

## Difficulté des listes thématiques d'un ouvrage bilingue selon la fréquence d'usage des mots

Alain Lortet, Université du Québec à Montréal, Canada

*Résumé : Les ouvrages de vocabulaires thématiques (OVT) sont formés de listes de mots dont le choix est souvent subjectif. Pourtant, la fréquence d'usage des mots permet de sélectionner le vocabulaire le plus courant qui est normalement appris en priorité. Dans la présente recherche, nous analysons un OVT, destiné aux apprenants de niveaux B1-B2 selon l'échelle du Cadre européen commun de référence pour les langues (CECRL), afin de vérifier la difficulté des listes thématiques de mots à l'aide de leur fréquence d'usage. Les résultats obtenus démontrent que la fréquence n'est aucunement prise en compte dans cet ouvrage et que le vocabulaire semble être choisi de manière subjective. Nous faisons alors plusieurs suggestions pour améliorer la réalisation des OVT.*

*Mots-clés : Fréquence des mots, Vocabulaire thématique, Ouvrage de vocabulaire*

*Abstract : Thematic Vocabulary Books consist in being lists of words whose choice is often subjective. However, the frequency of use of words makes it possible to select the most common vocabulary which is normally learned in priority. In this research, we analyze an Thematic Vocabulary Book, intended for B1-B2 level learners according to the Cadre européen commun de référence pour les langues (CECRL) scale, to check the difficulty of thematic lists of words with the help of their frequency of use. The results obtained show that the frequency is not taken into account in this book and that the vocabulary seems to be chosen subjectively. We then make several suggestions to improve the construction of Thematic Vocabulary Books.*

*Key words : Frequency of words, Thematic vocabulary, Vocabulary book*

### Introduction

Les listes de mots peuvent constituer un moyen rapide et efficace pour pallier un manque de vocabulaire chez un apprenant, aussi bien en langue première (L1) qu'en langue seconde (L2) ou en langue étrangère (LE) (Cobb et Horst, 2011). Elles peuvent être utilisées par l'enseignant pour la préparation d'un cours en L1 (Daigle, Anctil, Berthiaume et Bruneau, 2015) ou par l'apprenant pour la préparation d'examens de français en LE (Lindqvist et Ramnäs, 2016). Ces listes de mots peuvent être organisées par thèmes, comme dans les ouvrages de vocabulaires thématiques (OVT). Historiquement, ceux-ci ont même devancé les dictionnaires alphabétiques puisque leur principe d'ordonnement des mots par thèmes remonte à l'Antiquité (Ruchot, 2011). Leur réalisation implique différents choix, notamment ceux des thèmes, voire des sous-thèmes, et des mots constituant les listes thématiques.

### Les listes thématiques des ouvrages de vocabulaires thématiques

Les OVT se distinguent notamment par les thèmes (des concepts selon Ruchot) et les sous-thèmes qui les composent (Ruchot, 2011). Si certains thèmes sont récurrents dans les OVT et dans d'autres manuels de langue, d'autres sont moins systématiques. On retrouve ainsi fréquemment les thèmes de la *maison*, des *aliments*, de la *famille*, des *vêtements*, du *corps humain*, du *voyage* ou du *transport* (Havránková, 2015). D'autres thèmes sont rares comme la *religion*, le *bricolage* et le *jardinage*, le *ciel*, etc. (Havránková, 2015). L'organisation des thèmes diffère également d'un ouvrage à l'autre. Le thème des *achats* par exemple peut générer des sous-thèmes organisés d'après des noms (*commerce*, *magasinsboutiques*, *vêtements*, *supermarché*, *professions*, *nourriture*) ou des verbes (*acheter*, *vendre*) (Havránková, 2015). Le choix des thèmes et leur organisation ne font donc pas consensus entre les OVT.

Soulignons également que les mots contenus dans chaque thème diffèrent d'un manuel à l'autre, les listes pouvant se rallonger ou se raccourcir selon les auteurs. Le thème de la *famille*, par exemple, contient systématiquement les *membres*, mais occasionnellement les *événements familiaux*, voire les *familles traditionnelles et recomposées* (Havránková, 2015). Notons aussi que les mots ne sont souvent classés que dans un seul thème, de manière arbitraire, alors qu'ils pourraient l'être dans plusieurs (Ruchot, 2011). Il est également plus facile de classer des notions concrètes (les *parties du corps*, la *maison* ou les *vêtements*) que des notions abstraites (la *tristesse*, la *jalousie* ou l'*envie*) (Ruchot, 2011). La constitution des listes thématiques semble donc manquer de rigueur scientifique.

### Critères de sélection du vocabulaire

Le CECRL (2001) relève plusieurs critères de sélection du vocabulaire pour l'enseignement des langues, critères auxquels nous donnerons ici les noms de *spécifique* (mots ou expressions clés pour la réalisation de tâches communicatives issus de domaines thématiques ou en lien avec la culture des locuteurs de la L2), de *statistique* (selon la fréquence des mots issus de corpus), d'*authentique* (documents oraux ou écrits) et de *pragmatique* (non planifié, mais dépendant des besoins de l'apprenant lors de tâches communicatives). Notons également le critère de rentabilité/productivité correspondant à des termes génériques utiles pour obtenir des termes plus concrets ou spécifiques (Paccaud, 2011) et que nous qualifierons ici de critère *générique*.

Le critère *statistique* se distingue des autres du fait que l'exposition aux mots semble l'emporter sur l'enseignement des mots. En effet, les mots appris prioritairement ne sont pas forcément ceux enseignés en classe puisque le vocabulaire est également appris en dehors de la classe (Pignot-Shahov, 2014). L'une des conséquences est que les mots les plus fréquents sont appris plus tôt que les moins fréquents, aussi bien en L1 (Tomasello, 2003; Grossmann, 2011) qu'en L2 (Ellis, O'Donnell et Römer, 2015). Les mots fréquents sont même appris avant ceux enseignés en salle de classe lorsque des mots plus rares y sont montrés (Richards et Malvern, 2007) : « l'exposition au langage l'emporte sur les sélections thématiques particulières des manuels scolaires » (Pignot-Shahov, 2014, p. 124). Si l'on ajoute que le vocabulaire enseigné en classe est limité et dépend des choix pédagogiques (Havránková, 2015), la prise en compte de la fréquence des mots dans le choix du vocabulaire à enseigner devient alors primordiale.

### La fréquence d'usage des mots

La fréquence d'usage des mots est relativement célèbre depuis les travaux de Gougenheim (Benigno, Grossman et Kraif, 2015). Son *Français fondamental*, dont la première version est publiée en France, sous le titre *Le français élémentaire*, par le ministère de l'Éducation nationale (1954), contenait 1500 mots fréquents à l'oral (Cuq et Gruca, 2005). Comme le rappelle Havránková (2015), la fréquence a également été utilisée par plusieurs chercheurs ou enseignants à différentes époques. L'Abbé de l'Épée (1776) avait choisi 5400 mots censés être les plus importants afin de les enseigner à des sourds-muets. Il les avait segmentés en trois tranches de 1800 mots chacune correspondant à trois étapes d'apprentissage (Asmi, 2013). En 1924, le dictionnaire d'Henmon comportait les 3905 mots les plus fréquents d'un ensemble de 400 000 mots issus d'un corpus littéraire (Asmi, 2013). En 1928, l'institutrice Florence Baker s'est basée sur la fréquence et le vocabulaire observés dans sa classe de français pour établir un total de 400 mots censés être enseignés à des apprenants débutants et dont 93,4 % des mots sont inclus dans le *Français fondamental* (Huet, 2002). Un autre dictionnaire notable est celui réalisé par Juilland, Brodin et Davidovitch à partir de 500 000 mots (1970). Le lexicologue Brunet, en 2002, a retenu 1500 mots selon leur fréquence pour établir les programmes scolaires (Havránková, 2015). Notons aussi Manulex (Lété, Sprenger-Charolles et Colé, 2004) dont la base de données (BDD) comprend plus de 48 000 mots fréquents issus de 54 manuels scolaires. Ou encore, la BDD de lexique.org qui fournit la fréquence de 135 000 mots du français en se basant sur un corpus de livres du 20<sup>e</sup> siècle et de sous-titres de films du 21<sup>e</sup> siècle (New, Pallier, Ferrand et Matos, 2001).

Nous retrouvons la fréquence d'usage des mots comme principe de constitution de listes de mots à différentes échelles. En L1, la fréquence des mots a été utilisée au Québec pour constituer des listes orthographiques correspondant à chaque année du primaire (Daigle *et al.*, 2015). Ces listes de mots fréquents sont destinées à être complétées en fonction des thèmes abordés en classe (Daigle *et al.*, 2015). La fréquence est également utilisée dans les listes de l'examen universitaire du français LE en Suède appelé *Riksprovet*. Ces listes peuvent être complétées par des mots jugés utiles par les apprenants (Lindqvist et Ramnäs, 2016).

### Lacunes des ouvrages répertoriant du vocabulaire

Quel que soit le critère choisi pour la sélection du vocabulaire, le CECRL (2001) suggère de le mentionner. Ce n'est pourtant pas le cas dans de nombreux manuels comme l'a relevé Paccaud (2011). Cette auteure remarque également que le contenu des ouvrages qu'elle a analysé ne correspond pas aux publics ciblés lorsqu'il est défini : des ouvrages supposés tout public sont par exemple réservés aux étudiants. Dans la majorité des cas, le choix du vocabulaire des ouvrages est subjectif. Par exemple, parmi les sept manuels de référence qu'elle a passés en revue, Havránková (2015) relevait que la sélection du vocabulaire était purement subjective pour cinq d'entre eux tandis qu'un sixième se basait sur les occurrences des mots à l'intérieur

même du manuel. Seul le septième (Galison, 1971) était une réorganisation thématique du vocabulaire du *Français fondamental* de Gougenheim.

L'utilisation de la fréquence dans la constitution de listes de mots est parfois partielle, comme pour les listes orthographiques du primaire au Québec ou le *Riksprovet*. Notons également qu'aucune indication de difficulté n'est mentionnée pour les différents thèmes des OVT. Pourtant, la rareté de certains thèmes dans les OVT laisse penser qu'ils sont trop spécifiques et leurs mots plus difficiles à retenir du fait d'une moindre exposition à eux. Galison (1971) s'est appuyé sur la liste de mots fréquents de Gougenheim, établie à des fins d'enseignement du français en L2, pour établir une catégorisation par thèmes, mais cet ouvrage date de près de 50 ans. Rien n'indique donc que la sélection du vocabulaire dans les OVT actuels soit basée sur la fréquence.

## Objectif

Nous considérons que la fréquence est un critère crucial dans le choix des mots pour la constitution d'OVT. De plus, conformément à ce qui a été relevé précédemment, les mots fréquents ont tendance à être appris prioritairement par les débutants, la connaissance de mots rares reflétant alors un niveau plus avancé. Dans le présent travail, nous proposons d'analyser un OVT afin de vérifier si la fréquence ou la rareté des mots qui le composent sont conformes à la difficulté annoncée de l'ouvrage. Par conséquent, notre objectif est le suivant : vérifier la difficulté des listes thématiques d'un ouvrage bilingue à l'aide de la fréquence des mots qui les composent.

L'analyse d'un OVT permettra de proposer des pistes pour la réalisation d'ouvrages de ce type d'après des critères scientifiques afin de favoriser l'apprentissage du vocabulaire du français. Cette analyse conduira aussi à un nouvel usage de la fréquence dans l'étude du vocabulaire d'une langue en l'appliquant à des groupes de mots (des listes thématiques) plutôt qu'à des mots isolés.

## Méthodologie

Plusieurs critères ont été pris en compte pour le choix d'un OVT. Ces critères portent principalement sur le contenu. Toutefois, le premier critère concerne la date de publication. En effet, le vocabulaire évolue continuellement (des mots apparaissent, d'autres disparaissent). L'ouvrage doit donc être aussi récent que possible.

Un trop petit nombre de thèmes peut conduire à une quantité trop importante de mots par liste. Une division de ces thèmes en sous-thèmes peut se révéler utile au cas où ces listes thématiques seraient disparates en termes de nombre de mots. Il serait alors possible de prendre en considération des listes sous-thématiques d'un thème si leurs longueurs s'apparentaient à celles des autres listes thématiques.

Concernant le niveau de difficulté, les listes thématiques pouvant être utiles pour atteindre une masse critique de mots assez tôt dans l'apprentissage d'une langue (Cobb et Horst, 2011), nous avons ciblé les niveaux débutants et intermédiaires, entre A1 et B2 selon le CECRL (2001), pour le choix de l'ouvrage.

Les classes grammaticales devaient être mentionnées dans l'ouvrage pour plusieurs raisons. Leurs proportions diffèrent lorsque l'on segmente le vocabulaire en plusieurs zones de fréquence, par tranche de 500 mots par exemple (Lortet, à paraître). Elles limitent les confusions liées à la polysémie ou l'homonymie. Elles n'offrent pas toute la même facilité de classement : les verbes, notamment, sont difficiles à classer du fait de leur abstraction puisqu'ils peuvent désigner des mouvements, des gestes, des attitudes ou des activités (Ruchot, 2011). C'est parce qu'ils représentent la majorité des mots du français (Lortet, à paraître) et qu'ils sont plus faciles à classer que les noms ont été sélectionnés pour le présent travail.

Au final, nous avons choisi un OVT bilingue de l'éditeur Ophrys (Gosset, 2016), réputé pour ses livres scolaires et ses livres de langues. Il répondait aux différentes exigences énumérées plus haut : il est récent (2016), il est divisé en thèmes et en sous-thèmes, il est basé sur les niveaux du CECRL (B1-B2, soit intermédiaire) et il segmente les listes de mots par classes grammaticales. Autre avantage, le vocabulaire contenu dans cet OVT correspond au français de France, comme celui de la BDD de [lexique.org](http://lexique.org), ce qui a permis d'associer par paires les mots des deux sources en évitant le biais dû aux différences lexicales pouvant exister entre deux variantes dialectales. Dans le cas contraire, certains mots auraient pu être présents dans l'une des sources, mais pas dans l'autre.

Nous avons procédé en trois étapes pour évaluer la difficulté des listes. En premier lieu, nous avons récupéré la fréquence de chaque mot de l'OVT en les tapant dans un fréquencier en ligne, programmé à partir de la BDD de [lexique.org](http://lexique.org) et qui contient les mots de cette BDD ainsi que leur fréquence (Lortet, 2017).

Nous avons ensuite calculé la moyenne des fréquences des mots pour chaque liste. Cette moyenne est calculée par l'addition de l'ensemble des mots d'une liste thématique divisé par le nombre de mots de cette liste. La moyenne des fréquences par liste donne une idée de la difficulté de chaque liste par l'indication de la plus ou moins grande rareté des mots qui la composent.

Toujours à partir des fréquences des mots des listes thématiques, nous avons calculé l'écart-type. Celui-ci nous renvoie comme résultat le degré de dispersion des fréquences par rapport à la fréquence moyenne. Plus les fréquences sont éparpillées, plus l'écart-type est important.

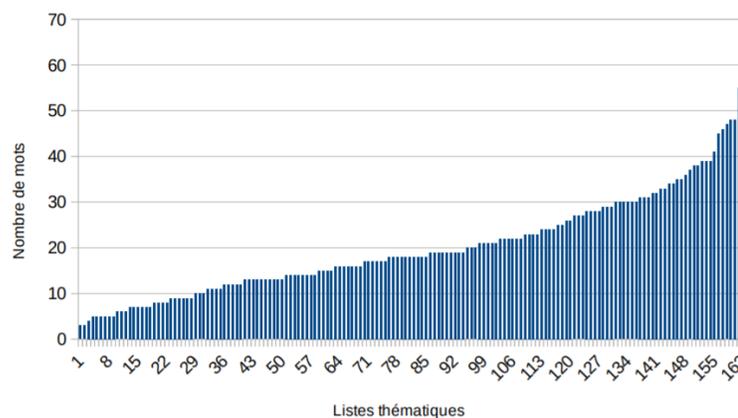
## Résultats et discussion

Comme mentionné dans l'objectif, nous sommes partis de la fréquence des mots de l'OVT afin d'obtenir la fréquence de leurs listes. Pour ce faire, nous avons recopié à l'ordinateur 163 listes thématiques de noms communs de l'ouvrage d'Ophrys (Gosset, 2016). Nous avons entré chaque mot dans le fréquencier qui les a cherchés dans la BDD de [lexique.org](http://lexique.org) afin de mesurer leur fréquence. Plus la valeur associée à la fréquence est basse, plus le mot est fréquent, puisque les mots sont classés selon une échelle ordinale. Les moyennes et les écart-types ont ensuite été calculés à partir de ces valeurs. Dans la présente section, nous exposons sous forme de graphiques les résultats des calculs effectués. Les résultats portent à la fois sur une seule dimension (longueur des listes, moyenne et écart-type de la fréquence des mots des listes thématiques) et sur deux dimensions (comparaison deux à deux des longueur, moyenne et écart-type). Dans chacune des figures de cette section, l'axe des abscisses indique les listes thématiques (163 au total) et l'axe des ordonnées indique, selon le cas, le nombre de mots, la moyenne ou l'écart-type des fréquences des mots de chaque liste.

### Dimension unique

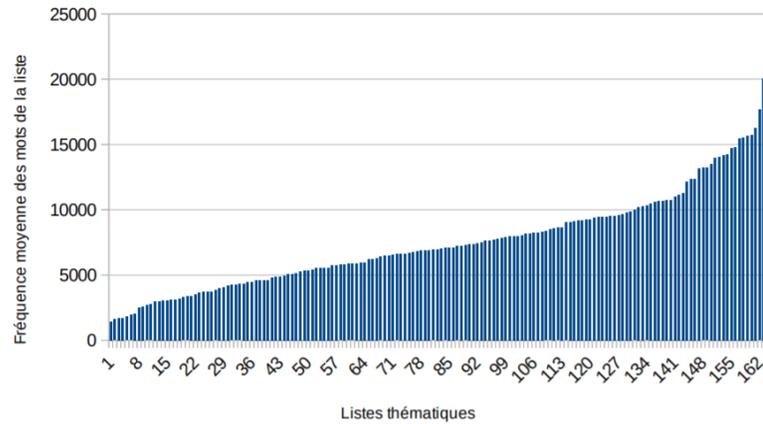
La première figure ci-dessous représente le nombre de mots par liste (la longueur des listes). Les listes de la figure ne sont pas classées selon leur ordre d'apparition dans l'OVT, mais selon leur nombre de mots. Comme nous pouvons le constater, il n'y a pas d'uniformité au niveau de la longueur des listes de l'OVT. Certaines listes n'en contiennent que 3, d'autres 62, en passant par pratiquement tous les nombres intermédiaires comme le montre la figure, même si les listes de 40 à 62 mots sont peu nombreuses (au nombre de 9).

Figure 1. Nombre de mots par liste



Après ordonnancement croissant des listes selon la fréquence moyenne des mots de chaque liste, nous obtenons le graphique suivant. Celui-ci nous montre une diversité totale des listes thématiques. En effet, les moyennes (en ordonnée) peuvent varier grandement d'une liste à l'autre.

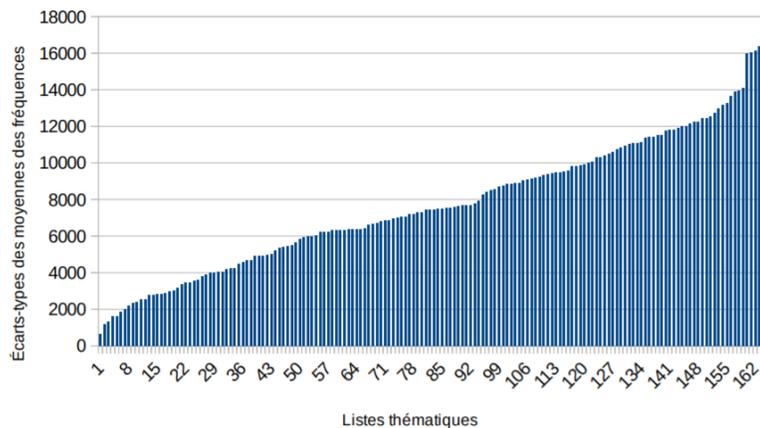
Figure 2. Tri croissant des moyennes des fréquences des mots des listes thématiques



Précisons que la difficulté de ces listes thématiques selon leur moyenne n'est pas reflétée dans leur ordre d'apparition dans l'ouvrage. Les premières listes présentées ne sont pas forcément les plus faciles au niveau de la rareté de leur vocabulaire, bien que les thèmes semblent les plus récurrents (le *corps humain*, présenté en premier dans l'ouvrage, est en 16<sup>e</sup> position pour la fréquence moyenne de ses mots).

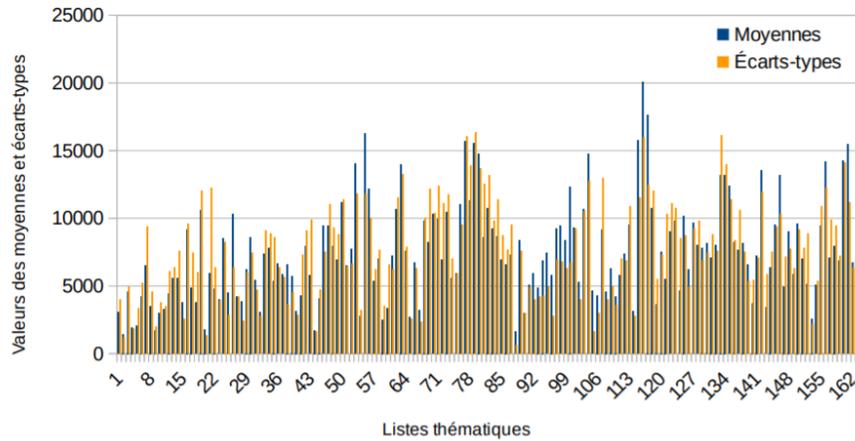
Après ordonnancement croissant des écarts-types de fréquences calculés pour chaque liste, nous observons également sur la figure ci-dessous un échelonnement croissant. La composition des listes thématiques est très diversifiée : les listes thématiques de l'ouvrage n'ont donc pas été constituées en tenant compte de la difficulté des mots. Certaines listes sont composées de mots aux fréquences assez semblables (la liste des *saisons* a un écart-type d'environ 633), d'autres ont des mots dont les fréquences peuvent être très disparates (la liste *Internet* a un écart-type d'environ 16 358).

Figure 3. Écarts-types des listes thématiques



L'ordre de présentation des listes thématiques dans l'OVT ne semble relié à aucun critère statistique. Comme le montre la figure ci-dessous où les traits verticaux représentent les listes dans leur ordre d'apparition dans l'OVT, la succession des moyennes et des écarts-types n'est ni croissante, ni décroissante. Il n'y a donc pas d'ordonnement de la difficulté des listes.

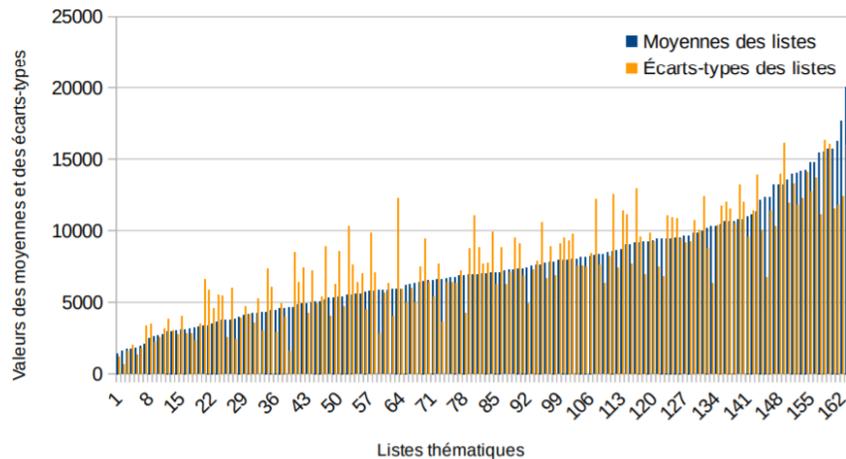
Figure 4. Ordre d'apparition des listes thématiques dans l'ouvrage



### Comparaisons

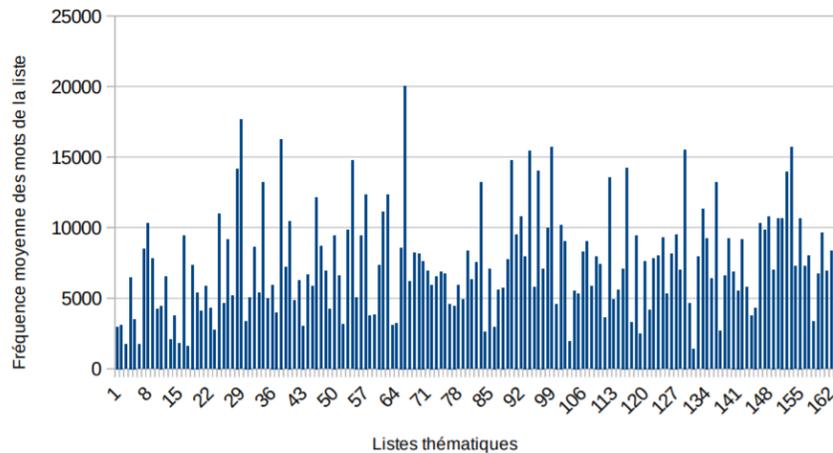
Nous pourrions penser que plus la moyenne d'une liste est élevée (donc composée de mots rares), plus l'écart-type est grand et regroupe donc des mots fréquents et rares. Or, il n'en est rien, comme l'indique la figure ci-dessous. Il ne semble pas y avoir de corrélation entre la moyenne et l'écart-type, signifiant qu'on peut rencontrer des mots plus ou moins rares dans les listes thématiques, que la moyenne soit élevée ou non.

Figure 5. Moyennes des listes thématiques rapportées aux écarts-types



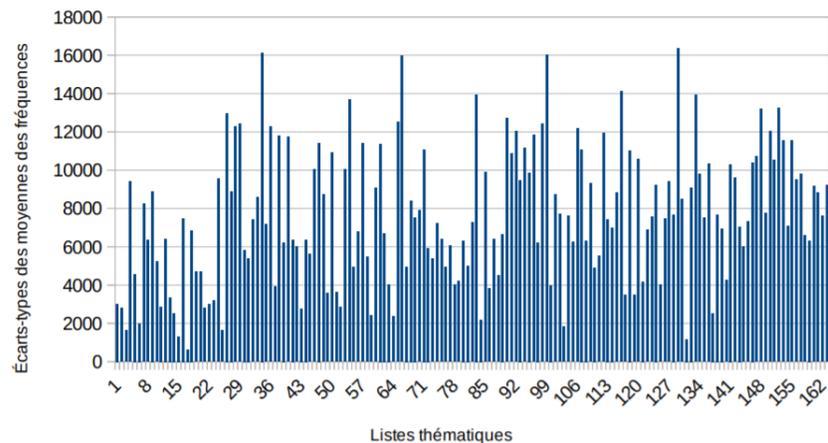
Il en est de même pour le rapport entre les moyennes et la taille des listes. Lorsque nous ordonnons les listes selon leur longueur comme dans la figure ci-dessous (de la plus petite, à gauche, à la plus grande, à droite), nous observons que les moyennes de ces listes sont très variables. Il n'y a donc pas de lien entre le nombre de mots et la rareté des mots : une liste peut très bien comporter de nombreux mots et avoir une moyenne des fréquences des mots plus faible que celle d'une liste comportant moins de mots.

Figure 6. Moyennes classées selon la taille des listes thématiques



De même, il n'apparaît aucune corrélation entre les écarts-types et la longueur des listes thématiques. Comme le montre la figure ci-dessous, où les listes sont classées du plus petit au plus grand nombre de mots, la fréquence entre les mots d'une liste peut être aussi bien réduite qu'importante, et ce sans aucun lien avec le nombre de mots.

Figure 7. Écarts-types des listes rapportées aux tailles des listes thématiques



## Conclusion

Nous avons analysé un OVT à l'aide de la fréquence d'usage des mots afin d'y déceler un choix du vocabulaire basé sur la fréquence des mots. Nous avons calculé, pour chaque liste : le nombre de mots, la moyenne des fréquences et l'écart-type des fréquences. Les résultats n'ont pas révélé d'indices du principe de fréquence d'usage dans le choix du vocabulaire et dans la constitution des listes thématiques. Nous pouvons en déduire que le choix du vocabulaire dans cet OVT est soit subjectif, soit lié à des critères inconnus, comme dans le cas des manuels analysés par Havránková (2015). Cela peut sembler être un recul par rapport à ce qu'avait initié Galisson (1971) avec sa catégorisation des mots du *Français fondamental*. Pour rappel, l'ouvrage analysé (Gosset, 2016) est censé être destiné aux apprenants de niveaux B1-B2 selon le CECRL. La diversité des fréquences des mots et des fréquences des listes thématiques va à l'encontre d'une telle classification et posture plutôt pour un ouvrage ne ciblant aucun public particulier, contenant des mots trop rares pour les apprenants débutants et trop évidents pour les apprenants avancés, un public de

niveau intermédiaire risquant de cumuler ces deux inconvénients. Relativisons en mentionnant les limites inhérentes à ce type d'analyse.

## Limites

L'usage de la fréquence des mots dans la constitution de listes peut conduire à exposer l'apprenant débutant à des mots présentant des difficultés notamment orthographiques. Cependant, les mots difficiles sont parfois indispensables dans l'usage de la langue. Les verbes irréguliers du 3<sup>e</sup> groupe du français sont par exemple surreprésentés parmi les mots très fréquents (Lortet, à paraître). Concernant les listes orthographiques établies au Québec pour le primaire, l'adverbe *beaucoup* est présenté en 1<sup>re</sup> année malgré la complexité de son écriture non phonétique, du fait qu'il est difficilement contournable. Le degré de difficulté des mots ne peut donc prévaloir sur la fréquence comme critère de sélection du vocabulaire.

Nous avons choisi comme ressource la BDD de [lexique.org](http://lexique.org) pour la quantité de mots qu'elle contient (135 000) et le fait qu'elle soit maintenue à jour (la dernière modification, du 14/12/2017, corrigeait 321 entrées). Cependant, les résultats seraient sans doute différents avec une autre BDD, même si une certaine similarité peut exister entre elles : rappelons par exemple que parmi les 400 mots fréquents de la liste de Baker, 93,14 % font également partie de celle du français fondamental de Gougenheim (Huet, 2002).

L'ouvrage choisi est bilingue, et seul le vocabulaire français a été analysé. Le risque est donc que la fréquence des mots anglais soit si différente de celle des mots français que nous n'aurions pas vu une objectivité basée sur la fréquence dans le choix des mots. Cependant, un ouvrage bilingue peut très bien être consulté par les apprenants des deux langues. Le problème de la fréquence d'usage du vocabulaire du français resterait alors posé pour un anglophone apprenant le français.

Certaines listes étaient extrêmement petites (trois mots) quand d'autres pouvaient atteindre 62 mots. Afin de rester fidèles au contenu de l'ouvrage, nous les avons gardées telles quelles. De plus, les deux solutions auxquelles nous avons pensé présentaient chacune un inconvénient. Nous aurions pu, par exemple, appliquer l'écart-type à la longueur des listes pour ne conserver que les deux tiers des listes, et donc éliminer les plus petites et les plus grandes. En réalisant ce calcul, nous avons obtenu une moyenne de 20 mots par liste et un écart-type de 11,44. Nous aurions alors pu ne garder que les listes contenant au minimum 8 mots (moyenne - écart-type) et au maximum 32 mots (moyenne + écart-type). Toutefois, nous n'aurions pas remarqué dans ces listes l'absence de corrélation entre leur taille d'une part et la moyenne et l'écart-type d'autre part.

Autre solution, nous aurions pu considérer les thèmes plutôt que les sous-thèmes lorsque les listes étaient trop courtes. En regroupant les listes portant sur les cinq sens, nous aurions obtenu une liste de 28 mots. À l'inverse, nous aurions pu segmenter les listes trop longues en différents thèmes, par exemple ne garder que les liens de parenté pour le thème de la *famille*. Dans tous les cas, la longueur des listes aurait été plus homogène, mais nous n'aurions pu conclure de l'absence de prise en compte de la fréquence des mots lors de la constitution des listes thématiques par l'auteur de l'ouvrage.

## Suggestions

À partir des résultats et des limites de l'analyse effectuée sur l'ouvrage d'Ophrys (Gosset, 2016), nous pouvons émettre plusieurs propositions pour la réalisation d'un OVT basé sur la fréquence d'usage des mots. L'ordonnement des listes thématiques pourrait se faire par difficulté croissante. L'apprenant pourrait ainsi suivre une progression dans la découverte de nouveaux mots. L'ordonnement selon la fréquence d'usage pourrait concerner également les mots des listes thématiques.

Nous avons proposé précédemment de définir des limites de longueur pour les listes en fonction de l'écart-type additionné ou soustrait de la moyenne. Les listes thématiques trop longues pourraient être segmentées d'au moins deux manières : selon la fréquence et selon les sous-thèmes. La liste thématique en tant que telle pourrait contenir les mots les plus fréquents tandis que le vocabulaire plus spécifique serait réparti en différentes listes sous-thématiques. Ces retouches pourraient constituer une étape lors de la réalisation de listes thématiques d'un OVT. Cette utilisation partielle de la fréquence se rapproche des listes orthographiques du Québec incitant les enseignants à ajouter du vocabulaire spécifique (Daigle *et al.*, 2015) ou des listes pour l'examen universitaire de français en Suède dont les deux tiers de mots fréquents sont

complétés d'un tiers de mots jugés utiles (Lindqvist et Ramnäs, 2016). Le vocabulaire supplémentaire serait ici suggéré par l'OVT.

Enfin, la fréquence d'usage des mots recensés reflète la difficulté de l'ouvrage. L'OVT analysé vise un niveau B1-B2. Constituer des listes conformes à un niveau précis est d'autant plus difficile que le CECRL ne propose pas de liste de mots, ce qui était cependant le cas dans sa version originale (Milton, 2006). Une solution consisterait à définir le nombre de mots à connaître pour chaque niveau et à utiliser ce nombre pour fixer une fréquence limite pour le choix des mots. Ainsi, le CECRL fixait à 2000 le nombre de mots à connaître en français au niveau B1 (Milton, 2006). Cependant, certains mots d'usage courant peuvent être situés des deux côtés de cette barrière. Par exemple, le mot *dimanche* est à la 866<sup>e</sup> place tandis que le mot *mercredi* est à la 4679<sup>e</sup> place selon [lexique.org](http://lexique.org) (Lortet, 2017). Pour pallier cela, nous proposons de définir les niveaux de l'apprenant non pas en fonction du nombre de mots à connaître, mais en fonction des fréquences moyennes des listes thématiques. La réalisation d'une telle échelle pourrait faire l'objet d'une recherche future.

## RÉFÉRENCES

- Asmi, J. (2013). *La didactique des langues et la notion du vocabulaire fondamental de Gougenheim*. Communication présentée au 37<sup>e</sup> Annual Congress of the American Romanian Academy, University of European Political and Economic Studies "Constantin Stere", Chisinau, Moldavie. Récupéré le 21 septembre 2018 de : <http://pubsys.ara-as.org/index.php/proc/article/download/1187/1059>
- Benigno, V., Grossmann, F. et Kraif, O. (2015). Les collocations fondamentales : une piste pour l'apprentissage lexical. *Revue française de linguistique appliquée* (1), 81-96. Récupéré le 21 septembre 2018 de : <https://www.cairn.info/revue-francaise-de-linguistique-appliquee-2015-1-page-81.htm>
- Cobb, T. et Horst, M. (2011). Does “word coach” coach words? *CALICO Journal*, 28(3), 639-661. Conseil de l'Europe (2001). *Cadre européen commun de référence pour les langues : apprendre, enseigner, évaluer*. Paris : Hatier. Récupéré de : <https://rm.coe.int/16802fc3a8>
- Cuq, J.-P. et Gruca, I. (2005). *Cours de didactique du français langue étrangère et seconde*. Grenoble : Presses universitaires de Grenoble.
- Daigle, D., Anctil, D., Berthiaume, R. et Bruneau, M. J. (2015). La liste orthographique: des mots à mémoriser ou des cibles lexicales à apprendre? *Québec français* (175), 36-37. Récupéré le 21 septembre 2018 de : <https://www.erudit.org/en/journals/qf/2015-n175-qf02472/81381ac.pdf>
- Galisson, R. (1971). *Inventaire thématique et syntagmatique du français fondamental*. Paris : Hachette-Larousse.
- Gosset, C. (2016). *Petit vocabulaire actuel de l'anglais*. Paris : Éditions Ophrys.
- Grossmann, F. (2011). Didactique du lexique: état des lieux et nouvelles orientations. *Pratiques. Linguistique, littérature, didactique*, (149-150), 163-183. Récupéré le 21 septembre 2018 de : <https://journals.openedition.org/pratiques/1732>
- Havránková, A. (2015). *Enseignement du vocabulaire* (Thèse de doctorat). Univerzita Karlova, Prague. Récupéré le 21 septembre 2018 de : [https://dspace.cuni.cz/bitstream/handle/20.500.11956/66732/DPTX\\_2012\\_2\\_11410\\_0\\_364202\\_0\\_136192.pdf?sequence=1](https://dspace.cuni.cz/bitstream/handle/20.500.11956/66732/DPTX_2012_2_11410_0_364202_0_136192.pdf?sequence=1)
- Henmon, V. A. C. (1924). *A French word book based on a count of 400,000 running words*. Madison, Wisconsin : University of Wisconsin.
- Huet, S. (2002). Les enseignements d'une dictée [html]. Récupéré le 21 septembre 2018 de : <https://www.samuelhuet.com/paid/44-polemos/218-enseignements-dictee5.html>
- Juilland, A. G., Brodin, D. R. et Davidovitch, C. (1970). *Frequency dictionary of French words*. La Haye : Mouton.
- L'Épée, C. M. de (1776). *Institution des sourds et muets par la voie des signes méthodiques*. Paris : Nyon l'aîné. Récupéré le 21 septembre 2018 de : [https://gallica.bnf.fr/ark:/12148/bpt6k1040582v\\_image](https://gallica.bnf.fr/ark:/12148/bpt6k1040582v_image)
- Lété, B., Sprenger-Charolles, L. et Colé, P. (2004). MANULEX: A grade-level lexical database from French elementary school readers. *Behavior Research Methods, Instruments, & Computers*, 36(1), 156-166. DOI : 10.3758/BF03195560
- Lindqvist, C. et Ramnäs, M. (2016). L'enseignement du vocabulaire à l'université. *Synergies Pays Scandinaves* (11/12), 55-64. Récupéré le 21 septembre 2018 de : [https://www.gerflint.fr/Base/Paysscandinaves11\\_12/lindqvist\\_ramnas%20.pdf](https://www.gerflint.fr/Base/Paysscandinaves11_12/lindqvist_ramnas%20.pdf)
- Lortet, A. (2017). *Fréquentier* [Application en ligne]. Récupéré le 21 septembre 2018 de : <http://frequencier.galexie.com>
- Lortet, A. (à paraître). Approche par fréquence pour l'enseignement du vocabulaire du français aux allophones. *Formation et profession*.
- Milton, J. (2006, juillet). French as a foreign language and the Common European Framework of Reference for Languages. In *Proceedings from the Crossing Frontiers: Languages and the International Dimension Conference* (pp. 1-6). Récupéré le 21 septembre 2018 de : [https://www.researchgate.net/profile/James\\_Milton4/publication/237584489\\_French\\_as\\_a\\_foreign\\_language\\_and\\_the\\_Common\\_European\\_Framework\\_of\\_Reference\\_for\\_Languages/links/0c96053bbb69a1ef56000000.pdf](https://www.researchgate.net/profile/James_Milton4/publication/237584489_French_as_a_foreign_language_and_the_Common_European_Framework_of_Reference_for_Languages/links/0c96053bbb69a1ef56000000.pdf)
- Ministère de l'Éducation nationale (1954). *Le Français élémentaire*. Paris : Centre national de documentation pédagogique.

- New B., Pallier C., Ferrand L. et Matos R. (2001). Une base de données lexicales du français contemporain sur Internet: LEXIQUE. *L'Année psychologique*, 101, 447-462. Récupéré le 21 septembre 2018 de :  
<http://www.lexique.org>
- Pignot-Shahov, V. (2014). Le développement lexical des apprenants de français langue étrangère d'une université britannique: une étude pilote. *Synergies Royaume-Uni et Irlande* (7). Récupéré de :  
[http://www.gerflint.fr/Base/RU-Irlande7/pignot\\_shahov.pdf](http://www.gerflint.fr/Base/RU-Irlande7/pignot_shahov.pdf)
- Richards, B. et Malvern, D. (2007). Validity and threats to the validity of vocabulary measurement. Dans Helmut Daller, James Milton & Jeanine Treffers-Daller (Eds.), *Modelling and assessing vocabulary knowledge* (pp. 79-92). Cambridge, UK : Cambridge University Press.
- Ruchot, T. (2011). Les vocabulaires thématiques français-russe et russe-français: panorama et réflexions sur un genre. *La Revue russe*, 37(1), 131-141. DOI : 10.3406/russe.2011.2471.
- Variance et écart-type (2013). Statistique Canada. Récupéré le 21 septembre 2018 de :  
<https://www.statcan.gc.ca/edu/power-pouvoir/ch12/5214891-fra.htm>