

UN CORPUS DE PRODUCTIONS ÉCRITES EN FRANÇAIS LANGUE ÉTRANGÈRE — MATÉRIAU POUR MIEUX COMPRENDRE LES CHOIX LEXICAUX D'APPRENANTS MULTILINGUES

The aim of this paper is to present a corpus containing 105 Swedish learners' written production in French and to discuss some possible didactic applications of such a corpus. Learners from four different grades (age 11-15) were invited to retell a short picture story in writing. One of the main characteristics of the corpus is that it is multilingual in nature, with many instances of influences from the learners' previously acquired languages. It is argued that this type of corpus can be useful in class, allowing for potential learning situations, linguistic development and metalinguistic reflection.

● **Christina Lindqvist**
| Université de
Göteborg et Østfold
university college



Christina Lindqvist est professeure de français à l'Université de Göteborg, Suède, et Østfold university college, Norvège. Ses recherches se situent dans les domaines de l'acquisition d'une troisième langue et de l'acquisition du vocabulaire chez les apprenants suédois du français. Elle enseigne la didactique des langues étrangères, la linguistique, la grammaire et le vocabulaire. Elle dirige des thèses de doctorat en didactique/apprentissage des langues étrangères.

Introduction

L'objectif de cet article est de présenter des recherches récentes menées sur un corpus de productions écrites d'apprenants suédois du français, en vue de mettre en évidence quelques applications didactiques possibles à partir de ce genre de corpus. Le corpus, qui a été constitué à des fins de recherche en premier lieu, et qui sera décrit plus amplement dans la section suivante, contient des narrations écrites à partir d'une série d'images racontant une petite histoire de deux enfants qui partent en pique-nique (voir l'annexe). Les apprenants sont des élèves du secondaire I (11 à 15 ans) en Suède. Nous allons montrer que ce type de corpus permet, entre autres, d'analyser les compétences linguistiques des apprenants à différents niveaux d'apprentissage et de mieux comprendre le processus d'apprentissage en se focalisant en premier lieu sur les aspects lexicaux et le multilinguisme des apprenants. Les résultats des recherches précédentes sur l'impact des langues déjà acquises sur le

français écrit nous permettront notamment de mettre en avant l'utilité d'un corpus d'apprenants multilingues dans le cadre de l'enseignement des langues.

Présentation du corpus

Le corpus consiste en 105 productions écrites d'élèves suédois, âgés de 11/12 à 14/15 ans, qui ont été récoltées dans quatre classes, de la 6^e à la 9^e année du système scolaire suédois. Il convient de noter qu'il s'agit de quatre classes différentes, ce qui veut dire que le corpus est de nature transversale et non longitudinale. Les élèves ont été invités à écrire un texte à partir d'une série d'images intitulée « The Dog Story », qui est l'histoire de deux enfants qui partent en pique-nique, mais lorsqu'ils arrivent dans la forêt, il s'avère que leur chien s'était caché dans leur panier et qu'il a mangé la nourriture qui s'y trouvait. Les élèves disposaient de 20 minutes pour cette tâche. Ils ne disposaient que d'un stylo et d'une feuille de papier, et n'étaient pas

Classe	Nombre d'élèves	Âge	Semestres d'études de français	Mots produits au total	Nombre min et max de mots	Moyenne de mots par texte	Langue maternelle	Langues étudiées ou parlées à la maison
6	17	11/12	1	503	7–74	26	Suédois	Anglais (17) Finnois (1) Norvégien (1) Polonais (1)
7	26	12/13	3	1835	18–164	66	Suédois	Anglais (26) Danois (1)
8	35	13/14	5	3516	34–197	100	Suédois	Anglais (35) Chinois (2) Espagnol (16) Allemand (1)
9	27	14/15	7	3066	62–263	114	Suédois	Anglais (27) Allemand (2) Espagnol (12) Chinois (4)

Tableau 1

Le corpus

autorisés à utiliser des dictionnaires, ni à collaborer. Conformément à la méthode pédagogique utilisée à l'école, ils devaient essayer de se concentrer sur la transmission du message et de communiquer de manière aussi détaillée que possible ce qu'ils voyaient sur les images. L'exemple 1 présente un texte écrit par un élève en 9^e. On voit que le texte est rédigé en français, mais contient des éléments d'autres langues comme « mountain » ou « años ».

Ex. 1. Texte d'un élève en 9^e

C'est un garçon et une fille et un chien et une mère dans la picture . Les enfants faisons sandwiches. Le chien manges les sandwiches. Puis les enfants allons en un picnic avec zero sandwiches. Les enfants allon en un en un mountain avec deux animeaux grande. Apres les enfants mangais les sandwiches il est zero sandwiches et un chien le chien mangais tout les sandwiches et les enfants mangais zero sandwiches.

PS. En le mountain il est trois très grand trees et le chien etais blanc et marron. Et les enfants etais sanq y sept años.

Le tableau 1 rend compte des caractéristiques du corpus. En ce qui concerne les connaissances d'autres langues des apprenants, ils avaient tous étudié l'anglais depuis l'âge de 7, 8 ou 9 ans, soit entre trois et huit ans lors de la collecte des données. Un trait caractéristique important des apprenants de ce corpus est donc

que l'anglais représente leur première et principale langue étrangère. De plus, d'autres langues étrangères sont aussi présentes, car en 8^e, il est possible d'opter pour une langue étrangère supplémentaire comme l'allemand ou l'espagnol. Certains élèves parlent aussi d'autres langues à la maison (p.ex. le norvégien ou le danois). Voilà pourquoi le nombre de langues supplémentaires augmente à partir de la 8^e année, contribuant au caractère multilingue du corpus.

Comme il ressort du tableau 1, il s'agit d'un corpus relativement restreint en termes de nombre total de mots et de participants. Néanmoins, il sert d'exemple d'un type de corpus d'apprenants qui pourrait facilement être conçu par le chercheur/enseignant à des fins différentes. En effet, la collecte des données écrites présente plusieurs avantages, dont le temps nécessaire à la collecte et au traitement des données, moindre que pour des données orales (cf. Tracy-Ventura & Paquot, 2020).

Résultats de recherche

Le corpus que nous venons de présenter permet d'analyser de nombreux aspects différents de l'interlangue des apprenants. Comme il ressort du tableau 1, une des caractéristiques des élèves de ce corpus est leur multilinguisme. En effet, alors qu'ils ont le suédois comme langue maternelle, et qu'ils ont tous commencé l'apprentissage de l'anglais bien avant celui du français, ils ont en outre des connaissances d'autres langues, telles

Ce type de corpus permet, entre autres, d'analyser les compétences linguistiques des apprenants à différents niveaux d'apprentissage et de mieux comprendre le processus d'apprentissage en focalisant en premier lieu sur l'aspect lexical et le multilinguisme des apprenants

que l'allemand, l'espagnol ou le chinois. Ainsi, les études menées sur ce corpus se sont surtout concentrées sur la manière dont les langues déjà acquises influencent l'écriture et ceci en examinant les choix lexicaux¹. L'analyse des choix lexicaux est efficace, car elle permet d'examiner plusieurs aspects de la compétence lexicale des apprenants en même temps (Jarvis & Pavlenko, 2008). Effectivement, le type de tâche exige que différents mots soient utilisés pour que le message soit transmis, ce qui implique qu'il est nécessaire d'avoir un vocabulaire assez étendu. Cependant, certains mots centraux du récit vont forcément manquer aux apprenants. Alors, que faire devant des problèmes de lexique? Les résultats des études antérieures montrent que plusieurs langues sont utilisées lors du processus d'écriture pour combler les lacunes lexicales. Ainsi, dans Lindqvist (2015), qui a uniquement inclus les apprenants ayant des connaissances en anglais à part le suédois, il s'est avéré que l'anglais était la source dominante du *transfert lexical*, c'est-à-dire les emplois lexicaux manifestant des influences d'autres langues que le français, et ceci dans toutes les classes. En effet, 70 % du total des occurrences de transfert provenaient de l'anglais. De plus, une différence qualitative a été notée dans le sens où les cas transférés du suédois étaient surtout des changements de langue non adaptés au français, p.ex. *berg* (fr. *montagne*, ang. *mountain*), *korg* (fr. *panier*, ang. *basket*), ou *borta* (fr. *disparu*, ang. *gone*). En revanche, ceux transférés de l'anglais étaient souvent des adaptations comme *basquette* (créée à partir du mot anglais *basket*, mot-cible français: *panier*) ou *travaille* (ang. *travel*, mot-cible français *vont*). Dans Lindqvist (2019) également, incluant cette fois tous les apprenants du corpus, les transferts lexicaux

provenaient des langues étrangères dans une plus large mesure que de la langue maternelle. Ainsi, alors que l'anglais dominait en tant que source d'influence, l'espagnol était aussi utilisé par un certain nombre d'apprenants.

Dans une étude visant à examiner d'autres aspects lexicaux à partir du même corpus, Lindqvist (2021) s'est focalisée sur certains items spécifiques dans la série d'images et sur la manière dont ils ont été produits par les apprenants dans leurs narrations écrites. Cette approche méthodologique a permis une analyse de la *profondeur* du vocabulaire, c'est-à-dire de l'aspect qualitatif des connaissances lexicales (à la différence de l'*étendue* du vocabulaire, qui concerne l'aspect quantitatif). Ainsi, l'analyse a porté sur l'orthographe, la relation forme-sens et les parties des mots, trois des aspects dans le modèle bien connu de Nation («How is the word written and spelled?», «What word form can be used to express this meaning?», «What word parts are needed to express the meaning?», Nation, 2020, p. 16). Les items inclus dans l'analyse étaient *mère*, *filles* et *garçon* (pour examiner l'orthographe), *chien* et *panier* (forme-sens), et *manger* et *dire* (parties des mots, soit morphèmes grammaticaux). Pour l'orthographe, les résultats ont mis en évidence un pattern assez variable en termes de difficulté et de degré d'exactitude. On note, par exemple, en principe les mêmes variantes d'orthographe de *mère* (*méré*, *mere*, *mère*, *maire*, *merè*, *mèrè*) dans toutes les classes, mais le degré d'exactitude est de plus en plus élevé à partir de la 7^e. Pour *filles*, en revanche, ce mot est correctement écrit sans exception en 6^e, alors que deux autres variantes apparaissent dans les autres classes (*fill*, *filles*). Concernant *garçon*, cette variante est accompagnée de *garcon* dans toutes les classes, tandis qu'il y a deux variantes supplémentaires en 6^e: *garquon*, *garzon*. Puis, pour l'analyse de la relation forme-sens, il est clair que le mot *panier* pose problèmes aux apprenants, résultant en l'emploi du suédois et de l'anglais (cf. plus haut), surtout différentes variantes du mot anglais *basket*: *basquette*, *basquet*, *bascet*, *baskuette*, *basquete*, *baskèt*, *baskèt*. Pour *manger*, le suédois (*äta*) et l'espagnol (*comer*) sont utilisés en 6^e et en 7^e, alors que les autres apprenants utilisent le mot cible sans exception. Enfin, pour les morphèmes grammaticaux, il s'est avéré que *dire* est très peu utilisé, mais

¹ Les aspects grammaticaux ont également été analysés dans certaines publications, cf. Lindqvist (2015 ; 2019).

que les formes sont toujours correctes. *Manger*, de l'autre côté, est employé dans toutes les classes avec une augmentation de flexions différentes et une maîtrise croissante au fur et à mesure des années.

Pour finir ce tour d'horizon des recherches menées sur le corpus en question, mentionnons aussi l'étude de Falk & Lindqvist (2022). Dans cette étude, les chercheurs examinent les attitudes d'enseignants vis-à-vis du multilinguisme en salle de classe des langues étrangères en Suède. Des extraits du corpus sont utilisés dans le but d'inviter les enseignants à réfléchir et à exprimer leurs attitudes sur la présence des langues déjà acquises dans les textes. Il s'est avéré que les attitudes divergent en fonction des langues : l'usage de l'anglais est généralement vu comme une stratégie efficace, contribuant à la compréhension du lecteur potentiel, tandis que l'emploi du suédois est considéré comme moins réussi.

Applications didactiques

Jusqu'ici, le corpus a surtout été exploré dans une perspective lexicale et translinguistique. En général, les résultats indiquent clairement que les langues déjà acquises sont utilisées dans une large mesure pour combler les lacunes lexicales. Un des résultats les plus intéressants d'un point de vue didactique est probablement que les langues étrangères sont souvent activées et employées de différentes manières lors de la rédaction d'un texte. Ainsi, à côté du suédois – la langue maternelle – les apprenants recourent à l'anglais ainsi qu'à l'espagnol à un haut degré. Notons que ces deux langues sont relativement proches du français, ce qui pourrait contribuer au fait qu'elles sont utilisées. En fait, les résultats d'une enquête utilisée dans l'étude de Lindqvist (2015) font ressortir que les apprenants perçoivent plus de similarités entre l'anglais et le français qu'entre le suédois et le français. Nous avons aussi pu constater que les enseignants suédois considèrent que l'emploi de l'anglais constitue une stratégie efficace de production écrite, alors qu'ils ne sont pas aussi positifs vis-à-vis du suédois. Ces résultats rappellent ceux obtenus dans le domaine des recherches sur l'acquisition d'une troisième langue, où maintes études ont montré que toutes les langues acquises peuvent impacter l'apprentissage d'une langue

étrangère supplémentaire. De surcroît, les chercheurs dans le domaine insistent souvent sur l'importance de reconnaître ce phénomène en salle de classe. De manière générale, une application didactique serait alors d'encourager l'emploi des langues déjà acquises. Les possibilités sont multiples, mais nous allons limiter notre discussion à l'utilité d'un corpus d'apprenants écrit en classe multilingue. Prenons à cette fin un exemple d'un texte écrit par un élève en 8^e.

Ex. 2. Texte d'un élève en 8^e

Une fille et un garçon adore sa mère. Ils sont manger avec sa mère. La filles c'est le soer de le garçon. Ils sont un chien qui s'apelle Pedde. Pedde aime manger beaucoup avec le garçon e la fille. Le garçon s'apelle Markus et la fille s'appelle Sofie. Sofie et Markus sur le **park**. Le **gras** c'est très verte et la souleille c'est très jaune. Ils sont une picnic. Dans le **basket** avec la **sallade** le chien Pedde être car Pede adore manger. Pede a mangé c'est tout dans le **basket**. Mais son bouteille au lau c'est n'ai pas manger. Sofie et Markus ne manger pas. Mais Sofie et markus sont son chien Pede. Sofie a cheveux blond ave une **dresse** blanche. Markus a cheveux noir ave une pulle blanche. Ils sont cinq ans et ils sont très gentille. Ils adore son chien et sa mère. Ils aimes le jour car c'est le très chaud et le soilleille brille.

Un des résultats les plus intéressants d'un point de vue didactique est probablement que les langues étrangères sont souvent activées et employées de différentes manières lors de la rédaction du texte.

Dans l'exemple 2, nous avons mis en caractères gras les mots manifestant des influences d'autres langues : *park* (ang./sué. *park*), *gras* (ang. *grass*, sué. *gräs*), *basket* (ang. *basket*), *sallade* (sué. *sallad*) et *dresse* (ang. *dress*). Ces exemples pourraient constituer la base d'une discussion en classe autour des différences et similarités entre différentes langues,

Ces exemples pourraient constituer la base d'une discussion en classe autour des différences et similarités entre différentes langues, par exemple au niveau de l'orthographe (*parc/park, sallad/salade*) et de la relation forme-sens (*dress/robe, basket/panier*).

par exemple au niveau de l'orthographe (*parc/park, sallad/salade*) et de la relation forme-sens (*dress/robe, basket/panier*). Ce genre de discussion pourrait contribuer à une situation d'apprentissage, non seulement pour ce qui est du développement linguistique mais aussi de la compétence métalinguistique et multilingue des apprenants. Ce corpus d'apprenants pourrait aussi être utilisé dans le cadre de l'écriture. En classe, l'enseignant pourrait choisir d'éviter de parler des influences d'autres langues comme une erreur et les soulever comme des exemples d'un trait caractéristique de la compétence linguistique de l'apprenant multilingue. En effet, le texte pourrait être traité en classe comme une première ébauche où l'apprenant a fait des efforts pour communiquer l'histoire en puisant dans son répertoire langagier contenant plusieurs langues, plutôt que d'abandonner devant des lacunes lexicales. Les élèves pourraient discuter des instances actuelles dans le texte et proposer des révisions, que ce soit au niveau du vocabulaire ou d'autres aspects de la langue. En résu-

mé, adopter une approche multilingue en salle de classe implique plusieurs avantages et l'emploi d'un corpus écrit d'apprenants multilingues peut s'avérer une ressource utile à cette fin.

Conclusion

Le but de cet article était de présenter un corpus d'apprenants contenant des productions écrites et les applications didactiques possibles de ce type de corpus. A partir des recherches précédentes sur le corpus, nous avons mis en exergue le caractère multilingue du corpus, permettant diverses applications didactiques en salle de classe. Nous avons notamment présenté les opportunités d'apprentissage et de réflexion métalinguistique à partir de textes contenant des influences de langues déjà acquises, trait inhérent au développement linguistique d'apprenants multilingues.

Bibliographie

Falk, Y. & Lindqvist, C. (2022). Teachers' attitudes towards multilingualism in the foreign language classroom: The case of French and German in the Swedish context. In: A. Krulatz, G. Neokleous & A. Dahl (Dir), *Theoretical and applied perspectives on teaching foreign languages in multilingual settings: pedagogical implications*. Bristol: Multilingual Matters, pp. 154-169.

Jarvis, S. & Pavlenko, A. (2008). *Crosslinguistic influence in language and cognition*. New York: Routledge.

Lindqvist, C. (2015). Do learners transfer from the language they perceive as most closely related to the L3? The role of psychotypology for lexical and grammatical cross-linguistic influence in French L3. In: G. De Angelis, U. Jessner & M. Kresic (Dir), *Multilingualism: Crosslinguistic influence in language learning*. London: Bloomsbury Publishing, pp. 231-251.

Lindqvist, C. (2019). Didactic challenges in the multilingual classroom. The case of French as a foreign language. In: M. J. Gutierrez-Mangado, M. Martínez Adrián, F. Gallardo Del Puerto (Dir), *Cross-Linguistic Influence: From Empirical Evidence to Classroom Practice*. Springer International Publishing Switzerland, pp. 87-99.

Lindqvist, C. (2021). Vocabulary knowledge in L3 French: A study of Swedish learners' vocabulary depth. *Languages* 6, no. 1: 26. <https://doi.org/10.3390/languages6010026>

Nation, P. (2020). The different aspects of vocabulary knowledge. In: S. Webb (Dir), *The Routledge handbook of vocabulary studies*. New York: Routledge, pp. 15-29.

Tracy-Ventura, N. & Paquot, M. (2020). Second language acquisition and corpora: An overview. In: N. Tracy-Ventura & M. Paquot (Dir), *The Routledge handbook of second language acquisition and corpora*. New York: Routledge, pp. 1-8.

Annexe

The Dog Story

