

Les corpus comme outils d'aide à la rédaction en langue étrangère

Eva Schaeffer-Lacroix

IUFM de Paris, école interne de Paris 4

elacroix@paris.iufm.fr

1. Introduction

Les corpus offrent un matériau linguistique qui peut servir dans des situations d'écriture diverses. Le scripteur cherche-t-il à rédiger un texte en tant que spécialiste d'un domaine ? Se contente-t-il de s'exprimer dans un registre peu élaboré, ou aspire-t-il à explorer les limites du dicible, à se comporter en poète ou en joueur ? Tout semble possible, du moment qu'il a l'occasion de s'appuyer, pour ce faire, sur un corpus dont le contenu reflète le discours qu'il choisit de tenir. Un tel corpus offre au scripteur des "modèles sociaux" ou des "exemples types" (Bronckart 1997), non dans le sens de modèles ou exemples prescrits, mais dans le sens de productions attestées.

Nous proposons de considérer les corpus comme outils d'aide à la rédaction en LE (Langue Étrangère). Les aides dont on dispose par ailleurs, à savoir les correcteurs orthographiques et syntaxiques, contribuent à surmonter des difficultés concernant le bas niveau. Concernant le haut niveau, il existe des logiciels qui soutiennent la planification et l'organisation des contenus. Selon Rastier (2004), les outils que propose la linguistique de corpus permettent de conjuguer les deux niveaux. Ils mettent en évidence le lien entre le plan du contenu d'un texte et le plan de l'expression.

La forme intérieure, loin d'être un mystère esthétique, est constituée par les régularités jusqu'à présent imperceptibles de la forme extérieure, celle de l'expression, que les moyens théoriques et techniques de la linguistique de corpus permettent à présent de mettre en évidence (Rastier, 2004).

Nous illustrerons dans cet article l'idée que l'utilisation de corpus, considérés comme une forme de base de données textuelles, peut être précieuse pour l'apprentissage de la rédaction, y compris pour le niveau d'apprentissage peu avancé à intermédiaire. Les arguments sont développés sur fond de toile d'une recherche menée avec plusieurs groupes d'adolescents français de 14-15 ans, apprenant l'allemand comme langue étrangère au sein d'un établissement secondaire d'enseignement (Schaeffer-Lacroix, 2009a).

Crinon, Legros et Marin (2002) nomment deux conditions qui déterminent la qualité de la production écrite : 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 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. Le développement d'une telle attitude dépend du type d'accompagnement qu'un scripteur expert offre à un scripteur novice.

Ces relations qui font que l'expert cheminera avec le novice sur un autre mode que celui de la correction ou remédiation ouvrent un espace de travail quasi singulier. L'expert cherchant à multiplier les pratiques d'écriture, à faire en sorte que le sujet écrivant travaille plusieurs textes dans un souci artisanal en prenant appui sur ses propres écrits, en empruntant ailleurs ce qui lui convient (Crinon, Legros et Marin, 2002).

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. Nous avons croisé ce concept de type constructiviste avec les apports de la psycholinguistique concernant la recherche d'informations (Rouet, Coutelet et Dinet, 2004) et avec la technique de recyclage d'éléments repérés dans les corpus.

Dans ce qui suit, nous développerons le rôle des corpus pour l'apprentissage de la rédaction en LE, en précisant en particulier les procédés de recherche d'information et l'exploitation intertextuelle des données.

2. Le rôle des corpus pour l'apprentissage de la rédaction

2.1. Ressources textuelles et modèles d'écriture

Un aperçu des modèles de la rhétorique classique et de la critique génétique permettra d'estimer l'apport éventuel de ces deux types de modèle pour évaluer le rôle des corpus, considérés comme une forme de ressource textuelle, pour l'apprentissage de la rédaction.

Le modèle de rédaction de la rhétorique classique est parfois associé à des termes comme "reproduction d'un modèle" et "centration sur le produit". Toutefois, la prescription de modèles et la linéarité du processus de rédaction ne sont pas des composantes obligatoires de la rhétorique. Ces caractéristiques sont peut-être davantage liées à certaines pratiques d'enseignement-apprentissage de la production écrite qui ont eu cours en France et en

Allemagne avant leur remise en cause autour des années 70 (Plane, 1996 ; Schenk, 1998). Il paraît judicieux de "revisiter" le modèle rhétorique pour la didactique de la production écrite en modifiant certains de ses aspects, comme par exemple la mise à disposition de ressources textuelles plus variées et en quantité plus importante. Avec l'émergence de bases de données pédagogiques (Develotte et Potolia, 2005), il semble intéressant d'enseigner la production écrite en mettant l'accent sur les genres textuels (Bronckart, 1997) et sur les techniques d'écriture (Mangenot, 1997 ; Bouchard, J.-P. Simon et Vourzay, 2004).

Le modèle de la génétique critique a été employé par Doquet-Lacoste (2004) dans un contexte d'enseignement/apprentissage de la production écrite en langue maternelle. Je retiens de cette approche l'intérêt pour les brouillons, donc les différentes versions produites avant la version considérée comme définitive. Le modèle de la génétique critique permet la prise en compte simultanée du processus d'écriture et des diverses variantes du produit. Selon Olive et Piolat (2003), un tel regard double semble prometteur.

2.2. Quels corpus pour l'apprentissage de la rédaction ?

Rastier (2004) préconise le texte comme unité minimale linguistique. Un projet d'écriture s'appuiera alors utilement sur un corpus qui regroupe des textes ayant des caractéristiques communes (par exemple, le genre textuel). Le corpus doit correspondre au thème ou à d'autres aspects du projet d'écriture. De tels corpus sont proposées – sous la dénomination de "base de données" – dans des projets décrits par Crinon, Legros et Marin (2002) pour le français, langue maternelle et par Mangenot (1997 ; 2000) pour le français, langue maternelle ou étrangère. Dans le contexte de l'enseignement-apprentissage de la production écrite en anglais, langue étrangère, Henry et Roseberry (1996 ; 2001 : 93-133) et Rohrbach (2003) font utiliser des corpus regroupant des textes qui représentent un même genre textuel. Lors du projet de Henry et Roseberry (2001), des apprenants adultes explorent des corpus afin de se préparer à la rédaction d'une lettre d'embauche et du support écrit d'un discours introduisant des conférenciers lors d'un colloque. À l'aide d'un concordancier, ils désignent la fréquence d'unités lexicales et de segments. Ils sont censés en tirer des conclusions susceptibles d'avoir des répercussions sur leur production écrite (Henry et Roseberry, 2001 : 100). La rédaction est exécutée à la fin du projet, donc après une phase d'exploration du corpus et après l'observation des caractéristiques des divers genres textuels qui y sont représentés. Nous retenons l'idée de faire observer les caractéristiques linguistiques et discursives des textes à l'aide d'outils d'interrogation de corpus. Henry et Roseberry n'évoquent toutefois pas explicitement le

procédé de rédaction basé sur l'exploitation de corpus. Il ne semble pas y avoir de va-et-vient entre les versions provisoires et le corpus avant finalisation du produit. Les paragraphes suivants évoqueront les actions possibles des apprenants vis-à-vis des données du corpus, en particulier lors de la révision de texte.

3. Que faire avec les données du corpus ?

3.1. Recherche d'informations

Crinon, Legros et Marin (2002 : 188) rappellent que la lecture joue un rôle important dans le processus d'écriture et de réécriture. Les textes sélectionnés pour soutenir un projet d'écriture peuvent être abordés de plusieurs manières. Il est possible d'en lire certains intégralement, ou bien, il est possible que le lecteur développe un regard sélectif afin d'y repérer les éléments qui paraissent pertinents en vue de la réalisation de son projet. Comme le précisent Rouet Rouet, Coutelet et Dinet (2004), le premier type de lecture mobilise la compétence de compréhension globale, tandis que le second type de lecture nécessite un regard qui cible les éléments susceptibles d'aider le lecteur à résoudre un problème. La recherche d'informations convient, à notre avis, davantage que la lecture exhaustive, à des apprenants d'allemand de niveau intermédiaire en situation de révision de texte. Elle a des chances de les soutenir encore mieux si la mise en forme matérielle du texte renforce la visibilité des éléments potentiellement pertinents. Les concordances sont un bon exemple d'une telle mise en forme matérielle. Elles permettent de présenter un mot clé ou un segment de façon particulièrement saillante, ceci grâce à sa position dans le document et grâce à sa mise en forme particulière (en gras, en rouge, en capitales, etc.). Les informations que les apprentis-scripteurs trouvent lors de la lecture peuvent ensuite être mises à profit lors de la phase de révision et lors du processus de replanification (Crinon, Legros et Marin, 2002).

3.2. Exploitation intertextuelle

Les textes ne sont jamais créés à partir de rien, mais ils s'insèrent dans d'autres textes écrits auparavant (Bouchard, J.-P. Simon et Vourzay, 2004), donc dans des intertextes. Suite à des importations volontaires ou involontaires d'unités ou de segments, de figures de style, de situations, etc., les intertextes peuvent laisser des traces dans les nouvelles productions. Le concept d'intertextualité évoque la dimension historique et sociologique des textes. Selon Bronckart (1997 : 111),

tout texte empirique est réalisé par emprunt à un genre, et il relève donc lui-même toujours d'un genre ; tout texte empirique procède cependant aussi d'une adaptation du genre-modèle aux valeurs attribuées par l'agent à sa situation d'action, et dès lors,

outre les caractéristiques communes au genre emprunté, il exhibe également des propriétés singulières, qui définissent son style propre.

En écrivant, nous reprenons dans une certaine mesure les caractéristiques des intertextes, mais nous les façonnons également en y ajoutant notre propre style, en proposant de nouveaux types de combinaisons de segments, d'éléments discursifs, etc. (Bronckart 1997 : 79). Teubert (2006) considère également qu'un nouveau texte est toujours joint à un discours existant. Il est une recombinaison et permutation de textes existants.

Dans le contexte de l'acquisition de la production orale, Aguado (2002 : 215) présente l'imitation, définie comme la reprise de ce que dit un interlocuteur, comme une preuve que l'attention de l'apprenant-locuteur est dirigée vers le support oral. Une telle attention peut également être mise à profit dans le contexte de l'apprentissage de la production écrite. Le scripteur qui a comme tâche l'identification d'unités ou de segments dans un texte afin d'en emprunter certains doit aller à la rencontre du texte. L'attention que l'apprenant accorde au support textuel est susceptible de l'aider à assimiler ou "re-traiter" au moins une partie de ce qu'il a observé et d'en retenir certains éléments. Doughty et Long (2003 : 63) décrivent la sélection de segments dans un corpus de textes à des fins d'exécution d'une tâche écrite comme une forme de "plagiat nécessaire". Une telle conception de l'écriture peut surprendre : de nombreux enseignants regrettent la tendance au plagiat de certains apprenants à qui Internet offre l'occasion d'obtenir des textes en faisant du "copier-coller" extensif au lieu d'en produire eux-mêmes. Toutefois, rappelons que l'imitation, aussi appelée mimésis, a pendant longtemps été un critère pour la classification générique de textes (Jenny, 2003). Selon Bouillaguet (1996 : 4), au 17^{ème} siècle,

[L'imitation était] une pratique courante et encouragée en Europe – la valeur des œuvres tenant non pas à une originalité à laquelle on ne croit guère, mais à l'habileté des auteurs à adapter un modèle ancien à une réalité nouvelle (Bouillaguet 1996 : 4).

À la fin du 19^{ème} siècle, Anatole France crée le lien entre une telle doctrine de l'imitation et le concept qui précède la théorie de l'intertextualité telle que nous la connaissons depuis le 20^{ème} siècle, grâce, entre autres, aux travaux de Julia Kristeva (Bouillaguet 1996 : 5). La "renaissance" du procédé d'imitation peut être particulièrement observée dans le contexte d'une écriture ludique (pastiche, parodie, collage, centon). L'intertextualité comme procédé littéraire est alors motivée par l'intention d'être drôle ou de vouloir créer un style incongru. Ce procédé se distingue radicalement de "l'imitation-culte" du 17^{ème} siècle (Bouillaguet 1996 : 5). L'encouragement à l'emprunt dans le cadre de l'enseignement/apprentissage de la production

écrite en LE n'est pas forcément motivé par l'intention de détourner des textes à des fins humoristiques. Toutefois, cet usage d'un intertexte n'est pas exclu. Citons l'exemple de deux apprenantes qui ont participé à notre recherche et qui ont tourné en dérision le titre d'une critique de film en remplaçant les cinq étoiles distribuées par l'auteur au film par deux étoiles.

Titre sur lequel Olivia et Géraldine se sont appuyées :

Diesmal haben alle 5 Rezensenten Recht!*

[Cette fois-ci, tous ceux qui ont donné cinq étoiles au film ont raison].

Nouveau titre proposé par Olivia et Géraldine :

Diesmal haben alle 2 Rezensenten Recht!*

[Cette fois-ci, tous ceux qui ont donné deux étoiles au film ont raison].

Selon St John et Cash (1995 : 50), le procédé de réutilisation de segments trouvés dans des intertextes peut améliorer la qualité de productions écrites dans le domaine de la syntaxe. Šimuněk (2007) a étudié la reprise de segments (comparables aux blocs lexicalisés, *exemplars*, formules) par des étudiants allemands, futurs enseignants d'anglais, lors d'un projet collaboratif impliquant des étudiants canadiens. Lors de ce projet, les apprenants semblent avoir eu assez peu recours au procédé de reprise de segments dans les courriels des partenaires canadiens. Parfois, des raisons affectives ont mené à une telle reprise (Šimuněk, 2007 : 295). Le manque de "popularité" du procédé de reprise d'éléments dans l'*input* peut s'expliquer, entre autres, par le fait qu'il ne correspond pas à une pratique scolaire répandue. Souvent, les scripteurs sont invités à formuler "avec leurs propres mots". Citons, à titre d'exemple, un extrait du programme d'enseignement pour l'allemand, palier 2, au collège (2007) : "Reformulation : L'élève sera capable de transmettre une information en utilisant ses propres mots." Dans un contexte de psychosociologie, Linard (1996) considère l'imitation de comportements humains comme un processus qui aide à intégrer des informations nouvelles.

Chez l'individu en développement, l'imitation permet de dériver activement sa structure du dedans d'une structure du dehors qui lui ressemble tout en étant autre (...). Cette imitation (...) ne peut en aucun cas se réduire à une simple copie-transfert entre imitant et imité. C'est une forme d'action auto-organisée et hétéro-régulée qui implique des capacités d'intégration et d'interprétation, totalement étrangères aux automates (Linard, 1996 : 97-98).

Ces paroles trouvent un écho en Vygotski, relu par Brossard (2004). Ce dernier explique que, selon Vygotski, l'imitation est fondamentale pour l'apprentissage ; elle n'est toutefois jamais

une reprise à l'identique : "l'activité en même temps qu'elle s'intériorise, elle se réorganise."
(Brossard, 2004 : 104).

4. Pour finir

Encourager les apprenants à identifier des unités dans le corpus et à les réutiliser afin de réviser leur production écrite est une méthode qui implique le traitement et parfois la modification des données identifiées. Ce sont des activités qui ont un coût cognitif plus ou moins élevé. À la lumière des perspectives développées par Aguado (2002 : 215), Linard (1996 : 97-100) et Vygotski (1934 : 353-355), cette méthode peut soutenir l'apprentissage de la production écrite en LE. Les scénarios pédagogiques impliquant des corpus doivent viser à rendre les apprenants co-responsables de leur travail de révision : ils les amèneront à réfléchir à des solutions en s'appuyant sur des données attestées. Ce type d'indépendance peut être considéré comme une des caractéristiques d'un apprenant-chercheur.

Pour finir, nous présenterons certains des avis des apprenants concernant l'exploitation de corpus pour l'apprentissage de la rédaction lors de notre recherche (Schaeffer-Lacroix, 2009b : 85) afin d'appuyer les arguments développés dans cet article.

ALEXA : Je suis satisfaite du projet, car j'ai vu l'allemand directement "à l'œuvre", j'ai pu découvrir et utiliser autrement la langue qu'en leçon (...).

CHLOÉ : Je suis assez satisfaite du projet, car il m'a permis de découvrir des moyens plus pratiques que le dico pour écrire un texte même si son usage doit quand même rester présent.

FERNAND : Je suis satisfait du projet, car je me sentais plongé dans le réel. Le fait d'écrire du texte et d'utiliser de nouveaux outils m'a beaucoup appris.

VICKY : Je suis satisfaite du projet, car on a travaillé en autonomie. Au début, j'étais angoissée car on se dirigeait dans des directions différentes tous. Mais à la fin, j'ai développé mes propres méthodes d'écriture en adaptant ce que je savais faire ou pas (...).

5. Références

- Aguado, K. (2002). *Imitation als Erwerbsstrategie. Interaktive und kognitive Dimensionen des Fremdspracherwerbs*. Habilitation à diriger des recherches. Disponible en ligne : <http://karin.aguado.de/publ/downloads/ImitationAlsErwerbsstrategie.pdf>
- Bouchard, R., Simon, J.-P. & Vourzay, M.-H. (2004). "L'intertextualité au service de l'interdiscursivité, étude de la production d'une synthèse de documents : à la recherche du "lieu commun". In *Actes du XXIVe colloque d'Albi, Langages et signification, l'intertextualité*, 2003. CALS/CPST. Toulouse : Presses de l'Université de Toulouse le Mirail. pp. 93-103.
- Bouillaguet, A. (1996). *L'écriture imitative. Pastiche, parodie, collage*. Paris : Éditions Nathan.

- Bronckart, J.-P. (1997). *Activité langagière, textes et discours. Pour un interactionnisme sociodiscursif*. Paris : Delachaux et Niestlé.
- Brossard, M. (2004). *Vygotski : lectures et perspectives de recherches en éducation. Suivi d'un inédit de L. S. Vygotski*. Traduit par Olga Anokhina et Michel Brossard. Lille : Presses Universitaires Septentrion.
- Crinon, J., Legros, D. & Marin, B. (2002). "Écrire et réécrire au cycle 3 : l'effet des mots clés sur la réécriture, avec et sans assistance informatique". In Plane, S. (dir.). *L'écriture et son apprentissage à l'école élémentaire. Repères*, n° 26-27 (2002/03). Paris : INRP (Institut National de Recherche Pédagogique). pp. 187-202.
- Develotte, C. & Potolia, A. (2005). "Les bases de données pédagogiques sur Internet : le cas du programme Éducasup". *Apprentissage des langues et systèmes d'information et de communication (ALSIC)*, vol. 8, n° 1. <http://alsic.revues.org/index370.html>
- Doquet-Lacoste, C. (2004). "Genèse des textes, genèse des scripteurs : regards sur la langue dans l'écriture". In Piolat, A. (dir.). *Écriture. Approches en sciences cognitives*. Aix-en-Provence : Publications de l'Université de Provence. pp. 233-250.
- Doughty, C. & Long, M. (2003). "Optimal psycholinguistic environments for distance foreign language learning". *Language Learning & Technology*, vol. 7, n° 3, pp. 50-80. <http://llt.msu.edu/vol7num3/doughty/default.html>
- Henry, A. & Roseberry, R. L. (1996). "A Corpus-based Investigation of the Language and Linguistic Patterns of One Genre and the Implications for Language Teaching". *Research in the Teaching of English*, n° 30. pp. 472-489.
- Henry, A. & Roseberry, R. L. (2001). "Using a small corpus to obtain data for teaching a genre". In Ghadessy, M. & Henry, A. & Roseberry, R. L. (dir.). *Small Corpus Studies and ELT*. Amsterdam/Philadelphia : John Benjamins Publishing Company. pp. 93-133. *Institut für Kommunikationswissenschaften*. Université de Bonn. <http://www.ikp.uni-bonn.de/forschung/computerlinguistik/computerlinguistik>
- Jenny, L. (2003). "Les genres littéraires". *Méthodes et problèmes*. Genève : Département de français moderne. <http://www.unige.ch/lettres/framo/enseignements/methodes/genres/>
- Johns, T. (1991). "Should you be persuaded - two samples of data-driven learning materials". In Johns, T. & King, P. (dir.) (1991). pp.1-13.
- Linard, M. (1996). *Des machines et des hommes*. Paris : L'Harmattan.
- Plane, S. (1996). "Écriture, réécriture et traitement de texte." In David, J. & Plane, S. (1996). pp. 37-77.
- Mangenot, F. (1997). "Aide à la rédaction à travers un dialogue 'ordinateur/apprenant'". *Revue de l'association Enseignement Public et Informatique* n° 87, septembre 1997. Paris : EPI. Disponible en ligne : <http://www.epi.asso.fr/revue/87/b87p149.htm>
- Olive, T. & Piolat, A. (2003). "Activation des processus rédactionnels et qualité des textes". In *Le Langage et l'Homme*, vol. 38, n° 2. pp. 191-206. Disponible en ligne : <http://cogprints.org/3605/1/OlivePiolatLH2003.pdf>
- Rastier, F. (2004). "Enjeux épistémologiques de la linguistique de corpus". *Texte !*, juin 2004. Rubrique Dits et inédits. http://www.revue-texto.net/1996-2007/Inedits/Rastier/Rastier_Enjeux.html
- Rohrbach, J.-M. (2003). "'Don't miss out on Göttingen's nightlife': Genreproduktion im Englischunterricht". *Praxis des Neusprachlichen Unterrichts*, 50. pp. 381-389.
- Rouet, J.-F., Coutelet, B. & Dinet, J. (2004). "La recherche d'informations dans les documents complexes : Processus cognitifs, apprentissage et développement". *Conférence invitée à la Journée d'Etudes sur le Traitement Cognitif des Systèmes d'Informations Complexes (JETCSIC)*, Genève, 17-18 juin 2004. Disponible en ligne : <http://tecfa.unige.ch/tecfa/jetcsic2004/papers/jfrjetcsic2004.pdf>

- Schaeffer-Lacroix, E. (2009a). *Corpus numériques et production écrite en langue étrangère. Une recherche avec des apprenants d'allemand*. Thèse de doctorat. Tome 1. Paris: Sorbonne nouvelle. <http://tel.archives-ouvertes.fr/tel-00439095/fr/>
- Schaeffer-Lacroix, E. (2009b). *Corpus numériques et production écrite en langue étrangère. Une recherche avec des apprenants d'allemand*. Thèse de doctorat. Tome 2. Paris: Sorbonne nouvelle. <http://tel.archives-ouvertes.fr/tel-00439095/fr/>
- Schenk, K. (1998). "Die allmähliche Verfertigung der Texte beim Schreiben. Vorschläge zur Strukturierung von Schreibsitzungen". *Beiträge zur Fremdsprachenvermittlung* 34 (1998). Landau, Konstanz : Institut für fremdsprachliche Philologien der Universität Koblenz-Landau, Universität Konstanz. pp. 98-117.
- Šimuněk, R. (2007). *Formelhafte Sprache und Internetprojekte. Zum Sprachgebrauch und -erwerb fortgeschrittener Fremdsprachenlerner*. Tübingen : Gunter Narr Verlag.
- St. John, E. & Cash, D. (1995). "German language learning via email: a case study". *ReCALL* vol. 7, n° 2. pp. 47-51.
- Teubert, W. (2006). "Korpuslinguistik, Hermeneutik und die soziale Konstruktion der Wirklichkeit". *Linguistik Online* 28, 3/06. pp. 41-60. http://www.linguistik-online.de/28_06/teubert.html
- Vygotski, L. (1934). *Pensée et langage*. Paris : La Dispute/SNÉDIT, 1997³.