

Se servir de corpus d'apprenants et de corpus d'experts pour former en langue étrangère

Eva Schaeffer-Lacroix (*Centre de Linguistique en Sorbonne, Université Paris Sorbonne*)

Les corpus numériques contiennent des exemples de formulation qui peuvent être considérés comme des reflets fiables de la langue en usage (Ellis & Wulff 2018). La confiance que l'on accorde à de telles ressources est motivée par deux raisons : s'il s'agit de corpus réflexifs, constitués de documents qui correspondent à un genre textuel identifié (Rastier 2004), leurs données illustrent un discours représentatif. De plus, les outils de corpus renseignent sur le type et le nombre de situations lors desquelles telle et telle formulation a été utilisée dans tel et tel contexte. Dans le contexte de l'enseignement-apprentissage des langues, le formateur ou la formatrice peut proposer à des scripteurs novices de comparer leurs formulations non acceptables, détectées à un stade initial de leur production écrite, à celles des auteurs des

textes d'un corpus réflexif, qui acquiert alors le statut de corpus d'experts. Les productions en langue étrangère peuvent être rassemblées dans un corpus d'apprenants, aussi appelé "corpus d'interlangue" (Jacques & Rinck 2017), renseignant sur les stades intermédiaires d'apprentissage d'une langue (Selinker 1972). Euler (2014) présente un cycle itératif qui décrit les liens qui peuvent être établis entre terrain, recherche fondamentale et

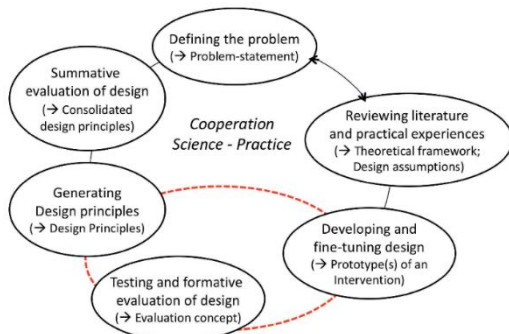


Figure 1 – Phases d'un processus de type DBR (Design-based research) (Euler 2014).

recherche en didactique. Je présenterai la façon dont un tel cycle peut prendre vie au sein d'un projet faisant appel à des corpus. En l'occurrence, quatre des six phases du modèle d'Euler sont prises en compte.

Définir l'obstacle

La création d'un scénario peut être motivée par des besoins constatés sur le terrain ; on vise alors l'amélioration d'une compétence en particulier. Pour donner un exemple, les apprenants d'allemand se heurtent fréquemment à la difficulté de choisir la préposition ou le préverbe qui convient pour un contexte donné. Cela peut être dû au fait qu'une même forme peut avoir plusieurs sens, comme *auf* [sur, à, etc.]. De plus, certains de ces éléments peuvent appartenir à plusieurs classes de mot : *auf dem Tisch* [sur la table] représente un usage prépositionnel de *auf*, tandis que dans *aufmachen* [ouvrir], cette forme a la fonction de préverbe.

Consulter la littérature scientifique et les expériences pratiques

Selon la complexité du fait linguistique que l'on souhaite aborder dans le scénario, le recours à la littérature scientifique et la prise en compte de projets existants peut être bénéfique.

Lors d'un projet antérieur, un groupe de futures documentalistes de niveau B1/B2 (intermédiaire) a été invité à écrire une scène complétant celles du script de la série télévisée *Weissensee* (Hess 2010). Les prépositions et préverbes spatiaux sont très présents dans les didascalies de scripts de film. Afin de pouvoir les aborder du point de vue de leur sens, j'ai présenté aux apprenant·e·s les concepts d'identification, différenciation et rupture, appliqués par Gilbert (2005) aux prépositions anglaises *in* [dans], *on* [sur] et *at* [à]. Pour réviser leur texte, les apprenantes se sont servies d'un corpus d'experts. J'ai analysé l'évolution des prépositions et préverbes au fil des différentes versions de texte (Schaeffer-Lacroix 2015). Les résultats obtenus sont compatibles avec ceux de Weber (2018), qui s'est servie du corpus d'apprenants *MERLIN-Korpus* pour analyser les erreurs dans le domaine des prépositions *an* [à, près de] et *auf* [sur], faites par des personnes non germanophones lors du passage de tests de certification en allemand.

Développement et ajustement du scénario

En m'appuyant sur les résultats obtenus pour le projet *Weissensee*, j'ai développé le projet *Audiodeskription*, visant la production d'une audiodescription, c'est-à-dire d'une piste audio complétant la piste son d'un document audiovisuel par des informations dont des personnes malvoyantes ou aveugles ont besoin pour comprendre l'action d'un film. Le projet *Audiodeskription* donne à des apprenant·e·s d'allemand de niveau B1/B2 l'occasion de s'entraîner à décrire des événements spatiaux et des événements de perception. Il peut s'agir, là aussi, d'observer les prépositions et les préverbes dans un corpus, mais d'autres faits linguistiques, comme les participes passés ou les mots composés, sont également à prendre en compte parce qu'ils sont, eux aussi, constitutifs pour le genre textuel "audiodescription". Le corpus d'experts *Buettenwarder*, composé de 69 manuscrits d'audiodescription produits pour la série télévisée *Neues aus Buettenwarder* (Eberlein 1997-2017), sert de lieu pour déterminer et explorer les faits linguistiques qui caractérisent l'audiodescription. Ce corpus a été déposé sur *Sketch Engine* (Kilgarriff *et al.* nd) et *TXM* (Heiden 2010) et, par la même occasion, étiqueté de façon automatique au niveau des classes de mots. Je présenterai des requêtes variées, portant sur des niveaux d'analyse aussi bien micro que macro, qui illustreront la convivialité et la performance des deux outils d'exploration de corpus.

Implémentation et évaluation formative du scénario

Le scénario est actuellement testé, sous forme de recherche-action, par un groupe de futur·e·s enseignant·e·s de documentation, lettres et musique. J'en présenterai les premiers résultats.

Références bibliographiques

- Eberlein, N. (1997-2017). Neues aus Büttengewässer. Série télévisée. Audiodescriptions créées par la chaîne télévisée NDR Fernsehen (Norddeutscher Rundfunk).
- Ellis Nick & Wulff Stefanie (2018). Usage-based approaches to second language acquisition. In D. Miller, F. Bayram, J. Rothman & L. Serratrice (dir.). *Bilingual Cognition and Language: The state of the science across its subfields*, 37–56.
- Euler, D. (2014). Design-Research – a paradigm under development. In D. Euler and P.F.E. Sloane (dir.). *Design-Based Research* (15-44). Stuttgart : Franz Steiner Verlag.
- Gilbert, E. 2005. Ébauche d'une formalisation des prépositions In, On et At. *Cycnos* 21|1. <http://revel.unice.fr/cycnos/index.html?id=11>
- Heiden, S. (2010). The TXM Platform: Building Open-Source Textual Analysis Software Compatible with the TEI Encoding Scheme (389–398). Presented at the 24th Pacific Asia Conference on Language, Information and Computation, Institute for Digital Enhancement of Cognitive Development, Waseda University. <https://halshs.archives-ouvertes.fr/halshs-00549764/document>
- Hess, A. 2010. *Weissensee*. Script d'une série télévisée. <http://www.stichwortdrehbuch.de/drehbuch>
- Jacques, M.-P. & Rinck, F. (2017). Un « corpus de littéracie avancée : résultat et point de départ. *Corpus* (16). <http://journals.openedition.org/corpus/2806>
- Kilgarriff, Adam, Rychly, Pavel and Pomikalek, Jan (nd). *Sketch Engine*. Corpus management system. <http://www.sketchengine.co.uk/>
- Rastier, F. (2004). Enjeux épistémologiques de la linguistique de corpus. *Texte !* http://www.revue-texto.net/Inedits/Rastier/Rastier_Enjeux.html
- Schaeffer-Lacroix, E. (2015). Impact de discussions métalinguistiques sur l'apprentissage de la production écrite en allemand, langue étrangère. *Linx. Revue des linguistes de l'université Paris X Nanterre*, (72), 233–237. <https://doi.org/10.4000/linx.1676>
- Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics* 10, 219-231. DOI : [10.1515/iral.1972.10.1-4.209](https://doi.org/10.1515/iral.1972.10.1-4.209)
- Weber, T. (2018). Grammatik und Lernerkorpora: Eine korpusorientierte Untersuchung von Präpositionalphrasen im deutschen MERLIN-Korpus. In E. Fuß, M. Konopka, B. Trawiński, U. H. Waßner (dir.). *Grammar and Corpora 2016* (415-424). Heidelberg : Heidelberg University Publishing. DOI : <https://doi.org/10.17885/heiup.361.509>

Informations concernant l'auteur

Eva Schaeffer-Lacroix est maître de conférences en études germaniques à Sorbonne Université / ESPÉ de Paris. Ses recherches portent sur l'usage des technologies pour l'enseignement-apprentissage des langues étrangères. Elle s'intéresse en particulier aux corpus numériques, à la production écrite et à la réflexion sur la langue.

Affiliation : Sorbonne Université / Espé de Paris.

Courriel : schaeffer-lacroix@orange.fr

Toile : <http://didaktik.hautetfort.com>

Adresse : laboratoire CeLiSo, Maison de la recherche, 28 rue Serpente, 75006 Paris, France.