



Approche par fréquence et par classes grammaticales pour l'enseignement du vocabulaire du français aux allophones

Alain Lortet 

Université du Québec à Montréal (Canada)

doi:10.18162/fp.2020.a201

CHRONIQUE • Recherche étudiante

Introduction

L'enseignement du vocabulaire du français aux allophones implique la sélection des mots à enseigner. Ces mots peuvent se présenter sous différentes formes (au pluriel ou conjugués par exemple) et faire l'objet de regroupements. La forme de base, non fléchie, du mot (masculin singulier pour les noms et adjectifs, infinitif pour les verbes) est la *lexie*. Ces lexies sont regroupables en familles de mots qui incluent les mots dérivés (comme *nourrir* et *nourriture*) ayant des classes grammaticales différentes (comme les verbes et les noms). Les mots à connaître pour atteindre les niveaux A1 à C2 définis par le Cadre européen commun de référence pour les langues (CECRL), outil de référence pour l'enseignement des langues étrangères, se comptent en centaines ou en milliers selon les niveaux.

L'enseignement du vocabulaire privilégiant les mots courants de la langue seconde ou étrangère peut avantager les apprenants. La compréhension d'un texte, par exemple, ne nécessite pas la connaissance de l'ensemble des mots qui le composent, mais des familles de mots les plus fréquentes (van Zeeland et Schmitt, 2012). L'approche par fréquence consiste à ordonner et à regrouper les mots à enseigner selon leur rareté et à considérer ou non les classes grammaticales selon qu'il s'agit de lexies ou de familles de mots.

Cependant, Meara relevait déjà en 2010 que diviser des mots par zones de fréquence (c'est-à-dire les 1000 mots les plus fréquents, puis les 1000 suivants, etc.) est pratique, mais ne prend pas en compte la difficulté des mots. Des verbes réguliers et irréguliers peuvent donc

se retrouver dans une même zone de fréquence. Les classes grammaticales, qui distinguent lexies et familles de mots comme indiqué plus haut, ne sont pas plus prises en considération. Cette omission aboutit à des familles de mots éparpillées sur plusieurs zones de fréquences, et donc à des regroupements de mots rares et courants ne favorisant pas la présentation des mots les plus fréquents aux apprenants allophones, ce qui faciliterait pourtant leur compréhension du français. Le choix des mots basé sur leur seule fréquence dans la langue ne tient donc pas compte des caractéristiques de ces mots.

Notre recherche : objectif et méthodologie

Notre recherche vise à affiner le choix du vocabulaire du français à enseigner en fonction du niveau des apprenants allophones en ne nous limitant pas à la seule fréquence du mot. Nous adoptons ici une approche à la fois par fréquence et par classes grammaticales en tenant compte des spécificités de ces dernières. Pour ce faire, nous avons établi des statistiques sur les classes grammaticales des mots par zones de fréquence. En considérant le nombre de mots à connaître selon le niveau des apprenants, nous suggérons ensuite quelles spécificités des classes grammaticales enseigner aux allophones selon leur niveau.

Nous avons créé une base de données à partir du corpus de mots du français de lexique.org (constitué de romans et de sous-titres de films récents). Nous avons analysé les proportions de chaque classe grammaticale ouverte (noms, verbes, adjectifs, adverbes) selon les zones de fréquence et des caractéristiques internes (les flexions et les régularités). Nous nous sommes limités aux 10 000 lexies et aux 30 000 mots fléchis les plus fréquents, les mots très rares n'étant pas en nombre suffisant pour donner du sens aux statistiques.

Nous avons choisi des multiples de 500 lexies pour délimiter les niveaux du CECRL afin de nous conformer aux données de la littérature. Nous considérons donc que le niveau A1 correspond à la connaissance de 500 lexies, le niveau A2 à 1000 lexies, le niveau B1 à 2000 lexies, le niveau B2 à 4000 lexies et les niveaux avancés à 8000 lexies et plus. Les niveaux A1 et A2 étant de 500 lexies seulement chacun, nous limitons la taille des zones de fréquence du vocabulaire à 500 lexies.

Spécificités des zones de fréquence selon les classes grammaticales

Les quatre classes grammaticales analysées sont plus proches les unes des autres en termes de quantité parmi les lexies les plus fréquentes (écart-type de 42,13 pour ces quatre classes dans les 500 mots les plus fréquents et de 71,91 pour les 500 suivants). Les noms augmentent ensuite nettement en quantité : ils forment 61 % du total des 47 342 lexies, contre 23 % pour les adjectifs, 11 % pour les verbes et 4 % pour les adverbes.

Les verbes sont plus nombreux que les adjectifs parmi les 3500 lexies les plus fréquentes. La proportion de verbes et d'adjectifs s'équilibre entre 3500 et 5000 lexies. Après ces 5000 lexies, les adjectifs prennent définitivement l'avantage.

Les adjectifs au pluriel sont peu nombreux parmi les mots fléchis fréquents. La proportion de noms au pluriel augmente dans les 2 premières zones de fréquence avant de se stabiliser.

Parmi les 5292 verbes recensés, 88 % (4639 verbes) sont du 1^{er} groupe. Parmi les 500 lexies les plus fréquentes, la quantité de verbes des 2^e et 3^e groupes (59 verbes) est proche de celle des verbes du 1^{er} groupe (62 verbes). La proportion des verbes pronominaux est plus forte entre les 500 et 1500 lexies les plus fréquentes (jusqu'à 61 % et 64 % des verbes).

La plupart des 1822 adverbes recensés (70 %) sont formés avec le suffixe « -ment ». Cependant, parmi les 1000 lexies les plus fréquentes, la grande majorité des adverbes sont irréguliers (77 sur 87).

À partir de ces résultats, nous suggérons des thèmes à aborder selon le niveau de l'apprenant. Le tableau ci-dessous récapitule ces suggestions.

Tableau 1

Suggestions d'enseignement des caractéristiques des classes grammaticales selon les niveaux du CECRL.

	Débutants		Intermédiaires		Avancés	
	A1	A2	B1	B2	C1	C2
Classes grammaticales	Toutes les classes	Priorité aux noms				
Noms	Noms singuliers	Noms au singulier et au pluriel				
Verbes	Verbes irréguliers	Verbes du 1 ^{er} groupe				
		Verbes pronominaux				
Adjectifs	Adjectifs singuliers	Adjectifs au singulier et au pluriel				
Adverbes	Adverbes irréguliers		Adverbes en « -ment »			
Verbes c. adjectifs	Priorité aux verbes			Équilibrer verbes et adjectifs		Priorité aux adjectifs

Signalons toutefois certaines limites à cette recherche. La fixation du nombre de mots par niveau est théorique et ne reflète pas forcément la réalité. De plus, la fréquence des mots évolue dans le temps (Meara, 2010), nécessitant un corpus le plus à jour possible.

Références

- Conseil de l'Europe. (2001). *Cadre européen commun de référence pour les langues : apprendre, enseigner, évaluer*. Repéré à <https://rm.coe.int/16802fc3a8>
- Meara, P. (2010). *EFL vocabulary tests* (2^e éd.). Swansea : ERIC Clearinghouse.
- 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(3-4), 447-462. <http://dx.doi.org/10.3406/psy.2001.1341>
- van Zeeland, H. et Schmitt, N. (2012). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics*, 34(4), 457-479. <http://dx.doi.org/10.1093/applin/ams074>

Pour citer cet article

Lortet, A. (2020). Approche par fréquence et par classes grammaticales pour l'enseignement du vocabulaire du français aux allophones. *Formation et profession*, 28(2), 110-112. <http://dx.doi.org/10.18162/fp.2020.a201>