



**HAL**  
open science

# Mots 'faciles' et mots 'difficiles' dans ReSyf : un outil pour la didactique du lexique mobilisant polysémie, synonymie et complexité

Núria Gala, Ludivine Javourey-Drevet

## ► To cite this version:

Núria Gala, Ludivine Javourey-Drevet. Mots 'faciles' et mots 'difficiles' dans ReSyf : un outil pour la didactique du lexique mobilisant polysémie, synonymie et complexité. LIDIL - Revue de linguistique et de didactique des langues, 2020, 62, <10.4000/lidil.8373>. <hal-03194331>

**HAL Id: hal-03194331**

**<https://hal.science/hal-03194331v1>**

Submitted on 9 Apr 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization



**Lidil**

Revue de linguistique et de didactique des langues

62 | 2020

**Recherches actuelles en didactique du lexique :  
avancées, réflexions, méthodes**

---

## Mots « faciles » et mots « difficiles » dans *ReSyf* : un outil pour la didactique du lexique mobilisant polysémie, synonymie et complexité

*“Easy” and “Difficult” Words in ReSyf: A Tool to Teach Vocabulary through the  
Notions of Polysemy, Synonymy and Complexity*

**Núria Gala et Ludivine Javourey-Drevet**

---



### Édition électronique

URL : <http://journals.openedition.org/lidil/8373>

DOI : 10.4000/lidil.8373

ISSN : 1960-6052

### Éditeur

UGA Éditions/Université Grenoble Alpes

### Édition imprimée

ISBN : 978-2-37747-226-0

ISSN : 1146-6480

### Référence électronique

Núria Gala et Ludivine Javourey-Drevet, « Mots « faciles » et mots « difficiles » dans *ReSyf* : un outil pour la didactique du lexique mobilisant polysémie, synonymie et complexité », *Lidil* [En ligne], 62 | 2020, mis en ligne le 03 novembre 2020, consulté le 05 novembre 2020. URL : <http://journals.openedition.org/lidil/8373> ; DOI : <https://doi.org/10.4000/lidil.8373>

---

Ce document a été généré automatiquement le 5 novembre 2020.

© Lidil

---

# Mots « faciles » et mots « difficiles » dans *ReSyf* : un outil pour la didactique du lexique mobilisant polysémie, synonymie et complexité

*“Easy” and “Difficult” Words in ReSyf: A Tool to Teach Vocabulary through the Notions of Polysemy, Synonymy and Complexity*

Núria Gala et Ludivine Javourey-Drevet

---

## **Remerciements**

La ressource *ReSyf* a été financée par l'Agence nationale de la recherche par le biais du projet Alector (ANR-16-CE28-0005) et par l'Agence nationale de la recherche belge (FNRS). Nous remercions vivement Mokhtar Boumedyen Billami (Aix Marseille Université) et Thomas François (Université catholique de Louvain) pour le travail de développement de la ressource. La version actuelle du site de consultation de *ReSyf* a été financée par ORTOLANG, nous remercions Dorian Ricci et Brayon Delmée (Université catholique de Louvain) pour leurs contributions.

## **1. Introduction**

- 1 Cet article envisage et discute diverses utilisations de la ressource *ReSyf*<sup>1</sup>, une base lexicale en ligne qui propose des synonymes désambiguïsés et gradués en fonction de leur difficulté en décodage et en compréhension de lecture en français (Billami, François & Gala 2018). La ressource a été construite semi-automatiquement à partir d'une liste initiale de synonymes issue de *JeuxDeMots* (Lafourcade, 2007). D'une part, la liste a été traitée afin de désambiguïser les sens des entrées et de filtrer les synonymes. D'autre part, une méthode de classement automatique a permis de graduer les synonymes d'une entrée en fonction de leur difficulté de lecture. Pour ce faire,

différents critères formels, statistiques et psycholinguistiques ont été pris en compte (François, Billami, Gala & Bernhard, 2016). Le résultat est une ressource téléchargeable à des fins de recherche (par exemple, en traitement automatique des langues) et consultable également en ligne à des fins pédagogiques. La navigation permet la recherche à partir d'un mot-clé, puis la navigation entre les synonymes des listes de résultat associées au mot-clé. De telles caractéristiques peuvent s'avérer un atout pour l'enseignement du lexique<sup>2</sup>, que ce soit à l'école élémentaire ou dans le cadre d'une aide à la lecture et à la compréhension de textes chez des enfants en difficulté.

- 2 Notre contribution s'organise en quatre parties. La première présente le contexte de notre travail, avec un focus sur la baisse observée des compétences en lecture chez des enfants francophones et un survol des méthodes et des ressources existant pour enseigner le lexique. Une deuxième partie s'attarde sur les travaux visant à identifier la complexité lexicale, c'est-à-dire estimer un niveau de difficulté de lecture et de compréhension associé à une unité lexicale donnée (principalement basés sur des listes de fréquence). La troisième partie décrit la ressource *ReSyf*, son contenu et les algorithmes mis en place pour désambiguïser les sens et graduer les synonymes en fonction de leur difficulté de lecture. Enfin, la dernière partie propose quelques pistes d'utilisation de la ressource dans un contexte pédagogique.

## 2. Contexte

### 2.1. L'acquisition de connaissances lexicales et son impact sur la lecture

- 3 Si on s'intéresse à des évaluations internationales de la compréhension de l'écrit, par exemple PIRLS 2016 (Mullis, Martin, Foy & Hopper, 2017), on constate que les résultats des élèves français d'environ 10 ans sont parmi les plus mauvais de l'Union européenne et baissent depuis une quinzaine d'années. Les élèves belges francophones sont passés en dessous de la moyenne des 500 points avec 497 points. Enfin, la situation n'est pas meilleure ailleurs dans la francophonie, par exemple au Québec où des « études récentes montrent que les élèves québécois sont parmi les plus faibles en lecture au Canada » (Berthiaume, Anctil & Daigle, 2015). D'après PISA 2018, évaluation internationale pour des élèves de 15 ans, on constate que les résultats sont stables depuis 2009 pour la France avec un score de 493 points en 2018, au-dessus de la moyenne des pays de l'OCDE située à 487 points. Environ 20 % des élèves favorisés, mais seulement 2 % des élèves défavorisés, sont parmi les élèves très performants en compréhension de l'écrit en France, pour des proportions respectives de 17 % et 3 % en moyenne dans les pays de l'OCDE. Cette situation préoccupante mérite que l'on s'y attarde.
- 4 La compréhension en lecture, qui intervient dans la construction des apprentissages dans un cadre premièrement scolaire, s'étend par la suite à tous les domaines de la connaissance et fait partie intégrante du quotidien : cela fait de la période scolaire un moment déterminant pour l'avenir des lecteurs<sup>3</sup>. Dès l'entrée à l'école maternelle, il est essentiel de repérer les élèves avec des fragilités au niveau des habiletés langagières à l'oral et qui pourraient développer des difficultés en raison de leur inexpérience de l'écrit. Le rôle que jouent les enseignants à cet égard est de la plus haute importance (Morin & Montésinos-Gelet, 2007).

- 5 La notion de lexique s'élabore sur la connaissance de la signification des mots (sémantique) et la connaissance de leur relation de sens avec d'autres mots. Le sens est construit par une exposition répétée en contexte, notamment lors de lectures, et par l'étude structurée du lexique. D'après Stanovich (2009), les enfants qui lisent moins acquièrent moins de vocabulaire et leur compréhension reste limitée, ils ne perçoivent pas l'intérêt de la lecture ni ne prennent du plaisir à lire. Cela est nommé l'« effet Mathieu ».
- 6 La théorie la plus influente de l'apprentissage de la lecture repose sur l'idée que les enfants s'appuient sur des compétences de décodage grapho-phonémique pour reconnaître des mots nouveaux. Selon le processus d'auto-apprentissage, chaque décodage réussi d'un mot inconnu à l'écrit mais stocké dans le lexique mental permet d'acquérir une orthographe spécifique du mot qui est le fondement de la reconnaissance orthographique des mots. Ainsi, le décodage grapho-phonémique agit comme un mécanisme d'auto-apprentissage (Écalle & Magnan, 2015 ; Fayol & Jaffré, 1999 ; Ziegler & Goswami, 2005). Pour le prouver de façon quantitative, Ziegler, Perry et Zorzi (2014) ont implémenté le principe du « décodage et auto-apprentissage » dans un modèle computationnel qui permet de simuler numériquement ce processus d'apprentissage. Notons qu'un tel modèle computationnel de lecture est plausible sur le plan biologique et développemental. Ce modèle a été utilisé pour expliquer comment des déficits – visuels et phonémiques – affectent le développement orthographique de lecteurs dyslexiques. Il peut modéliser des différences interindividuelles, différentes trajectoires d'apprentissage et même les résultats probables d'une intervention ciblée (Perry, Zorzi & Ziegler, 2019 ; Ziegler, Perry & Zorzi, 2020).
- 7 L'objectif majeur de l'enrichissement du lexique est l'accès à la compréhension et à la production orale et écrite. La compréhension d'un texte écrit est le produit multiplicatif de deux variables indépendantes : la reconnaissance des mots écrits et la compréhension orale (Sprenger-Charolles, Desrochers & Gentaz, 2018). Ces deux activités, identification et compréhension orale, sont donc indispensables à la compréhension écrite et sont d'importance égale (Hoover & Gough, 1990 ; Goigoux, Cèbe & Pironom, 2016 ; Lervåg, Hulme & Melby-Lervåg, 2018 ; Castles, Rastle & Nation, 2018). L'acquisition du langage oral, dont le vocabulaire fait partie, est essentielle. *ReSyf* peut soutenir l'apprentissage du vocabulaire dans le cadre d'un enseignement méthodique et systématique.

## 2.2. Quelques méthodes et ressources pour l'enseignement du lexique

- 8 D'après Grossmann & Calaque (2000) et Grossmann (2011), le vocabulaire est enseigné à l'école d'un point de vue :
  - sémantique : étude du sens des unités lexicales et des relations qu'elles entretiennent avec d'autres unités dans leurs champs lexicaux (synonymie, antonymie, hyperonymie, hyponymie) et sémantiques (polysémie) ;
  - morphologique : étude des morphèmes qui composent les unités lexicales (leur construction par dérivation, par composition, les familles de mots) ;
  - étymologique : étude des origines des unités lexicales, des variations éventuelles de sens à travers le temps, des influences d'autres langues, etc.

- 9 Dans la pratique, l'apprentissage du vocabulaire peut être *incident* (explication du sens au sein d'une activité de lecture) ou *explicite* (appuyé sur l'appropriation de notions métalexicales), les deux activités étant complémentaires et à la fois nécessaires (Grossmann, 2011). Dans cette perspective, les ressources lexicales sont indispensables (Tremblay & Polguère, 2014). Pour un meilleur apprentissage du vocabulaire, les outils utilisés en classe doivent idéalement, d'après Cellier (2011), rendre compte de la structure du lexique (un réseau d'unités interreliées) et doivent faciliter la mémorisation et la réactivation des mots du vocabulaire de l'apprenant (en particulier grâce aux relations entre les mots). Cependant, force est de constater que les ressources numériques accessibles restent encore très proches des ressources classiques en format papier, mise à part certains dispositifs ludiques variés comme ceux du *Centre collégial de développement de matériel didactique* du Québec<sup>4</sup> et le *Programme de formation de l'école québécoise*<sup>5</sup> (Garcia-Debanc, Masseron & Ronveaux, 2013 ; Tremblay, Lefrançois & Lombard, 2013) ou la ressource *Crisco*<sup>6</sup>, dictionnaire de synonymes permettant un aperçu de la fréquence et une visualisation en 3D de synonymes.
- 10 Par ailleurs, si la notion de réseau lexical est exploitée dans des exemples concrets, comme le réseau lexical du français (Lux-Pogodalla & Polguère, 2011), ou dans *JeuxDeMots*<sup>7</sup> (Lafourcade, 2007), l'usage de ce type de ressources par les enseignants est encore très peu répandu (le premier n'est pas encore disponible en ligne et le second, un jeu sérieux permettant de construire un réseau lexico-sémantique en langue française, est plutôt destiné aux développements en traitement automatique des langues).
- 11 Outre la notion de système lexical (et de réseau lexical), il nous semble que les ressources utilisées en milieu pédagogique doivent s'appuyer sur la notion de polysémie et de niveau de complexité de lecture et de compréhension d'un mot. Si la tradition lexicographique intègre les variantes ou extensions de sens dans les ressources classiques (la polysémie est une notion clé, les variantes de sens sont identifiées dans les dictionnaires), la notion de complexité n'a longtemps été appréhendée que par le biais de listes de fréquence (les mots ne sont ni désambiguïsés ni reliés entre eux).
- 12 Enfin, les outils didactiques d'entraînement à la compréhension en lecture tels que *Lectorino & Lectorinette* (Goigoux & Cèbe, 2013), mais aussi *Gafi* (Bentolila, 1994), proposent un travail sur le lexique avec des situations de production orales et écrites pour multiplier et réitérer l'utilisation du vocabulaire appris dans différents contextes.
- 13 *ReSyf* est un outil qui peut étayer des activités élaborées par les professeurs pour compléter celles déjà disponibles. Les situations de jeux ne sont pas à négliger, car ce sont des occasions d'une répétition ludique et créative. Il existe des lotos de mots, des mots croisés, des devinettes, des charades, etc. Les jeux de société renforcent l'intérêt lexical : le jeu du prince des mots tordus, *Imagidés*, *Scrabble Flash*, *La Course aux mots*, *Dixit*, etc. En appui sur des jeux lexicaux, l'enseignant doit organiser, faire des liens, catégoriser, restructurer, expliquer sans cesse ce qui a été acquis au niveau du vocabulaire, car l'acquisition demeure un processus dynamique (Anctil, 2015).

### 3. Les listes de fréquence

- 14 Si l'utilisation des listes de mots dans l'enseignement du lexique est ancienne (Boulanger, 2003), l'introduction d'informations quantitatives associées aux unités lexicales date du début du  $xx^e$  siècle en milieu anglosaxon. Le psychologue et professeur d'anglais E. L. Thorndike pose les bases avec *The Teachers' Book of Words* (1921) ; d'autres auteurs comme C. K. Ogden avec *Basic English* (1930) utilisent l'information statistique à des fins pédagogiques, avec l'idée que plus un mot est fréquent, plus il est pertinent à apprendre. En France, des listes comme *Le français fondamental* (Gougenheim, 1958) ou les *Listes orthographiques de base du français* (Catach, 1984) sont des ressources dans ce même esprit. L'idée de définir un vocabulaire de base, ou fondamental, était « d'établir des priorités, de manière à sérier les difficultés et à programmer "scientifiquement" les apprentissages » (Grossmann, 2011).
- 15 Cependant, si cette approche n'est pas exempte de critiques (problème de la représentativité des corpus et de leur taille, entre autres), avec l'arrivée des corpus numériques, de nombreuses ressources ont été créées pour le français avec des méthodes de la statistique lexicale. Ces ressources intègrent des fréquences d'usage obtenues à partir des corpus. Dans le domaine de la psycholinguistique, *Brulex* (Content, Mousty & Radeaux, 1990) ou *Lexique 3* (New, Pallier, Ferrand & Matos, 2001) sont deux exemples notables. En didactique du lexique, la base lexicale *Manulex*<sup>8</sup> (Lété, Sprenger-Charolles & Colé, 2004) fournit des informations statistiques en combinant les critères de fréquence et de dispersion dans des manuels scolaires pour 23 900 lemmes. Quant à *FLELex*<sup>9</sup> (François, Gala, Watrin & Fairon, 2014), elle a été construite dans le même esprit : en utilisant des manuels d'enseignement du français langue étrangère. Les deux lexiques sont accessibles librement en ligne et construits à partir de corpus identifiés dans un niveau. Pour *Manulex*, ce sont les niveaux scolaires (cours préparatoire, cours élémentaire 1, cours élémentaire 2 et cours moyen ensemble) ; pour *FLELex*, ce sont les niveaux établis par le *Cadre européen commun de référence pour les langues* (CECRL). Le résultat est, pour *Manulex*, un fichier (matrice téléchargeable en format tableur) où à chaque mot sont associées différentes valeurs en tenant compte d'un niveau scolaire (par exemple, les fréquences estimées d'usage pour le nom *pirate* sont de 47,46<sup>10</sup> au CP, 17,37 en CE1 et 68,39 dans les trois niveaux suivants du primaire ; la fréquence globale est de 57,58). Pour *FLELex*, le résultat est affiché dans une interface où il est possible de visualiser la fréquence d'apparition d'une unité lexicale en colonnes (éventuellement, on peut comparer une unité lexicale à une autre) dans chacun des six niveaux du CECRL (A1 à C2). Dans *FLELex*, le nom *pirate* est présent en A2 et absent dans les autres niveaux. Dans ces deux ressources, les unités lexicales sont désambiguïsées grammaticalement : on peut chercher *pirate* comme nom ou comme adjectif (dans *Manulex*, la fréquence estimée d'usage globale est alors de 12,01 ; dans *FLELex*, l'adjectif *pirate* est inexistant). Toutefois, il n'y a pas de désambiguïsation sémantique, c'est-à-dire que la nuance entre *pirate* (au sens de « marin ») ou *pirate* (« escroc ») est rendue impossible.
- 16 Bien que la notion de complexité lexicale ne soit explicite dans aucune des ressources mentionnées, il est possible d'utiliser les fréquences obtenues sur corpus comme indicatif du type de lexique qu'un élève d'un niveau identifié pourra rencontrer pendant son apprentissage dans ce même niveau. À partir de là, il est possible d'extrapoler et de faire l'hypothèse que les unités lexicales des niveaux supérieurs sont des unités « complexes » pour le niveau considéré.

## 4. Une ressource lexicale avec des synonymes gradués en fonction de leur difficulté

- 17 Dans une perspective systémique qui intègre les considérations mentionnées précédemment (notion de réseau lexical, encodage explicite de la polysémie, dépassement des listes de fréquence pour identifier le niveau de complexité des mots), nous avons créé la ressource *ReSyf*. Cet outil permet une navigation dans des réseaux d'unités lexicales désambiguïsées et graduées en fonction de leur difficulté (voir plus bas pour les critères de difficulté). Il s'agit d'une proposition innovante avec des applications en didactique du lexique L1 ou L2, dont nous présentons les principales caractéristiques dans la suite de cet article.

### 4.1. Création de *ReSyf* : méthodologie

- 18 À partir des données de *JeuxDeMots* (Lafourcade, 2007), nous avons extrait une liste d'unités lexicales reliées entre elles par la relation de synonymie, une relation lexicale d'équivalence entre deux sens. La synonymie absolue étant rare, on considère qu'un synonyme est une unité lexicale avec une valeur sémantique suffisamment proche pour qu'elle puisse en remplacer une autre dans un contexte donné (Polguère, 2002). Par exemple, le nom *pirate* peut être remplacé par *corsaire*, *écumeur* ou *flibustier* dans le sens lié à la mer, ou par *escroc* ou *arnaqueur* dans le sens lié à des activités illicites. Comme *JeuxDeMots* est une plateforme participative en ligne sous forme de « jeu sérieux » (Lafourcade, Le Brun & Joubert, 2015), ce sont des humains qui, initialement, rentrent les synonymes ou les idées associées à partir d'un mot cible. Selon la consigne donnée, les internautes ne font pas la différence ni de catégorie grammaticale (*pirate* pouvant être un nom ou un adjectif) ni de sens associé au mot cible. Le résultat pour un mot cible est donc une liste d'unités lexicales liées par une relation lexico-sémantique, validée avec des algorithmes permettant de filtrer des éventuelles erreurs et de donner des poids à des propositions majoritaires (plus un mot est proposé par différents internautes, plus fort est son poids dans le réseau lexical).
- 19 Les avantages d'utiliser les données de *JeuxDeMots* au détriment d'autres ressources disponibles en ligne pour le français sont nombreuses. Premièrement, ce sont des humains qui ont fait le choix de proposer des synonymes associés à un mot cible, ce qui permet une grande variété des données (à la différence de beaucoup de bases lexicales existantes, *JeuxDeMots*, et donc *ReSyf*, contiennent des expressions polylexicales, par exemple, *écumeur de mer* comme synonyme de *pirate*). Deuxièmement, les unités lexicales se voient attribuer un poids en fonction de leur popularité dans les propositions, ce qui permet des traitements automatiques sur les mots les plus fiables, c'est-à-dire ceux ayant un poids plus fort. Ces deux éléments garantissent la qualité de la ressource. Enfin, un mécanisme de tri permet de raffiner les sens et de faire, ainsi, un premier tri entre les différents sens associés à une unité lexicale. Dans le cas de *pirate*, le *Diko* de *JeuxDeMots*<sup>11</sup> propose quatre sens possibles : « illicite », « escroc », « informatique » et « marin ».
- 20 À partir d'une liste initiale de mots cibles avec leurs synonymes, et en appliquant des algorithmes de désambiguïsation à base de connaissances provenant de *JeuxDeMots*, nous avons procédé à un enrichissement automatique des listes de synonymes,

considérés à partir de ce stade comme des mots-sens et non plus comme des mots-forme (Billami, 2018). Notre but était d'aboutir à une liste la plus précise possible d'unités lexicales désambiguïsées (mots-sens) et avec une granularité sémantique optimale par rapport au public visé par la ressource (des apprenants de français, principalement des enfants, mais aussi des adultes illettrés ou apprenants de français langue étrangère). Certains raffinements du *Diko* de *JeuxDeMots* étant considérés comme trop subtils, certains sens ont été regroupés avec une méthode de comparaison de sens dans des signatures lexicales (Billami, 2018). C'est le cas, par exemple, des sens « illicite » et « escroc » pour *pirate* : dans *ReSyf*, ils apparaissent sous un seul même sens de *pirate*.

- 21 Pour chaque mot-sens et sa liste (ou vecteur) de synonymes, un modèle statistique permettant d'attribuer des rangs de 1 à  $n$  aux synonymes en fonction de leur difficulté a été appliqué, avec 1 étant le synonyme le plus facile de la liste (François, Billami, Gala & Bernhard, 2016). Ce modèle utilise un ensemble de variables linguistiques et psycholinguistiques reconnues comme ayant un impact dans la difficulté de lecture (Gala, François, Bernhard & Fairon, 2014). Parmi ces variables, la fréquence d'un mot est déterminante, mais ce n'est pas la seule variable. D'autres caractéristiques comme la longueur de l'unité, l'irrégularité de la consistance phonème-graphème, une structure syllabique complexe (différente de consonne-voyelle ou voyelle), la présence de plusieurs morphèmes (mots dérivés ou composés), la polysémie, un nombre important de voisins orthographiques, etc., sont des facteurs qui rajoutent de la difficulté lors de la lecture (Gala, François, Bernhard & Fairon, 2014). Le modèle statistique développé par les auteurs permet de graduer les synonymes dans un vecteur. Il est supervisé et de type *pairwise* (il compare des paires mot-cible/mot-sens synonymes). Nous avons entraîné ce modèle avec 19 038 mots provenant de la ressource *Manulex*, en prenant en compte les différentes variables que nous avons définies et à l'aide de l'algorithme *SVMRank* (Herbrich, Graepel & Obermayer, 2000). Le résultat est une ressource de 57 589 entrées distribuées tel que le montre le tableau 1.

Tableau 1. – Distribution des entrées dans *ReSyf*.

	Noms	Verbes	Adjectifs	Adverbes	Total
Unités non polylexicales	21.495	5.065	7.635	1.105	35.300
Unités polylexicales, locutions	16.111	5.102	662	414	22.289
Nombre d'unités lexicales polysémiques (UPo)	6.737	1.779	1.691	126	10.333
Nombre d'unités lexicales monosémiques (UMo)	30.869	8.388	6.606	1.393	47.256
Moyenne de synonymes (UPo)	12,95	17,97	16,93	6,16	14,39
Moyenne de synonymes (UMo)	4,19	6,86	9,27	4,71	5,39
Moyenne de sens (UPo)	2,95	3,03	2,65	2,25	2,9

- 22 Le nombre d'unités monosémiques est presque cinq fois supérieur aux unités polysémiques (cela vient directement des données de *JeuxDeMots*). Avec la mise en place de notre approche de désambiguïsation, la moyenne des sens pour les unités polysémiques est de trois sens environ par entrée. En ce qui concerne la distribution des catégories lexicales, 65,30 % des entrées de *ReSyf* sont des Noms, 17,65 % sont des Verbes et 14,41 % sont des Adjectifs. Les Adverbes ne comptent que pour 2,64 % du total des entrées. Les unités polysémiques sont celles qui proposent le plus de synonymes,

12,95 en moyenne pour les Noms. Pour ce qui est des unités polylexicales, l'algorithme de graduation ne permet pas de considérer l'unité en tant que telle : la graduation est faite en calculant une moyenne des valeurs obtenues pour les mots-sens la composant. Par exemple, pour *écumeur de mer*, les valeurs d'*écumeur* et de *mer* seront considérées. On observe, sous l'article de *pirate* au sens de « marin », qu'*écumeur de mer* est considéré comme plus facile que *marin*, ce qui n'est pas tout à fait exact. Le biais vient de l'unité *mer*, qui est courte et fréquente et qui fausse le calcul de la valeur de difficulté<sup>12</sup>.

Figure 1. – Résultat de la recherche à partir de 'pirate' dans l'interface en ligne de ReSyf.



- 23 La figure 1 montre le résultat de la recherche via l'interface web pour l'unité *pirate* au complet. L'entrée est présentée selon deux catégories : en tant qu'adjectif (avec les sens « illicite », « escroc » et « marin ») et en tant que nom (avec les sens « escroc », « informatique » et « marin »). Pour le sens « illicite », *pirate* est considéré comme un terme plus simple. Parmi la liste de synonymes, il est par ailleurs possible de cliquer sur l'unité lexicale *illicite* et d'accéder à une autre page avec une liste de 14 synonymes pour cet adjectif : *illicite* s'y trouve en neuvième position. D'après la ressource, *secret* ou *interdit* sont des synonymes plus simples à lire et à comprendre par des enfants en école élémentaire pour cette notion-là.
- 24 Comme on peut le voir, l'adjectif *pirate* est considéré comme plus facile à lire et à comprendre dans les sens « escroc » et « illicite », au détriment d'autres formes comme *arnaqueur* et *illicite*. Dans le cas du nom *pirate* au sens de « marin », le nom *marin* est considéré comme plus simple, au détriment d'autres formes comme *corsaire*, *écumeur*, *boucanier* ou *flibustier*.

## 4.2. Évaluation des données de *ReSyf*

- 25 Les algorithmes de désambiguïsation sémantique appliqués aux données de *ReSyf* ont été évalués sur un échantillon de données manuellement désambiguïsées provenant de *JeuxDeMots* (Billami, 2018). Quant à l'évaluation du modèle d'ordonnement des synonymes, il a été évalué à l'aide d'un jeu de données de référence obtenu grâce à une consultation menée auprès de 40 annotateurs<sup>13</sup> qui devaient classer du plus simple au plus complexe une liste de 36 ensembles de synonymes choisis aléatoirement parmi les données obtenues de *JeuxDeMots* (François, Billami, Gala & Bernhard, 2016). Pour les 36 sens évalués, nous avons obtenu un accord inter-annotateur moyen ( $\alpha$  de Krippendorff) de 0,4. Cette valeur est plutôt faible (un accord parfait se rapprocherait de 1), mais il faut tenir compte de la subjectivité de la tâche (ordonner des unités lexicales en fonction de la perception de leur difficulté de lecture et de compréhension). Si on s'intéresse aux tris eux-mêmes, le résultat est plus satisfaisant, car dans 83,3 % des cas, les ensembles de synonymes sont triés de façon identique ou inversés d'un seul rang par rapport à ce que notre modèle statistique propose. Par exemple, la majorité d'annotateurs propose l'ordre *maigre, squelettique, osseux* là où notre modèle propose *osseux* avant *squelettique*. En nombre de synonymes, 91,04 % sont triés dans le même ordre par les annotateurs et le modèle automatique. Ces résultats sont très satisfaisants, même si des améliorations sont encore nécessaires dans les filtres pour éviter certains mots-sens, notamment pour des unités polylexicales, comme mentionné plus haut.
- 26 En effet, lorsque l'on consulte la ressource, les résultats obtenus pour un mot cible ne se révèlent pas toujours didactiquement pertinents. À l'origine, *ReSyf* a été construit dans un objectif de traitement automatique de la langue pour servir de base de connaissances dans un système de simplification de textes. Bien que le public cible de *ReSyf* ait toujours été le public apprenant, les données lexicales à la base (c'est-à-dire les sens) proviennent des propositions faites par des humains dans le cadre d'un jeu en ligne (cf. *JeuxDeMots*, section 4.1). Si plusieurs joueurs proposent le mot *vautour* comme synonyme de *pirate* dans le sens de « personne avide et malhonnête », ce sens sera considéré et traité comme une unité lexicale valable dans la liste de mots associée au mot-sens *pirate* (« escroc »). À ce jour, les synonymes liés à des registres de langue ou à des domaines de spécialité ne sont pas signalés dans *ReSyf*; nous travaillons à les identifier au sein d'une liste de synonymes, par exemple *ecchymose* dans les synonymes de *bleu* (« ecchymose »). Dans ces cas, ce sera à l'enseignant de proposer et de choisir les mots pertinents dans le cadre d'une recherche donnée. Ainsi, il pourra expliquer les différents sens de *pirate* (« escroc ») et évoquer également la polysémie du mot *vautour*. Dans le cas des unités polylexicales, des travaux sont en cours pour améliorer l'identification de ces unités en tant que telles et ne pas calculer leur indice de difficulté (de 1 à  $n$ ) en moyennant les indices des mots qui le composent. Dans des cas comme *écumeur de mer* ou *bras droit*, l'ordonnement proposé par *ReSyf* est discutable, car il y a dans la série des synonymes plus simples (respectivement, *pirate* ou *aide*), mais ce choix de moyennner fonctionne bien dans de nombreux autres cas comme *bras de fauteuil*, considéré plus facile qu'*accoudoir* et *accotoir*, ou *cheval de bois* considéré plus facile que *tréteau* et *chevalet*. Une fois de plus, ce sera à l'enseignant de pointer et d'expliquer les différents mots et nuances aux apprenants dans un souci d'accompagnement dans l'utilisation de la ressource. Enfin, l'usage qui pourra être fait variera en fonction du cycle d'enseignement, du niveau et des objectifs visés. En cycle

secondaire, et dans le cas d'enfants en difficulté de lecture et de compréhension, le travail à mettre en place avec la ressource sera très différent d'un travail en cycle 3 où l'accent pourra être mis dans les nuances de sens et les unités lexicales plus complexes. Dans tous les cas, c'est bien évidemment la pédagogie qui prime, la technologie n'est là qu'en appui et en complément.

## 5. ReSyf : quels usages possibles pour l'enseignant ou pour l'apprenant ?

- 27 L'accès de plus en plus facile au numérique au sein de l'école peut aider les enseignants, en fournissant aux élèves des occasions de travailler individuellement et d'augmenter leur niveau d'intérêt pour les textes écrits. Le numérique permet de varier les dispositifs d'enseignement. Cependant, sans toutefois se substituer aux activités pédagogiques de l'enseignant, il les complète, particulièrement dans les tâches d'apprentissage qui nécessitent une remédiation individuelle ou un grand nombre de répétitions (Lassault & Ziegler, 2018). Une ressource comme *ReSyf* peut trouver toute sa place dans des contextes divers d'enseignement du vocabulaire : elle permet de sélectionner des mots qui facilitent la compréhension du texte.
- 28 L'enseignant peut aussi utiliser cet outil numérique afin d'explicitier le vocabulaire dans des situations didactiques variées. Il peut proposer différentes tâches à l'élève pour étudier le vocabulaire comme :
- établir des liens (créer des associations, recourir à un arbre sémantique, définir le contexte des nouveaux mots) ;
  - distinguer les sens en fonction des catégories du discours des mots (noms, adjectifs, etc.) ;
  - découvrir les différentes acceptions de sens des unités polysémiques (*pirate, cour, bleu*, etc.) ;
  - insister sur le fait que la désambiguïsation des unités lexicales en général, et particulièrement les homophones, nécessitent un recours au contexte (*la voile/le voile*).
- 29 L'enseignement du vocabulaire à l'aide d'un contexte favorise une meilleure compréhension en lecture (Cèbe & Goigoux, 2015). Les enseignants doivent prendre conscience de stratégies métacognitives qu'ils ont automatisées afin de pouvoir les modéliser auprès de leurs élèves. Ceux-ci apprendront en reproduisant consciemment les étapes des stratégies modélisées par les enseignants afin d'identifier les mots, de comprendre le sens d'un mot inconnu (vocabulaire), de saisir le sens du texte et de développer l'écriture (Barth, 2011 ; Bianco, Lima & Sylvestre, 2004).
- 30 À ce jour, *ReSyf* a été utilisé dans les écoles et dans des cabinets d'orthophonistes pour préparer le travail avec les apprenants. Ainsi, *ReSyf* a permis de simplifier le vocabulaire d'un corpus (textes littéraires et documentaires scientifiques) dans le cadre de tests de lecture visant à évaluer l'effet de la simplification sur des lecteurs âgés de 7 et 9 ans (projet ANR *Alector*<sup>14</sup>). Concrètement, 79 textes différents ont été simplifiés manuellement en utilisant *ReSyf* et *Manulex* comme sources de graduation de la complexité du lexique des corpus. Ces textes, intégrés dans un dispositif numérique (tablette), ont été lus dans 26 classes entre le CE1 et le CM1, soit par presque 600 élèves entre 7 et 9 ans.
- 31 Les résultats de ces tests montrent que la simplification du texte, lorsqu'elle est effectuée à trois niveaux (lexical, syntaxique et discursif), améliore la vitesse de lecture et la compréhension des contenus (Gala, Tack, Javourey-Drevet, François & Ziegler,

2020). Des analyses en cours montrent que ces améliorations se produisent également lorsque l'on simplifie uniquement le lexique des textes littéraires. L'utilisation de *ReSyf* s'avère ainsi déterminante lors de la création des textes adaptés (simplifiés). La simplification consiste à réduire la complexité linguistique tout en conservant le sens original du texte (Siddharthan, 2014). Elle entraîne des modifications, en particulier au niveau lexical et morphosyntaxique, l'objectif étant de restaurer le plaisir de la lecture et une stimulation à l'exposition aux textes écrits.

- 32 Les exploitations possibles de *ReSyf* dans le domaine de la didactique du lexique et de la remédiation ouvrent des possibilités nouvelles pour les apprentissages lexicaux. Bien entendu, *ReSyf* n'est qu'un outil possible en complément d'un travail sur le vocabulaire en contexte. Il n'en reste pas moins que les notions mobilisées par la ressource — polysémie, synonymie et complexité — s'avèrent fondamentales pour l'enrichissement du répertoire lexical des apprenants.

---

## BIBLIOGRAPHIE

ANCTIL, Dominic. (2015). Un meilleur enseignement lexical pour une plus grande appropriation de la langue. Dans Service de la langue française et Conseil de la langue française et de la politique linguistique (dir.), *S'approprier le français. Pour une langue conviviale. Actes du colloque de Bruxelles (2013)* (p. 101-117). Louvain-la-Neuve : De Boeck Supérieur.

BARTH, Britt-Mari. (2011). *Le savoir en construction*. Paris : Retz.

BENTOLILA, Alain. (1994). *Gafi, méthode de lecture pour le cours préparatoire*. Paris : Nathan.

BERTHIAUME, Rachel, ANCTIL, Dominic & DAIGLE, Daniel. (2015). *L'enseignement du vocabulaire auprès d'élèves de 4<sup>e</sup> année du primaire en contexte de classe ordinaire : évaluation d'un dispositif d'enseignement en fonction des caractéristiques des élèves* (Rapport n° 2015-LC-187666). Montréal : Université de Montréal.

BIANCO, Maryse, LIMA, Laurent & SYLVESTRE, Emmanuel. (2004). Comment enseigner les stratégies de compréhension en lecture. Dans E. Gentaz & P. Dessus (dir.), *Comprendre les apprentissages* (p. 48-68), Paris : Dunod.

BILLAMI, Mokhtar Boumedyen. (2018). *Désambiguïsation sémantique dans le cadre de la simplification lexicale : contributions à un système d'aide à la lecture pour des enfants dyslexiques et faibles lecteurs* (Thèse de doctorat). Aix Marseille Université, Marseille.

BILLAMI, Mokhtar Boumedyen, FRANÇOIS, Thomas & GALA, Núria. (2018). *ReSyf: A French Lexicon with Ranked Synonyms*. Communication présentée au 27th International Conference on Computational Linguistics (COLING-2018), Santa Fe.

BOULANGER, Jean-Claude. (2003). *Les inventeurs de dictionnaires : de l'eduba des scribes mésopotamiens au scriptorium des moines médiévaux*. Ottawa : Les Presses de l'Université d'Ottawa.

CASTLES, Anne, RASTLE, Kathleen & NATION, Kate. (2018). Ending the Reading Wars: Reading Acquisition from Novice to Expert. *Psychological Science in the Public Interest*, 19(1), 5-51.

- CATACH, Nina. (1984). *Listes orthographiques de base du français. Les mots les plus fréquents et leurs formes flechies les plus frequentes*. Paris : Nathan.
- CÈBE, Sylvie & GOIGOUX, Roland. (2015). Vocabulaire et compréhension de textes : et si nous ne laissons à l'élève que la charge qui lui revient ? *Spirale - Revue de recherches en éducation*, 55, 119-136.
- CELLIER, Micheline. (2011). Des outils pour structurer l'apprentissage du vocabulaire. *Eduscol*. Disponible en ligne sur <[http://cache.media.eduscol.education.fr/file/Dossier\\_vocabulaire/57/6/Micheline\\_Cellier\\_111202\\_C\\_201576.pdf](http://cache.media.eduscol.education.fr/file/Dossier_vocabulaire/57/6/Micheline_Cellier_111202_C_201576.pdf)> (consulté en janvier 2019).
- CONTENT, Alain, MOUSTY, Philippe & RADEAUX, Monique. (1990). Brulex. Une base de données lexicale informatisée pour le français écrit et parlé. *L'Année psychologique*, 90, 551-566.
- ÉCALLE, Jean & MAGNAN, Annie. (2015). *L'apprentissage de la lecture et ses difficultés* (2<sup>e</sup> éd.), Paris : Dunod.
- FAYOL, Michel & JAFFRÉ, Jean-Pierre. (1999). L'acquisition/apprentissage de l'orthographe. *Revue française de pédagogie*, 126(1), 143-170.
- FRANÇOIS, Thomas, BILLAMI, Mokhtar Boumedyén, GALA, Núria & BERNHARD, Delphine. (2016). *Bleu, contusion, ecchymose* : tri automatique de synonymes en fonction de leur difficulté de lecture et compréhension. Dans *Actes de la conférence conjointe JEP-TALN-RECITAL 2016. Vol. 2 : TALN* (p. 15-29). Paris.
- FRANÇOIS, Thomas, GALA, Núria, WATRIN, Patrick & FAIRON, Cédric. (2014). FLELex: A Graded Lexical Resource for French Foreign Learners. Dans N. Calzolari et coll. (dir.), *The Proceedings of the Ninth International Conference on Language Resources and Evaluation - LREC'14* (p. 3766-3773). Reykjavik : ELRA.
- GALA, Núria, FRANÇOIS, Thomas, BERNHARD, Delphine & FAIRON, Cédric. (2014). Un modèle pour prédire la complexité lexicale et graduer les mots. Dans *Actes de la 21<sup>e</sup> conférence sur le Traitement automatique des langues naturelles - TALN'2014* (p. 91-102). Marseille.
- GALA, Núria, TACK, Anaïs, JAVOUREY-DREVET, Ludivine, FRANÇOIS, Thomas & ZIEGLER, Johannes C. (2020). Alector: A Parallel Corpus of Simplified French Texts with Alignments of Misreadings by Poor and Dyslexic Readers. Dans N. Calzolari et coll. (dir.), *Proceedings of the 12th Language Resources and Evaluation Conference - LREC 2020* (p. 1356-1361). Marseille : ELRA.
- GARCIA-DEBANC, Claudine, MASSERON, Caroline & RONVEAUX, Christophe (dir.). (2013). *Enseigner le lexique*. Namur : Presses universitaires de Namur.
- GOIGOUX, Roland & CÈBE, Sylvie. (2013). *Lectorino & Lectorinette. Apprendre à comprendre les textes narratifs*. Paris : Retz.
- GOIGOUX, Roland, CÈBE, Sylvie & PIRONOM, Julie. (2016). Les facteurs explicatifs des performances en lecture-compréhension à la fin du cours préparatoire. *Revue française de pédagogie*, 196, 67-84.
- GOUGENHEIM, Georges. (1958). *Dictionnaire fondamental de la langue française*. Paris : Didier.
- GROSSMANN, Francis. (2011). Didactique du lexique : état des lieux et nouvelles orientations. *Pratiques*, 149-150, 163-183. <<https://doi.org/10.4000/pratiques.1732>>.
- GROSSMANN, Francis & CALAQUE, Élisabeth (dir.). (2000). Enseignement/apprentissage du lexique. *Lidil*, 21.

- HERBRICH, Ralf, GRAEPEL, Thore & OBERMAYER, Klaus. (2000). Large Margin Rank Boundaries for Ordinal Regression. Dans M. I. Jordan, Y. LeCun & S. A. Solla (dir.), *Advances in Neural Information Processing Systems* (p. 115-132). Cambridge : MIT Press.
- HOOVER, Wesley A. & GOUGH, Philip B. (1990). The Simple View of Reading. *Reading and Writing*, 2(2), 127-160.
- LAFOURCADE, Mathieu. (2007). Making People Play for Lexical Acquisition with the JeuxDeMots Prototype. Dans *SNLP'07 : 7th International Symposium on Natural Language Processing* (p. 7-14). Pattaya.
- LAFOURCADE, Mathieu, LE BRUN, Nathalie & JOUBERT, Alain. (2015). *Games with a Purpose (GWAPS)*. Wiley.
- LASSAULT, Julie & ZIEGLER, Johannes C. (2018). Les outils numériques d'aide à l'apprentissage de la lecture. *Langue française*, 199(3), 111-121.
- LERVÅG, Arne, HULME, Charles & MELBY-LERVÅG, Monica. (2018). Unpicking the Developmental Relationship between Oral Language Skills and Reading Comprehension: It's Simple, but Complex. *Child Development*, 89(5), 1821-1838.
- LÉTÉ, Bernard, SPRENGER-CHAROLLES, Liliane & COLÉ, Pascale. (2004). Manulex: A Grade-Level Lexical Database from French Elementary-School Readers. *Behavior Research Methods, Instruments and Computers*, 36, 156-166.
- LUX-POGODALLA, Veronika & POLGUÈRE, Alain. (2011). Construction of a French Lexical Network: Methodological Issues. Dans *Proceedings of the First International Workshop on Lexical Resources, WoLeR 2011. An ESSLLI 2011 Workshop* (p. 54-61). Ljubljana.
- MORIN, Marie-France & MONTÉSINOS-GELET, Isabelle. (2007). Effet d'un programme d'orthographe approchées en maternelle sur les performances ultérieures en lecture et en écriture d'élèves à risque. *Revue des sciences de l'éducation*, 33(3), 663-683.
- MULLIS, Ina V. S., MARTIN, Michael O., FOY, Pierre & HOOPER, Martin. (2017). *PIRLS 2016 International Results in Reading*. Boston College, TIMSS & PIRLS International Study Center. Disponible en ligne sur <<https://timssandpirls.bc.edu/isc/publications.html>>.
- NEW, Boris, PALLIER, Christophe, FERRAND, Ludovic & MATOS, Rafael. (2001). Une base de données lexicales du français contemporain sur Internet : Lexique 3. *L'Année psychologique*, 101, 447-462.
- OGDEN, Charles Kay (1930). *Basic English: A General Introduction with Rules and Grammar*. Londres : Paul Treber & Co.
- PERRY, Conrad, ZORZI, Marco & ZIEGLER, Johannes C. (2019). Understanding Dyslexia Through Personalized Large-Scale Computational Models. *Psychological Science*, 30(3), 386-395.
- POLGUÈRE, Alain. (2002). *Notions de base en lexicologie*. Observatoire de linguistique Sens-Texte. Disponible en ligne <[www.etudier.com/dissertations/Notions-De-Base-En-Lexicologie/487680.html](http://www.etudier.com/dissertations/Notions-De-Base-En-Lexicologie/487680.html)> (consulté le 5 octobre 2020).
- SIDDHARTHAN, Advaith. (2014). A Survey of Research on Text Simplification. *ITL - International Journal of Applied Linguistics*, 165(2), 259-298.
- SPRENGER-CHAROLLES, Liliane, DESROCHERS, Alain & GENTAZ, Édouard. (2018). Apprendre à lire-écrire en français. *Langue française*, 199(3), 51-67.
- STANOVICH, Keith E. (2009). Matthew Effects in Reading: Some Consequences of Individual Differences in the Acquisition of Literacy. *Journal of Education*, 189(1-2), 23-55.

THORNDIKE, Edward. (1921). *The Teacher's Word Book*. Teachers College, Columbia University.

TREMBLAY, Ophélie, LEFRANÇOIS, Pascale & LOMBARD, Vincent. (2013). Des listes de mots pour les trois cycles du primaire au Québec. Dans C. Garcia-Debanc, C. Masseron & C. Ronveaux (dir.), *Enseigner le lexique, objets et pratiques* (p. 65-83). Namur : Presses Universitaires de Namur.

TREMBLAY, Ophélie & POLGUÈRE, Alain. (2014). Une ontologie linguistique au service de la didactique du lexique. Dans F. Neveu, P. Blumenthal, L. Hriba, G. Gerstenberg, J. Meinschaefer & S. Prévost (dir.), *4<sup>e</sup> Congrès mondial de linguistique française (CMLF 2014)* (p. 1173-1188). Berlin : EDP Sciences.

ZIEGLER, Johannes C. & GOSWAMI, Usha. (2005). Reading Acquisition, Developmental Dyslexia, and Skilled Reading Across Languages: A Psycholinguistic Grain Size Theory. *Psychological Bulletin*, 131(1), 3-29. <<https://doi.org/10.1037/0033-2909.131.1.3>>.

ZIEGLER, Johannes C., PERRY, Conrad & ZORZI, Marco. (2014). Modelling Reading Development through Phonological Decoding and Self-Teaching: Implications for Dyslexia. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1634). <<https://doi.org/10.1098/rstb.2012.0397>>.

ZIEGLER, Johannes C., PERRY, Conrad & ZORZI, Marco. (2020). Learning to Read and Dyslexia: From Theory to Intervention through Personalized Computational Models. *Current Directions in Psychological Sciences*, 29(3), 293-300.

## NOTES

1. Voir <<https://cental.uclouvain.be/resyf/>>.
2. Dans cet article, comme dans Grossmann (2011), nous utiliserons principalement le terme *lexique* lorsque nous aborderons des notions générales (enseignement du lexique, didactique du lexique). Le terme *vocabulaire* peut être utilisé dès lors qu'on fait référence spécifiquement à « la part du lexique instanciée par un individu ou un groupe social » (p. 163), principalement dans le cadre de son apprentissage à l'école primaire.
3. DEPP, Note d'information, n° 20.20, juin 2020. Disponible en ligne sur <[www.education.gouv.fr/journee-defense-et-citoyennete-2019-plus-d-un-jeune-francais-sur-dix-en-difficulte-de-lecture-303885](http://www.education.gouv.fr/journee-defense-et-citoyennete-2019-plus-d-un-jeune-francais-sur-dix-en-difficulte-de-lecture-303885)>.
4. Voir <[www.ccdmd.qc.ca/fr/jeux\\_pedagogiques/?id=81](http://www.ccdmd.qc.ca/fr/jeux_pedagogiques/?id=81)>.
5. Voir <[www.franqus.ca/MELS/liste\\_orthographique/outil\\_de\\_recherche/](http://www.franqus.ca/MELS/liste_orthographique/outil_de_recherche/)>.
6. Voir <[http://crisco.unicaen.fr/accueil-crisco-863157.kjsp\[...\]](http://crisco.unicaen.fr/accueil-crisco-863157.kjsp[...].)>.
7. Voir <[www.jeuxdemots.org/diko.php](http://www.jeuxdemots.org/diko.php)>.
8. Voir <[www.manulex.org](http://www.manulex.org)>.
9. Voir <<http://cental.uclouvain.be/flelex/>>.
10. Dans *Manulex*, les fréquences d'occurrences des mots ont été calculées à partir d'un corpus de 54 manuels scolaires (1,9 millions de mots). Nous montrons la fréquence estimée d'usage pour un million de mots (U) qui tient compte de la fréquence brute avec un ajustement avec la dispersion des occurrences à travers les manuels.
11. Le *Diko* est la ressource consultable en ligne (<[www.jeuxdemots.org/diko.php](http://www.jeuxdemots.org/diko.php)>) ; *JeuxDeMots* est la plateforme permettant d'acquérir les données en production participative (*crowdsourcing*) sous forme de jeu sérieux (*Game with a purpose* ou *GWAP*).
12. Des travaux sont en cours pour mettre en place d'autres heuristiques afin de mieux traiter les unités polylexicales (à commencer par les identifier en tant que telles dans les corpus).

13. Les annotateurs étaient des étudiants universitaires de deuxième et troisième cycle et des enseignants-chercheurs majoritairement francophones natifs ou avec un niveau de français C1 ou C2 (selon l'échelle du CECRL).

14. Voir <<https://alectorsite.wordpress.com/>>.

---

## RÉSUMÉS

Les ressources lexicales sont indispensables à l'enseignement explicite du lexique (Tremblay & Polguère, 2014). Pour un meilleur apprentissage, idéalement les outils utilisés en classe doivent rendre compte de la structure du lexique (un réseau d'unités interreliées) et doivent faciliter la mémorisation et la réactivation des mots du vocabulaire de l'apprenant (en particulier grâce aux relations entre les mots). Toutefois, force est de constater que, outre des dispositifs ludiques variés, les ressources numériques existantes sont encore très proches des ressources classiques. La notion de réseau lexical est peu exploitée et, globalement, peu d'enseignants connaissent des dispositifs issus de la recherche. Dans cet article, nous présentons *ReSyf* (Billami, François & Gala, 2018), une ressource lexicale accessible en ligne qui intègre les notions de polysémie, de synonymie et de complexité de lecture. Les exploitations possibles de *ReSyf* dans le domaine de la didactique du lexique et de la remédiation orthophonique ouvrent des possibilités nouvelles pour les apprentissages lexicaux en complément d'un travail sur le vocabulaire en contexte. Dans cette proposition, nous décrivons brièvement la ressource et une expérience de simplification de textes à l'aide de *ReSyf*. Ce dispositif a permis de mettre en place des tests de lecture avec du matériel adapté qui s'est avéré une aide temporaire efficace pour l'apprentissage de la lecture et du vocabulaire des apprenants, tous profils confondus.

Lexical resources are essential to explicit teaching of vocabulary (Tremblay & Polguère, 2014). To enable stronger vocabulary learning, the tools used in the classrooms should show the structure of the lexicon (a network of lexical units) and assist the learners to memorize and to reactivate the meaning of the words that they already know (i.e. using the lexical relations among the words). However, in addition to a varied set of playful propositions, existing digital resources are still similar to classical devices. The notion of lexical network is not, or hardly ever exploited and very few teachers are aware of propositions coming from research. In this paper, we present *ReSyf* (Billami, François & Gala, 2018), a lexical resource available online based on the notions of polysemy, synonymy and reading complexity. *ReSyf* offers new possibilities for teaching vocabulary at school or speech therapy training, alongside with working with the vocabulary in context. In this proposition, we briefly describe the resource and an experience of text simplification using *ReSyf*. This experience enabled to propose reading tests with adapted reading material resulting in an efficient temporary device to improve reading, comprehension, and vocabulary learning for a variety of French learners.

## INDEX

**Mots-clés** : ressource lexicale, polysémie, synonymie, lecture, vocabulaire

**Keywords** : lexical resource, polysemy, synonymy, reading, vocabulary

## AUTEURS

### **NÚRIA GALA**

Aix Marseille Université, Laboratoire Parole et Langage (LPL), CNRS UMR 7309

### **LUDIVINE JAVOUREY-DREVET**

Aix Marseille Université, Laboratoire de Psychologie cognitive (LPC), CNRS UMR 7290, et  
laboratoire Apprentissage, Didactique, Évaluation, Formation (ADEF), UR 4671