une scène du script de film, a été présente en toile de fond. Les séances à distance ont été entièrement dédiées à des discussions métalinguistiques, menées à partir d'énigmes linguistiques ou de passages que nous avons relevés dans les premières versions de textes produits par les apprenantes. Pour communiquer à distance, nous nous sommes servies de la fonction « commentaire » de la plateforme *Edmodo* (Borg & O'Hara, 2008), des fonctions audio et clavardage (chat) de *Skype* (Microsoft Corp, 2012) et du traitement de texte collaboratif en ligne *TitanPad* (Renner, 2010). Un entretien filmé avec une didacticienne extérieure a clos le projet.

Le projet *Weissensee*, exécuté en 2012 avec trois futures enseignantes documentalistes inscrites en master 1 enseignement de l'IUFM de Paris, a bénéficié des compétences avancées des participantes dans le domaine de l'usage d'outils numériques et de banques de données en ligne. Les étudiantes ont réservé un accueil très favorable à l'idée d'explorer le corpus du projet (correspondant au script de la série *Weissensee*, Hess, 2010) à l'aide de Sketch Engine, et la prise en main de cet outil a été très rapide. Les répercussions positives de l'usage de cet outil, soutenu par la méthode d'observation collaborative de la langue appelée « discussion métalinguistique », ont pu être démontrées dans Schaeffer-Lacroix (2015b ; 2016c). Ce projet a alimenté mes activités de recherche pendant plusieurs années. Il est mentionné dans cinq publications, dont un ouvrage (Schaeffer-Lacroix, 2019b). Deux de mes articles sont entièrement construits sur les données de ce projet (Schaeffer-Lacroix, 2015b ; 2016c). J'ai présenté ses résultats à six conférences ayant eu lieu à Aix-la-Chapelle (Allemagne), Bruxelles (Belgique), Hamar (Norvège), Le Havre, Nancy et Nanterre (voir [Annexe 4](#)).

5.1.2 *Weissensee Transfer*

Le projet *Weissensee Transfer* a été conçu comme une suite à donner au projet *Weissensee*. L'idée était de faire de futures enseignantes du secondaire les ambassadrices de l'usage de corpus numériques dans le secondaire. Le public, huit étudiantes en master 1 Enseignement²⁸ (parcours allemand) et une étudiante Erasmus germanophone (future enseignante de français et de latin), a été invité à construire, en groupes de trois, un scénario d'enseignement de l'allemand faisant appel à des outils numériques. Une des activités développées par les petits groupes devait inclure l'exploration du corpus *Weissensee* (v. paragraphe [5.1.1](#)) à l'aide de Sketch

²⁸ Le groupe étant composé de sept femmes et d'un homme, le féminin « l'emporte » donc sur le masculin ici.

Engine. Le tableau 5 contient la feuille de route du module auquel était adossé ce projet.

Tableau 5 – Synopsis du module *Weissensee Transfer*.

Thème	Information	À faire :
Public	8 étudiantes germanistes, inscrites en Master 1 Enseignement + une étudiante Erasmus	
Période	19 mars – 23 avril 2013	
Lieu	IUFM de Paris, Batignolles, salle informatique	Voir si la salle informatique est réservée / disponible.
Volume horaire	6 séances à 2h30 = 15 heures	
Équipement	Ordinateurs, vidéoprojecteur, TBI (utilisé uniquement si un groupe le souhaite), dictaphone de l'IUFM (audio et vidéo)	
Formes de travail	En groupe entier ; par groupes de 3 à 4	
Modalités présence/distance	15 heures en présentiel. Échanges entre les cours possibles à l'aide de <i>Moodle</i> et <i>Edmodo</i>	
Outils de communication	<i>Moodle</i> - Dépôt de ressources (enseignante) - Dépôt de devoirs (apprenantes) - <i>Forum</i> (enseignante et apprenantes). - <i>Chat</i> pour le tutorat. <i>Edmodo</i> comme espace où les étudiantes deviennent des enseignantes potentielles. Elles y déposeront des ressources, elles créeront des activités (quiz, sondage, etc.).	
Outil de corpus	Sketch Engine	
Outils pour la recherche	Traces écrites sur forums <i>Moodle</i> et sur <i>Edmodo</i> ; enregistrements audio pour chaque groupe (avec le dictaphone IUFM et un ou deux autres dictaphones). Caméra pour les entretiens finaux.	Trouver un ou deux dictaphones.
Corpus	<i>Weissensee</i> (6 épisodes, 2010) <i>Weissensee</i> bilingue (1 ^{er} épisode) deTenTen (Sketch Engine)	Revoir la traduction (faire relire par un francophone).
Évaluation institutionnelle	Note de groupe pour le scénario - échanges en présentiel (enregistrements audio) et sur forum ;	Fixer les critères d'évaluation.

	<ul style="list-style-type: none"> - éléments du scénario rendus au milieu du cours ; - produit final. <p>Note individuelle :</p> <p>présentation d'un chapitre de Grosbois (2012) à l'oral ou à l'écrit.</p> <p>Sans note, avec retour verbal de ma part :</p> <p>journal de bord</p> <ul style="list-style-type: none"> - tenu pendant le stage d'observation du mois de février 2013, afin de pouvoir répondre à la question : « Qu'est-ce qui peut être un obstacle à l'apprentissage ? » ; - tenu individuellement pour illustrer le procédé de création de séquence. 	
Utilisation des données pour la recherche	Une décharge à faire signer.	Choisir/adapter un modèle de décharge.
Entretiens	Un entretien à la fin de la dernière séance.	

Le projet *Weissensee Transfer* s'est déroulé de façon satisfaisante. Toutefois, sa valorisation scientifique n'a pas été aussi aisée que celle du projet *Weissensee*. Son exécution a été suivie de la phase mouvementée de réforme de la formation et des concours d'enseignement de 2013. Cette période a eu raison des ressources vitales dont j'aurais eu besoin pour analyser les données dans la foulée et pour rédiger des publications et des propositions de communication qui auraient pu s'en alimenter. Comme précisé dans Schaeffer-Lacroix (2015a), l'époque était pourtant propice au thème des enseignantes utilisatrices de corpus numériques pour enseigner. En 2013, la revue *ReCALL* a publié un appel à publication d'articles mettant à l'honneur Tim Johns, figure phare de ce domaine ; le volume qui y correspond, *Researching uses of corpora for language teaching and learning*, a paru en 2014 (Boulton, Pérez-Paredes, 2014). En 2015, la revue LLT (*Language Learning & Technology*) a lancé un appel concernant le thème *Corpora in Language Learning and Teaching*, paru en octobre 2017 (Vyatkina, Boulton, 2017). J'ai, malgré les conditions difficiles, fait une proposition de publication dans *ReCALL* avec les données de *Weissensee Transfer*, mais elle a été refusée. C'est seulement en 2016 que j'ai repris l'exploitation approfondie des données, ce qui a mené, par la suite, à leur présentation à deux conférences, une lors de l'événement TaLC12 (Teaching and language corpora conference) à Gießen (Allemagne) et une au congrès RANACLÈS à Paris (voir [Annexe 4](#)). De plus, j'ai eu l'occasion d'intégrer ces données dans un chapitre de l'ouvrage édité pour Routledge en 2019 par Peter Crosthwaite (Schaeffer-Lacroix, 2019a). Ce volume présente des recherches dans le

domaine de l'interrogation de corpus par ou pour de jeunes apprenants (Crosthwaite, 2019).

5.1.3 *Kommas*

Tout comme le projet *Weissensee*, le projet *Kommas* [Virgules] a visé à faire vivre l'idée que les apprenants d'une langue étrangère peuvent être amenés à observer la langue en adoptant des méthodes qui ressemblent à celles utilisées par les experts de la langue concernée, à condition qu'on leur donne l'occasion de réfléchir à plusieurs, à partir de données mises en corpus, à un obstacle linguistique qu'ils ont besoin de surmonter. Voici le résumé de l'article s'appuyant sur les données de ce projet (Schaeffer-Lacroix, 2018d).

Le concept d'apprenant chercheur attribue à l'apprenant les traits d'un linguiste cherchant à comprendre le fonctionnement de la langue. Souvent mentionné dans le contexte de l'utilisation des corpus pour l'enseignement-apprentissage d'une langue étrangère, ce concept nécessite de la part des apprenants la capacité de formuler une question portant sur un fait linguistique et de transformer cette question en une requête sur corpus. Afin d'observer la façon dont de telles compétences peuvent se mettre en place, une expérience a été proposée à deux groupes de futurs enseignants d'allemand ayant produit des écrits académiques pendant un module de didactique. L'analyse d'erreurs appliquée à leurs textes a motivé le choix de proposer à ce public un travail de conceptualisation de la virgule. Lors d'un atelier de formation linguistique et technique d'une durée de 6 heures, les étudiants avaient l'occasion d'observer la virgule dans un corpus d'experts. Ils étaient regroupés en binômes afin d'être incités à une interaction verbale pendant leur travail d'exploration. Pour l'un des groupes, une approche déductive a été choisie et pour l'autre une approche inductive. L'analyse des données a permis de déterminer trois types d'exploration du corpus par les apprenants et d'établir un lien entre leurs modes d'investigation et le traitement spécifique auquel ils ont été soumis. Les interactions verbales et les films d'écrans contiennent des indices selon lesquels l'approche déductive a, davantage que l'approche inductive, mené à des démarches s'apparentant à une enquête scientifique.

Les données du projet *Kommas* ont été récoltées en 2016 lors d'un atelier de prise en main de Sketch Engine, qui était la version aboutie de plusieurs tentatives d'intégrer l'exploration de corpus dans un de mes cours dispensés à l'Espé de l'académie de Paris, dédié aux méthodes et pratiques en didactique de l'allemand. Les membres de deux groupes de futures enseignantes inscrites en master 1 MEEF parcours allemand devaient rédiger, en allemand ou français langue étrangère, leurs « fruits de lecture » d'un des textes faisant l'objet des séances du module et les déposer sur un espace d'écriture collaborative en ligne. De plus, chacun d'entre eux était invité à écrire un commentaire se référant à un des textes déposés par les pairs. J'ai annoté les productions en vue de les faire réviser après la tenue de l'atelier Sketch Engine. Cet atelier a combiné la prise en main technique et le thème des virgules en allemand. Il se trouve que tous les textes déposés en contenaient trop ou pas assez, ou bien les virgules étaient insérées à des endroits non pertinents. J'ai donc estimé que ce thème représentait un obstacle devant être surmonté par tous les membres du groupe. Les étudiantes ont eu l'occasion d'explorer, par

groupes de deux, les virgules dans un corpus de textes comparables, rédigés par un groupe de germanistes d'une promotion antérieure. Les écrans d'ordinateur ont été filmés à l'aide de Screencast-O-Matic, et j'ai enregistré les interactions verbales lors des explorations de corpus, mais aussi pendant le reste de l'atelier. Une bonne partie des enregistrements et l'intégralité des entretiens post-recherche ont été transcrites à l'aide de l'outil oTranscribe (Bentley, [sans date]). Les données issues de ce projet ont pu être valorisées à l'occasion d'une communication au colloque JLC (Journées internationales de la linguistique de corpus) à Grenoble en 2017 et à l'aide d'une publication dans la revue *Lidil* (Schaeffer-Lacroix, 2018d).

5.1.4 *Audiodeskription*

Le projet *Audiodeskription* a été développé en 2018 pour un public de huit futurs enseignants spécialistes d'une autre discipline que l'allemand (documentation, histoire, lettres, éducation musicale, premier degré), inscrits en master 1 MEEF de l'Espé de l'académie de Paris (maintenant appelé « Inspé ») afin de valider un niveau B2 (Conseil de l'Europe, 2001 ; 2018). L'idée était de donner aux apprenants accès à des ressources culturelles et de leur faire créer des productions verbales permettant de les entraîner le plus possible aux cinq activités langagières certifiantes (compréhension de l'écrit et de l'oral, production écrite et orale, interaction verbale). J'ai sélectionné comme support principal le film *Gundermann* (Dresen, 2018). Ce film, primé par le prix allemand *Goldene Lola* en 2019, retrace la vie intense et brève d'un célèbre chanteur engagé et ouvrier qui a marqué le paysage culturel et politique de la RDA (République Démocratique d'Allemagne).

Chacun des apprenants avait à créer un bout de l'audiodescription (AD) du générique du film *Gundermann*²⁹. Pour ce faire, il était prévu que le groupe explore le corpus Buettenwarder, contenant les scripts d'audiodescription de 69 épisodes de la série *Neues aus Büttewarder* (Eberlein, 1997).

Le tableau 6 présente les activités prévues, les compétences visées et les méthodes et/ou outils envisagés lors de la planification du projet.

²⁹ <https://www.kino.de/film/gundermann-2018/>

Tableau 6 – Planification du projet Audiodescription.

Activité	Compétence	Méthode et/ou outil
Chercher à comprendre le contenu du générique	Compréhension de l'oral	Ne pas s'arrêter à chaque détail
Découper le générique en plusieurs scènes	Compétences littéraires et techniques	Regarder le générique plusieurs fois. Délimiter les scènes en insérant des marqueurs à l'aide de VideoAnt
Répartir les scènes sur les apprenants	Compétence sociale : prendre ensemble des décisions	Outil de sondage Tricider pour décider qui prendra en charge quelle scène
Déterminer le contenu du script d'audiodescription	Empathie : présentation du générique sans montrer l'image	En binômes : une personne a vu le générique, l'autre non. La deuxième pose à la première des questions afin de comprendre l'essentiel de l'action du générique (Qui ou quoi est visible ? Où ? Quelles actions ? Quel aspect ?).
	Interaction orale : « De quelles informations a-t-on besoin pour arriver à suivre le générique sans le voir ? »	
Rédiger le script d'audiodescription	Production écrite	Linguee, DeepL, Pons
	Révision de texte et littératie numérique	deSkell, Sketch Engine
Mettre en voix le script	Oral en continu (phonologie, prosodie, accent de mot)	Naturalreaders
Enregistrer le script	Oral en continu, orthoépie	Online voice recorder
Implémenter le script sur YouDescribe	Littératie numérique	YouDescribe

Le projet n'a pas permis d'aller aussi loin que je l'aurais souhaité en termes d'utilisation de corpus pour soutenir le travail de conceptualisation et la révision des versions provisoires des textes produits par les apprenants. Toutefois, une bonne partie des étapes mentionnées dans le tableau 6, destinées à faire créer l'audiodescription d'un bout du générique de *Gundermann*, ont été prises en compte pendant l'exécution de ce module d'allemand.

Nous nous sommes servis de VideoAnt (ANON., [sans date]), un outil permettant d'annoter des passages d'un film et de commenter les annotations ajoutées. Le groupe s'en est servi pour découper le générique en plusieurs parties (v. Figure 35) et pour rédiger et réviser les scripts AD qui s'y réfèrent.

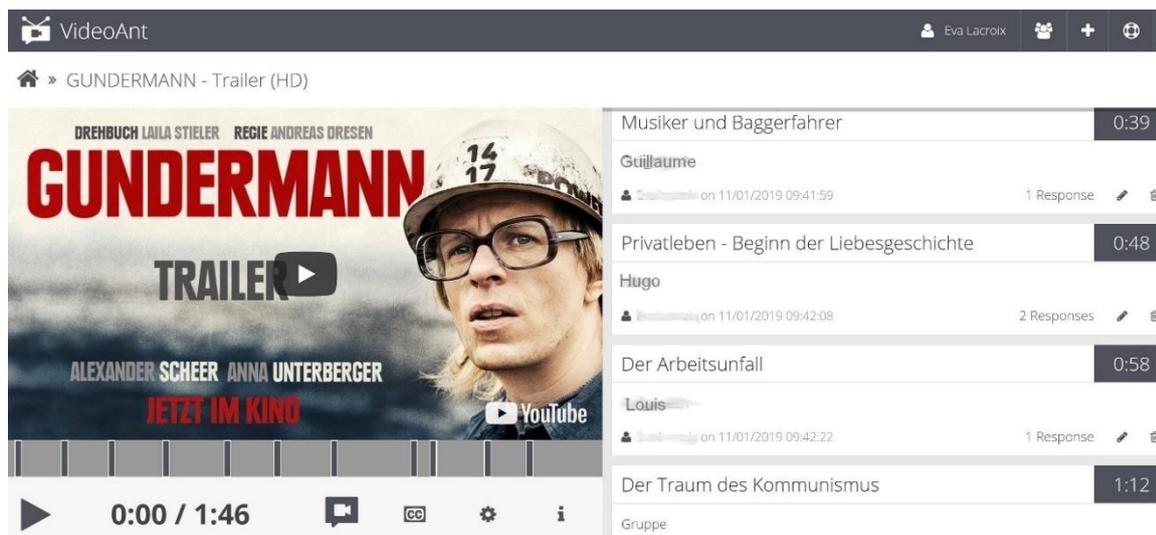


Figure 35 – VideoAnt pour annoter le générique de Gundermann.

Le tableau 7 présente le découpage en dix parties fait en fonction des choix des membres du groupe. La colonne de gauche correspond aux titres formulés par les apprenants. La colonne de droite contient ma traduction de ces titres (abréviations : « Sek. » veut dire « Sekunden », "sec." veut dire "secondes »).

Tableau 7 – Distribution de passages à audiodécrire.

GUNDERMANN – Trailer

1.10417 Sek.: Im Konzertsaal (Clara)
 9.52123 Sek.: GG und Conny im Restaurant (Clara)
 17.8812 Sek.: Im Konzertsaal (Kollektiv)
 27.5329 Sek.: GGs SED-Kandidatur (Aldina)
 38.5003 Sek.: Musiker und Baggerfahrer (Guillaume)
 47.4207 Sek.: Privatleben und Beginn der Liebesgeschichte (Hugo)
 57.5624 Sek.: Der Arbeitsunfall (Louis)
 71.7357 Sek.: Der Traum des Kommunismus (Kollektiv)
 75.161 Sek.: Im Stasiarchiv (Inga)
 81.3113 Sek.: Die Identitätskrise (Didier)
 92.3596 Sekunden: Spiegelbild (Christine)

GUNDERMANN – Générique

1 10417 sec. : Dans la salle de concert (Clara)
 9,52123 sec. : GG et Conny au restaurant (Clara)
 17,8812 sec. : Dans la salle de concert (collectif)
 27 5329 sec. : Candidature pour le SED (Aldina)
 38 5003 sec. : Musicien et conducteur de pelle mécanique (Guillaume)
 47,4207 sec. : Vie privée - Début de l'histoire d'amour (Hugo)
 57,5624 sec. : L'accident du travail (Louis)
 71 7357 sec. : Le rêve du communisme (collectif)
 75 161 sec. : Dans les archives de la Stasi (Inga)
 81 3113 sec. : La crise d'identité (Didier)
 92 3596 sec. : Image miroir (Christine)

Les étudiants ont chacun choisi un ou deux passages pour en faire l'audiodescription. À des fins de formation, deux des passages ont été annotés de façon collective avant de traiter les autres passages de façon individuelle. Christine, la documentaliste du groupe, a produit le texte suivant pour décrire la scène finale, appelée « Spiegelbild » [Image en miroir].

Gundermann und Conny sind in einem blauen Badezimmer. Sein Oberkörper ist nackt. Er guckt in den Spiegel. Conny trägt einen Pulli und einen Mantel. Sie steckt hinter ihm ihre Haare hoch. Er macht eine Grimasse.

[Gundermann et Conny dans une salle de bain bleue. Son buste à lui est nu. Il se regarde dans le miroir. Conny porte un pull et un manteau. Placée derrière lui, elle relève ses cheveux. Il fait une grimace.



Figure 36 – YouDescribe pour publier une audiodescription en ligne.

Le produit final a été implémenté en ligne, en version privée, à l'aide de l'outil de création d'audiodescriptions YouDescribe (Miele, 2019). La figure 36 présente un extrait de la production de Christine publiée à l'aide de cet outil.

Le projet *Audiodeskription* a souffert de l'instabilité du groupe. Deux des personnes ont quitté le cursus universitaire avant la fin, et les deux futurs enseignants de musique du groupe avaient des engagements professionnels les empêchant parfois de venir en cours. Toutefois, les retours des apprenants permettent de penser que le module a globalement donné satisfaction au groupe. Il a offert à des personnes ayant un niveau de départ très varié des activités pouvant correspondre à leurs besoins d'apprentissage individuels. En vue d'alimenter une prochaine édition de ce module, j'ai créé des activités supplémentaires faisant appel au corpus *Buettenwarder*. Ces activités s'appuient sur les besoins dans le domaine de la morphologie observés dans les données du projet *Audiodeskription*. Elles figurent dans Schaeffer-Lacroix (soumis), un article qui est en train d'être expertisé en vue de sa publication dans la revue *InJAL (International Journal of Applied Linguistics)*.

5.1.5 Cahier des clauses techniques particulières

L'idée d'une collaboration avec Isabelle Salengros Iguenane, un des membres du groupe RIDL (v. 3.7.2), a mûri pendant plusieurs années avant de se concrétiser en la présentation d'une conférence à deux voix au colloque PERL le 13 décembre 2019 (voir [Annexe 4](#)) et en la publication d'un article dans la revue *Le langage et l'homme* (Salengros Iguenane, Schaeffer-Lacroix, 2020). Toutes deux spécialistes de production écrite en langue étrangère, l'une pour

l'allemand, l'autre pour le français langue étrangère, nous avons réussi à conjuguer l'expertise d'Isabelle Salengros Iguenane en FOS (français sur objectifs spécifiques), allant de pair avec une connaissance approfondie du milieu des grandes écoles, et mon expertise en linguistique de corpus. J'ai eu l'occasion de côtoyer le milieu des grandes écoles grâce à une vacation en allemand langue étrangère en 2015 à l'ENSAE (École nationale de la statistique et de l'administration économique). Lors de ce cours, les étudiants ont fait appel à Sketch Engine (Kilgarriff et al., [sans date]) pour la révision de texte.

Pour un public d'élèves-ingénieurs allophones inscrits à l'École des Ponts ParisTech, Isabelle Salengros Iguenane a conçu et exécuté plusieurs fois un module d'enseignement-apprentissage dédié à la production d'un écrit professionnel, le CCTP (Cahier des clauses techniques particulières). Nous nous sommes intéressées à la possibilité d'intégrer l'exploration d'un corpus réflexif dans ce module afin d'aider les apprenants à comprendre les spécificités de ce type d'écrit et afin de leur donner l'occasion de réviser de façon éclairée leurs erreurs en français langue étrangère. Pour ce faire, nous avons constitué un corpus contenant des CCTP représentant plusieurs corps de métier.

Tableau 8 – Constitution du corpus CCTP.

Corps de métier	Nombre de documents	Nombre de mots
Aménagement	2	13 836
		4 032
Charpente	3	9 595
		10 481
		4 325
Géotechnique	2	3 310
		1 248
Gros œuvre	5	12 514
		18 442
		20 062
		50 476
		15 116
Passerelle	2	25 142
Total	14	188 579

Dans Salengros Iguenane et Schaeffer-Lacroix (2020), nous présentons notre mode d'analyse qui croise les apports d'une lecture linéaire, d'une analyse discursive et sémantique à l'aide de

Tropes (Molette, Landré, 1995) et d'une analyse quantitative à l'aide de Sketch Engine (Kilgarriff et al., [sans date]). Ce travail a mené au développement d'un scénario reposant sur les activités suivantes, impliquant une exploration du corpus CCTP :

- activités exploratoires ;
- identification de faits linguistiques exprimant l'effacement de l'agent ;
- identification de faits linguistiques à caractère anaphorique ;
- identification de faits linguistiques ayant une valeur déontique (obligation, nécessité, devoir) ;
- gloser à partir d'une forme ou d'une formule.

L'une des activités d'aide à la rédaction d'un CCTP se concentre sur la voix passive, très fréquente dans le corpus, permettant de formuler de façon neutre et objective.

La figure 37 présente une des occurrences obtenues pour la requête portant sur le lemme « être » suivi d'un participe II à une distance de 0 à 3 mots, menant majoritairement à des résultats à la voix passive.

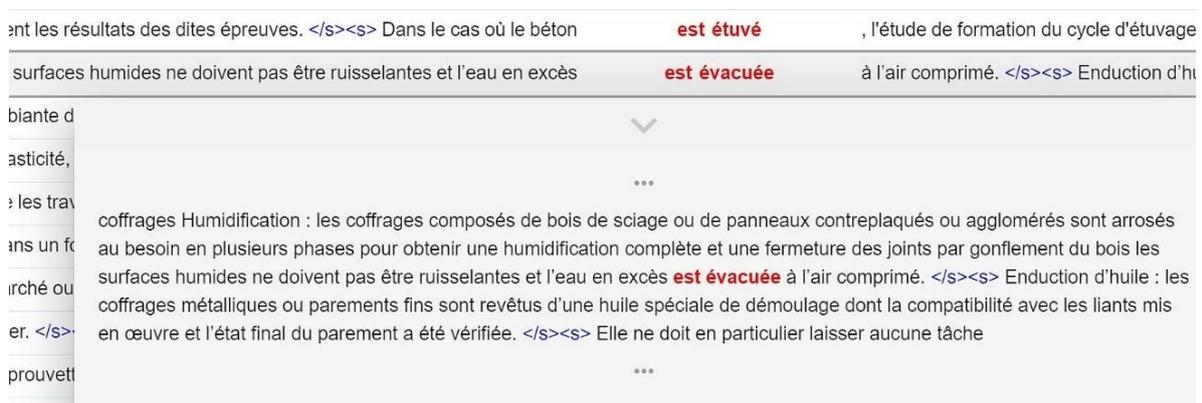


Figure 37 – La voix passive pour être neutre et objectif.

Le repérage du sujet du groupe verbal à la voix passive, en l'occurrence le groupe nominal « l'eau », est susceptible de faire comprendre aux élèves-ingénieurs la raison d'être de la terminaison féminine de la forme au participe II faisant partie de ce groupe verbal. Un clic sur le pivot (l'élément central, faisant objet de la requête) permet d'élargir le contexte, si nécessaire. Un exercice tel que celui présenté dans le Tableau 9 (sur feuille ou en ligne) peut compléter le travail de réflexion mené en commun (appelé ci-dessus « gloser ») en vue de stabiliser les apprentissages.

Tableau 9 – Feuille de travail avec lignes de concordance.

Les serrures électromécaniques de sécurité	sont prévu__	pour donner un niveau de sécurité élevé à l'exploitant.
Cette protection doit avoir	été appliqué__	avant mise en place.
L'eau en excès	est évacué__	à l'air comprimé.

D'autres types d'activités sont présentés dans Salengros Iguenane et Schaeffer-Lacroix (2020). À l'occasion de la prochaine édition du module en printemps 2020 à l'École des Ponts ParisTech, les activités sur corpus seront implémentées et documentées à l'aide d'une recherche-action lors de laquelle le corpus CCTP servira à faire observer par un groupe d'élèves-ingénieurs les caractéristiques de ce type d'écrit professionnel et à réviser les premières versions des textes qu'ils auront produits.

5.1.6 *Kleinanzeigen*

Le projet *Kleinanzeigen* [Petites annonces] sera exécuté début juin 2020, ensemble avec Caroline Rakotondraivo (professeure agrégée au collège Lamartine, Paris et formatrice académique) et Claire Laurin (professeure agrégée du second degré (PRAG) à l'Inspé et au lycée Jacques Decour, Paris), toutes les deux membres du département d'allemand de l'Inspé de l'académie de Paris. Ce projet s'adresse à un public d'élèves en classe de 4^e LV1 (quatrième année d'apprentissage de l'allemand). Son descriptif a été validé par Peter Steck, inspecteur d'académie (Paris) et inspecteur pédagogique régional (Île-de-France) d'allemand, ainsi que par les deux membres de la direction du collège Lamartine, lieu de l'implémentation du projet. Une rencontre avec le principal adjoint a eu lieu en décembre 2019, et les parents des élèves concernés sont invités à assister à une réunion d'information début février 2020.

Leçons tirées du passé

La conception du projet *Kleinanzeigen* vise à tirer les enseignements des recherches-action exécutées dans le passé. Les facteurs qui ont entravé le bon déroulement de l'un ou l'autre des projets précédents étaient le faible nombre des participants, l'instabilité du nombre des participants présents, les aléas de la technique (p. ex., enregistrements audio et/ou vidéo incomplets ou détériorés) et des choix non pertinents au niveau de la conception du projet. Les retours réservés des chercheurs ayant expertisé certains de mes articles m'ont fait comprendre

que parfois, je manquais d'une vision suffisamment claire par rapport aux questions de travail et aux méthodes et activités à mettre en place, susceptibles de permettre d'y répondre. Je pense en particulier à un article dédié au projet *Filmlokal* exécuté en classe de première au lycée du parc de Vilgénis à Massy, portant sur l'écriture d'un mini-script de film en allemand langue étrangère à l'aide d'un corpus de scripts de film. Ce texte a été refusé par les experts du comité scientifique du *Language Resources and Evaluation Journal* en raison de l'absence d'une question de recherche clairement identifiée. Pour deux autres articles que j'ai écrits, dont Schaeffer-Lacroix (soumis), des modifications majeures ont été demandées afin que l'agencement de leur contenu corresponde mieux aux exigences des organes de publication visés. Il arrive aussi qu'un article soit refusé par une revue, mais accepté par une autre. C'était le cas pour l'article que j'ai co-rédigé avec Elsa Chachkine et Françoise Demaizière (Chachkine et al., 2013), qui a été refusé par les experts du comité scientifique de *RDLC* (Recherches en didactique des langues et des cultures). Après une légère réécriture, notre article a été accepté pour publication dans *Linguistik Online*.

Afin d'éviter les écueils mentionnés ci-dessus, j'ai pris le temps de développer le projet *Kleinanzeigen* et d'attendre le bon moment pour pouvoir l'implémenter avec un groupe classe paraissant stable. Une autre condition nécessaire était le nombre d'élèves suffisamment grand, permettant de diviser la classe en un groupe recherche et un groupe témoin. De plus, j'ai fait le choix de veiller à ce que les données récoltées puissent être explorées à l'aide de l'outil de statistique R, dont l'intérêt pour l'analyse de données issues de recherches en didactique m'a été démontré lors des formations continues que j'ai eu l'occasion de suivre (voir paragraphe [2.2.1](#) et [Annexe 1](#)). Ce faisant, je poursuis une méthode d'analyse quantitative adoptée pour la première fois dans Schaeffer-Lacroix (2018d).

Motivation pour exécuter le projet

Les recherches-action dans une classe de langue du secondaire portant sur des faits linguistiques identifiés sont rares, et plus rares encore sont celles et planifiées et exécutées de façon conjointe par des chercheurs, enseignants et inspecteurs. Il semble important de documenter de façon précise et scientifiquement adéquate l'usage de ressources et outils numériques pour les besoins de l'enseignement-apprentissage d'une langue étrangère, en l'occurrence l'adjectif et les formules en allemand. Les apports des corpus numériques pour un public jeune n'ont pas encore été suffisamment précisés dans la littérature scientifique. De plus, on dispose encore de peu d'analyses en didactique des langues faisant appel à des outils de statistique textuelle.

Description du projet

Le projet s'intéresse à la façon dont les ressources et outils de corpus peuvent soutenir l'enseignement-apprentissage d'une langue étrangère, en l'occurrence l'allemand en classe de quatrième au collège. À l'occasion d'un projet d'écriture d'une annonce en ligne pour vendre, échanger ou donner un objet, l'équipe de recherche souhaite observer comment les apprenants s'emparent des ressources et outils proposés et dans quelle mesure ils ou elles en tirent profit pour concevoir, rédiger et réviser leur texte. Nous pensons que l'exploration d'un corpus numérique peut soutenir, davantage que la consultation d'un site d'annonces en ligne, la conceptualisation de faits linguistiques connus pour être appris tardivement. En principe, nous souhaitons faire explorer un corpus réflexif contenant uniquement des annonces. À ce stade, nous ne sommes pas certaines de pouvoir en disposer (v. Constitution du corpus). En attendant, nous comptons faire utiliser deSkell (ANON., 2019b), une interface de corpus créée pour des besoins d'apprentissage de l'interface de Sketch Engine. Nous avons choisi de concentrer nos recherches et observations sur les terminaisons des adjectifs (indiquant le genre, le nombre et le cas), repérées par (Diehl et al., 2000, p. 198) comme contenus d'apprentissage particulièrement complexes, acquis tardivement. De plus, nous comptons observer dans quelle mesure l'usage d'outils de corpus favorise, davantage que le travail avec un site d'annonces, le repérage et l'emploi pertinent de formules contenant des adjectifs, p. ex. « in neuwertigem Zustand » [dans un état neuf/comme neuf].



Figure 38 – Interface de deSkell.

[neuf + état
4 Il se trouve dans un état neuf.
5 Le jean a été porté, mais il est comme neuf.]

Les deux hypothèses suivantes cadrent le projet (mais l'équipe reste ouverte par rapport à d'autres apprentissages ou évolutions constatés pendant le projet).

H1 : La manipulation des données d'un corpus numérique aide, mieux que la consultation d'un site Internet tel que kleinanzeigen.de, la conceptualisation de l'adjectif par rapport à ses trois fonctions possibles (v. Tableau 10).

Tableau 10 – Fonctions de l'adjectif.

Fonction	Exemple	Traduction
épithète	<i>ein tolles Spiel</i>	un jeu génial
épithète	<i>tolle Spiele</i>	des jeux géniaux
attribut du sujet	<i>Das Spiel ist toll.</i>	Ce jeu est génial .
adverbe	<i>ein toll gemachtes Spiel</i>	un jeu génialement bien conçu

Indices pour H1 : traces d'une activité d'apprentissage invitant les élèves à identifier les fonctions des adjectifs dans les annonces (enregistrements audio des séances entières ainsi que des interactions verbales entre binômes d'élèves) ; traces des actions numériques (écrans d'ordinateur filmés) ; informations recueillies à l'aide des entretiens post-recherche.

H2 : La manipulation des données du corpus aide, mieux que la manipulation des données du site Internet, la conceptualisation de l'adjectif par rapport à ses terminaisons (motivées par la fonction, le genre, le nombre et le cas).

Indices pour H2 : scores des terminaisons pertinentes obtenus dans des textes contenant des mots inconnus (genre indiqué), proposés lors d'activités d'entraînement ; différence entre le pré-test et le post-test.

H3 : Nous nous attendons à ce que le répertoire linguistique du groupe expérimental se distingue de celui du groupe témoin en ce qui concerne le choix de l'adjectif. Il sera plus étendu, plus varié et plus pertinent.

Indices pour H3 : nombre et variété des adjectifs pertinents utilisés ; présence d'adjectifs pertinents peu fréquents ; différence entre le pré-test et le post-test.

Gestion du projet

Le projet *Kleinanzeigen* est géré à l'aide de Trello (Atlassian, 2010), une plateforme que l'équipe du parcours allemand a eu l'occasion d'utiliser et de faire utiliser dans le contexte de l'enseignement au niveau du master MEEF. Intuitif, ludique et esthétique, Trello correspond aux attentes que j'ai par rapport à un outil soutenant aussi bien le partage de documents que l'échange entre participants d'un cours ou entre partenaires d'un projet.



Figure 39 – Trello pour gérer le projet Kleinanzeigen.

Constitution du corpus

Une difficulté non résolue à ce jour est la récolte de données pour constituer un corpus réflexif contenant uniquement de petites annonces. Techniquement, il est facile et rapide de créer un tel corpus, par exemple à l'aide de l'outil BootCat (Baroni, Bernardini, 2004), accessible via Sketch Engine. Cet outil récupère de façon automatique le contenu de pages Internet ciblées, voire de sites entiers. Toutefois, je me pose des questions concernant la légalité d'une telle démarche. Est-ce que les petites annonces sont des « œuvres de l'esprit » ou pas ? Si oui, il faut obtenir l'accord des auteurs des textes ou du propriétaire du site sur lequel ils sont publiés s'il a obtenu, de sa part, l'accord des personnes ayant composé les textes. Des questions de ce type sont discutées dans Schaeffer-Lacroix (2019d).

Valorisation de la recherche

Un premier pas vers la valorisation de cette recherche a été fait en proposant une communication à AILA World Congress of Applied Linguistics, qui aura lieu à Groningen (Pays-Bas) en août 2020. Notre proposition ayant pour titre « Designing a sales ad writing project for young learners of German: Do corpus tools make a difference? » a été acceptée.

5.2 Corpus et linguistique

Noms communs de personnes dans le langage administratif suisse

Entre mars 2015 et octobre 2016, j'ai obtenu, en parallèle à mon emploi permanent à l'Espé de l'académie de Paris, un poste à temps partiel (20%) à l'université de Genève. Cet emploi a eu

comme objectif de me faire participer, en tant que linguiste de corpus multilingue (allemand, français, italien), à un projet soutenu par le FNS (Fonds national suisse), intitulé *Politique linguistique et usage de la langue dans la Suisse plurilingue: noms communs de personnes dans le langage administratif (2013-2016)*. La participation à ce projet m'a permis de m'insérer dans une équipe internationale, composée de membres allemands, italiens et suisses. Ayant pour tâche d'explorer, en fonction des questions de recherche de l'équipe, le corpus multilingue constitué avant mon arrivée (Elmiger, 2015), j'ai pu mettre en pratique les compétences développées grâce à mon travail d'apprentissage comme autodidacte et grâce aux formations continues dont j'ai pu bénéficier (voir paragraphe 2.2.1). J'ai particulièrement apprécié le thème engagé du projet, dont le descriptif est fourni sur la page d'accueil du site du département de langue et de littérature allemandes de l'université de Genève³⁰.

POLITIQUE LINGUISTIQUE ET USAGE DE LA LANGUE DANS LA SUISSE PLURILINGUE: NOMS COMMUNS DE PERSONNE DANS LE LANGAGE ADMINISTRATIF

Depuis plus de trente ans, les liens entre langage et sexes sont sujet de controverse. Le présent projet de recherche, financé par le Fonds National Suisse, se propose d'étudier comment la langue administrative au sein de la Confédération a évolué au fil des années et dans quelle mesure les langues officielles de la Suisse divergent sur ce sujet.

Le projet étudie la mise en œuvre concrète de la revendication visant la parité linguistique dans l'usage administratif, dans les différentes langues nationales. Pour ce faire, un corpus de textes est établi, constitué de textes issus de trois périodes différentes (témoignant de l'évolution des directives sur l'application du principe de l'égalité des sexes). Ce corpus fera l'objet d'analyses quantitatives (utilisation de méthodes de linguistique des corpus) et qualitatives (relevant de la linguistique du texte). Les directives légales varient selon le type de texte à rédiger, ainsi que la structure linguistique même des langues administratives en Suisse, le principe d'égalité s'exprime différemment selon les contextes.

Après la ratification de l'article constitutionnel sur l'égalité entre femmes et hommes (1981), la question d'une dénomination paritaire des femmes et hommes a pris de l'importance dans les débats politiques au niveau fédéral. Bien que des propositions dans ce sens soient apparues dès 1991 pour les langues administratives – allemand, français et italien –, celles-ci n'ont acquis un statut de directive de plein droit que pour l'allemand, en 1993. Depuis 2010, la loi sur les langues promeut la rédaction non sexiste dans les trois langues officielles, mais les pratiques réelles des administrations, fédérales ou locales, n'ont pas encore été étudiées de manière systématique.

Pour étudier l'évolution des textes officiels dans les différentes langues, le projet est articulé selon un axe diachronique: sur trois périodes différentes, du début de la discussion politique à aujourd'hui. Le projet comprend trois problématiques: quelles sont les directives institutionnelles et comment ont-elles évolué dans l'intervalle étudié? Comment ces directives sont-elles appliquées par les personnes impliquées dans la rédaction et traduction de textes et quelles attitudes peuvent être observées? Quelle variation synchronique et diachronique peut s'observer dans divers genres textuels et dans les différentes langues

³⁰ <https://www.unige.ch/lettres/alman/fr/recherche/sprachpolitik/>

officielles?

En rapport avec le débat sur la parité linguistique entre femmes et hommes, le projet vérifiera quelles sont les conséquences observables de ces mesures, quelles différences présentent les quatre langues nationales et quelles sont les réactions à une telle politique sociolinguistique dans un contexte plurilingue. De plus, le projet aura aussi un intérêt pratique: en dépouillant les différents guides, propositions et recommandations et en vérifiant leur efficacité, on pourra inférer des recommandations pour une rédaction paritaire s'appliquant à plusieurs langues.

Durée du projet: 1er mars 2013 - 30 septembre 2016³¹

Direction: Prof. Kirsten Adamzik (Université de Genève)

Corequérant: Prof. Iwar Werlen (Université de Berne)

Le rapport final, rédigé en allemand, est suivi d'un condensé d'informations en français et italien. Ci-dessous, je rappelle le volet français résumant les résultats obtenus grâce aux analyses sur corpus (Elmiger et al., 2017b, p. 164).

À quel point la formulation non sexiste est-elle liée aux noms communs de personne ? La linguistique de corpus en permet-elle une recherche efficace ?

Dans cette étude, nous avons voulu examiner la formulation non sexiste par une approche de linguistique de corpus. Cette approche a mené à des résultats absolument intéressants (cf. partie 2), tout en affichant ses limites. Mises à part les difficultés dans la définition et le répertoriage des noms communs de personne de sorte qu'ils puissent être cherchés de manière ciblée puis quantifiés (cf. introduction, chapitre 1.5), les noms communs de personne qui ont principalement pu être étudiés sont simples, en cela qu'ils ne sont constitués que d'un seul mot ou alors qu'ils contiennent un trait d'union. Une continuation impliquant des recherches supplémentaires serait néanmoins nécessaire à une description plus étoffée de la formulation non sexiste. Les mots composés, notamment dans le domaine des noms communs de personne ou collectifs, n'ont été jusqu'à présent que mal étudiés, ce qui est entre autres dû au fait que des phénomènes variés compliquent la définition de telles dénominations composées : il faut d'une part tenir compte de la fréquence des associations lexicales et d'autre part du fait que certaines désignations complexes sont consacrées d'un point de vue terminologique. Il est question en outre de la différence entre des attributs qui définissent le nom commun de personne et ceux qui le décrivent, etc. L'utilisation de noms communs de personne en tant que parties de mots composés – notamment dans l'allemand *leser·freundlich* (agréable à lire) ou *Bürger·beteiligung* (participation citoyenne) – demeure également un sujet en grande partie inexploré, tant dans les langues synthétiques (telles que l'allemand) que dans les langues analytiques (telles que les langues romanes). Finalement, l'objectivité et l'accessibilité à la recherche par des méthodes de linguistique de corpus peuvent être interrogées relativement à certaines stratégies concernant la formulation non sexiste. Il s'agit tantôt de phénomènes sémantiques complexes (p. ex. les connotations positives ou négatives), tantôt du remplacement de noms communs de personne (utilisées de manière générique) par d'autres constructions (formes tendant à l'abstraction, constructions passives, pronominalisation, etc.). Comment peut-on traiter l'« évidence négative », c'est-à-dire le fait qu'un texte évite l'utilisation de certaines formes problématiques ? Il faudrait bien alors tenter d'étudier de manière qualitative la conception, la rédaction, la correction et la traduction de textes, afin de pouvoir reconstituer les procédés rédactionnels et les changements de formulation en comparant diverses versions des mêmes textes.

³¹ Le projet a été prolongé jusqu'au 31 octobre 2016.

5.2.1 Constitution de l'équipe de recherche

J'ai intégré le projet FNS en cours pour une période de 20 mois. Avant mon arrivée, l'équipe avait recruté deux personnes qui n'ont pas pu rester jusqu'au bout du projet. Ce fait m'a aidé à prendre conscience de la fragilité potentielle d'une équipe de recherche et de la nécessité de prendre le temps pour choisir ses collaborateurs, que, dans l'idéal, on aura rencontrés dans d'autres contextes avant le début du projet. Deux exemples liés à d'autres projets corroborent cette constatation : en printemps 2019, j'ai conçu, ensemble avec des membres de quatre établissements supérieurs différents, dont un en Allemagne, un projet appelé *Norme(s) grammaticale(s) à la lumière des corpus numériques : apprentissage, adaptation et dépassement*, que nous avons proposé au CIERA (Centre interdisciplinaire d'études et de recherches sur l'Allemagne), dans le cadre de l'appel à projets PFR (Programme de formation-recherche). Il s'est avéré difficile de créer une synergie au sein de l'équipe. Les disponibilités et les attentes par rapport aux modes d'exécution du projet étaient très inégales, ce qui est peut-être une des raisons pour lesquelles le projet n'a pas été retenu. Un deuxième exemple reflétant les difficultés qui peuvent se présenter est celui du projet qui devait être soumis par une équipe dirigée par Michael Beißwenger (Universität Duisburg-Essen) et Ciara Wigham (université Clermont Auvergne) dans le cadre des actions COST (*European Cooperation in Science and Technology*). Leur projet, dans lequel je devais être impliquée, portait sur la coopération entre des spécialistes des langues étrangères à l'université et un partenaire industriel. Il a été abandonné en octobre 2019, à un stade très proche du but, en raison de nouvelles responsabilités administratives très lourdes qu'a dû endosser Michael Beißwenger.

5.2.2 Conception de la recherche

La conception du projet FNS a prévu trois volets, évoqués dans l'extrait de fiche de poste ci-dessous.

Gegenstand ist die seit über dreissig Jahren kontrovers diskutierte sprachliche Gleichbehandlung der Geschlechter. Das Projekt befasst sich mit der Frage, wie sich (sprach)politische Massnahmen (Gleichstellungsartikel, Sprachengesetz) in einem föderalistischen, mehrsprachigen Bundesstaat konkret auf den amtlichen Sprachgebrauch in den vier Schweizer Landessprachen auswirken und welche Vorgaben (z. B. Leitfäden) bei der Texterstellung und -übersetzung beachtet werden (müssen). Dafür wird ein diachrones Korpus verschiedener Textsorten erstellt, das sowohl quantitativ (korpuslinguistisch) als auch qualitativ (textlinguistisch) ausgewertet wird; daneben werden Interviews mit verschiedenen Schlüsselpersonen durchgeführt.

[La recherche porte sur l'égalité linguistique des sexes, qui fait l'objet d'un débat controversé depuis plus de trente ans. Le projet cherche à analyser dans quelle mesure les instruments de la politique (linguistique), dont les articles sur l'égalité et la loi sur les langues, d'un État

fédéral multilingue ont un impact concret sur l'utilisation des langues officielles dans les quatre langues nationales suisses et quelles directives (par ex. les guides de formulation) doivent être respectées pour la rédaction et la traduction des textes. Pour ce faire, un corpus diachronique de différents types de textes est produit, qui est évalué à la fois quantitativement (corpus linguistique) et qualitativement (texte linguistique) ; en outre, des entretiens sont menés avec diverses personnes clés.]³²

Il s'agissait de faire des analyses empiriques appliquées à un grand corpus diachronique trilingue interrogeable à l'aide de CQPweb (Lancaster University, [sans date]).

Tableau 11 – Composition du corpus Bundesblatt/Feuille fédérale/Foglio federale.

Bundesblatt	Feuille fédérale	Foglio federale
203 585 806 tokens	239 125 036 tokens	85 223 085 tokens
65 671 textes	62 366 textes	21 400 textes

Ce corpus contient des documents publiés par la Chancellerie suisse entre 1849 et 2014, composés de genres textuels administratifs variés. Sa composition est décrite dans Schaeffer-Lacroix (2016b).

Les trois volets du corpus du projet contiennent l'intégralité des feuilles fédérales publiées depuis 1849 en allemand et français et depuis 1971 en italien (Chancellerie fédérale, 1849-2014). Elles paraissent à un rythme hebdomadaire. On y trouve, entre autres, les rapports du Conseil fédéral aux Chambres, les arrêtés et les lois qui ont été votés par les Chambres, les décisions de la Chancellerie fédérale en matière d'initiatives populaires ou de référendums et les notifications des unités de l'administration fédérale et des tribunaux (Chancellerie fédérale, 2015).

De plus, des méthodes d'analyse de linguistique textuelle (Adamzik, Alghisi, 2015), de linguistique générale (Elmiger, 2014) et de lexicologie (Schaeffer-Lacroix, 2016b) ont été appliquées aux données.

Un troisième volet a complété ces recherches, s'intéressant aux discours des personnes responsables de la rédaction de textes administratifs pendant la période couverte par le projet. Des entretiens ont été menés et transcrits par Verena Tunger (Elmiger et al., 2017b ; 2019).

5.2.3 Valorisation de la recherche

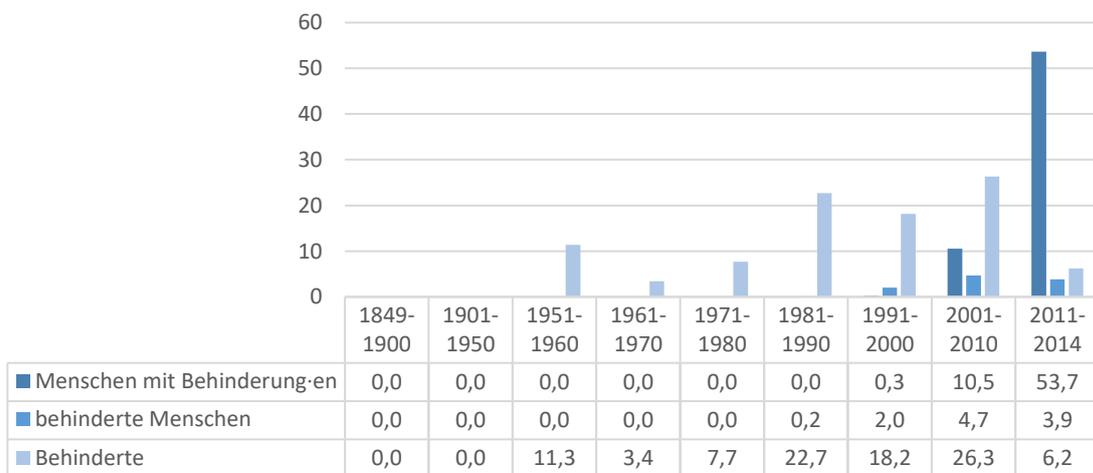
Le projet FNS m'a appris à penser le travail de recherche en termes de rentabilité pour la diffusion de la recherche : il s'est avéré important de prendre des notes et rédiger de façon régulière en ayant en tête la possibilité de transformer les écrits en contenus d'articles, de conférences et du rapport final. De plus, j'avais un rendez-vous hebdomadaire par

³² Traduction assistée par www.DeepL.com/Translator (version gratuite).

visioconférence avec Daniel Elmiger, l'un des porteurs du projet, ce qui m'a aidée à travailler de façon très régulière. Ces modes d'organisation ont permis à l'équipe d'être très productive au niveau de la diffusion des résultats. En janvier 2016, j'ai participé à un panel de l'équipe proposé au colloque Vals-Alsa à Genève, et en juin 2016, j'ai présenté des résultats du projet à JADT (Journées internationales d'analyse statistique des données textuelles) à Nice (voir [Annexe 4](#)). En lien avec le projet suisse, j'ai rédigé ou corédigé cinq publications (Schaeffer-Lacroix, 2016b ; Elmiger et al., 2017a ; Alghisi et al., 2017 ; Elmiger et al., 2017b ; 2019).

5.2.4 Perspectives ouvertes par la participation au projet FNS

La participation au projet suisse a eu des répercussions positives sur mon évolution de chercheuse. La collaboration étroite avec Daniel Elmiger m'a aidée à progresser dans le domaine de la linguistique allemande et française et de publier un article (relu par Daniel Elmiger et Verena Tunger) qui relève de ce domaine (Schaeffer-Lacroix, 2016b). De plus, j'ai appris à maîtriser CQPweb (Lancaster University, [sans date]) et j'ai perfectionné mes techniques d'interrogation de corpus.



Graphique 6 – Menschen mit Behinderung-en, behinderte Menschen, Behinderte.

Les analyses portant sur l'identification de noms de personnes m'a rendue sensible non seulement à la cause des femmes, mais aussi à la cause des personnes porteuses de handicap. Le graphique 6 retrace un des résultats que nous avons obtenus dans ce domaine (Elmiger et al., 2017a, p. 51). On peut constater que dans le *Bundesblatt*, depuis 1991 (et encore plus depuis 2011), le nom « Behinderte » [handicapés] est secondé, voire remplacé, par les mots composés

« behinderte Menschen » [humains³³ handicapés] et encore plus souvent par « Menschen mit Behinderungen » [humains porteurs de handicap]. Dans le rapport final du projet, la poursuite des recherches dans le domaine du handicap est suggérée de façon explicite (Elmiger et al., 2017a, p. 164).

Il serait certainement intéressant de comparer la question de l'égalité des genres dans le langage avec d'autres thématiques linguistiques, telles que l'intelligibilité du langage administratif (par exemple les textes « en langue facile à lire ») ou les dénominations de personnes avec un handicap : trouve-t-on dans ces domaines plus ou moins d'harmonie entre l'interprétation d'une institution et celle des autres ?

Cette prise de conscience par rapport aux désignations de personnes en situation de handicap m'a dirigée vers un domaine de recherche supplémentaire que je cultive depuis août 2017, à savoir la mise en corpus de scripts d'audiodescription dans le but d'en faire un outil d'enseignement-apprentissage des langues étrangères (Schaeffer-Lacroix, soumis) (v. paragraphe 5.1.4) et un support pour la traduction (v. section 5.3).

5.3 Corpus et traduction

Les corpus numériques au service de la traduction d'audiodescriptions

5.3.1 Mise en place d'une collaboration scientifique franco-allemande

Mon intérêt pour les questions inclusives a motivé mon choix de fonder ma demande de CRCT (Congé pour recherches ou conversions thématiques) pour l'année universitaire 2019/2020 sur ce thème. J'ai obtenu un congé d'une durée d'un semestre, ce qui m'a permis de mettre en place une collaboration scientifique franco-allemande portant sur la traductibilité de scripts d'audiodescription. En juillet 2019, j'ai pris contact avec Nathalie Mälzer, professeure des universités à l'*Institut für Übersetzungswissenschaft und Fachkommunikation* de l'université de Hildesheim (Allemagne), connue pour son expertise en audiodescription et en traduction. Sa réaction positive m'a encouragée à faire une demande de bourse dans le cadre du programme DAAD³⁴ – Sorbonne Université « Missions de recherche pour enseignants-chercheurs de Sorbonne Université (Volet 2) », qui m'a été accordée pour un séjour de recherche au sein de

³³ La traduction n'est pas heureuse (« humains » est masculin), mais elle permet de rester près du sens du nom « Menschen ».

³⁴ DAAD (Deutscher akademischer Austauschdienst) [German Academic Exchange Service]

cet institut pendant le mois de janvier 2020. Le programme du séjour inclut ma participation à des réunions des membres de l'institut et à des séminaires sur l'audiodescription et sur la communication spécialisée. J'aurai l'occasion d'assister, au *Theaterhaus Hildesheim*, à une pièce de théâtre avec mise à disposition de lunettes intelligentes, qui sont « des dispositifs informatiques portés comme des lunettes traditionnelles qui utilisent le principe de réalité augmentée pour afficher des informations en temps réel devant le champ de vision de l'utilisateur »³⁵. De plus, il est prévu que je rencontre des représentants du *Blindenverein Hannover* [association des personnes aveugles de Hanovre]. Des visites des entreprises *Video to voice* et *ARD text* à Berlin et du département de l'accessibilité du NDR (Norddeutscher Rundfunk) à Hambourg sont planifiées. Ce sont des établissements spécialisés en production d'audiodescriptions.

Ce séjour doit également servir à affiner notre projet de recherche. Nous avons décidé de mettre en place une recherche s'intéressant au rôle des corpus numériques pour le développement d'une méthode de traduction semi-automatique d'audiodescriptions de séries télévisées, en allemand et en français dans un premier temps et impliquant d'autres langues dans l'avenir. Dans le paragraphe [5.3.2](#), je reprends un choix parmi les éléments soumis au DAAD lors de ma demande de bourse.

5.3.2 Description du projet

Le projet *Les corpus numériques au service de la traduction d'audiodescriptions* (en anglais référencé sous *TVseriesAD Translation Project*) vise à optimiser et à rationaliser la création semi-automatique d'audiodescriptions multilingues à l'aide de corpus comparables.

Contexte du projet

L'équipe de recherche Paris-Hildesheim considère la traduction augmentée³⁶ d'audiodescriptions (AD), faisant appel aux corpus numériques comparables et à des outils de

³⁵ Définition trouvée sur Wikipedia : https://fr.wikipedia.org/wiki/Lunettes_intelligentes

³⁶ Définition du terme de "traduction augmentée" par Lommel : "A form of human translation carried out within an integrated technology environment that provides translators access to sub-segment, adaptive machine translation (MT) and translation memory (TM), terminology lookup, and automatic content enrichment (ACE) to aid their work, and that automates project management, file handling, and other ancillary tasks." <https://www.thelanguageoflocalization.com/2018/05/09/term-of-the-week-augmented-translation/>

traduction assistée par ordinateur, comme l'un des moyens de rendre les audiodescriptions de documents multimédias (p. ex. les séries télévisées) accessibles à un large public tout en maîtrisant les coûts et la qualité de la production. L'équipe s'attend à ce que l'utilisation de corpus AD comparables (c'est-à-dire des corpus dont les données contiennent des données dans au moins deux langues différentes, s'inscrivant dans le même genre textuel) et l'utilisation d'outils de traduction qui fonctionnent avec des modèles de réseaux neuronaux (p. ex. DeepL GmbH, 2017) soutiennent de façon satisfaisante les tâches de localisation (adaptation culturelle et prise en compte des besoins des utilisateurs).

Questions de recherche abordées

Quel rôle les corpus numériques peuvent-ils jouer dans le processus de création de scripts d'audiodescription ? Dans quelle mesure contribuent-ils, en combinaison avec des outils de traduction assistée par ordinateur (TAO), à relever les défis culturels, linguistiques, techniques et pragmatiques de la traduction de scripts d'audiodescriptions ? Peut-on et doit-on répondre à la demande de certains usagers qui souhaitent que l'audiodescription soit chargée d'émotion (par opposition à une audiodescription à teneur objective)³⁷ ? Si oui, comment ? Dans quelle mesure l'émotion peut-elle être identifiée par des outils de corpus et/ou de TAO ?

Résultats escomptés

Nous comptons analyser des corpus AD dans le but d'identifier les propriétés linguistiques récurrentes, régulières et normalisées des scripts d'audiodescriptions. De tels éléments ont des chances de pouvoir être automatisés et rationalisés pendant le processus de traduction. Le projet devra en plus permettre d'évaluer dans quelle mesure l'intervention humaine a un rôle à jouer pour traduire et adapter des éléments culturels, émotionnels ou créatifs.

Éléments étudiés

Nous comptons mener des recherches linguistiques et technologiques sur le thème de la traduction augmentée de scripts AD. Une recherche empirique sur l'acceptation des audiodescriptions par les utilisateurs est également envisageable. Nous nous concentrons sur des corpus comparables en allemand ou en français qui contiennent des scripts

³⁷ "Literal description is too boring. I don't like boring voices, so I need emotional connection as well." (Cavallo & Fryer, 2018, S. 21).

d'audiodescriptions de séries télévisées. Ces corpus seront soumis à une analyse linguistique et préparés comme modèles pour entraîner un outil de traduction assistée par ordinateur (cf. projet européen *Sumat* (Sous-titres comme données modèles pour améliorer la traduction assistée par ordinateur)³⁸). Le projet pourrait évoluer vers une REI (*recherche exploratoire et innovation*³⁹) et alors impliquer un partenaire industriel. Ce dernier aurait pour tâche d'améliorer ou de développer un outil de traduction de type NMT (*neuronal machine translation*) adapté aux besoins que nous aurons identifiés.

Étapes de travail

1. Recueillir des informations sur les projets de recherche déjà réalisés, couvrant deux ou trois des thèmes prioritaires que nous avons choisis : audiodescription, corpus, traduction. Choisir les résultats sur lesquels nous pouvons nous appuyer.

2. Dresser l'état de l'art dans le domaine de la création et/ou la traduction d'audiodescriptions, faisant appel aux corpus et/ou aux outils de traduction automatique.

- Feuille Excel partagée dans laquelle nous listons et commentons des publications portant sur l'audiodescription, les corpus et la traduction ;

- groupe *Zotero* « Audiodescription ».

3. Explorer les pratiques existantes

Prise de contact avec les institutions culturelles et/ou les entreprises suivantes : *NDR* (Hambourg), *ARD text* et *Video to voice* (Berlin), *Eclair-digital* (collaboration avec *Arte* et *Netflix*⁴⁰) : interviews sur les pratiques, méthodes et outils existants. Quel rôle jouent les outils numériques lors de la création et la traduction de scripts d'audiodescription ?

4. Analyse contrastive d'audiodescriptions en allemand et en français

- Création de corpus AD allemands et français de séries télévisées ; collecte, traitement et

³⁸ <https://cordis.europa.eu/project/id/270919/fr>

³⁹ Cf. le programme de recherche ANR "Laboratoires communs organismes de recherche publics - PME / ETI (LabCom)" <https://anr.fr/fr/detail/call/laboratoires-communs-organismes-de-recherche-publics-pme-eti-labcom-6/>

⁴⁰ <https://eclair.digital/fr/partenariats-certifications>) ; <https://np3.netflixstudios.com/audio-description>

- annotation des données ; analyse des particularités linguistiques et culturelles des scripts AD ;
- identification des éléments difficiles à appréhender à l'aide d'un outil de traduction de type NMT et proposer des solutions ;
 - traduction semi-automatique d'un échantillon de scripts AD allemand/français – français/allemand ;
 - recherche empirique pour tester la pertinence de nos traductions des scripts AD.

État actuel de la recherche et l'état de l'art

Depuis 2007, le thème de l'audiodescription a suscité des recherches dans diverses disciplines (santé, sciences économiques et sociales, linguistique appliquée, didactique, arts, sciences et technologies). Les publications proviennent principalement d'Asie, d'Europe de l'Est et d'Europe du Sud (notamment d'Italie et d'Espagne). Le secteur du cinéma et de la pédagogie du musée est bien couvert. Il existe encore peu de littérature sur les audiodescriptions de séries télévisées. Dans le secteur de la traduction et dans celui de la linguistique des corpus, les combinaisons linguistiques « allemand-français/français-allemand » n'ont pas encore été beaucoup documentées. Il est prévu de repérer et d'évaluer des publications scientifiques contenant au moins deux des thèmes suivants : audiodescription, traduction (machine et/ou humaine), linguistique du corpus.

Théories/travaux de recherche sollicités

Notre projet s'inspire de la proposition de Jankowska, Milc, & Fryer de (2017) de traduire les scripts d'audiodescription directement en langue étrangère à partir de scripts existants, créés dans la langue du document, au lieu d'en produire de nouveaux de façon indépendante. Nous souhaitons combiner cette procédure avec des techniques de traduction automatique, en particulier celles faisant appel à des modèles de réseaux neuronaux (Bentivogli et al., 2018 ; Popović, 2018). Ces techniques pourraient être améliorées à l'aide du contenu de corpus comparables (Rapp, Sharoff, & Zweigenbaum, 2016). La description linguistique des corpus peut s'inspirer des travaux de Braun (2007, 2011), Arma (2011) et de Salway (2007) ainsi que des méthodes et résultats de Reviere (2018) correspondant à une recherche empirique basée sur des corpus néerlandais. Nos recherches sur l'acceptation des traductions obtenues par les utilisateurs se baseront, entre autres, sur des questionnements formulés dans Mälzer-Semlinger (2012). Notre projet s'inspirera de la méthode adoptée dans Mälzer et Wünsche (2018) pour des recherches portant sur l'acceptation par des enfants de sous-titres de programmes télévisés.

Collecte des données

Il est prévu de collecter et d'explorer des corpus de scripts d'audiodescriptions. Nous disposons de scripts d'audiodescriptions de trois séries télévisées allemandes pour lesquels nous avons négocié les droits (librement utilisables pour l'enseignement et la recherche) : *Neues aus Büttenwarder* (Eberlein, 1997), *Dahoam is dahoam* (ANON., 2007) et les six épisodes de la mini-série *Eden* (Berger et al., 2019). *Eden* a été tourné en français et synchronisé en allemand et en d'autres langues. Les audiodescriptions ont été créées pour le français et l'allemand. Ces données serviront pour faire des analyses linguistiques permettant d'identifier les différences entre les deux langues et pour définir les défis à relever soit par la machine, soit par des humains, lors de traductions de scripts d'audiodescriptions pour l'allemand et le français. Nous avons l'intention d'acquérir d'autres ensembles de données, en particulier des scripts AD de séries ayant un intérêt pour un public au niveau international. En outre, il est prévu de collecter les données empiriques suivantes, qui feront l'objet d'une évaluation qualitative et quantitative : entretiens oraux avec des experts en AD ; questionnaires sur la pratique de la création d'audiodescriptions et de la traduction de scripts AD ; recherche empirique sur les traces numériques que laissent des utilisateurs et sur leurs réactions verbales face aux traductions ADs produites par l'équipe Paris-Hildesheim.

Planning du projet

Tableau 12 – Planning de TVseriesAD Translation Project.

Période	Actions
À partir d'août 2019	Collecte de références bibliographiques pour dresser l'état de l'art de la recherche
8.-9.10. 2019	Participation d'Eva Schaeffer-Lacroix au META Forum <i>European Language Grid</i> , Bruxelles
17.10.2019	Participation d'Eva Schaeffer-Lacroix à l'atelier Digit_Hum, « Les humanités numériques en langues », École Normale Supérieure, Paris
3.-31.1.2020	Séjour de recherche d'Eva Schaeffer-Lacroix à l'Université de Hildesheim
Février ou mars 2020	Rencontre avec l'équipe AD de <i>Arte</i> pour échanger autour des pratiques de production d'audiodescriptions pour la mini-série <i>Eden</i> .
Février - mars 2020	Formulation de la proposition de recherche
Mars 2020	Soumissions envisagées pour financer le projet recherche : <u>DFG-ANR SHS 2020</u> <u>ELG Pilot Projects (European Language Grid)</u>

5-6 juin 2020	Présentation d'une conférence au colloque <i>3rd Swiss Conference on Barrier-free Communication</i> à Winterthur
7-8 septembre 2020	Présentation d'une conférence au colloque UCCTS 2020 Using Corpora in Contrastive and Translation Studies (6th edition) à Bolzano
26-27 janvier 2021	ARSAD 2021 Advanced Research Seminar on Audio Description à Barcelone (Espagne)
Janvier 2021	Tester les outils et les méthodes de traduction
Février - avril 2021	Traduction automatique et humaine d'un échantillon de données
Mai 2021	Recherche impliquant le terrain (actions et réactions d'un échantillon d'utilisateurs par rapport aux traductions de nos traductions de scripts d'audiodescription)
Avril 2021 - février 2022	Rédaction d'articles et présentations à des conférences
Mars 2022	Organiser une conférence
Avril - décembre 2022	Publication d'articles
(En fonction des calendriers institutionnels)	Contributions à des maquettes de masters portant sur l'inclusion et/ou l'audiodescription
Janvier à mars 2023	Rédaction du rapport de recherche

5.3.3 Constitution de l'équipe de recherche

Le noyau de l'équipe est représenté par Nathalie Mälzer et moi-même. L'implication dans le projet de Saskia Josephine Schulz (spécialiste d'inclusion) et de Sylvia Jaki (spécialiste de linguistique générale et outillée et des sciences des médias), toutes les deux de l'université de Hildesheim, est également prévue. En lien avec le *TVseriesAD Translation Project*, j'ai accepté de co-diriger, ensemble avec Nathalie Mälzer, le mémoire de master 2 de Léana Debris, inscrite en MuM (Medientext und Medienübersetzung) [Sciences et traduction des médias] à l'université de Hildesheim. Ce mémoire porte sur les caractéristiques linguistiques des audiodescriptions en langue française et de leur traduction en allemand de la mini-série *Eden* (Berger et al., 2019). Ses résultats seront intégrés dans notre projet. De plus, nous avons l'intention de confier à un stagiaire en licence ou master de traduction multilingue des tâches d'annotation à l'aide de XML-Tei (Burnard, 2014) des données que nous avons récoltées. De notre côté, nous offrirons à ce stagiaire la possibilité de s'insérer dans un projet de recherche international correspondant à sa spécialisation.

5.3.4 Gestion du projet

En termes de communication, de planification et du développement d'une vision générale, les expériences que j'ai pu faire lors du projet FNS *Noms communs de personnes dans le langage administratif suisse* m'ont été très utiles. L'un des modes de gestion du FNS que je souhaite reproduire est la régularité des échanges entre membres d'une équipe de recherche, avec des dates fixées à l'avance et une définition claire des tâches à traiter par chacun et chacune. La gestion du projet est, comme pour le projet *Kleinanzeigen*, assistée par l'outil *Trello*.

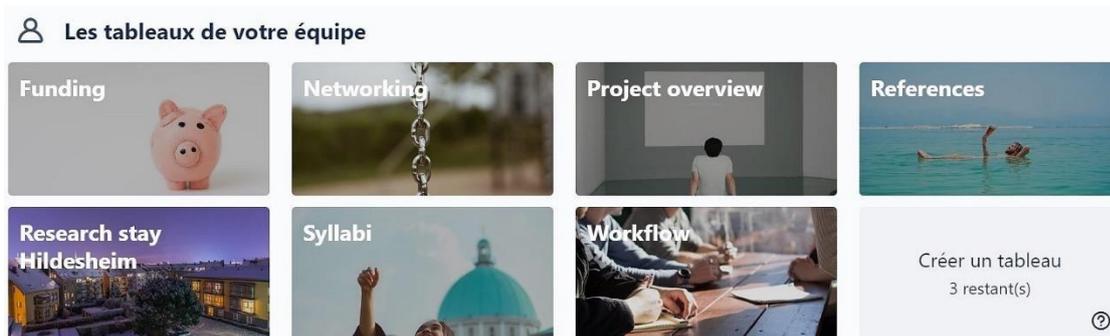


Figure 40 – Trello pour gérer le TVseriesAD Translation Project.

Les fruits de nos recherches seront documentés dans un fichier Excel partagé et dans notre espace groupe Zotero. Nous les discuterons lors de nos rencontres en présentiel et à distance et dans notre espace de gestion de projets Trello. Nos résultats seront présentés lors de conférences et ils mèneront, nous le souhaitons, à des publications.

5.3.5 Constitution du corpus

La première étape du projet a été la récolte de données pour la constitution du corpus. En novembre 2019, Nathalie Mälzer a pris contact avec son réseau scientifique « inclusion » en vue d'obtenir des scripts d'audiodescription d'une série susceptible d'intéresser un public international. Son initiative a mené à l'établissement d'une collaboration avec Philip Klenk, *Beaufragter für Barrierefreiheit* [responsable des questions d'accessibilité] de la chaîne SWR (Südwestrundfunk) à Stuttgart, qui a proposé la mise à disposition des scripts d'audiodescription de la version allemande de la mini-série *Eden* (Berger et al., 2019), tournée en français. De plus, il nous a mis en relation avec l'équipe de la chaîne de télévision *Arte* responsable des scripts d'audiodescription de la version originale française de cette série. Une rencontre entre l'équipe de *Arte* et des représentantes de notre équipe (Nathalie Mälzer, Saskia Josephine Schulz et moi-même) est en train d'être organisée.

5.3.6 Récolte de fonds

En octobre 2019, j'ai assisté au META Forum à Bruxelles⁴¹. Cet événement a renseigné les participants sur les possibilités d'utilisation de la plateforme *European Language Grid* (ELG Consortium, 2019), financée par le programme européen de recherche Horizon 2020. Un appel à projets sera ouvert en mars 2020⁴² pour subventionner des recherches d'une durée de 9 à 12 mois, commençant en juillet 2020. Nous comptons répondre au volet b) de à cet appel, qui concerne la subvention de recherches dans le domaine de développements informatiques s'appuyant sur des technologies mises à disposition sur la plateforme *European Language Grid*. De plus, nous avons l'intention de répondre à l'appel à projets franco-allemand en sciences humaines et sociales DFG-ANR SHS 2020, ouvert depuis le 12 décembre 2019. Il semble également justifié de solliciter une subvention de la part du consortium Corli, qui offre tous les ans des financements servant à rémunérer des vacataires spécialistes de création et de structuration de corpus.

5.3.7 Valorisation de la recherche

Un premier pas vers la dissémination de nos recherches a été fait par notre proposition de communication soumise pour l'événement *3rd Swiss Conference on Barrier-free Communication*, qui aura lieu début juin 2020 à Winterthur (Suisse). Deux autres événements paraissent centraux pour notre projet, à savoir la conférence *UCCTS 2020 | Using Corpora in Contrastive and Translation Studies (6th edition)* à Bolzano (Italie) en septembre 2020 et la conférence *ARSAD 2021 | Advanced Research Seminar on Audio Description*, qui aura lieu à Barcelone (Espagne) fin janvier 2021.

⁴¹ <https://www.european-language-grid.eu/meta-forum-2019/>

⁴² <https://www.european-language-grid.eu/open-calls/>

6. Les évolutions visées

Comment et vers où évoluer dans les années à venir ? Je souhaite continuer à travailler sur des thèmes s'intéressant aux besoins de personnes et continuer à découvrir des outils numériques permettant de soutenir la recherche et la pratique dans le domaine des langues étrangères.

Une des évolutions que je poursuis est la création et la structuration de corpus contenant au moins un million de mots, voire beaucoup plus. Cela nécessite la collaboration avec des équipes de spécialistes en linguistique ainsi que l'échange avec des spécialistes en TAL et en statistique. Cette évolution est déjà en marche : des contacts avec des interlocuteurs et interlocutrices de chaînes de télévision allemandes m'ont permis de récolter des données dans le domaine de l'audiodescription, données qui sont en cours de structuration et de préparation pour leur mise en corpus. La collaboration avec un stagiaire que nous comptons recruter, suivant de préférence le parcours *Traitement numérique multilingue* à l'Institut National de Langues et Civilisations Orientales de Paris, pourra également faire avancer les choses dans ce domaine.

Une deuxième évolution désirable serait l'exploration plus intense du champ de la médiation, dans le sens qui est attribué à ce terme dans le volume complémentaire du CECR (Conseil de l'Europe, 2018, p. 34) :

la médiation combine la réception, la production et l'interaction. Par ailleurs dans de nombreux cas, on ne se sert pas d'une langue uniquement pour communiquer un message, mais plutôt pour développer une idée par le moyen de ce qu'on appelle la 'mise en langage' (parler d'une idée et ainsi exprimer ses pensées : 'languaging' en anglais) ou pour susciter la compréhension et la communication.

Plus loin (*ibidem*), il est précisé que le champ d'intervention de la médiation dépasse celui de la communication interlangue, ce qui rappelle le domaine des besoins dits « spécifiques ».

- ▶ Faire que la communication soit possible entre des personnes qui, pour une raison quelconque, ne peuvent pas communiquer directement.
- ▶ Agir comme intermédiaire entre des interlocuteurs incapables de se comprendre directement, en principe des locuteurs de langues différentes (mais pas nécessairement).

La médiation peut s'inscrire dans le domaine des productions audiovisuelles, qui offrent un éventail d'activités oscillant entre traduction, description et commentaire : création de sous-titres, d'audiodescriptions, synopses ou critiques de films.

Un deuxième domaine qui m'attire en raison de son fort potentiel sociétal et éducatif est celui de l'inclusion, que déjà abordé sous l'angle de l'audiodescription et que j'aimerais poursuivre en m'intéressant au thème du langage simplifié. Ce dernier commence à être pris en compte par

les administrations gouvernementales dans des pays francophones et germanophones, comme en témoignent le guide d'accompagnement du référentiel général d'accessibilité pour les administrations (Gouvernement de l'état français, 2017) et le guide *Leichte Sprache* publié par le Bundes-Ministerium für Arbeit und Soziales (2014). La réécriture, une des activités pratiquées dans ce secteur, peut être effectuée par des auteurs, mais aussi par de futures enseignantes et par des chercheuses. Là aussi, les outils numériques peuvent avoir un rôle à jouer.

Conclusion

Dans cette synthèse, je me suis intéressée à la didactique des langues, la linguistique de corpus et la traduction, en y appliquant la grille de lecture « outillage ». Ces trois domaines semblent tous avoir besoin de justifier leur raison d'être par rapport à d'autres domaines ou disciplines universitaires établis. Dans l'acte d'enseigner ou de traduire et dans le geste d'exploration de données de corpus, le vécu de l'humain est pris en compte ; ses gestes et paroles y prennent corps. Ce lien avec « le réel » rend les objets d'étude complexes et parfois difficiles à saisir. En même temps, il offre une densité toujours surprenante et jubilatoire. J'ai tenté de montrer que le « penchant pratique » de ces trois domaines ne les empêche pas de jouer un rôle dans le monde de la recherche.

Il est intéressant de voir que la situation de la didactique des langues, telle qu'elle est présentée dans l'introduction de ce rapport de synthèse, ressemble en certains points à la description de la situation de la traduction proposée par Boisseau et Chuquet (2009).

La situation de la traduction à la croisée de disciplines diverses, les nombreuses compétences requises des traducteurs, la variété toujours plus grande des genres que l'on traduit – de l'ouvrage de sciences humaines au site internet, de la série télévisée aux œuvres majeures de la littérature, du mode d'emploi au texte juridique... – rendent sans doute illusoire et impossible toute approche globale et généralisante de cette activité protéiforme et des produits qu'elle engendre. Chacun l'abordera donc sous un angle qui lui est propre, à l'aune de ses intérêts et de ses compétences, en privilégiant la pratique ou la théorie, sans toutefois que ces dernières puissent être dissociées ; et quelle que soit la perspective adoptée, elle ne peut se concevoir qu'en interaction et en complémentarité avec d'autres approches.

Sur sa page personnelle de chercheur en linguistique computationnelle et en cognition⁴³, Carstensen cite un passage de Manning (2015) dans lequel la prise en compte de la perspective de l'apprentissage humain est présentée comme primordiale pour le développement d'outils technologiques. Il ne faut jamais perdre de vue le fait que ces derniers ont pour fonction d'aider les personnes.

“It would be good to return some emphasis within NLP [Natural Language Processing, K.-U.C.] to cognitive and scientific investigation of language rather than almost exclusively using an engineering model of research. [...] I would encourage everyone to think about problems, architectures, cognitive science, and the details of human language, how it is learned, processed, and how it changes, rather than just chasing state-of-the-art numbers on a benchmark task.”

[« Il serait utile de mettre l'accent, dans le cadre du traitement du langage naturel, sur l'étude

⁴³ <https://www.kai-uwe-carstensen.de/#about>

cognitive et scientifique du langage plutôt que d'utiliser presque exclusivement un modèle issu de l'ingénierie de la recherche. J'encourage chacun à s'intéresser aux problèmes, architectures, sciences cognitives et aux détails du langage humain et à la façon dont il est appris, traité et modifié, plutôt que de se contenter de faire la course aux chiffres les plus récents pour évaluer une tâche étalon ».]

Il semble nécessaire et prometteur de cultiver le dialogue entre les chercheurs et chercheuses représentant des domaines scientifiques aussi variés. Sans vouloir tomber dans les travers d'une hiérarchisation stérile, je propose de considérer les didacticiennes des langues comme partenaires privilégiées pour stimuler et alimenter un tel dialogue. Elles sont formées pour chercher à écouter les personnes avec qui elles sont en contact et pour se mettre à leur portée ; elles sont habituées à réduire à l'essentiel l'information qu'elles distribuent et à l'adapter aux besoins et connaissances des destinataires. Comme le disait un des chercheurs ayant participé à la formation à l'outil TXM offerte à l'EHESS en avril 2018, une des compétences principales des didacticiennes est de « savoir expliquer des choses compliquées d'une façon qui les fait paraître simples ». J'en retiens l'idée que mener un dialogue compréhensible, transparent et ouvert peut aller de pair avec une activité de recherche menée avec sérieux, précision et engagement.

Références

1FINDR, 2019. *Elsevier* [en ligne]. 2019. Disponible à l'adresse :

<https://1findr.1science.com/home/>.

ABNEY, Steven, 2011. Data-Intensive Experimental Linguistics. In : *Linguistic Issues in Language Technology* [en ligne]. 16 novembre 2011. Vol. 6, n° 0.

[Consulté le 7 septembre 2019]. Disponible à l'adresse :

<https://journals.linguisticsociety.org/elanguage/lilt/article/view/2578/2542.html>

ADAMZIK, Kirsten et ALGHISI, Alessandra, 2015. Normen für den behördlichen Sprachgebrauch in der Schweiz. In : *Bulletin suisse de linguistique appliquée (Bulletin VALS/ASLA)*. 2015. Vol. 3, n° Numéro spécial 2015, p. 119-135.

ALEXANDER, Bryan, ASHFORD-ROWE, Kevin, BARAJAS-MURPHY, Noreen, DOBBIN, Gregory, KNOTT, Jessica, MCCORMACK, Mark, POMERANTZ, Jeffrey, SEILHAMER, Ryan et WEBER, Nicole, 2019. EDUCAUSE Horizon Report: 2019 Higher Education Edition. In : *EDUCAUSE publications* [en ligne]. 2019. Disponible à l'adresse : <https://library.educause.edu/resources/2019/4/2019-horizon-report>.

ALGHISI, Alessandra, ELMIGER, Daniel, SCHAEFFER-LACROIX, Eva et TUNGER, Verena, 2017. « KünstlerInnen », « Mitarbeiter(innen) » und « Vertreter/-innen »: Sprachnormabweichende Formen in Schweizer Behördentexten. In : ELMIGER, Daniel, RACINE, Isabelle et ZAY, Françoise (éd.), *Bulletin Vals-Asla*. 2017. Vol. Processus de différenciation: des pratiques langagières à leur interprétation sociale, n° 1, p. 189-201.

AMADIEU, Franck et TRICOT, André, 2014. *Apprendre avec le numérique : mythes et réalités*. Paris : Retz. ISBN 978-2-7256-3320-6.

ANON., 2004. *Photofiltre 7* [en ligne]. S.l. : s.n. Disponible à l'adresse :

<http://www.photofiltre.com>.

ANON., 2007. Dahoam is Dahoam. [en ligne]. Série télévisée. 2007 à aujourd'hui. Disponible à l'adresse : <https://www.br.de/br-fernsehen/sendungen/dahoam-is-dahoam/index.html>.

ANON., 2008. *Edmodo* [en ligne]. S.l. : Borg, Nicolas and O'Hara, Jeff.

[Consulté le 31 juillet 2018]. Disponible à l'adresse : <https://www.edmodo.com/>.

ANON., 2013. *Arrêté du 27 août 2013 fixant le cadre national des formations dispensées au sein des masters « métiers de l'enseignement, de l'éducation et de la formation » - Article 19* [en ligne]. 27 août 2013. S.l. : s.n. [Consulté le 25 octobre 2019]. Disponible à l'adresse : https://www.legifrance.gouv.fr/affichTexte.do;jsessionid=8FA0371F0DA0FCB8A52DFA8F8639117E.tplgfr36s_1?cidTexte=LEGITEXT000027907924&dateTexte=20200901.

ANON., 2019a. Club Corpus CeLiSo | Centre de linguistique en Sorbonne, Université Paris-Sorbonne. In : *Carnet Hypothèses* [en ligne]. 2019. [Consulté le 22 novembre 2019]. Disponible à l'adresse : <https://clubcorpus.hypotheses.org/>.

ANON., 2019b. *deSkill* [en ligne]. S.l. : Lexical Computing. Disponible à l'adresse : <https://www.sketchengine.eu/deskell-german-corpus/>.

ANON., 2019c. Mahara de l'INSPE de Paris. [en ligne]. 2019. [Consulté le 2 décembre 2019]. Disponible à l'adresse : <https://amahara.espe-paris.fr/>.

ANON., 2019d. Plateforme Moodle de l'INSPE de l'Académie de Paris 2019-2020. [en ligne]. 2019. [Consulté le 2 décembre 2019]. Disponible à l'adresse : <https://moodle2019.espe-paris.fr>.

ANON., 2019e. *YouTube Studio – Applications sur Google Play* [en ligne]. S.l. : s.n. [Consulté le 21 novembre 2019]. Disponible à l'adresse : <https://play.google.com/store/apps/details?id=com.google.android.apps.youtube.creator&hl=fr>.

ANON., 2020. TechSmith Camtasia | Video editing software. In : *TechSmith* [en ligne]. 2020. [Consulté le 2 décembre 2019]. Disponible à l'adresse : <https://www.techsmith.fr/camtasia.html>.

ANON., [sans date]. *Diigo (Digest of Internet Information, Groups and Other Stuff)*. S.l. : Diigo, Inc.

ANON., [sans date]. *Feedly-Where readers become leaders*. [en ligne]. S.l. : DevHD. Disponible à l'adresse : <https://feedly.com/i/welcome>.

ANON., [sans date]. Framapad - Éditeur de texte collaboratif. [en ligne]. [Consulté le 2 décembre 2019 c]. Disponible à l'adresse : <https://framapad.org/>.

ANON., [sans date]. *Google forms* [en ligne]. S.l. : Google drive.
[Consulté le 30 novembre 2019 d]. Disponible à l'adresse :
https://docs.google.com/forms/d/e/1FAIpQLSfExHuE5YIOxLw7zq4G1JWeD7FFbtcWDTJ33mbZyJolwN--jg/viewform?usp=embed_facebook.

ANON., [sans date]. R: The R Project for Statistical Computing. [en ligne].
[Consulté le 31 juillet 2018 e]. Disponible à l'adresse : <https://www.r-project.org/>.

ANON., [sans date]. *ScreenCast-O-Matic* [en ligne]. S.l. : Big Nerd Software.
[Consulté le 31 juillet 2018 f]. Disponible à l'adresse : <https://screencast-o-matic.com>.

ANON., [sans date]. *The future of forms is here / Typeform* [en ligne]. S.l. : Typeform.
[Consulté le 30 novembre 2019 g]. Disponible à l'adresse : <https://www.typeform.com/>.

ANON., [sans date]. *VideoANT* [en ligne]. University of Minnesota : College of Education and Human Development. [Consulté le 9 avril 2014 h]. Video Annotation Tool [Academic Technology Services, UMN]. Disponible à l'adresse : <http://ant.umn.edu/>.

ANON., [sans date]. *VLC* [en ligne]. S.l. : VideoLAN non-profit organization.
[Consulté le 1 novembre 2019 i]. Disponible à l'adresse :
<https://www.videolan.org/vlc/features.html>.

ARMA, Saveria, 2011. *The Language of Filmic Audio Description: a Corpus-Based Analysis of Adjectives*. Thèse de doctorat [en ligne]. Université de Naples. 29 novembre 2011.
[Consulté le 29 mai 2018]. Disponible à l'adresse : <http://www.fedoa.unina.it/8740/>.

ATLASSIAN, 2010. Trello. In : *Trello* [en ligne]. 2010. [Consulté le 16 septembre 2019].
Disponible à l'adresse : <https://trello.com>.

AUDACITY, 2007. *Enregistreur éditeur audio* [en ligne]. S.l. : s.n.
[Consulté le 16 décembre 2019]. Disponible à l'adresse : <https://audacity.fr/>.

BARONI, Marco et BERNARDINI, Silvia, 2004. BootCaT: Bootstrapping Corpora and Terms from the Web. In : *Proceedings LREC*. 2004. p. 4. Disponible à l'adresse :
http://sslmit.unibo.it/~baroni/publications/lrec2004/bootcat_lrec_2004.pdf

BENTIVOGLI, Luisa, BISAZZA, Arianna, CETTOLO, Mauro et FEDERICO, Marcello, 2018. Neural versus phrase-based MT quality: An in-depth analysis on English–German and English–French. In : *Computer Speech & Language*. mai 2018. Vol. 49, p. 52-70. DOI 10.1016/j.csl.2017.11.004.

BENTLEY, Elliot, [sans date]. *oTranscribe* [en ligne]. S.l. : MuckRock foundation. [Consulté le 17 mai 2019]. Disponible à l'adresse : <https://otranscribe.com/>.

BERGER, Edward, MUELLER-STÖFEN, Nele et WENDT, Marianne, 2019. *Eden* [en ligne]. 2019. [Consulté le 20 novembre 2019]. Disponible à l'adresse : <http://www.imdb.com/title/tt10107210/>.

BERGER, Emanuele, CRESCENTINI, Alberto, GALEANDRO, Cristina et MAINARDI CROHAS, 2010. La triangulation au service de la recherche en éducation. Exemples de recherche dans l'école obligatoire — AREF 2010. In : *Actes du congrès de l'Actualité de la recherche en éducation et en formation (AREF)* [en ligne]. Université de Genève : s.n. 2010. [Consulté le 28 novembre 2019]. Disponible à l'adresse : <https://plone.unige.ch/aref2010/communications-orales/premiers-auteurs-en-b/La%20triangulation.pdf/view>.

BLUME, Bob, 2019. Unterricht digital. In : *Bob Blume* [en ligne]. 2019. [Consulté le 14 décembre 2019]. Disponible à l'adresse : <https://bobblume.de/>.

BOCH, Françoise et FRIER, Catherine (éd.), 2015. *Écrire dans l'enseignement supérieur : Des apports de la recherche aux outils pédagogiques*. Grenoble : UGA Éditions. ISBN 978-2-84310-307-0.

BOISSEAU, Maryvonne et CHUQUET, Hélène, 2009. Présentation : Linguistique et traduction : réflexions théoriques et applications. In : *Revue française de linguistique appliquée*. 2009. Vol. XIV, n° 1, p. 5. DOI 10.3917/rfla.141.0005.

BOULTON, Alex et PÉREZ-PAREDES, Pascual, 2014. ReCALL special issue: Researching uses of corpora for language teaching and learning Editorial Researching uses of corpora for language teaching and learning. In : *ReCALL*. mai 2014. Vol. 26, n° 2, p. 121-127. DOI 10.1017/S0958344014000068.

- BRAUN, Sabine, 2007. Audio Description from a discourse perspective: a socially relevant framework for research and training. In : *Linguistica Antverpiensia, New Series – Themes in Translation Studies* [en ligne]. 2007. Vol. 0, n° 6. [Consulté le 10 septembre 2019]. Disponible à l'adresse : <https://lans-tts.uantwerpen.be/index.php/LANS-TTS/article/view/197>.
- BRAUN, Sabine, 2011. Creating Coherence in Audio Description. In : *Meta: Journal des traducteurs*. 2011. Vol. 56, n° 3, p. 645. DOI 10.7202/1008338ar.
- BUNDES-MINISTERIUM FÜR ARBEIT UND SOZIALES (éd.), 2014. *BMAS - Leichte Sprache - Ein Ratgeber* [en ligne]. 2014. S.l. : s.n. [Consulté le 17 mai 2019]. Disponible à l'adresse : <https://www.bmas.de/DE/Service/Medien/Publikationen/a752-leichte-sprache-ratgeber.html>.
- BURNARD, Lou, 2014. *What is the Text Encoding Initiative? - OpenEdition Press* [en ligne]. S.l. : s.n. [Consulté le 16 avril 2014]. Disponible à l'adresse : <http://books.openedition.org/oep/426>.
- CARSTENSEN, Kai-Uwe, 2017. *Sprachtechnologie - Ein Überblick* [en ligne]. <http://www.kai-uwe-carstensen.de/> : s.n. Disponible à l'adresse : <http://kai-uwe-carstensen.de/Publikationen/Sprachtechnologie.pdf>.
- CASTELLOTTI, Véronique, 2010. Attention ! Un plurilinguisme peut en cacher un autre. Enjeux théoriques et didactiques de la notion de pluralité. In : *Recherches en didactique des langues et des cultures* [en ligne]. 1 avril 2010. Vol. 7, n° 1. DOI 10.4000/rdlc.2056. Disponible à l'adresse : <http://journals.openedition.org/rdlc/2056>.
- CAVALLO, Amelia et FRYER, Louise, 2018. *EXTANT Integrated Access Inquiry 2017-18* [en ligne]. Scientific report. S.l. Disponible à l'adresse : https://www.extant.org.uk/docs/uploads/Extant_Integrated_Access_Report_2018_Full_Length.docx.
- CHACHKINE, Elsa, DEMAIZIÈRE, Françoise et SCHAEFFER-LACROIX, Eva, 2013. Pour un apprenant réfléchissant. In : *Linguistik Online* [en ligne]. 1 juillet 2013. Vol. 60. [Consulté le 6 novembre 2019]. DOI 10.13092/lo.60.1204. Disponible à l'adresse : <https://bop.unibe.ch/linguistik-online/article/view/1204>.

CHAMBERS, Angela et LE BARON, Florence, 2006. *Chambers-Le Baron Corpus of Research Articles in French* [en ligne]. 12 décembre 2006. S.l. : s.n.

[Consulté le 22 novembre 2019]. Disponible à l'adresse :

<https://ota.bodleian.ox.ac.uk/repository/xmlui/handle/20.500.12024/2527>.

CHARLES, Maggie, 2016. Using a Corpus as a Resource for Writing Information for Students. In : *Research supervision pages* [en ligne]. 2016. [Consulté le 22 novembre 2019].

Disponible à l'adresse :

https://www.academia.edu/36960993/Using_a_Corpus_as_a_Resource_for_Writing_Information_for_Students.

CHARLIRELLE, 1975. *Behind the words: 6e : livre de l'élève*. Paris : O.C.D.L : Hatier. ISBN 978-2-218-03275-2.

CHENG, Winnie, WARREN, Martin et XUN-FENG, Xu, 2003. The language learner as language researcher: putting corpus linguistics on the timetable. In : *System*. juin 2003.

Vol. 31, n° 2, p. 173-186. DOI 10.1016/S0346-251X(03)00019-8.

CHINI, Danielle, 2009. Linguistique et didactique : où en est-on ? Quelle place pour une approche conceptualisante de la construction de la langue dans la perspective actionnelle ? In : *Recherches en didactique des langues et des cultures* [en ligne]. 1 octobre 2009. Vol. 6, n° 2.

[Consulté le 28 février 2019]. DOI 10.4000/rdlc.1958. Disponible à l'adresse :

<http://journals.openedition.org/rdlc/1958>.

COLLECTIF, 2019. Éduveille | Autour des recherches en éducation et formation. In : *Éduveille | Autour des recherches en éducation et formation* [en ligne]. 2019.

[Consulté le 30 novembre 2019]. Disponible à l'adresse : <https://eduveille.hypotheses.org/>.

CONSEIL DE L'EUROPE (éd.), 2001. *Un cadre européen commun de référence pour les langues: apprendre, enseigner, évaluer*. Paris : Didier. ISBN 978-2-278-05813-6.

CONSEIL DE L'EUROPE (éd.), 2018. *Cadre européen commun de référence pour les langues : apprendre, enseigner, évaluer. Volume complémentaire avec de nouveaux descripteurs*. Strasbourg : Conseil de l'Europe. Division des Politiques éducatives. Service de l'Éducation.

CORPORATION FOR DIGITAL SCHOLARSHIP, [sans date]. *Zotero / Your personal research assistant* [en ligne]. S.l. : Corporation for Digital Scholarship.

[Consulté le 30 novembre 2019]. Disponible à l'adresse : <https://www.zotero.org/>.

CROSTHWAITE, Peter (éd.), 2019. *Data-Driven Learning for the Next Generation: Corpora and DDL for Pre-tertiary Learners*. London ; New York, NY : Routledge. ISBN 978-1-138-38801-7. P128.C68

DALMAS, Martine, 2019. Französische Sichtweisen und folgenschwere Konsequenzen. In : STICKEL, Gerhard (éd.), *Deutsch von außen*. ebook. S.l. : De Gruyter. p. 113-129. ISBN 978-3-11-062269-0.

DAS PORTAL FÜR DIE KORPUSRECHERCHE IN DEN TEXTKORPORA DES INSTITUTS FÜR DEUTSCHE SPRACHE, [sans date]. COSMAS II - Startseite. [en ligne]. [Consulté le 8 décembre 2019]. Disponible à l'adresse : <https://www.ids-mannheim.de/cosmas2/>.

DAUTEL, Klaus, [sans date]. Grammatik in Bildern. [en ligne].

[Consulté le 4 novembre 2019]. Disponible à l'adresse :

<https://www.zum.de/Faecher/D/BW/gym/grammatik/index.htm>.

DEEPL GMBH, 2017. *DeepL Translator* [en ligne]. S.l. : s.n.

[Consulté le 28 novembre 2018]. Disponible à l'adresse : <https://www.deepl.com/translator>.

DEMAIZIÈRE, Françoise et BERTHOUD, Anne-Claude, 1998. Des formes linguistiques pour communiquer : vers de nouveaux rapports entre linguistique et didactique. In : SPRINGER, Claude (éd.), *Les linguistiques appliquées et les Sciences du langage* [en ligne].

Strasbourg : Université Strasbourg 2. 1998. p. 164-178. Disponible à l'adresse :

<https://edutice.archives-ouvertes.fr/edutice-00000375/document>. edutice

DEMAIZIÈRE, Françoise et NARCY-COMBES, Jean-Paul, 2007. Du positionnement épistémologique aux données de terrain. In : *Recherches en didactique des langues et des cultures* [en ligne]. 21 juin 2007. Vol. 4, n° 4. [Consulté le 29 novembre 2019].

DOI 10.4000/rdlc.4850. Disponible à l'adresse : <http://journals.openedition.org/rdlc/4850>.

DIEHL, Erika, CHRISTEN, Helen, LEUENBERGER, Sandra, PELVAT et STUDER, Thérèse, 2000. *Grammatikunterricht, alles für der Katz? Untersuchungen zum Zweitspracherwerb Deutsch*. Tübingen : Niemeyer. Reihe Germanistische Linguistik, 220. ISBN 978-3-484-31220-3. PF3066 .G683 2000

DRESEN, Andreas, 2018. *Gundermann* [en ligne]. 23 août 2018. S.l. : s.n. Disponible à l'adresse : <https://www.imdb.com/title/tt7205942/>.

DRUIDE INFORMATIQUE INC., 2019. Antidote. [en ligne]. 2019. [Consulté le 5 décembre 2019]. Disponible à l'adresse : <https://www.antidote.info/fr/>.

EBERLEIN, Norbert, 1997. *Neues aus Büttengewässer* [en ligne]. Série télévisée. Disponible à l'adresse : https://www.ndr.de/mediathek/mediatheksuche105_broadcast-129.html.

ELG CONSORTIUM, 2019. European Language Grid. [en ligne]. 2019. [Consulté le 1 novembre 2019]. Disponible à l'adresse : <https://www.european-language-grid.eu/>.

ELMIGER, Daniel, 2014. Cachez ces doublons que je ne saurais voir: les doubles formes féminine et masculine dans le langage administratif suisse. In : *Cahiers de linguistique (Genres, Langues et Pouvoirs)*. 2014. Vol. 40, n° 1, p. 155-170.

ELMIGER, Daniel, 2015. *Les corpus Feuille fédérale / Bundesblatt / Foglio fédérale. V. 1.2* [en ligne]. 2015. S.l. : Université de Genève. [Consulté le 11 juin 2019]. Disponible à l'adresse : <https://archive-ouverte.unige.ch/unige:80593>.

ELMIGER, Daniel et MATTHEY, Marinette, 2017. De L2 à L3: Didactique du plurilinguisme et théories de l'acquisition. In : LAURENS, Véronique et VÉRONIQUE, Daniel (éd.), *Le français dans le monde - Recherches et applications*. 2017. Vol. 61, p. 136-145.

ELMIGER, Daniel, SCHAEFFER-LACROIX, Eva et TUNGER, Verena, 2019. La rédaction non-sexiste en Suisse: pluralité des discours et des pratiques. In : RABATEL, Alain et ROSIER, Laurence (éd.), *Le Discours et la Langue*. 2019. Vol. Les défis de l'écriture inclusive, n° 111, p. 125-150.

ELMIGER, Daniel, TUNGER, Verena et SCHAEFFER-LACROIX, Eva, 2017a. *Geschlechtergerechte Behördentexte: linguistische Untersuchungen und Stimmen zur Umsetzung in der mehrsprachigen Schweiz : Forschungsbericht*. Genève : Université de Genève. ISBN 978-1-365-70544-1.

ELMIGER, Daniel, TUNGER, Verena et SCHAEFFER-LACROIX, Eva, 2017b. Geschlechtergerechte Sprache in Schweizer Behördentexten: Möglichkeiten und Grenzen einer mehrsprachigen Umsetzung. In : REISIGL, Martin et SPIESS, Constanze (éd.), *Sprache und Geschlecht*. Duisburg : OBST (Osnabrücker Beiträge zur Sprachtheorie). 90. p. 61-90. ISBN 978-3-95605-035-0.

ERICSSON, K. Anders et SIMON, Herbert A., 1987. Verbal reports on thinking”, in C. Faerch and G. Kasper (eds.), *Introspection in Second Language Research*. Clevedon, Avon, Multilingual Matters, p. 24–54. Ericsson K. Anders and Simon Herbert, 1987, “. In : FAERCH, Claus et KASPER, Gabriele (éd.), *Introspection in Second Language Research*. Clevedon, Avon : Multilingual Matters. p. 24-54. ISBN 978-0-585-23528-8.

EULER, Dieter, 2017. Design principles as bridge between scientific knowledge production and practice design. In : *EDeR. Educational Design Research*. 26 janvier 2017. Vol. 1, n° 1, p. 1-17. DOI 10.15460/eder.1.1.1024.

FUCHS, Catherine, 2014. Le tournant quantitatif en TAL et en linguistique. In : *L'Information Grammaticale*. 2014. n° 0, p. 8–13. DOI 10.2143/IG.142.0.3038752.

GAUTIER, Laurent et BACH, Matthieu, 2017. La terminologie du vin au prisme des corpus oraux de dégustation/présentation (français-allemand) : entre émotions, culture et sensorialité. In : *Éla. Études de linguistique appliquée*. 2017. Vol. N° 188, n° 4, p. 485-509. DOI 10.3917/ela.188.0485.

GLYNN, Dylan et ROBINSON, Justyna A. (éd.), 2014. *Corpus Methods for Semantics: Quantitative studies in polysemy and synonymy* [en ligne]. Amsterdam : John Benjamins Publishing Company. [Consulté le 13 septembre 2019]. Human Cognitive Processing. ISBN 978-90-272-2397-5. Disponible à l'adresse : <http://www.jbe-platform.com/content/books/9789027270337>.

GOMRINGER, Eugen, 1997. *theorie der konkreten poesie: texte und manifeste 1954 - 1997*. Wien : edition splitter. Werke, Band 2. ISBN 978-3-901190-34-6.

GOOGLE, 2020. *Google Docs* [en ligne]. S.l. : Google. [Consulté le 5 décembre 2019]. Disponible à l'adresse : <https://docs.google.com/document/u/0/>.

GOUVERNEMENT DE L'ÉTAT FRANÇAIS (éd.), 2017. *Guide d'accompagnement - RGAA 3 2017 / Les documents de référence du S.I. de l'État* [en ligne]. 2017. S.l. : s.n. [Consulté le 17 mai 2019]. Disponible à l'adresse : <http://references.modernisation.gouv.fr/referentiel/guide-accompagnement-RGAA.html>.

GROSBOIS, Muriel, 2012. *Didactique des langues et technologies: de l'EAO aux réseaux sociaux*. Paris : Presses de l'Université Paris-Sorbonne. ISBN 978-2-84050-848-9.

GUILLEMIN-FLESCHER, Jacqueline, 1981. *Syntaxe comparée du français et de l'anglais: problèmes de traduction*. Paris : Ophrys. ISBN 978-2-7080-0502-0. PC2099 .G78

HABERT, Benoît, 2004. Outiller la linguistique : de l'emprunt de techniques aux rencontres de savoirs. In : *Revue française de linguistique appliquée*. 2004. Vol. Vol. IX, n° 1, p. 5-24.

HAWKINS, Eric W., 1987. *Awareness of language: an introduction*. Rev. ed. Cambridge [Cambridgeshire] ; New York : Cambridge University Press. ISBN 978-0-521-33762-5. LB1575.8 .H38 1987

HEIDEN, Serge, MAGUÉ, Jean-Philippe et PINCEMIN, Bénédicte, 2010. TXM : Une plateforme logicielle open-source pour la textométrie - conception et développement. In : *Proceedings of 10th International Conference Journées d'Analyse statistique des Données Textuelles* [en ligne]. S.l. : Edizioni Universitarie di Lettere Economia Diritto. 9 juin 2010. p. 1021-1032. [Consulté le 30 novembre 2018]. Disponible à l'adresse : <https://halshs.archives-ouvertes.fr/halshs-00549779/document>.

HESS, Annette, 2010. Weissensee | Verband Deutscher Drehbuchautoren e.V. (VDD). [en ligne]. S.l. 2010. [Consulté le 28 novembre 2018]. Disponible à l'adresse : <https://drehbuchautoren.de/podcast/2010-09-14/weissensee>.

JANITZA, Jean et TRAVERS, Monique, 2003. *Babel en éducation: linguistique allemande et didactique des langues*. Asnières : Institut d'Allemand d'Asnières. Publications de l'Institut d'Allemand, 33. ISBN 978-2-910212-26-1.

JANKOWSKA, Anna, MILC, Michał et FRYER, Louise, 2017. Translating audio description scripts... into English. In : *THE SLOVAK ASSOCIATION FOR THE STUDY OF ENGLISH*. Košice, 2017. p. 1-15. Disponible à l'adresse : <http://www.skase.sk/>

JOHNS, Tim, 1997. Kibbitzing one-to-ones. Web version of notes for presentation at BALEAP. In : *BALEAP meeting on Academic Writing*. University of Reading. novembre 1997.

KAUFMANN, Jean-Claude et SINGLY, François de, 2014. *L'entretien compréhensif*. Paris : A. Colin. ISBN 978-2-200-25990-7.

KECK, Casey et KIM, YouJin, 2014. *Pedagogical grammar*. Amsterdam ; Philadelphia : John Benjamins Publishing Company. ISBN 978-90-272-1217-7. P53.412 .K43 2014

KERBRAT-ORECCHIONI, Catherine, 2012. Le contexte revisité. In : *Corela. Cognition, représentation, langage* [en ligne]. 2 avril 2012. n° HS-11. [Consulté le 22 mars 2018]. DOI 10.4000/corela.2627. Disponible à l'adresse : <http://journals.openedition.org/corela/2627>.

KILGARRIFF, Adam, RYCHLÝ, Pavel et POMIKÁLEK, Jan, [sans date]. *Sketch Engine / language corpus management and query system* [en ligne]. S.l. : Lexical Computing. [Consulté le 31 juillet 2018]. Disponible à l'adresse : <https://www.sketchengine.eu/>.

KRAVCHENKO-BIBERSON, Olga, 2012. Le modèle cognitif et la TOE : deux points de vue sur l'identité sémantique des unités polysémiques. In : *Corela* [en ligne]. 2 avril 2012. n° HS-11. [Consulté le 15 octobre 2019]. DOI 10.4000/corela.2024. Disponible à l'adresse : <http://journals.openedition.org/corela/2024>.

LACROIX, Eva, 2005. *L'allemand, les TICE et les tâches*. [en ligne]. Mémoire de DEA en didactique des langues et cultures. Paris : Paris 3 Sorbonne nouvelle. [Consulté le 5 décembre 2019]. Disponible à l'adresse : http://didatic.net/article.php3?id_article=71.

LANCASTER UNIVERSITY, [sans date]. CQPweb. [en ligne].

[Consulté le 8 décembre 2019]. Disponible à l'adresse : <https://cqpweb.lancls.ac.uk/>.

LECHNER, Renée, HENNING, Urs et MÜLLER, Emil, 2019. Web2 Unterricht. [en ligne].

2019. [Consulté le 30 novembre 2019]. Disponible à l'adresse : <https://web2-unterricht.ch/>.

MÄLZER, Nathalie et WÜNSCHE, Maria, 2018. *Inklusion am Theater* [en ligne]. Berlin,

Bern, Bruxelles, New York, Oxford : Peter Lang D. [Consulté le 4 septembre 2019].

ISBN 978-3-631-74157-3. Disponible à l'adresse :

<https://www.peterlang.com/view/title/65911>.

MÄLZER-SEMLINGER, Nathalie, 2012. Narration or description: What should audio description "look" like? In : PEREGO, Elisa (éd.), *Emerging topics in translation: audio description* [en ligne]. S.l. : EUT Edizioni Università di Trieste.

[Consulté le 5 septembre 2019]. ISBN 978-88-8303-347-6. Disponible à l'adresse :

<https://www.openstarts.units.it/handle/10077/6359>.

MANNING, Christopher D., 2015. Computational Linguistics and Deep Learning. In :

Computational Linguistics. décembre 2015. Vol. 41, n° 4, p. 701-707.

DOI 10.1162/COLI_a_00239.

MAYAFFRE, Damon, 2002. Les corpus réflexifs : entre architextualité et hypertextualité. In :

Corpus [en ligne]. 15 novembre 2002. n° 1. [Consulté le 26 octobre 2018]. Disponible à

l'adresse : <http://journals.openedition.org/corpus/11>.

MIELE, Joshua, 2019. YouDescribe. In : *The Smith Kettlewell Eye Research Institut*

[en ligne]. 2019. [Consulté le 4 octobre 2019]. Disponible à l'adresse :

<https://www.ski.org/project/youdescribe#targetText=YouDescribe%20is%20a%20free%2C%20web,audio%20description%20to%20YouTube%20content>.

MINISTÈRE DE L'ÉDUCATION NATIONALE ET DE LA JEUNESSE - DIRECTION

GÉNÉRALE DE L'ENSEIGNEMENT SCOLAIRE, [sans date]. Boîte à outils - Internet

responsable. In : *Eduscol* [en ligne]. [Consulté le 2 décembre 2019]. Disponible à l'adresse :

<https://eduscol.education.fr/internet-responsable/ressources/boite-a-outils.html>.

MOCHET, Marie-Anne et COSTE, Daniel (éd.), 2005. *Plurilinguisme et apprentissages: mélanges Daniel Coste*. Lyon : ENS Éd. Hommages. ISBN 978-2-84788-088-5.

NAVAMUEL, Fidel, 2019. Les Outils Tice. In : *Les Outils Tice* [en ligne]. 2019. [Consulté le 30 novembre 2019]. Disponible à l'adresse : <https://outilstice.com/>.

PIEKLARZ-THIEN, Magdalena, 2018. Von der disziplinären Abhängigkeit zum gleichberechtigten Miteinander? Eine Diskussion über das sensible Verhältnis von Sprachwissenschaft und Sprachdidaktik am Beispiel der Germanistischen Linguistik und der DaF-Didaktik. In : *Glottodidactica. An International Journal of Applied Linguistics*. 10 septembre 2018. Vol. 45, n° 1, p. 117-132. DOI 10.14746/gl.2018.45.1.08.

PINCEMIN, Bénédicte, 2012. Atelier d'initiation à TXM. In : *Textométrie* [en ligne]. 27 septembre 2012. [Consulté le 14 décembre 2019]. Disponible à l'adresse : http://textometrie.ens-lyon.fr/html/enregistrement_atelier_initiation_TXM_fr.html.

POPOVIĆ, Maja, 2018. Language-related issues for NMT and PBMT for English–German and English–Serbian. In : *Machine Translation*. septembre 2018. Vol. 32, n° 3, p. 237-253. DOI 10.1007/s10590-018-9219-5.

POTHIER, Maguy, 2001. Formation à la recherche et recherche de la formation en didactique. In : *Éla. Études de linguistique appliquée*. 2001. Vol. n° 123-124, n° 3, p. 385. DOI 10.3917/ela.123.0385.

RAPP, Reinhard, SHAROFF, Serge et ZWEIGENBAUM, Pierre, 2016. Recent advances in machine translation using comparable corpora. In : *Natural Language Engineering*. juillet 2016. Vol. 22, n° 4, p. 501-516. DOI 10.1017/S1351324916000115.

REDACTED, [sans date]. *Easily blur/pixelate/black out parts of an image online*. [en ligne]. S.l. : s.n. [Consulté le 5 décembre 2019]. Disponible à l'adresse : <https://marky.space/redacted/>.

REINMANN, Gabi, 2019. Hochschuldidaktik. [en ligne]. 2019. [Consulté le 30 novembre 2019]. Disponible à l'adresse : <https://gabi-reinmann.de/>.

REVIERS, Nina, 2018. Studying the language of Dutch audio description: An example of a corpus-based analysis. In : *Translation and Translanguaging in Multilingual Contexts*. 24 avril 2018. Vol. 4, n° 1, p. 178-202. DOI 10.1075/ttmc.00009.rev.

RÜHL, Marco, 2000. *Linguistique pour germanistes: Une tentative de médiation entre la tradition française et la tradition allemande de l'étude de la langue allemande* [en ligne]. Lyon : ENS Éditions. [Consulté le 9 septembre 2019]. ISBN 978-2-902126-83-5. Disponible à l'adresse : <http://books.openedition.org/enseditions/1648>.

SALEM, André, 2009. Lexi-Co. [en ligne]. 2009. [Consulté le 3 décembre 2019]. Disponible à l'adresse : <http://www.lexi-co.com/>.

SALENGROS IGUENANE, Isabelle et SCHAEFFER-LACROIX, Eva, 2020. Le cahier des clauses techniques particulières (CCTP) : spécificités d'un écrit professionnel. In : RAUS, Rachele et FROELIGER, Nicolas (éd.), *Le Langage et l'Homme*. paraître 2020. Vol. La terminologie et l'enseignement du français de spécialité, n° 54/2, p. 135-156.

SALWAY, Andrew, 2007. A Corpus-based Analysis of Audio Description. In : CINTAS DÍAZ, Jorge, ORERO, Pilar et REMAEL, Aline (éd.), *Media for All: Subtitling for the Deaf, Audio Description, and Sign Language*. Amsterdam : Rodopi. p. 151–174.

SCHAEFFER-LACROIX, Eva, soumis. Integrating corpus-based audio description tasks into an intermediate-level German course. In : BRUTT-GRIFFLER, Janina, PERRIN, Daniel et ZAKI, Mai (éd.), *International Journal of Applied Linguistics (InJAL)*. Vol. Corpus-based language teaching and learning: applications and implications.

SCHAEFFER-LACROIX, Eva, 2006. L'aide à la rédaction en allemand au collège et à l'université : techniques de production écrite et corpus pédagogiques. In : *ATALA 6 - Aide à la rédaction écrite* [en ligne]. Paris : s.n. 3 juin 2006. Disponible à l'adresse : <http://didaktik.hautetfort.com/media/02/02/2314131106.pdf>.

SCHAEFFER-LACROIX, Eva, 2009a. *Corpus numériques et production écrite en langue étrangère. Une recherche avec des apprenants d'allemand* [en ligne]. Thèse de doctorat. Paris : Université de la Sorbonne nouvelle - Paris III. [Consulté le 11 juin 2019]. Disponible à l'adresse : <https://tel.archives-ouvertes.fr/tel-00439095>.

SCHAEFFER-LACROIX, Eva, 2009b. Les blogues en cours d'allemand. In : AURIA, Frédéric (éd.), *e nouveau bulletin de l'A.D.E.A.F. (Association de Développement de l'Enseignement de l'Allemand en France)*. 2009. Vol. Décembre 2009, n° 106, p. 35-38.

SCHAEFFER-LACROIX, Eva, 2010a. Analyse de Dispositifs médiatisés en langues et accompagnement-tutorat - Association des Chercheurs et Enseignants Didacticiens des Langues Étrangères. In : *ACEDLE* [en ligne]. 2010. [Consulté le 26 octobre 2019]. Disponible à l'adresse : http://acedle.org/old/spip.php?article2798&var_recherche=schaeffer-lacroix.

SCHAEFFER-LACROIX, Eva, 2010b. Évolution du système linguistique d'apprenants d'allemand à l'aide de corpus numériques. L'exemple de l'adjectif. In : *Alsic. Apprentissage des Langues et Systèmes d'Information et de Communication* [en ligne]. 28 février 2010. n° Vol. 13. [Consulté le 29 juin 2018]. DOI 10.4000/alsic.1678. Disponible à l'adresse : <http://journals.openedition.org/alsic/1678>.

SCHAEFFER-LACROIX, Eva, 2011. Création de corpus réfléchis pour l'enseignement de la production écrite en allemand, langue étrangère. In : ESTÈVE, Isabelle, FAURE, Auriane, GUITTON, Aurélie, MOUT, Tiphaine, SOUQUE, Agnès et TOUTATI, Aïcha (éd.), *Autour des langues et du langage n° 2 : perspective pluridisciplinaire*. Grenoble : PUG. p. 173-180. ISBN 978-2-7061-1671-1.

SCHAEFFER-LACROIX, Eva, 2012. En quête de « cyber-profs » socio-constructivistes. In : *Nouveaux Cahiers d'Allemand*. 2012. n° 2012/2, p. 217-222.

SCHAEFFER-LACROIX, Eva, 2014. Utiliser des corpus numériques avec un public Lansad. In : *Alsic* [en ligne]. 7 mai 2014. n° Vol. 17. [Consulté le 26 octobre 2019]. DOI 10.4000/alsic.2720. Disponible à l'adresse : <http://journals.openedition.org/alsic/2720>.

SCHAEFFER-LACROIX, Eva, 2015a. Analyse de trois systèmes de gestion de corpus pour l'enseignement-apprentissage des langues étrangères. In : *Alsic. Apprentissage des Langues et Systèmes d'Information et de Communication* [en ligne]. 3 février 2015. n° Vol. 18, n° 1. [Consulté le 9 octobre 2018]. Disponible à l'adresse : <http://journals.openedition.org/alsic/2852>.

SCHAEFFER-LACROIX, Eva, 2015b. Impact de discussions métalinguistiques sur l'apprentissage de la production écrite en allemand, langue étrangère. In : *Linx. Revue des linguistes de l'université Paris X Nanterre*. 1 septembre 2015. n° 72, p. 233-237. DOI 10.4000/linx.1676.

SCHAEFFER-LACROIX, Eva, 2016a. Éditorial. In : *Alsic. Apprentissage des Langues et Systèmes d'Information et de Communication* [en ligne]. 28 décembre 2016. n° Vol. 19, n° 2. [Consulté le 26 octobre 2019]. Disponible à l'adresse : <http://journals.openedition.org/alsic/2970>.

SCHAEFFER-LACROIX, Eva, 2016b. Mais où sont donc les femmes ? À la recherche de formes doubles dans un corpus de textes administratifs suisses. In : SALEM, André et FLEURY, Serge (éd.), *JADT2016 : ACTES / PROCEEDINGS / jadt.org* [en ligne]. Nice : s.n. 7 juin 2016. [Consulté le 1 novembre 2019]. Disponible à l'adresse : <http://lexicometrica.univ-paris3.fr/jadt/jadt2016/>.

SCHAEFFER-LACROIX, Eva, 2016c. Talking about German verb particles identified in concordance lines – from spontaneous to expert-like metatalk. In : *Language Awareness*. 2 avril 2016. Vol. 25, n° 1-2, p. 127-143. DOI 10.1080/09658416.2015.1122023.

SCHAEFFER-LACROIX, Eva, 2017. Céline Poudat et Frédéric Landragin, Explorer un corpus textuel : Méthodes – pratiques – outils. Louvain-la-Neuve, De Boeck Supérieur, 2017, 240 p. In : *Lidil. Revue de linguistique et de didactique des langues* [en ligne]. 1 novembre 2017. n° 56. [Consulté le 26 octobre 2019]. Disponible à l'adresse : <http://journals.openedition.org/lidil/4720>.

SCHAEFFER-LACROIX, Eva, 2018a. Initiation à TXM – Site compagnon de formations universitaires. In : *Carnet Hypothèses* [en ligne]. 2018. [Consulté le 22 novembre 2019]. Disponible à l'adresse : <https://initxm.hypotheses.org/>.

SCHAEFFER-LACROIX, Eva, 2018b. Les outils de corpus pour repenser la grammaire. In : *Les Langues Modernes n° 3/2018*. Grammaire ? Vous avez dit grammaire ? Représentations et pratiques enseignantes. 2018. p. 61-70.

SCHAEFFER-LACROIX, Eva, 2018c. *Préparer un corpus pour TXM* [en ligne]. janvier 2018. S.l. : HAL. [Consulté le 22 novembre 2019]. Disponible à l'adresse : <https://hal.archives-ouvertes.fr/hal-01676053>.

SCHAEFFER-LACROIX, Eva, 2018d. Questionner des corpus pour questionner la langue. L'exemple des virgules en allemand. In : *Lidil. Revue de linguistique et de didactique des langues* [en ligne]. 2 novembre 2018. n° 58. [Consulté le 20 décembre 2018]. Disponible à l'adresse : <http://journals.openedition.org/lidil/5515>.

SCHAEFFER-LACROIX, Eva, 2019a. Barriers to trainee teachers' corpus use. In : CROSTHWAITE, Peter (éd.), *Data-Driven Learning for the Next Generation: Corpora and DDL for Pre-tertiary Learners*. London ; New York, NY : Routledge. p. 47-64. ISBN 978-1-138-38800-0. P128.C68 D38 2020

SCHAEFFER-LACROIX, Eva, 2019b. *Corpus et didactique de l'allemand - La langue à bras-le-corps*. Limoges : Lambert-Lucas. ISBN 978-2-35935-276-4.

SCHAEFFER-LACROIX, Eva, 2019c. Enseignements universitaires et recherches scientifiques. In : *Blogue professionnel* [en ligne]. 2019. Disponible à l'adresse : <http://didaktik.hautetfort.com/>.

SCHAEFFER-LACROIX, Eva, 2019d. Les corpus web à travers le prisme de l'ALMT. In : POUDAT, Céline, WIGHAM, Ciara et LIÉGEOIS, Loïc (éd.), *Corpus*. à paraître 2019. Vol. 21. Corpus complexes Traitements, standardisation et analyse des corpus de communication médiée par les réseaux.

SCHAEFFER-LACROIX, Eva, MAGNUS, Gilbert et GRASS, Thierry, 2020. Usages variés de l'outil d'exploration de corpus TXM dans le contexte de l'allemand. In : TERRISSE, Bénédicte, LAMBERTZ-POLLAN, Ruth, ROUDAUT, Maiwenn, DAHM, Johannes, FORBRIG, Karsten, FRADIN, Clément et ZIELINSKI, Gwenaëlle (éd.), *Machines / Maschinen. Les machines dans l'espace germanique : de l'automate de Kempelen à Kraftwerk*. Presses Universitaires de Rennes. Rennes : s.n.

SCHILLER, Anne, TEUFEL, Simone et STOCKERT, Christine, 1995. *Vorläufige Guidelines für das Tagging deutscher Textcorpora mit STTS* [en ligne]. 1995. S.l. : s.n. Disponible à l'adresse : https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/stts_guide.pdf.

SCHMALE, Günter, 2015. Korpusbasierung, Multimodalität, Interaktivität, Konstruktionsorientierung – Prinzipien einer zeitgemäßen linguistischen Pragmatik und Fremdsprachendidaktik. In : *Studia Germanica Gedanensia*. 15 septembre 2015. n° 33, p. 84-108. Disponible à l'adresse :

<https://czasopisma.bg.ug.edu.pl/index.php/SGG/article/view/1415>

SCHMID, Helmut, 1994. Probabilistic Part-of-Speech Tagging Using Decision Trees. In : *Proceedings of International Conference on New Methods in Language Processing* [en ligne]. Manchester, UK : s.n. 1994. Disponible à l'adresse : <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/tree-tagger1.pdf>.

SCHMIDT, Thomas, 2012. EXMARaLDA and the FOLK tools – two toolsets for transcribing and annotating spoken language. In : DECLERCK, Thierry, CHOUKRI, Khalid et CALZOLARI, Nicoletta (éd.), *Proceedings of the 8th International Conference on Language Resources and Evaluation: May 23-25, 2012*. Istanbul (Turquie) : s.n. 2012. p. 236-240.

SCHMIDT, Thomas, 2019. *EXAKT (EXMARaLDA Analysis- and Concordance Tool)* [en ligne]. S.l. : s.n. [Consulté le 2 décembre 2019]. Disponible à l'adresse : <https://exmaralda.org/en/exakt-en/>.

SCHMIDT, Thomas et WÖRNER, Kai, 2013. *EXMARaLDA* [en ligne]. Oxford : Oxford University Press. [Consulté le 17 mai 2019]. Disponible à l'adresse : <http://oxfordhandbooks.com/view/10.1093/oxfordhb/9780199571932.001.0001/oxfordhb-9780199571932-e-030>.

SILVERMAN, David, 2017. *Doing qualitative research*. Fifth edition. London Thousand Oaks, California : SAGE Publications Ltd. ISBN 978-1-4739-6698-7.

STRANOVSKÁ, Eva, FICZERE, Anikó, HVOZDÍKOVÁ, Silvia et HOCKICKOVÁ, Beáta, 2019. Learning Experience as a Factor of Motivation in Lower-Secondary School in Foreign Language Learning. In : *Problems of Education in the 21st Century*. 21 juin 2019. Vol. 77, n° 3, p. 437-448. DOI 10.33225/pec/19.77.437.

TANGUY, Ludovic et FABRE, Cécile, 2014. Évolutions de la linguistique outillée. In : *L'Information Grammaticale*. 2014. n° 0, p. 14–22. DOI 10.2143/IG.142.0.3038753.

VÉRONIQUE, Daniel, 2010. Sur l'acquisition des langues étrangères : entre le nomologique et l'actionnel. In : *Le français dans le monde - Recherches et applications*. 2010. n° 48, p. 76-85. Disponible à l'adresse : http://fipf.org/sites/fipf.org/files/09037120_076-134.pdf

VYATKINA, Nina et BOULTON, Alex, 2017. Corpora in Language Learning and Teaching. In : *Language Learning & Technology*. 2017. Vol. 21, n° 3, p. 1-8. DOI 10125/44750.

WEISSMANN, Dirk, 2012. L'horizon utopique d'une totalité des langues et cultures Plurilinguisme et écriture plurilingue chez Ilma Rakusa. In : *Germanica*. 2012. n° 51, p. 149-163. Disponible à l'adresse : <https://www.cairn.info/revue-ela-2012-3-page-313.htm>

Liste des principaux acronymes et sigles

ACEDLE	Association des chercheurs et enseignants didacticiens des langues étrangères
AD	Audiodescription
AGES	Association des germanistes de l'enseignement supérieur
Alsic	Apprentissage des langues et systèmes d'information et de communication
ANR	Agence nationale des recherches
ASH	Adaptation scolaire et scolarisation des élèves handicapés
ATALA	Association pour le traitement automatique des langues
CAPES	Certificat d'aptitude au professorat de l'enseignement du second degré
CCTP	Cahier des clauses techniques particulières
CECRL	Cadre européen commun de référence pour les langues Colloque international des jeunes chercheurs en didactique des langues et en linguistique
CEDIL	
CeLiSo	Centre de linguistique en Sorbonne
Cléo	Centre pour l'édition électronique ouverte
CNAM	Conservatoire national des arts et métiers
CNRS	Centre national de la recherche scientifique
Corli	Consortium corpus, langues et interactions
COST	European Cooperation in Science and Technology
CQL	Corpus query language
DAAD	Deutscher akademischer Austauschdienst
DaF	Deutsch als Fremdsprache
DEA	Diplôme d'études approfondi
DH	Digital Humanities
DNL	Discipline non linguistique
EHESS	École des hautes études en sciences sociales
EIAH	Environnements informatiques pour l'apprentissage humain
ELG	European Language Grid
Espé	École supérieure du professorat et de l'éducation
FLE	Français langue étrangère
FNS	Fonds national suisse

FOS	Français sur objectifs spécifiques
FSTG	Formation des professeurs stagiaires
Inspé	Institut national supérieur du professorat et de l'éducation
IUFM	Institut universitaire de la formation des maîtres
JADT	Journées internationales d'analyse statistique des données textuelles
JLC	Journées internationales de la linguistique de corpus
Lansad	Langues pour spécialistes d'autres disciplines
LEA	Langues étrangères appliquées
LLCE	Langues, littératures et civilisations étrangères
MEEF	Métiers de l'enseignement, de l'éducation et de la formation
MOOC	Massive open online course
NMT	Neuronal machine translation
PERL	Pôle d'élaboration de ressources linguistiques
PRAG	Professeur agrégé de l'enseignement supérieur
RANACLÈS	Rassemblement national des centres de langues de l'enseignement supérieur
REI	Recherche exploratoire et innovation
SkELL	Sketch Engine for Language Learning
T.I.C.	Technologies de l'information et de la communication
TAL	Traitement automatique des langues
TaLC	Teaching and language corpora conference
TAO	Traduction assistée par ordinateur
TEI	Text-encoding initiative

Liste des figures

Figure 1	La marguerite de la didactique.
Figure 2	Mots-clés « intercultural, language learning » (1findr).
Figure 3	Mots-clés « translation, language learning » (1findr).
Figure 4	Le DaF comme point de départ des questionnements scientifiques.
Figure 5	Quelques repères chronologiques de l'évolution du TAL.
Figure 6	Étiquetage de « ohne » [sans] par TreeTagger.
Figure 7	Étiquettes erronées de TreeTagger.
Figure 8	Traduction automatique avec YoutubeStudio.
Figure 9	Plan d'un mémoire de master 2 Enseignement en didactique de l'allemand.
Figure 10	Suivi du mémoire de Kim Geheré à l'aide de Google documents.
Figure 11	Suivi du mémoire à l'aide de Trello.
Figure 12	Contenus du master « Recherches en didactique » de l'Inspé de Bretagne.
Figure 13	Initxm, site compagnon d'une formation pour doctorants.
Figure 14	Questionnaire avant la formation à TXM.
Figure 15	Lieux de communication.
Figure 16	Application du correcteur Antidote à une proposition de communication.
Figure 17	Insertion d'un document multimédia dans un article publié en ligne.
Figure 18	Moteur de recherche 1findr.
Figure 19	Outil de curation Feedly.
Figure 20	Outil de curation Diigo.
Figure 21	Blogue professionnel académique.

Figure 22	Entraide dans le groupe Facebook TXM.
Figure 23	Blogue pédagogique.
Figure 24	Konkrete Poesie.
Figure 25	Dessiner des chiens spécialisés.
Figure 26	Page d'accueil d'Edmodo.
Figure 27	Formulaire d'autorisation d'utilisation des données.
Figure 28	EXMARaLDA pour transcrire l'entretien du projet <i>Weissensee</i> .
Figure 29	Transcription d'un fichier audiovisuel avec oTranscribe.
Figure 30	Annotation d'un script d'audiodescription à l'aide de XML.
Figure 31	Fenêtre surgissante pour une étiquette TreeTagger (TXM).
Figure 32	Analyse quantitative des types de requêtes (projet <i>Kommas</i>).
Figure 33	Corpus Aufsätze sur Sketch Engine.
Figure 34	Comparaison de l'usage de la virgule en LE et LM (TXM).
Figure 35	VideoAnt pour annoter le générique de <i>Gundermann</i> .
Figure 36	<i>YouDescribe</i> pour publier une audiodescription en ligne.
Figure 37	La voix passive pour être neutre et objectif.
Figure 38	Interface de <i>deSkill</i> .
Figure 39	Trello pour gérer le projet <i>Kleinanzeigen</i> .
Figure 40	Trello pour gérer le TVseriesAD Translation Project.

Liste des graphiques

Graphique 1	Mes activités scientifiques en didactique et linguistique outillée.
Graphique 2	Freins perçus par rapport à l'usage du numérique en cours d'allemand.
Graphique 3	Thèmes des interactions verbales du projet <i>Weissensee</i> .
Graphique 4	Évolution de l'emploi de « an » par Cornélia (projet <i>Weissensee</i>).
Graphique 5	Termes « compliqué, problème, difficile » (<i>Weissensee Transfer</i>).
Graphique 6	Menschen mit Behinderung·en, behinderte Menschen, Behinderte.

Liste des tableaux

Tableau 1	Mes activités en lien avec des colloques de didactique en France.
Tableau 2	Mes activités en lien avec des revues de didactique en France.
Tableau 3	Entretiens menés entre 2005 et 2020.
Tableau 4	Données récoltées.
Tableau 5	Synopsis du module <i>Weissensee Transfer</i> .
Tableau 6	Planification du projet <i>Audiodeskription</i> .
Tableau 7	Distribution de passages à audiodécrire.
Tableau 8	Constitution du corpus <i>CCTP</i> .
Tableau 9	Feuille de travail avec lignes de concordance.
Tableau 10	Fonctions de l'adjectif.
Tableau 11	Composition du corpus Bundesblatt/Feuille fédérale/Foglio federale.
Tableau 12	Planning de TVseriesAD Translation Project.

Annexes

Annexe 1 – Formations continues suivies

1er juillet 2006 : Atelier d'initiation à Sketch Engine. Formateur : Adam Kilgarriff. TaLC7 (Teaching and language corpora conference), université Paris Diderot.

12 mars 2010 : Atelier d'initiation à COSMAS II (*Corpus Search, Management and Analysis System*) proposé par l'IDS (Institut für Deutsche Sprache) à Mannheim.

5-16 juillet 2010 : *CLARA Summer School on Advanced Resource Creation, Archiving and Usage*. Formateurs : Peter Wittenburg, Alexander König, Paul Trilsbeek, Hedda Lausberg, Onno Crasborn, Costanza Navarretta, Asli Özyürek, Paul Boersma, Stefano Masneri, Han Sloetjes, Thomas Schmidt, Taras Zakharko, Andreas Witt, Menzo Windhouwer, Jacquelyn Ringersma, Huib Verweij, Andre Moreira, Dieter van Uytvanck, Peter Withers, Dirk Roorda, Henk Harmsen. Max-Planck-Institut, Nijmegen (Hollande).

22 février 2011 : Tutorium der Sektion Computerlinguistik. Korpuslinguistik mit Online-Ressourcen — eine interaktive Einführung für Linguisten [Atelier « Linguistique de corpus basée sur des ressources en ligne - une introduction interactive pour linguistes »]. Formatrices : Stephanie Dipper et Heike Zinsmeister. Formation aux interfaces de IMS Open Corpus Workbench, CQPweb et DWDS. 33. Jahrestagung der DGfS (Deutsche Gesellschaft für Sprachwissenschaft), Georg-August-Universität Göttingen (Allemagne).

9-10 juin 2011 : Formation LODEL utilisateurs (logiciel d'édition électronique). Formateur : Pierre Mounier. Organisée par le Cléo (Centre pour l'édition électronique ouverte), Paris.

Mai - juin 2014 : Mooc « Monter un Mooc de A à Z » (plateforme FUN).

2012/2013 - Séminaire doctoral « Informatique pour la recherche. Approches textométriques, ergonomies numériques ». Formateur : Jean-Marc Leblanc. Université Paris-Est Créteil Val de Marne.

2-3 décembre 2014 : Ateliers TXM : « Initiation à TXM » (Formatrice : Bénédicte Pincemin) et « Préparation de corpus et import dans TXM » (Formateur : Serge Heiden), ENS Lyon.

26-29 mai 2015 : Initiation à l'édition électronique et la pratique de l'encodage XML/TEI (formateurs : Lou Burnard, Emmanuel Château). École nationale des chartes.

22 et 29 janvier 2016 et 25 mars 2016 : Ateliers AntConc et TXM, organisés par le laboratoire CeLiSo (Centre de linguistique en Sorbonne). Formateurs : Daniel Henkel, Pierre Labrosse et Kim Oger.

19-24 juin 2016 : École thématique CNRS *MISAT* - Méthodes informatiques et statistiques en analyse des textes. Équipe : Jules Duchastel, Damon Mayaffre, Serge Heiden, Lou Burnard, Jean-Marie Viprey, Ludovic Lebart, Bénédicte Pincemin, Pierre Ratinaud, Margareta Kastberg, Virginie Lethier. Besançon.

20 juillet 2016 : Introduction to statistics for linguistics with R. Formateur : Christoph Wolk. Pre-conference workshop TaLC12.

22-23 février 2017 : Formation LODEL utilisateurs (logiciel d'édition électronique). Formatrices : Mélanie Carmona et Élodie Picard. Organisé par OpenEdition. Paris.

19-20 juin 2017 et 11 septembre 2017 : « Initiation à R ». Formateur : Dylan Glynn. Formation organisée par le laboratoire CeLiSo (Sorbonne Université).

18 juillet 2018 : Using corpora to teach sociolinguistics: A practical workshop. Introduction into #LancsBox. Formateur et formatrices : Vaclav Brezina, Dana Gablasova et Irene Marín Cervantes. Pre-conference workshop, *TaLC13*, Cambridge (Angleterre).

Annexe 2 – Expertise scientifique

Membre du comité scientifique de revues

Alsic (Apprentissage des Langues et Systèmes d'Information et de Communication)

ÉLIS (Échanges Linguistiques en Sorbonne)

Linguistik online

Phrasis (Associazione Italiana di Fraseologia & Paremiologia)

RPPLSP (Recherche et pratiques pédagogiques en langues de spécialité)

Revue *CORPUS*, n° 21 : « Corpus complexes. Traitements, standardisation et analyse des corpus de communication médiée par les réseaux ». Sous la direction de Céline Poudat, Ciara R. Wigham et Loïc Liégeois.

Membre du comité scientifique de colloques

2ème colloque du PERL : Entre présence et distance. Enseigner et apprendre les langues à l'université à l'ère numérique. Colloque International organisé par le PERL, Université de Paris, IPGP (Institut de Physique du Globe de Paris), 12-13-14 décembre 2019.

VIe Congrès International de Phraséologie et Parémiologie, Université de Roma Tre (Italie), 11-13 septembre 2019.

23e séminaire de L'École doctorale francophone des Pays de Visegrád : « Entre tradition et innovation dans les Études françaises en Europe centrale », Bratislava (Slovaquie), 14-16 novembre 2019.

VIe Congrès International de Phraséologie et Parémiologie, Université de Roma Tre (Italie), 11-13 septembre 2019.

8e Conférence internationale Pragmatique et communication interculturelles. Université de Chypre, Aglantzia, 8-19 juin 2018.

Aranea 2018, Corpus web - un outil accessible à l'apprentissage des langues. Conférence organisée par la Chaire UNESCO de l'Université Comenius de Bratislava (Slovaquie). 23-24 novembre 2018.

Congrès Entre présence et distance - Accompagner et découvrir des pratiques pédagogiques en langue à l'université. 16-18 novembre 2017, Université Sorbonne Paris Cité.

CEDIL 2014 ; CEDIL 2018 : « Qu'a changé le numérique ? Perspectives linguistiques et didactiques » (Colloque international des étudiants chercheurs en didactique des langues et en linguistique).

Annexe 3 – Événements scientifiques animés ou (co-)organisés

Congrès international Eurocall, 25-27 août 2021, CNAM, Paris.

Modération de la table ronde « Linguistique et variation » du 7 juin 2018 lors du 51e Congrès de l'Association des Germanistes de l'Enseignement Supérieur « Voyages, migrations, mobilités », 7-9 juin 2018, Institut Goethe, Paris.

XXIV^e congrès du Rassemblement National des Centres de Langues de l'Enseignement Supérieur (RANACLES), 24-26 novembre 2016, Espé de l'académie de Paris – École interne de l'Université Paris-Sorbonne.

2ème responsable de la section *Reale und virtuelle Lernumgebungen* [Environnements d'apprentissage réels et virtuels]. Colloque IDT (Internationale Deutschlehrertagung) 2013, Bolzano (Italie). Blogue faisant suite au colloque.

Journée d'étude *Korpusbasierte Sprachreflexion* [Corpus et réflexion sur la langue], Paris-Sorbonne, Centre Malesherbes, le 16 juin 2012, organisée en coopération avec Prof. Dr. Martine Dalmas (Paris-Sorbonne). Matin : exposés. Après-midi : initiation à l'exploitation de corpus numériques pour la réflexion sur la langue.

Journée d'étude de l'ACEDLE (Association des Chercheurs et Enseignants Didacticiens des Langues) du 4 juin 2010 à IUFM de l'académie de Paris. Avec Muriel Grosbois.

Journée d'étude de l'ACEDLE *Recherches émergentes en didactique des langues* du 12 juin 2009 à IUFM de l'académie de Paris, organisée en coopération avec Muriel Grosbois.

Annexe 4 – Communications

Conférences invitées⁴⁴

2018 :

(C-KEY). « Contours et potentiel des corpus web pour l'enseignement-apprentissage d'une langue étrangère ». Conférence internationale *ARANEA 2018* : « Les corpus web comme instruments de formation linguistique ». Faculté de lettres et sciences humaines de l'Université Comenius de Bratislava (Slovaquie), 23-24 novembre 2018.

(Invitation par un laboratoire de recherche externe). « Se servir des corpus d'apprenants et des corpus d'experts pour former en langue étrangère ». Journée d'étude « Lier recherche et formation professionnelle : des corpus à l'interface ». Maison des Sciences de l'Homme, Clermont-Ferrand, 16 novembre 2018.

(Invitation à une journée d'étude sans sélection). « Corpus complexes et Apprentissage des Langues Médié par les Technologies (ALMT) – quoi, pourquoi, comment ? ». Traitements et

⁴⁴ Le classement dans cette section s'inspire du Guide des produits de la recherche. Sous-domaine : « Esprit humain, langage, éducation ». Panel disciplinaire : « Linguistique » (Hcéres).

standardisation des corpus multimodaux et web 2.0 – Une journée hommage à Thierry Chanier. Grands moulins, Université Paris Diderot Paris 7, 25 mai 2018.

2014 :

(Invitation à une journée d'étude sans sélection). « Utiliser les corpus avec des apprenants Lansad ». Séminaire Le numérique pour l'enseignement et l'apprentissage des langues pour non-spécialistes, organisé par le comité de rédaction *Alsic* et l'Université Paul Valéry, Montpellier 3, sur initiative de la Mission numérique pour l'enseignement supérieur (Mines - MESR). Université Paul Valéry, Montpellier, 12-13 décembre 2013. Toutes les communications du séminaire peuvent être visionnées sur Canal-U.

2009 :

(Invitation par un laboratoire de recherche externe). « Entre perception et réflexion. Enseigner-apprendre une langue étrangère ». Intervention lors du séminaire doctoral *Approches cognitives en linguistique* dirigé par Martine Dalmas. Université Paris 4, 30 octobre 2009.

(Invitation par un laboratoire de recherche externe). « Mener une recherche-action en cours d'allemand ». 2ème journée du séminaire *Langage, société, communication, didactique FLM/FLS/FLE et situations de francophonie. Faire de la recherche-action aujourd'hui en sciences humaines : fondements épistémologiques, concepts et méthodes*. Université Cergy-Pontoise, 20 novembre 2008.

Colloques à portée internationale⁴⁵

2020 :

(Proposition soumise avec Nathalie Mälzer et Saskia Josephine Schulz). “Purpose and features of the TVseriesAD corpus”. 3rd Swiss Conference on Barrier-free Communication. ZHAW Zürcher Hochschule für Angewandte Wissenschaften, Winterthur, 5-6 juin 2020.

(Proposition acceptée). “Designing a sales ad writing project for young learners of German: Do corpus tools make a difference?”. AILA World Congress of Applied Linguistics. Groningen (Pays-Bas), 9-14 août 2020.

⁴⁵ Définition retenue : un colloque à portée internationale possède un comité de lecture majoritairement composé de chercheurs internationaux.

2019 :

Avec Isabelle Salengros Iguenane. « Structurer un module de FOS (Français sur objectifs spécifiques) dédié à un écrit professionnel à l'aide d'analyses du corpus CCTP (Cahier des clauses techniques particulières) ». 2ème colloque du PERL : Entre présence et distance. Enseigner et apprendre les langues à l'université à l'ère numérique. Organisé par le PERL, Université de Paris & USPC. Paris, 12-13-14 décembre 2019.

2018 :

“Language learning from an audio description corpus”. *TaLC13* 2018, Faculty of Education, Cambridge (Angleterre), 18-21 juillet 2018.

2017 :

« Apprivoiser les virgules en allemand – Une approche sur corpus ». 9èmes Journées internationales de la Linguistique de Corpus (JLC2017). Université Grenoble-Alpes, 3-6 juillet 2017.

2016 :

“Which corpus management systems for non-specialists in computer science? The case of French trainee teachers of German”. *TaLC12* (12th Teaching and Language Corpora Conference). Université de Gießen (Allemagne), 20-23 juillet 2016.

« Mais où sont donc les femmes ? À la recherche de formes doubles dans un corpus de textes administratifs suisses ». JADT 2016 (Journées internationales d'Analyse statistique des Données Textuelles). Espace séminaire Le Saint Paul, Nice. 7-10 juin 2016.

„Umsetzung sprachlicher Vorgaben im Bereich der geschlechtergerechten Personenbezeichnungen in Schweizer Behördentexten“. Beitrag für das Panel „Vielfalt in der Einheit: Schweizerische Verwaltungssprache“ an der Tagung der VALS-ASLA (Vereinigung für angewandte Linguistik in der Schweiz) 2016. Université de Genève (Suisse), 20–22 janvier 2016.

2015 :

« Le commentaire collectif numérique comme activité d'enseignement-apprentissage d'une langue étrangère ». Colloque international interdisciplinaire *Le commentaire : du manuscrit à la toile. Modes d'interventions et dispositifs techniques*. Université libre de Bruxelles (Belgique), 18-20 mars 2015.

2014 :

“Talking about German verb particles and prepositions identified in concordance lines”. 12th International Conference - Association for Language Awareness. Hedmark University College, Hamar (Norvège), 1-4 juillet 2014.

2013 :

« Énonciation, interaction et corpus au service de l'apprentissage de l'allemand, langue étrangère ». Colloque Cultures de recherche en linguistique appliquée. Nancy, 14-16 novembre 2013.

„Korpusbasierte Sprachrätsel im Fremdsprachenunterricht“. 2. Sektionentagung der Gesellschaft für Angewandte Linguistik e.V. Rheinisch-Westfaelische Technische Hochschule, Aix-la-Chapelle (Allemagne, 19-20 septembre 2013.

2012 :

« Observer les caractéristiques linguistiques de genres textuels pour réviser des textes en allemand, langue étrangère. Une approche basée sur des corpus en ligne. Colloque international Former à l'écrit universitaire : un terrain pour la linguistique ?. Organisé par l'Université de Paris Ouest Nanterre La Défense et par l'Université de Chicago. Nanterre, 22-23 novembre 2012.

“Corpus-based metalinguistic discussions designed as writing pre-tasks”. TaLC10 (10th Teaching and Language Corpora Conference). Université de Varsovie (Pologne), 11-14 juillet 2012.

« Pour un apprenant réfléchissant », présentation à trois voix avec Elsa Chachkine et Françoise Demaizière. Colloque *Acedle 2012 - Recherches en didactique des langues - Apprendre les langues autrement*. Organisé par le Centre International des Langues, Université de Nantes, 6-9 juin 2012.

2011 :

„Methodisch-didaktisch motivierter Einsatz von Online-Tools zur Vorbereitung der Zertifizierung CLES2 im Bereich Deutsch als Fremdsprache“. Sektion 5: Medien und E-Learning (Pr.Dr. Langner). 1. Saarbrücker Fremdsprachentagung: Hochschulischer Fremdsprachenunterricht - Anforderungen - Ausrichtung - Spezifik [1er congrès des langues étrangères de l'université de Saarbruck : enseignement des langues étrangères - besoins -

tendances - situations spécifiques], Université de Saarbruck (Allemagne), 4-5 novembre 2011.

“Foreign language text revision within a top-down and bottom-up framework. Taking keywords in context as a starting-point for language observation”. 33ième congrès annuel de la DGfS (Deutsche Gesellschaft für Sprachwissenschaft), Göttingen : Text: Strukturen und Verarbeitung. Arbeitsgemeinschaft 5: Prozesse der Textproduktion in der Schule: Strukturen und Verarbeitung aus sprachdidaktischer und psycholinguistischer Perspektive [Le texte : structures et traitement. Processus de production de texte en milieu scolaire]. Université de Göttingen (Allemagne), 23-25 février 2011.

2010 :

« Qu’est-ce qui rend les corpus pédagogiques ? ». Eurocall 2010 (Colloque international : Les structures en réseaux et les communautés virtuelles dans l’apprentissage de la langue médiatisé par les TIC). Université de Bordeaux, 8-11 septembre 2010.

« Création de corpus réfléchis pour l’enseignement de la production écrite. L’exemple de l’allemand, langue étrangère ». *CEDIL 2010* (Colloque international des jeunes chercheurs En Didactique des langues et en Linguistique). Laboratoire LIDILEM, Université Stendhal, Grenoble, 29 juin-2 juillet 2010.

“Constructivist writing tasks based on corpora and concordance lines”. *IVACS 2010* (The Fifth Inter-Varietal Applied Corpus Studies Conference on Connecting Corpus Linguistics). Université d’Edinburgh (Angleterre), 18-19 juin 2010.

2008 :

« Faire pratiquer les corpus au collège lors d’un projet d’écriture en allemand ». 16ème congrès de RANACLÈS (RASsemblement NAtional des Centres de Langues de L’Enseignement Supérieur). Université de Nice Sophia-Antipolis. *Didactique et pratique des langues, quelles priorités ?*. Nice, 20-22 novembre 2008.

« Corpus et concordances au service de nouvelles pratiques de formulation en L2 ». Colloque *ACEDLE* (Association des Chercheurs et Enseignants Didacticiens des Langues Étrangères) 2008 - *Recherches en didactique des langues*. Strasbourg, 17-19 janvier 2008.

2007 :

« Exploitation de corpus lors de la production d'écrits électroniques en L2 ». 5^{ème} Journées de la Linguistique de Corpus. Université de Bretagne Sud, Lorient, 13-15 septembre 2007.

Colloques et journées d'étude à portée nationale

« Caractéristiques multimodales de corpus pédagogiques ». Journée d'étude de l'axe 3 « didactique » du laboratoire CeLiSo (Centre de Linguistique en Sorbonne) : « La multimodalité dans l'enseignement-apprentissage des langues : quels enjeux ? ». Espé de Paris, 7 décembre 2018.

Avec Thierry Grass et Gilbert Magnus. « Usages variés de l'outil d'exploration de corpus TXM dans le contexte de l'allemand ». Congrès *Les 50 ans de l'Association des Germanistes de l'Enseignement Supérieur (AGES)*. Université de Nantes, 8-10 juin 2017.

2014 :

« Exploiter les corpus allemands pour soutenir le développement d'habiletés cognitives ». Journée d'étude du GERALS (Groupe d'Etude et de Recherche en ALlemand de Spécialité) - *Les corpus et l'allemand de spécialité*. Université Le Havre, 7 novembre 2014.

2016 :

« Les futurs enseignants de langues face aux corpus numériques ». XXIV^e congrès RANACLÈS. 24-26 novembre 2016, Espé de l'académie de Paris – École interne de l'Université Paris-Sorbonne.

2008 :

« L'intertextualité comme procédé d'écriture en L2 ». Journée d'études *ACEDLE* (Association des Chercheurs et Enseignants Didacticiens des Langues Étrangères) : *Recherches émergentes en didactique des langues*. IUFM de Paris, 13 juin 2008.

2007 :

« Production écrite en L2 à l'aide de concordances. Création d'un dépliant touristique en allemand ». 7^{ème} colloque *UNTELE* (Usages des Nouvelles Technologies dans l'Enseignement des Langues Étrangères) : *La communication transculturelle, le réseau global, et l'acquisition des langues secondes*. Université de technologie Compiègne, 31 mars 2007.

2006 :

« Les corpus pédagogiques au service de l'apprentissage de la production écrite de l'allemand au collège et à l'université ». Journée d'étude *ATALA* (Association pour le Traitement Automatique des Langues) : *Aide à la rédaction écrite*. Paris, 3 juin 2006.

Annexe 5 – Engagements pour des sociétés savantes

ACEDLE (Association des chercheurs et enseignants didacticiens des langues étrangères)

Adalsic (Association de la revue ALSIC)

AGES (Association des Germanistes de l'Enseignement Supérieur)

ALA (Association for Language Awareness)

Corli (Consortium Huma-Num Corpus, Langues et Interactions)

ELN (European Literacy Network, COST Action IS1401)

GERALS (Groupe d'étude et de recherche en allemand de spécialité)

Annexe 6 – Corpus mentionnés dans le rapport de synthèse

Aufsaetze. Corpus d'experts et corpus d'apprenants contenant des résumés et commentaires d'articles pédagogiques.

Buettenwarder. Corpus contenant 69 manuscrits d'audiodescriptions de la série télévisée *Neues aus Büttewarder* (Eberlein, 1997-2017).

Bundesblatt/Feuille fédérale/Foglio federale. Corpus contenant l'intégralité des feuilles fédérales publiées depuis 1849 en allemand et français et depuis 1971 en italien par la Chancellerie fédérale entre 1849 et 2014.

Eden. Corpus contenant les scripts AD de la mini-série *Eden* (Berger et al., 2019).

Instruktionen. Corpus contenant les sous-corpus suivants : *Buettenwarder*, *Heidiromane*⁴⁶, *Schleier*⁴⁷ et *Weissensee*.

⁴⁶ http://www.deutschestextarchiv.de/book/show/spyri_heidi_1880

⁴⁷ <http://www.gutenberg.org/ebooks/31842>

TVseriesAD corpus. Corpus contenant le corpus Buettenwarder, Dahoam et Eden.

Weissensee. Corpus contenant le script de la première saison de la série *Weissensee* ==> Hess 2010.

Annexe 7 – Traductions

2018 – Création des sous-titres allemands du court-métrage *Radio Lune* de Pierre Gaffié.

2018 – Traduction de l’anglais vers le français de l’interface de *Sketch Engine*.

2017 – Contribution à la traduction de l’anglais vers le français du thème Wordpress *Dynamic News* de ThemeZee.

1995-1998 - Traduction de l’allemand vers le français de la quatrième de couverture et du sommaire de livres appartenant aux rubriques psychologie, spiritualité, bien-être. Éditions Dervy, Librairie de Médicis (groupe Albin Michel).

1993 – Contribution à la création d’entrées pour le *Dictionnaire de poche Français-Allemand*. Paris : Larousse-Bordas.

Annexe 8 – Direction de mémoires de master 2

Cette liste informe sur ma direction de mémoires master enseignement (allemand) ou master recherche (études germaniques) puis master MEEF parcours allemand. « FSTG » veut dire « formation de stagiaires ». Cette formation concerne les personnes ayant fait un master 1 dans un autre domaine que l’enseignement. Un des mémoires s’inscrit dans le master 2 MuM (Medientext und Medienübersetzung) [Sciences et traduction des médias] de l’université de Hildesheim (Allemagne). Je le co-dirige avec Nathalie Mälzer, professeure des universités. Les quatre premiers mémoires sont en cours de réalisation, et les autres ont été soutenus. Afin d’informer et d’accompagner les étudiants et étudiantes, j’ai mis en place d’un espace numérique qui est régulièrement réactualisé⁴⁸. Le master 2 enseignement est abrégé en "Ens.".

⁴⁸ <https://frama.link/allemandmemoires>

Année	Prénom NOM	Master 2	Titre du mémoire
2019-2020	Léana DEBRIS	MuM	Caractéristiques linguistiques des audiodescriptions en langue allemande et française de la série Eden
2019-2020	Solène BONNEREAU	MEEF	Apports du jeu Classcraft pour l'apprentissage institutionnel de l'allemand
2019-2020	Kim GEHERE	MEEF	Quelle intégration en classe d'allemand des élèves en situation de handicap ?
2019-2020	Véronique LEBRUN	MEEF	La pré-tâche comme élément de différenciation pédagogique en allemand langue étrangère
2018-2019	Fabiola CLAIN	MEEF	Le rôle du plurilinguisme dans la compréhension de l'écrit
2018-2019	Laura GERHARD	MEEF	Wortschatzarbeit (DaF-Bereich) außerhalb des Unterrichts mithilfe des Online-Lerntools Quizlet [Le lexique en allemand langue étrangère appris en dehors de la classe à l'aide de l'outil d'apprentissage Quizlet]
2018-2019	Maiwenn MÉNEC	MEEF	Corps, mouvement et apprentissage
2016-2017	Lydia PINTART	MEEF	Lernen durch Lehren (LdL) [Apprendre en enseignant] – Une “ancienne” méthode qui mérite du renouvellement
2016-2017	Delphine PUCHEU	MEEF	L'apprentissage de l'écrit au cycle terminal : comment encourager la lecture suivie ?
2015-2016	Elisabeth ARDIN	FSTG	La place de l'enregistrement vocal en cours de langue vivante – La réalisation d'une pièce radiophonique en allemand
2015-2016	Julie COULOMBEL	MEEF	L'utilisation des contes en cours d'allemand au lycée – Théorie, pratique et résultats
2015-2016	Caroline PUAUD	FSTG	Schachnovelle [Le joueur d'échecs] – De la lecture littéraire à l'adaptation filmique
2015-2016	Larissa WAIDOSCH	FSTG	Les clichés

2014-2015	Heidi SIMÉANT	MEEF	Pistes pour une interaction spontanée fluide en cours d'allemand
2014-2015	Ria VARGA	MEEF	Motivation des élèves par des projets collectifs
2014-2015	Marion WIMMER	FSTG	War Erziehung früher anders?
2012-2013	Harilandy RAKOTORA-HALAHY	Ens.	Analyse d'erreurs syntaxiques : La place des verbes
2012-2013	Gabriel RAUSCHERT	Ens.	L'enseignement de la DNL (Discipline non linguistique)
2012-2013	Pierre-François SMETS	Ens.	Le support audiovisuel au service de la compréhension de l'oral dans un contexte institutionnel – le cas de la langue étrangère allemande en Palier 2
2011-2012	Laurence BENAZET	Ens.	L'exploitation d'un texte écrit en classe : différentes méthodologies d'approche
2011-2012	Dorothea BLUMENBERG	Ens.	L'école publique d'excellence – L'école publique parfois « plus égale » que d'autres
2011-2012	Aliette DELORME	Ens.	L'espace scolaire, un espace-ressource
2011-2012	Sarah DJADAOUJEE	Ens.	La place de la religion en cours d'allemand, langue étrangère, au sein d'un enseignement public
2011-2012	Alice MAURY	Ens.	La place du jeu en cours de langue
2011-2012	Mélanie MUSTO	Ens.	Les groupes de compétences dans l'apprentissage des langues vivantes
2011-2012	Svetlana NIKIFORA	Ens.	Tâches pour les élèves artistiques, les futurs musiciens, chanteurs, danseurs, pendant l'apprentissage de la langue
2011-2012	Nadine SCHWEITZER	Ens.	L'importance de la prise de note
2011-2012	Christina ZIPPER	Ens.	La motivation dans le cadre de l'enseignement de l'allemand

Résumé: *Recherches et pratiques outillées en allemand et français*

Ce rapport de synthèse s'intéresse au rôle occupé par les outils numériques pour la mise en synergie de la didactique des langues, la linguistique de corpus et la traduction. Le point de vue central est celui de la didactique, considérée comme une praxéologie faisant appel à des domaines de référence qui varient en fonction des contextes et questions de recherche. Dans le premier chapitre, les aspects théoriques et pratiques des entrées « didactique, linguistique, TAL et traduction » sont abordés. Dans la section 1.1, la didactique des langues est considérée comme le dénominateur commun entre la linguistique et le domaine pratique représenté par l'enseignement d'une langue étrangère. La section 1.2 s'intéresse au TAL (traitement automatique des langues), compris comme un domaine de recherche qui a comme pendant pratique la linguistique outillée. La linguistique de corpus y est présentée comme dénominateur commun entre les deux domaines. La section 1.3 présente la traductologie comme le versant scientifique d'un domaine dont les pendants pratiques sont le secteur de l'enseignement de la traduction et la pratique professionnelle de la traduction. Le deuxième chapitre est dédié au rôle que jouent la recherche et l'outillage dans le secteur de la formation. Le premier volet s'intéresse à la formation initiale des enseignants et des didacticiens. Il plaide pour une prise en compte d'activités de recherche au sein de cursus universitaires. Le deuxième volet présente les formations continues dans le domaine des outils numériques et de la linguistique outillée, qui peuvent intéresser des didacticiens, enseignants, linguistes et traducteurs. Le troisième chapitre retrace mon parcours de chercheuse, qui est mis en perspective à l'aide d'un aperçu des organes français de la recherche en didactique des langues. J'y présente les revues et colloques qui me paraissent centraux pour ce secteur, et je mentionne mon implication dans ce domaine. De plus, on trouve dans ce chapitre un bilan des activités de recherche que j'ai menées pendant ma carrière universitaire. Le quatrième chapitre est dédié à la méthodologie appliquée à des recherches impliquant le terrain. J'y décris la façon dont les outils numériques sont mis au service de l'observation, de la récolte et de l'analyse de données. Les outils de corpus sont mentionnés dans la quatrième, cinquième et sixième section de ce chapitre, en particulier en lien avec des procédés d'annotation et d'analyse quantitative et en lien avec les corpus d'experts et les corpus d'apprenants. Le cinquième chapitre présente les projets outillés que j'ai conçus et/ou auxquels j'ai participé. Il est divisé en trois sections, faisant toutes appel aux corpus numériques. La première est adossée à la didactique des langues, la deuxième à la linguistique et la troisième à la traduction. Le sixième chapitre propose des pistes pour des actions et évolutions futures. La prise en compte de besoins sociétaux sera accentuée grâce à la poursuite de ma volonté d'insertion dans le secteur de l'inclusion par le biais de recherches portant sur des outils numériques développés pour répondre à des besoins dits « spécifiques ».

Abstract: *Digital tools for research and practice in German and French*

This habilitation thesis focuses on the role played by ICT (Information and Communication Technologies) in bringing together applied linguistics, corpus linguistics, and translation. The central point of view is that of applied linguistics, considered as a praxeology involving reference fields that vary according to one's research context and questions. In the first chapter, theoretical and practical aspects of the topics "applied linguistics, linguistics, automatic language processing and translation" are addressed. In section 1.1, the field of applied linguistics is considered as the common denominator between linguistics and foreign language teaching. Section 1.2 focuses on automatic language processing, understood as a research field for which digital linguistics represents the practical counterpart. Corpus linguistics is portrayed as a common denominator between these two fields. In section 1.3, translation science is presented as the scientific branch of a field whose practical counterparts are the translation education sector and professional translation. The second chapter is dedicated to the role played by research and tools in the training sector. Section 2.1 focuses on the initial training of teachers and applied linguists. It argues for research activities to be taken into account in university courses. Section 2.2 presents workshops in the field of digital tools and ICT-supported linguistics that may be of interest to applied linguists, teachers, general linguists and translators. The third chapter retraces my career as a researcher, which is put into perspective with an overview of the French research bodies in applied linguistics. In this part, I present the journals and the conferences that I consider central to this sector, and I mention my involvement in this field. In addition, this section provides an overview of the research activities I have carried out during my academic career. The fourth chapter is dedicated to the methodology applied to research involving fieldwork. In this part, I explain how digital tools can be used for observation, data collection and analysis. Corpus-related topics are mentioned in the fourth, fifth and sixth sections of this chapter: annotation, quantitative analysis processes, expert corpora and learner corpora. The fifth chapter presents the ICT projects I have designed and/or participated in. The first is related to applied linguistics, the second to linguistics and the third to translation. The sixth chapter suggests ways forward for future actions and developments. The consideration of societal needs will be further accentuated thanks to a targeted evolution towards the inclusion sector, represented by digital tools developed to meet so-called special needs.