

**UNIVERSITÉ SORBONNE NOUVELLE – PARIS 3**

ÉCOLE DOCTORALE 268

Thèse en sciences du langage

**Eva SCHAEFFER-LACROIX**

**CORPUS NUMÉRIQUES  
ET PRODUCTION ÉCRITE  
EN LANGUE ÉTRANGÈRE**

UNE RECHERCHE AVEC DES  
APPRENANTS D'ALLEMAND

**Thèse dirigée par Françoise DEMAIZIÈRE et André SALEM**

Soutenue le 28 novembre 2009

**Jury :**

Johannes Angermüller, professeur (Johannes-Gutenberg-Universität Mainz), rapporteur

Martine Dalmas, professeur (université Paris 4), rapporteur

Françoise Demaizière, ingénieur de recherche HDR (université Paris 7), experte invitée

Dominique Macaire, maître de conférence HDR (université Paris 3, IUFM d'Aquitaine, Bordeaux 4)

André Salem, professeur émérite (université Paris 3), directeur de thèse

## Résumé

Cette thèse en didactique des langues étrangères montre le potentiel des corpus pour l'apprentissage de l'allemand au collège. Conformément aux objectifs poursuivis dans une recherche-action, elle renseigne sur l'évolution des acteurs impliqués dans un dispositif spécifique, en tenant compte de facteurs multiples, tels que le type d'*input*, les outils, les modes de travail - en présentiel et à distance - et la relation pédagogique. Le scénario pédagogique est basé sur l'exploitation de corpus spécialisés pour la création de textes du domaine du tourisme. Le projet et les outils proposés ont soutenu le développement de modes de lecture qui ont favorisé la recherche ciblée d'informations. Le recyclage d'éléments repérés dans l'*input* a mené, pour une partie des apprenants, à l'obtention de productions écrites dont les caractéristiques linguistiques et discursives sont proches de celles attestées dans les corpus. L'écriture intertextuelle a permis aux participants, à des degrés variables, de se décentrer de leur langue maternelle et de progresser dans l'apprentissage de l'altérité. L'observation de formes linguistiques à l'aide de lignes de concordances et l'intégration de collocations identifiées dans l'*input* ont contribué au développement du système linguistique, en particulier dans le domaine de la flexion des adjectifs.

Le premier chapitre de la thèse est consacré aux corpus et concordances. Il les distingue d'objets qui leur sont proches, et il évoque les objectifs liés au travail sur corpus. J'y propose un état des lieux concernant les ressources et outils pour la langue allemande. Le deuxième chapitre s'intéresse aux recherches dans le domaine de l'apprentissage et de l'enseignement des langues étrangères. Il présente une palette de points de vue oscillant entre le bord linguistique et le bord culturel, et il décrit des méthodologies d'enseignement/apprentissage qui sont en accord avec ma recherche. Le troisième chapitre traite de la production écrite. J'y présente, entre autres, des modèles d'écriture et le rôle de l'*input* et de l'*output* dans la production écrite. La dernière partie de ce chapitre s'interroge sur le lien entre les outils d'aide à la rédaction et les ressources pédagogiques. Le quatrième chapitre reprend des pistes proposées dans les trois chapitres précédents afin d'évaluer le potentiel des corpus et concordances pour l'apprentissage de la production écrite en LE (Langue Étrangère). L'analyse d'activités existantes, basées sur des concordances, et la présentation d'une activité que j'ai développée terminent ce chapitre. Le chapitre 5 situe ma recherche par rapport à d'autres types de recherche. Il présente les conditions de mise en place d'une recherche-action

dans un établissement du secondaire. J'y décris mon scénario, les outils que j'ai proposés aux apprenants, ainsi que les activités préparatoires et d'entraînement. Le sixième chapitre propose l'analyse des données rassemblées lors de ma recherche. J'y étudie le discours et les actions des apprenants vis-à-vis des corpus et concordances et j'observe l'évolution de certaines caractéristiques de leurs productions écrites. Dans ce chapitre, je m'intéresse également aux traces qui permettent d'avancer que mon projet à amené les apprenants à intégrer de nouveaux aspects linguistiques et culturels dans leur système de représentation de la langue allemande. Le chapitre 7 regroupe les éléments de réponse à mes questions qui débouchent, dans le chapitre 8, sur l'esquisse de chantiers de recherche futurs dans le domaine de l'apprentissage de LE à l'aide de corpus.

Jusqu'à présent, peu d'études ont décrit les effets de l'exploitation des corpus lors de projets d'écriture en langue étrangère auxquels participent des scripteurs de niveau intermédiaire (Braun, 2007). Les données empiriques attestant de l'utilisation des corpus par les apprenants d'allemand, langue étrangère, sont particulièrement rares. Les matériaux d'apprentissage, basés sur des corpus, font également défaut pour cette langue (Lüdeling et Walter, 2009 : 7). La quasi-absence de méthodes et de matériaux conçus pour l'allemand a motivé mon choix de m'inspirer de recherches menées en didactique du français écrit, langue maternelle, concernant l'utilisation de bases de données textuelles comme aide à la rédaction. Crinon, Legros et Marin (2002b : 187-188) nomment deux conditions qui déterminent la qualité de la production écrite en LM (Langue Maternelle) : il convient de mettre à disposition des apprenants un nombre important de textes. Ces textes doivent être compatibles avec ceux que les apprenants sont censés produire, c'est-à-dire qu'ils doivent appartenir au même genre textuel que celui représenté dans la base de données. Les corpus semblent répondre à cette double demande. Crinon, Legros et Marin (2002b : 187) plaident pour une nouvelle attitude de consultation de textes, celle de "chercheur d'informations". Ils dressent un parallèle entre cette attitude et une démarche d'écriture intertextuelle : les scripteurs sont encouragés à chercher des éléments pertinents dans les bases de données afin d'enrichir leur production écrite. Jorro (2004 : 157) précise que le développement d'une telle attitude dépend du type d'accompagnement qu'un scripteur expert offre à un scripteur novice. Dans le domaine de la linguistique de corpus, le concept d'apprenant-chercheur, présenté, entre autres, dans les travaux de Johns (1991), fait écho à ce qui précède. J'ai croisé ce concept de type constructiviste avec les apports de la psycholinguistique concernant la recherche

d'informations (Rouet *et al.*, 2004) et avec la technique de recyclage d'éléments repérés dans les corpus.

Mes questions de recherche sont les suivantes.

Quels effets peuvent être constatés quand on propose à des apprenants adolescents, ayant un niveau intermédiaire en langue étrangère, un scénario de type constructiviste qui conçoit les corpus spécialisés comme aide à la rédaction ? Peut-on observer l'émergence d'une attitude d'apprenant-chercheur, susceptible de soutenir l'exploitation de l'*input* lors de la rédaction ? Le scénario permet-il d'obtenir des productions dont les caractéristiques linguistiques et discursives reflètent les normes attestées dans les corpus ? Le projet, aide-t-il les apprenants à se décentrer de leur langue maternelle et de leurs représentations de la LE ? Mène-t-il au développement de leur système linguistique ?

Afin de pouvoir apporter des éléments de réponse à ces questions, j'ai mis en place une recherche-action d'une durée de trois ans. Les constantes ont été l'âge du public (à part pour un des groupes classe) et l'utilisation de corpus spécialisés et de lignes de concordances lors de l'exécution de tâches de production écrite.

Mon public s'est composé de quatre groupes d'apprenants, ayant entre 14 et 15 ans et d'un groupe d'apprenants de 10 à 11 ans. Lors de l'exécution des projets, trois des groupes classe étaient en quatrième année d'apprentissage, un groupe en troisième année et un autre en première année. Les séances en présentiel ont eu lieu dans le collège où je travaille comme enseignante d'allemand titulaire. En complément des séances en présentiel, les apprenants avaient l'occasion de travailler à distance et de me contacter en dehors des cours en présentiel s'ils avaient besoin d'aide. L'objet de l'enseignement/apprentissage du projet a été la révision individualisée des textes rédigés par les apprenants. J'ai proposé comme méthode de travail la sélection de segments pertinents dans un *input* écrit (corpus et concordances), correspondant au types de textes que les apprenants ont été censés produire, et le réemploi (modifié ou non modifié) des segments lors de la révision. Les références et méthodologies principales qui sous-tendent ma recherche-action sont les suivantes : la recherche d'informations (Tricot, 2007 ; Rouet *et al.*, 2004), la théorie du repérage de Schmidt (identification de collocations et traitement en LE), l'approche par tâches (R. Ellis, 2003), l'idée que les connaissances se

construisent individuellement, mais aussi à l'aide d'autres personnes (Piaget, 1969 ; Bruner, 1983), le *data driven approach* (Johns, 1991) et la conceptualisation (Chini, 2004).

Ma recherche s'est déroulée en trois phases. La première, la phase d'exploration, a eu lieu en 2006/07. Elle a servi à observer la façon dont les treize apprenants du groupe révisaient leurs productions écrites lors du projet *Création de dépliants touristiques*. À partir des données obtenues lors de cette phase, j'ai développé le scénario pour *Projekt Prospekt*, projet principal de ma recherche, que j'ai soumis à un groupe de 17 apprenants en 2007/08. En 2008/09, j'ai créé un scénario qui a exploité les éléments de réponse à mes questions que j'ai obtenus lors des deux premières phases. J'ai proposé à 33 apprenants, répartis en deux groupes classe, un projet d'écriture d'annonces comme celles que l'on rédige pour trouver un correspondant. Certaines des références qui ont sous-tendu mes actions ont progressivement perdu de leur pertinence. D'autres se sont avérées opératoires au fil de la recherche.

J'ai conçu les trois productions écrites du projet principal, *Projekt Prospekt*, à savoir les "Informations pratiques", la "Visite guidée" et le "Commentaire dans un livre d'or électronique", en me basant sur les critères du DCL (Diplôme de Compétence en Langue à usage professionnel, 2002). Je n'ai pas utilisé ces critères à des fins d'évaluation. Par contre, ils ont ouvert des pistes pour l'analyse des productions des apprenants.

À défaut de pouvoir m'appuyer sur des corpus allemands existants dans les domaines concernés par les tâches, j'ai créé des mini-corpus pédagogiques (Braun, 2005), à partir de données trouvées sur la Toile pour lesquelles j'ai obtenu les droits d'auteur. J'ai ensuite soumis les données des corpus au concordancier *Text-based concordances (v2)* de Cobb (nd). J'ai publié les corpus et leurs concordances sur un de mes blogues pédagogiques afin que les apprenants y aient accès pendant et en dehors des séances d'allemand.

Lors de l'analyse des données, je me suis intéressée au discours des apprenants concernant les outils et méthodes, à leur évolution en termes d'apprentissage et aux caractéristiques des productions écrites obtenues. J'ai fait des relevés à l'aide de collections d'outils de statistique textuelle (*Lexico3* de Fleury, Lamalle, W. Martinez et Salem, 2003 et *AntConc3.2.1w* d'Anthony, 2007). J'ai analysé la fréquence lexicale dans mon journal de bord, dans les entretiens oraux qu'une enquêtrice a fait avec les apprenants de la phase principale de la

recherche, et dans les productions écrites des apprenants. J'ai localisé les passages co-présents dans les corpus de référence et les corpus d'apprenants. J'ai mesuré la fréquence des prises de parole dans les entretiens.

Pour certains relevés, j'ai fait appel à des outils proposés par le traitement de texte *Word 2007*. À l'aide de la fonction "Statistiques", j'ai calculé le nombre de mots des productions des apprenants. J'ai comparé les différentes versions des trois productions écrites pour le projet principal *Projekt Prospekt* à l'aide de la fonction "Comparer" du menu "Révision" du traitement de texte *Word 2007*. Cette fonction permet de localiser et de compter les insertions, suppressions, déplacements, les modifications concernant la mise en forme et les commentaires ajoutés. J'ai observé les modifications dans le domaine de l'adjectif (absence ou présence de terminaisons, place de l'adjectif) et de la syntaxe du groupe verbal et nominal afin de rendre compte du degré de complexification des productions écrites des apprenants en fonction de leur utilisation ou non-utilisation des corpus et concordances.

Le bilan de ma recherche est le suivant.

- Les corpus donnent accès, davantage que les dictionnaires, à des collocations qui sont attestées pour un genre textuel donné.
- Les corpus et concordances peuvent permettre d'observer des normes linguistiques et discursives de genres textuels ce qui contribue au développement d'un regard synthétique sur les données en LE.
- Les corpus et concordances peuvent permettre d'obtenir des rédactions qui sont suffisamment conformes aux normes linguistiques et discursives attestées pour le genre textuel qui correspond au texte.
- Les corpus et concordances peuvent contribuer à rendre les apprenants plus indépendants dans leur travail de révision.
- Les corpus et concordances peuvent permettre de progresser dans le domaine de la flexion des adjectifs, de la ponctuation et des mots composés.

Mon analyse des données a permis de retracer certaines évolutions du système linguistique des apprenants. Mes études concernant les adjectifs et la syntaxe lors de la rédaction de textes ayant une longueur de 60 à 120 mots attestent davantage d'une évolution de la flexion des

adjectifs que d'une évolution dans le domaine de la syntaxe. Certains apprenants ont inséré des segments repérés dans l'*input* sans que cela fasse sens dans le nouveau contexte. J'en conclus qu'ils n'ont pas traité les données repérées et que, par conséquent, il n'y a pas eu apprentissage de la LE. Le repérage doit aller de pair avec un travail de réflexion métalinguistique, menant à la conceptualisation des phénomènes linguistiques observés. L'analyse des travaux de deux apprenants en particulier permet de dire qu'une telle réflexion peut mener à une conceptualisation dans le domaine des mots composés. Une des analyses illustre les apports du scénario en termes de prise en compte de normes discursives concernant la ponctuation. Des recherches complémentaires sont nécessaires dans le domaine de la conceptualisation à l'aide de lignes de concordance.

J'ai amené les apprenants à réfléchir à des solutions possibles en s'appuyant sur des données attestées. Ce type d'indépendance peut être considéré comme une des caractéristiques d'un apprenant-chercheur. L'accompagnement offert aux apprenants lors de l'utilisation des outils peut jouer un rôle crucial pour le développement d'une telle attitude. Au fil du projet, j'ai développé une façon d'accompagner les apprenants susceptible de les aider à aller vers les corpus et concordances en ayant conscience de ce qu'ils y cherchaient, au lieu de les pousser à répliquer ma démarche de recherche.

L'exploitation directe, donc "indépendante", des corpus et concordances par des apprenants de niveau intermédiaire dépend, entre autres, du type de ressources et d'outils d'interrogation dont on dispose. J'ai trouvé que pour l'allemand, dans ces domaines, le choix est moins large que pour l'anglais et que l'emploi d'outils conçus pour cette langue n'est pas forcément une solution acceptable. Des recherches complémentaires et des développements d'outils et de ressources sont nécessaires. Ces recherches pourront se baser sur l'utilisation des concordanciers *AntConc3.2.1w* (Anthony, 2007) et *Lexico3* (Fleury, Lamalle, W. Martinez et Salem, 2003) et sur l'emploi de la version franco-allemande de Linguee (Fink et Frahling, 2008) dont j'attends le développement futur.

Ma recherche a consisté à mettre en place un scénario pédagogique qui a fait appel à des corpus afin de soutenir l'apprentissage de la production écrite en LE. L'analyse de mes données a fait ressortir que les corpus et les outils d'interrogation ne représentent pas un kit d'apprentissage "prêt à l'emploi". Leur emploi nécessite un investissement en termes de temps

et d'activité cognitive. Pourtant - ou justement en raison de cela -, le potentiel des corpus pour l'enseignement/apprentissage de LE est réel. Les productions peu nativisées qu'ont produites les apprenants ayant réussi à se familiariser avec les outils et les méthodes de travail en témoignent tout particulièrement. J'ai pu constater que l'exploitation des corpus n'est réservée ni aux adultes, ni aux apprenants ayant un niveau avancé en LE : les apprenants de niveau intermédiaire en tirent profit, à condition d'avoir accès à des ressources, des outils et un guidage adéquats.