

Élaboration d'une liste de vocabulaire dans un contexte FLE

Elaboration of a vocabulary list in a French as a foreign language context

Christina Lindqvist

Université de Göteborg

christina.lindqvist@sprak.gu.se

ORCID : 0000-0003-1590-1437

Mårten Ramnäs

Université de Göteborg

marten.ramnas@sprak.gu.se

ORCID : 0000-0002-2531-0019

Abstract

In this article, we discuss the elaboration and use of vocabulary lists aimed for learners of French as a foreign language. These lists are commonly based on corpora, which, in the ideal case, are representative, relevant and large. As for the English language, this kind of corpora has been available for a long time (e.g. the COCA and the BNC). Vocabulary lists, which are often used in learning contexts, have been based on these corpora. The situation is, however, less favorable when it comes to the French language, with fewer corpora meeting the mentioned criteria, and, thus, fewer possibilities to create vocabulary lists that are useful for learners. In this contribution, we present work that has been done in order to create a vocabulary list, *Riksprovsordlistan*, containing about 4,000 words and used at all Swedish universities. The discussion focuses on methodological challenges such as choice of counting unit – lemma vs. word family –, the role of frequency, thematic vocabulary, as well as characteristics of written vs. spoken corpora.

Keywords: vocabulary lists, learning, French as a foreign language, corpora

1. INTRODUCTION

Dans cet article, nous discuterons de l'élaboration et l'utilisation des listes de vocabulaire dans l'enseignement des langues en prenant comme exemple l'enseignement du français langue étrangère (FLE) dans le contexte universitaire suédois, où nous menons un projet consacré au développement de l'enseignement du vocabulaire depuis quelques années. En effet, comme nous l'avons fait remarquer ailleurs (Lindqvist & Ramnäs, 2017 ; 2020), l'enseignement du vocabulaire a reçu relativement peu d'attention dans les universités suédoises historiquement, surtout en comparaison avec l'enseignement de la grammaire. De ce fait, selon notre expérience, les étudiants sont en principe censés apprendre le vocabulaire seuls, en dehors de la salle de classe. En outre, et d'autant plus important dans le cadre de cette contribution, il y a eu, jusqu'aux années 2010, un manque de matériaux appropriés à l'apprentissage du vocabulaire. Ainsi, il n'y avait, par exemple, pas de liste de vocabulaire pertinente pour les apprenants, ce qui a, bien évidemment, rendu l'apprentissage individuel plus difficile. Cependant, la situation s'est améliorée ces dernières années. Il existe maintenant une liste de vocabulaire commune contenant environ 4 000 mots, *Riksprovsord-listan*, valable pour tous les étudiants en première année de français en Suède, quelle que soit l'université. Cette liste est à l'origine basée sur un corpus du français parlé, *Corpaix* (Campione, Véronis & Deulofeu, 2005), et a, par la suite, été modifiée de différentes manières, comme nous le verrons dans les sections suivantes. L'objectif de cet article est de soulever des problèmes liés à l'élaboration de cette liste de vocabulaire. Plus précisément, nous nous posons la question de savoir si elle est appropriée et utile dans une perspective d'apprentissage.¹ Les étudiants sont censés apprendre les mots grâce à divers méthodes et exercices proposés dans les cours. Cependant, cet article ne s'intéressera pas aux méthodes d'enseignement.

L'enseignement du vocabulaire basé sur des listes tirées de corpus importants a une longue tradition dans le cadre de l'enseignement de l'anglais (Dang, 2020). Ceci est sans doute lié au fait qu'il existe des bases de données établies dans cette langue depuis longtemps. On pourrait, par exemple, mentionner le COCA (*Corpus of Contemporary English*) et le BNC (*British National Corpus*). À partir de ces types de bases de données, il est possible de créer des listes selon la fréquence des mots dans la langue en général (cf. Aston & Burnard, 1998 ; Davies & Gardner, 2010). La fréquence est depuis des décennies considérée comme le facteur le plus important pour

¹ Il convient de signaler que la liste en question a été constituée à partir de la fréquence de mots isolés dans un corpus (voir plus loin), qui ne contenait pas d'expressions figées. Ainsi, tout en admettant que celles-ci sont fort importantes et intéressantes du point de vue de l'acquisition aussi bien que de l'enseignement, nous ne les incluons pas dans notre discussion. Pour la fréquence des collocations françaises, voir Benigno, Grossmann & Kraif (2015). Pour une analyse approfondie de l'acquisition des expressions figées, soit des *séquences préfabriquées*, d'apprenants suédophones de français, nous renvoyons le lecteur à la thèse de doctorat de Fanny Forsberg (2006).

l'apprentissage du vocabulaire en langue étrangère (Cobb, 2013 ; Milton, 2009), dans le sens que l'on a tendance à d'abord apprendre les mots les plus fréquents et ensuite les mots les moins fréquents dans la langue cible. Les mots les plus fréquents sont alors considérés comme constituant le vocabulaire de base ou fondamental (Laufer & Nation, 1995). Voilà donc la raison pour laquelle les listes de vocabulaire sont souvent des listes de fréquence, fondées sur des corpus de la langue cible. La plupart des chercheurs sont d'accord pour dire qu'il est approprié de faire une distinction entre les 2 000 mots les plus fréquents – à considérer comme les mots de haute fréquence, et ainsi constituant le vocabulaire de base – et les autres mots moins fréquents d'une langue donnée (Dang, 2020, p. 297 ; Laufer & Nation, 1995). Comme le constatent Siepmann, Bürgel & Diwersy (2016, p. 1) et Cobb (2014, p. 275), il y a moins de corpus récents et accessibles pour ce qui est de la langue française par rapport à la langue anglaise, mais nous pouvons mentionner *A Frequency Dictionary of French* (Lonsdale & Le Bras, 2009), *Le Corpus de référence du français contemporain* (CRFC) (Siepmann, Bürgel & Diwersy, 2016), ou encore le site *Lexique* (www.lexique.org), qui contient un certain nombre de bases de données. Signalons aussi les listes plus anciennes *Dictionnaire fondamental de la langue française* (Gougenheim, 1958), *Le français fondamental* (Gougenheim, Michéa, Rivenc & Sauvageot, 1967), ou *Les fréquences d'utilisation des mots en français écrit contemporain* (Baudot, 1992).

Afin de proposer des listes utiles aux apprenants, il est souhaitable que celles-ci soient basées sur des corpus remplissant certains critères : ils doivent être représentatifs, pertinents et de grande taille (Dang, 2020). En d'autres termes, une liste utile contient les mots qui représentent la langue utilisée par les locuteurs natifs, dans le plus de contextes différents possible, à l'écrit comme à l'oral. Les mots doivent aussi être pertinents et utiles pour les apprenants. Pourtant, ceci n'est pas toujours le cas, au moins dans le contexte suédois. Au contraire, à notre connaissance, les listes qui ont traditionnellement été utilisées dans les universités suédoises ont souvent été « établies en se basant sur l'intuition et non sur des critères de sélection scientifiques fruits de recherches dans le domaine de l'apprentissage du vocabulaire » (Lindqvist & Ramnäs, 2017, p. 60). Or, ceci a commencé à changer ces dernières années. Dans cette contribution, nous présenterons et examinerons d'un œil critique le travail visant à créer une liste de vocabulaire français destinée aux apprenants suédois en première année à l'université, en nous rapportant aux critères mentionnés ci-dessus.

2. LA CRÉATION D'UNE LISTE DE VOCABULAIRE DESTINÉE AUX APPRENANTS SUÉDOIS

À notre université, ainsi que dans toutes les universités en Suède où le français est enseigné, les étudiants en première année ont depuis une dizaine d'années recours à une liste de vocabulaire contenant à l'heure actuelle environ 4 000 mots, *Riksprov-*

sordlistan. Vers la fin du premier semestre, les étudiants doivent tous passer un test de grammaire et de vocabulaire, commun à toutes les universités. Pour ce qui est de la partie vocabulaire, les mots testés sont sans exception tirés de la liste. Dans les sections suivantes, nous allons examiner l'élaboration et l'utilité de cette liste, tout en nous référant aux notions centrales évoquées en introduction : taille, représentativité et pertinence.

La première version de la liste² a été créée dans le cadre d'un programme de recherche à l'Université de Stockholm, *High-level proficiency in second language use*, à savoir du sous-projet *Aspects of the advanced learner's lexicon* (cf. p. ex. Lindqvist, Bardel & Gudmundson, 2011). L'objectif principal de ce projet était d'examiner la compétence lexicale chez les apprenants suédois avancés de français et d'italien³. Se focalisant sur la richesse lexicale dans la langue parlée des apprenants, il était nécessaire de pouvoir la comparer à la langue parlée des locuteurs natifs, c'est-à-dire à un corpus de référence. L'idée de départ était de suivre de près la méthode *Lexical Frequency Profile* élaborée par Laufer & Nation (1995) pour l'anglais écrit. Cette méthode permet de diviser les mots de la production d'un apprenant dans différentes zones de fréquence, l'hypothèse étant qu'une proportion relativement élevée de mots peu fréquents serait signe d'un lexique riche des apprenants (cf. le critère de la fréquence évoqué plus haut). Il est donc important de souligner que la liste originale n'était pas initialement destinée à l'utilisation dans le cadre de l'enseignement. Nous y reviendrons plus loin.

La liste originale a été fondée sur le *Corpaix*, qui avait été constitué à l'Université de Provence (Campione, Véronis & Deulofeu, 2005), et qui avait aussi été utilisé afin de définir une liste de vocabulaire élémentaire par Tidball & Treffers-Daller (2008) dans le contexte britannique. Ce corpus, entièrement basé sur la langue parlée, contient environ un million de mots, produits par des locuteurs natifs de français lors de différentes tâches, p. ex. des entretiens et des conversations traitant de différents thèmes comme les voyages, les souvenirs personnels, la politique et les expériences professionnelles. Les enregistrements ont tous été effectués dans des contextes naturels, non formels, et de manière relativement spontanée, à la maison, au travail, dans un magasin, dans un bureau, etc. (Pallaud & Henry, 2004). Une liste présentant le rang et la fréquence des formes des mots (les mots-types) était accessible en ligne à l'époque. Seuls les mots-types comptant plus de dix occurrences figuraient dans la liste. Dans le cadre du projet mentionné plus haut, la liste a été lemmatisée par le moyen du logiciel *TreeTagger* (Schmid, 1994 ; 1995), et puis la fréquence de chaque lemme a été calculée à l'aide du logiciel *WordSmith* (Scott, 2004). La liste finale contient 2 746 lemmes (cf. Lindqvist, Gudmundson & Bardel, 2013, pour plus de détails).

² La liste complète n'a pas été publiée, mais une liste des 100 mots les plus fréquents est présentée en annexe dans Lindqvist (2021).

³ Nous n'aborderons pas la partie du projet consacrée à la langue italienne dans cette contribution. Pour plus d'informations sur le corpus italien, voir Lindqvist, Gudmundson & Bardel (2013).

Hormis l'utilisation dans le cadre du projet de recherche, cette liste a donc été adoptée dans l'enseignement du vocabulaire à l'université en Suède. Ceci n'était pas prévu, puisque la liste avait été élaborée dans un contexte scientifique délimité, avec un but spécifique ; or, comme nous l'avons signalé plus haut, il y avait un manque de matériaux sur lesquels les apprenants pouvaient travailler le vocabulaire. Étant donné l'importance de la fréquence des mots pour l'apprentissage, on pourrait postuler que la liste a servi à constituer un vocabulaire de base, ainsi qu'un certain nombre de mots moins fréquents, couvrant presque les 3 000 mots les plus fréquents de la langue parlée (cf. Dang, 2020). Pourtant, on pourrait, au moins en partie, questionner sa pertinence quant aux critères évoqués en introduction. Nous reviendrons sur ces critères dans la section suivante.

3. LA LISTE ORIGINALE FACE AUX CRITÈRES DE TAILLE, DE REPRÉSENTATIVITÉ ET DE PERTINENCE

Pour commencer, nous constatons que la liste originale basée sur *Corpaix* est fondée sur la fréquence des mots dans la langue française, ce qui correspond bien à l'état des lieux de la recherche sur l'apprentissage du vocabulaire, où les chercheurs insistent sur le rôle primordial du facteur de la fréquence (Milton, 2009 ; Dang, 2020). En même temps, on pourrait se demander si elle est représentative de la langue française en général. Nous rappelons que le corpus de départ, *Corpaix*, était d'une taille relativement restreinte : un million de mots. De plus, les données ont été collectionnées dans le sud de la France, et sur la langue parlée uniquement, ce qui nous mène à questionner la pertinence de la liste dans un contexte d'apprentissage comme le nôtre. En fait, nous avons récemment fait une comparaison entre la liste du *Corpaix* et la liste des 5 000 mots les plus fréquents élaborée par Lonsdale & Le Bras (2009 ; cf. Lindqvist & Ramnäs, 2020). Cette liste, *A Frequency Dictionary of French*, a été conçue à partir d'un corpus de 23 millions de mots, incluant des données de la langue parlée et écrite provenant d'un grand nombre de sources et genres différents. Notre comparaison a montré que la liste originale basée sur *Corpaix* contient un certain nombre de mots peu fréquents dans la langue française en général, et qui ne font pas partie de la liste des 5 000 mots les plus fréquents de Lonsdale & Le Bras (2009), comme par exemple *bêche*, *bûche*, *jonquille*, *ouate*, *radoter* et *joufflu*. Par la suite, nous avons fait une recherche sur le site *Compleat lexical tutor* (créé par Cobb, www.lextutor.ca), qui contient les 25 000 mots les plus fréquents dans le corpus utilisé par Lonsdale & Le Bras. Il s'avère en effet que ces mots apparaissent dans des zones de fréquences très basses : *bêche* (16K⁴), *bûche*

⁴ La lettre *K* correspond à 1 000 et est habituellement utilisée en faisant référence aux zones de fréquence. Ainsi, la zone de fréquence 1K inclut les 1 000 mots les plus fréquents dans la langue en question, et ainsi de suite.

(13K), *jonquille* (23K), *ouate* (20K), *radoter* (20K), *joufflu* (21K). Selon nous, ces résultats nous permettent d'avancer que ces mots ne devraient pas faire partie des mots utiles à connaître pour un étudiant suédois de français en première année. En revanche, nous avons aussi pu constater dans une étude antérieure (Lindqvist, 2010), que des mots comme *couteau*, *bière* et *pharmacie* sont peu fréquents dans le corpus *Corpaix*, des mots qui seraient d'autant plus utiles à connaître pour les apprenants, déjà à un stade initial de l'acquisition. Ces mots sont aussi relativement peu fréquents dans le corpus de Lonsdale & Le Bras, d'après une recherche sur lextutor.ca, appartenant aux zones 4K (*couteau* et *bière*) et 7K (*pharmacie*). En effet, cette problématique n'est pas nouvelle, car Gougenheim, Michéa, Rivenc & Sauvageot (1967, p. 145) ont évoqué la notion de *vocabulaire disponible* pour les mots « d'une fréquence faible et peu stable, qui sont toutefois des mots usuels et utiles ». Cette catégorie de mots inclut ainsi des mots utiles, connus par tous les locuteurs natifs d'une langue donnée, mais qui apparaissent néanmoins rarement dans la langue en général. Donc, l'idée selon laquelle la fréquence est le seul critère pour l'établissement d'une liste de vocabulaire a longtemps été remise en question, et nos exemples semblent appuyer cette critique. En effet, d'autres critères devraient être considérés, comme nous le verrons dans ce qui suit.

Il convient, par exemple, d'aborder le rôle de la fréquence vis-à-vis des mots dits thématiques. Cette catégorie contient des mots que l'on retrouve habituellement dans les manuels scolaires, qui sont souvent structurés autour de thèmes, p. ex. la nourriture, la maison, les vêtements, etc. (Milton, 2009). Souvent, ce genre de mots ne sont pas fréquents dans la langue en général, ce qui fait que les apprenants vont rencontrer des mots peu fréquents en début d'apprentissage. En même temps, les mots thématiques sont utiles et nécessaires à connaître. Ainsi, cette catégorie ressemble à la catégorie de vocabulaire disponible (*cf.* ci-dessus), mais elle est définie différemment, à partir de matériaux particuliers, utilisés dans un contexte restreint, et non basée sur l'utilisation dans la langue en général. En fait, on pourrait se demander si, dans un contexte d'apprentissage formel, où l'input des élèves est souvent limité aux manuels scolaires et à la langue utilisée par l'enseignant, la fréquence des mots diffère de celle de la langue en général. Comme nous l'avons fait remarquer ailleurs (Lindqvist, 2021, p. 87) : « Dans une perspective didactique, il serait peut-être intéressant de créer des listes de fréquence à partir de manuels pour avoir une image représentative de la fréquence des mots dans ce genre de texte spécifique, pour un public consacré ». Il semblerait effectivement que les mots qui sont fréquents et utiles pour les apprenants débutants ne font pas forcément partie des mots les plus fréquents de la langue en général. Ainsi, on pourrait considérer l'élaboration de ce genre de liste comme un complément aux listes de fréquence traditionnelles, comme celle de Lonsdale & Le Bras (2009). Ceci dans le but de pouvoir présenter aux élèves un vocabulaire de base utile et pertinent.

Ces défauts de la liste originale basée sur *Corpaix*, dans une perspective pédagogique, ont aussi été notés par les enseignants des différentes universités suédoises. Voilà pourquoi la liste a connu quelques modifications d'après des discussions menées dans le groupe des enseignants. Ont été ajoutés des mots censés être utiles pour les apprenants, ainsi que des expressions figées. Les ajouts comptent environ 1 300 mots, ce qui fait que *Riksprovsordlistan* contient à l'heure actuelle environ 4 000 mots au total. Or, même si la liste correspond maintenant mieux aux besoins des apprenants, les ajouts n'ont pas été faits selon des critères scientifiques, mais plutôt en utilisant la méthode « intuitive » des enseignants, ce qui n'est pas une démarche souhaitable, comme nous l'avons évoqué plus haut.

Une autre question centrale concerne l'unité de comptage. Contrairement à beaucoup de listes de fréquence sur la langue anglaise, la liste française contient les lemmes les plus fréquents. D'ailleurs, c'est aussi le cas de la liste de Lonsdale & Le Bras (2009) et *Riksprovsordlistan*. Par contre, les listes anglaises sont, dans la plupart des cas, basées sur des familles de mots, celles-ci contenant un mot de base et toutes ses formes fléchies et dérivées (Nation, 2001). Nous admettons que le concept de famille de mot semble offrir plusieurs avantages (cf. Ramnäs & Lindqvist, 2021). Lindqvist, Gudmundson & Bardel (2013), pour leur part, expliquent le choix d'utiliser le lemme en tant qu'unité de comptage dans leur projet sur la richesse lexicale en production orale, où elles ont développé un logiciel similaire au *Lexical Frequency Profile* (cf. plus haut), en disant qu'en production orale, il n'est pas sûr que l'apprenant connaissant un mot membre soit à même de produire aussi tous les autres membres de la famille. Cette problématique est par ailleurs exemplifiée dans une perspective pédagogique dans Ramnäs & Lindqvist (2021) : le mot *invisibilité* appartient à la zone de fréquence 19K dans le corpus de Lonsdale & Le Bras (2009), alors que *visible*, qui serait le mot de base de la famille, figure dans la zone de fréquence 3K. Le choix de l'unité de comptage influe sur l'évaluation de la compétence lexicale des apprenants. Ainsi, l'utilisation de formes lexicales rares comme *invisibilité*, qui appartient à une famille de mot fréquente, indiquerait un lexique riche, alors qu'au fond la famille de mots en question ferait partie du vocabulaire de base. Le mot *invisibilité* appartient-il à un vocabulaire de base ou pas ?

Dans une perspective didactique d'enseignement, cela pourrait être intéressant de proposer une liste de familles de mots pour la langue française. Ramnäs & Lindqvist (2021) discutent les avantages et les inconvénients de différentes unités de comptage et plaident en faveur des familles de mots dans un contexte d'enseignement. Selon les chercheurs, cette unité présente des avantages surtout quant aux connaissances réceptives des apprenants. D'abord, la charge d'apprentissage (*learning burden*) des apprenants serait réduit, et l'apprentissage plus efficace, puisqu'ils auraient recours à plus de formes d'une même unité lexicale en même temps. De même, Cobb (2013 : 80) soutient que la notion de famille de mots est « based on learning principles rather

than linguistic or computational principles ». De plus, l'utilisation des familles de mots permettrait à l'enseignant de rapprocher les domaines de grammaire et de vocabulaire dans le sens où il faudra focaliser les aspects morphologiques dans le cadre du vocabulaire, en enseignant les suffixes et les préfixes, par exemple.⁵

4. CONCLUSION

Dans cet article, nous avons examiné l'utilité d'une liste de vocabulaire, *Riksprovordlistan*, dans l'enseignement du français langue étrangère en Suède, tout en relevant les défis associés à ce travail. Ces défis sont liés à la fois au choix du corpus, à l'unité de comptage et à la constitution de la liste. Nous nous sommes posé la question de savoir si la liste est appropriée et utile pour les étudiants auxquels elle est destinée. Quant au corpus, nous avons pu constater que celui qui a été utilisé au départ avait quelques limites, étant de taille relativement restreinte, incluant la langue orale uniquement et se concentrant sur une zone géographique spécifique de la France. Ainsi, il n'est probablement pas complètement représentatif de la langue française en général. Même si les mots de très haute fréquence (pronoms, déterminants, prépositions, verbes auxiliaires) y apparaissent, le corpus ne permet pas d'engendrer une liste de vocabulaire fondamentale et pertinente pour les apprenants. Dans l'ensemble, nous dirions donc que la liste en question n'est pas tout à fait utile pour être employée dans un contexte d'apprentissage. Les ajouts qui ont été faits remédient certes à quelques-uns des défauts, mais le fait qu'ils ont été faits de manière intuitive et subjective, comme expliqué ci-dessus, constitue un problème. Dans l'avenir, nous nous proposerons d'élaborer une nouvelle liste de vocabulaire basée sur un corpus plus représentatif de la langue française orale et écrite en général, qui soit de plus grande taille. De plus, nous voyons l'intérêt de présenter une liste contenant des familles de mots, permettant aux étudiants d'apprendre plus de mots de manière plus efficace. Enfin, comme évoqué ailleurs (Ramnäs & Lindqvist, 2021), une liste de familles de mots pourrait présenter des avantages dans une perspective didactique. En utilisant une telle liste, les enseignants pourraient faire des liens entre l'enseignement du vocabulaire et celui de la grammaire, puisqu'il sera alors logique de focaliser sur les affixes dérivationnels.

⁵ Nous renvoyons le lecteur à Ramnäs et Lindqvist (2021) pour une discussion plus approfondie.

BIBLIOGRAPHIE

- Aston, G. & Burnard, L. (1998). *The BNC handbook : exploring the British National Corpus with SARA*. Edinburgh : Edinburgh University Press.
- Baudot, J. (1992). *Les fréquences d'utilisation des mots en français écrit contemporain*. Montréal : Presses de l'Université de Montréal.
- Benigno, V., Grossmann, F. & Kraif, O. (2015). Les collocations fondamentales : une piste pour l'apprentissage lexical. *Revue française de linguistique appliquée*, 1, 81-96.
- Campione, E., Véronis, J. & Deulofeu, J. (2005). The French corpus. In E. Cresti & M. Moneglia (eds.), *C-ORAL-ROM : Integrated reference corpora for spoken romance languages* (pp. 111-133). Amsterdam : John Benjamins.
- Cobb, T. (2013). Frequency 2.0. Incorporating homoforms and multiword units in pedagogical frequency lists. In C. Bardel, C. Lindqvist & B. Laufer (ed.), *L2 vocabulary acquisition, knowledge and use : new perspectives on assessment and corpus analysis* (pp. 79-108). Amsterdam : Eurosla.
- Cobb, T. (2014). A resource wish-list for data-driven learning in French. In H. Tyne, V. André, A. Boulton, C. Benzitoun & Y. Greub (dir.), *French through corpora : Ecological and data-driven perspectives in French language studies* (pp. 255-290). Newcastle : Cambridge Scholars.
- Dang, T.N.Y. (2020). Corpus-based word lists in second language vocabulary research, learning, and teaching. In S. Webb (eds.), *The Routledge Handbook of Vocabulary Studies* (pp. 288-303). New York : Routledge.
- Davies, M. & Gardner, D. (2010). *Frequency dictionary of contemporary American English : Word sketches, collocates, and thematic lists*. New York : Routledge.
- Forsberg, F. (2006). *Le langage préfabriqué en français parlé L2 : Étude acquisitionnelle et comparative*. (Thèse de doctorat). Université de Stockholm.
- Gougenheim, G. (1958). *Dictionnaire fondamental de la langue française*. Paris : Didier.
- Gougenheim, G., Michéa, R., Rivenc, P. & Sauvageot, A. (1967). *L'Élaboration du français fondamental (1er degré)*. Paris : Didier.
- Laufer, B. & Nation, P. (1995). Vocabulary size and use : Lexical richness in L2 written production. *Applied Linguistics*, 16 (3), 307-322.
- Lindqvist, C. (2010). La richesse lexicale dans la production orale de l'apprenant avancé de français. *La revue canadienne des langues vivantes/The Canadian Modern Language Review*, 66 (3), 393-420.
- Lindqvist, C. (2021). L'apprentissage du vocabulaire en langue étrangère. In P. Leclercq, A. Edmonds & E. Sneed German (eds.), *Introduction à l'acquisition des langues étrangères* (pp. 79-92). Louvain-la-Neuve : De Boeck Supérieur.
- Lindqvist, C., Bardel, C. & Gudmundson, A. (2011). Lexical richness in the advanced learner's oral production of French and Italian L2. *International Review for Applied Linguistics in language teaching*, 49 (3), 221-240.
- Lindqvist, C., Gudmundson, A. & Bardel, C. (2013). A new approach to the measuring of vocabulary in L2 oral production. In C. Bardel, C. Lindqvist & B. Laufer (eds.), *L2 vocabulary acquisition, knowledge and use. New perspectives on assessment and corpus analysis* (pp. 109-126). Amsterdam : Eurosla.
- Lindqvist, C. & Ramnäs, M. (2017). L'enseignement du vocabulaire à l'université. *Synergies pays scandinaves*, 11-12, 55-64.
- Lindqvist, C. & Ramnäs, M. (2020). Vokabulärundervisning i franska – då, nu och i framtiden. In A. Romeborn & E. Bladh (eds.), *Romanistiken i Sverige. Tradition och förnyelse* (pp. 265-280). Göteborg : Acta Romanica Gothoburgensis/Kriterium.
- Lonsdale, D. & Le Bras, Y. (2009). *A Frequency Dictionary of French*. New York : Routledge.
- Milton, J. (2009). *Measuring second language vocabulary acquisition*. Bristol : Multilingual Matters.
- Nation, I.S.P. (2001). *Learning vocabulary in another language*. Oxford : Oxford University Press.

- Pallaud, B. & Henry, S. (2004). Amorces de mots et répétitions : des hésitations plus que des erreurs en français parlé. In G. Purnelle, C. Fairon & A. Dister (dir.), *Le poids des mots. Actes des 7es Journées internationales d'Analyse statistique des Données Textuelles* (pp. 848-858). Louvain-la-Neuve : Presses universitaires de Louvain.
- Ramnäs, M. & Lindqvist, C. (2021). Le concept de famille de mot appliqué au FLE. *Langues modernes*, 3, 45-53.
- Schmid, H. (1994). Probabilistic part-of-speech tagging using decision trees. *Proceedings of International Conference on New Methods in Language Processing* (pp. 44-49). Manchester, UK.
- Schmid, H. (1995). Improvements in part-of-speech tagging with an application to German. *Proceedings of the ACL SIGDAT-Workshop* (pp. 47-50). Dublin, Ireland.
- Scott, M. (2004). *WordSmith tools version 4*. Oxford : Oxford University Press.
- Siepmann, D., Bürgel, C. & Diwersy, S. (2016). Le Corpus de référence du français contemporain (CRFC), un corpus massif du français largement diversifié par genres. *SHS Web of conferences*, 27. https://www.shs-conferences.org/articles/shsconf/pdf/2016/05/shsconf_cmlf2016_11002.pdf [accès : 16.12.2022].
- Tidball, F. & Treffers-Daller, J. (2008). Analysing lexical richness in French learner language : What frequency lists and teacher judgment can tell us about basic and advanced words. *French Language Studies*, 18 (3), 299-313.