



HAL
open science

Modèles adaptatifs pour prédire automatiquement la compétence lexicale d'un apprenant de français langue étrangère

Anaïs Tack, Thomas François, Anne-Laure Ligozat, Cédric Fairon

► To cite this version:

Anaïs Tack, Thomas François, Anne-Laure Ligozat, Cédric Fairon. Modèles adaptatifs pour prédire automatiquement la compétence lexicale d'un apprenant de français langue étrangère. JEP-TALN-RECITAL 2016, Jan 2016, Paris, France. hal-01631772

HAL Id: hal-01631772

<https://hal.science/hal-01631772v1>

Submitted on 10 Nov 2017

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.

Modèles adaptatifs pour prédire automatiquement la compétence lexicale d'un apprenant de français langue étrangère

Anaïs Tack¹ Thomas François^{1,3} Anne-Laure Ligozat² Cédric Faison¹

(1) Université catholique de Louvain, IL&C, CENTAL, Place Blaise Pascal 1, 1348 Louvain-la-Neuve, Belgique

(2) LIMSI, CNRS et ENSIIE, rue John Von Neumann, 91400 Orsay, France

(3) Chargé de recherches FNRS

anaïs.tack@uclouvain.be, thomas.francois@uclouvain.be, annlor@limsi.fr,
cedrick.faison@uclouvain.be

RÉSUMÉ

Cette étude examine l'utilisation de méthodes d'apprentissage incrémental supervisé afin de prédire la compétence lexicale d'apprenants de français langue étrangère (FLE). Les apprenants ciblés sont des néerlandophones ayant un niveau A2/B1 selon le Cadre européen commun de référence pour les langues (CECR). À l'instar des travaux récents portant sur la prédiction de la maîtrise lexicale à l'aide d'indices de complexité, nous élaborons deux types de modèles qui s'adaptent en fonction d'un retour d'expérience, révélant les connaissances de l'apprenant. En particulier, nous définissons (i) un modèle qui prédit la compétence lexicale de tous les apprenants du même niveau de maîtrise et (ii) un modèle qui prédit la compétence lexicale d'un apprenant individuel. Les modèles obtenus sont ensuite évalués par rapport à un modèle de référence déterminant la compétence lexicale à partir d'un lexique spécialisé pour le FLE et s'avèrent gagner significativement en exactitude (9%-17%).

ABSTRACT

Adaptive models for automatically predicting the lexical competence of French as a foreign language learners

This study examines the use of supervised incremental machine learning techniques to automatically predict the lexical competence of French as a foreign language learners (FFL). The targeted learners are native speakers of Dutch having attained the A2/B1 proficiency level according to the Common European Framework of Reference for Languages (CEFR). Following recent work on lexical proficiency prediction using complexity indices, we elaborate two types of models that adapt to feedback disclosing the learners' knowledge. In particular, we define (i) a model that predicts the lexical competence of learners having the same proficiency level and (ii) a model that predicts the lexical competence of one particular learner. The obtained models are then evaluated with respect to a baseline model, which predicts the lexical competence based on a specialised lexicon for FFL, and appear to gain significantly in accuracy (9%-17%).

MOTS-CLÉS : prédiction lexicale, modèles adaptatifs, apprentissage incrémental, FLE.

KEYWORDS: lexical prediction, adaptive models, incremental learning, FFL.

1 Introduction

La problématique de la prédiction de la difficulté de textes pour un allophone intéresse un certain nombre de chercheurs qui font dialoguer le traitement automatique du langage (TAL) et l'apprentissage des langues assisté par ordinateur (ALAO) (Heilman *et al.*, 2007; François & Fairon, 2013; Pilán *et al.*, 2014). Pouvoir effectuer ce type de prédiction est important dans un contexte ALAO, car présenter des textes inadéquats à un apprenant est problématique. Des textes trop simples vont limiter l'apprentissage, tandis que des textes trop ardues risquent de conduire à une mauvaise compréhension du contenu et à une perte de motivation. Or, on sait que la lecture joue un rôle crucial dans l'acquisition incidente du lexique, pour autant que s'opère une bonne compréhension du texte (Krashen, 1989). Pour ce faire, il faut que la majorité des mots figurant dans un texte, c'est-à-dire entre 95% et 98%, soient connus par l'apprenant (Hu & Nation, 2000; Laufer & Ravenhorst-Kalovski, 2010). Afin de pouvoir sélectionner de manière automatisée des matériaux appropriés qui facilitent l'apprentissage, nous pensons qu'il peut être utile d'utiliser des techniques issues du TAL pour évaluer la complexité lexicale d'un texte pour un apprenant en particulier.

Reste à définir ce qui est visé par le terme de complexité lexicale. De nombreuses études ont cherché à identifier, dans la nature même des mots, des caractéristiques qui les rendent plus complexes à traiter pour le cerveau humain. La psychologie cognitive a ainsi montré qu'un nombre important de ces caractéristiques – tels la fréquence lexicale, le nombre de sens, le nombre de voisins orthographiques, l'âge d'acquisition, etc. – influencent la vitesse de traitement des mots¹. En TAL, quelques travaux se sont nourris de ces connaissances pour implémenter des modèles visant à prédire automatiquement le niveau de difficulté d'un mot sur la base de variables psycholinguistiques (Shardlow, 2013; Gala *et al.*, 2014). Cependant, ces tentatives, basées sur bon nombre de caractéristiques lexicales, ne dépassent que difficilement une baseline reposant uniquement sur la fréquence des mots.

Une autre approche de la complexité lexicale inverse le problème et se focalise sur l'apprenant plutôt que sur les mots. Elle consiste à modéliser le lexique d'un apprenant ou d'un groupe homogène d'apprenants et à définir comme difficiles les mots supposés inconnus par le modèle. Cette approche requiert cependant de définir ce qu'est la compétence lexicale d'un apprenant. Cette notion est loin d'être évidente, car il y a plusieurs éléments à prendre en compte lorsqu'on veut mesurer automatiquement les connaissances lexicales d'un apprenant (Gyllstad, 2013). En effet, une dimension du vocabulaire L2 communément envisagée est la quantité de mots qu'un apprenant a acquis jusqu'à présent (*taille du vocabulaire*). Une deuxième perspective (*profondeur du vocabulaire*) mesure à quel point l'apprenant a intériorisé ces mots. Ainsi, l'acquisition du vocabulaire concerne l'intériorisation progressive de trois propriétés d'un mot : (i) sa forme phonétique et graphique, (ii) son sémantisme et (iii) son usage en contexte (Nation, 2001). Une dernière dimension concerne la *fluidité lexicale*, c'est-à-dire la facilité avec laquelle l'apprenant utilise et comprend le vocabulaire acquis en production et en réception. Enfin, il importe également de prendre en compte la façon même de développer une telle compétence, qui s'avère particulièrement transitoire et incrémentale (Schmitt, 1998).

Les études qui se sont penchées sur la prédiction de la compétence lexicale d'un apprenant d'un point de vue TAL restent encore peu nombreuses. Une tentative a été faite dans le cadre du projet FLELex (François *et al.*, 2014), dans lequel a été produit, à l'exemple de la ressource Manulex pour le français langue maternelle (Lété *et al.*, 2004), une ressource lexicale graduée pour le français langue étrangère (FLE). La ressource FLELex² décrit, pour environ 15 000 mots de base du français, leur

1. Le lecteur peut consulter la synthèse de Ferrand (2007) pour des détails sur ces études.

2. La ressource est disponible sur le site <http://cental.uclouvain.be/flelex/>.

distribution sur les 6 niveaux du Cadre européen commun de référence pour les langues (CECR), allant des niveaux élémentaires (A1 et A2) aux niveaux intermédiaires (B1 et B2) et avancés (C1 et C2). Tack *et al.* (2016) se sont basés sur cette ressource pour prédire les mots connus d'un apprenant en transformant la distribution en un niveau unique et en comparant le niveau du mot au niveau CECR de l'apprenant. Néanmoins, un modèle prédictif qui se fonde sur un lexique gradué par niveau du CECR tel que FLELex semble fournir une image restreinte de la compétence lexicale d'un apprenant. D'abord, vu que les prédictions ne prennent en compte que le niveau de maîtrise de l'apprenant, le modèle ne rend pas compte des différences possibles entre apprenants du même niveau. Ensuite, puisque les prédictions sont dérivées d'une connaissance symbolique, a priori, de la compétence lexicale (c'est-à-dire à partir de manuels scolaires de FLE), le modèle n'intègre pas de données pour modéliser a posteriori la compétence lexicale d'un apprenant. Enfin, étant donné que les prédictions du modèle restent constantes, le modèle prédictif résultant n'est pas adaptatif et ne rend pas compte de la nature incrémentale de l'acquisition du vocabulaire L2.

Or, dans le domaine de l'apprentissage des langues intelligemment assisté par ordinateur (ALIAO), plusieurs travaux ont abouti à la création de tuteurs intelligents s'adaptant à l'activité d'un apprenant dans un environnement d'apprentissage. Par exemple, les systèmes tutoriels développés par Stockwell (2007) et par Chen & Chung (2008) tracent l'activité au sein des exercices de vocabulaire afin de déterminer les difficultés lexicales individuelles d'un apprenant donné et de lui présenter de nouveaux exercices appropriés. Ainsi, les systèmes tutoriels intelligents se caractérisent principalement par l'intégration d'un modèle de l'utilisateur qui enregistre les données sur l'activité d'un usager. Afin de concevoir un tel modèle (aussi appelé *modèle de l'apprenant* dans un contexte L2), de nombreux tuteurs intelligents se fondent sur des méthodes d'apprentissage incrémental (Licchelli *et al.*, 2004). L'usage des méthodes incrémentales est indispensable lorsque l'on veut « tracer un phénomène qui évolue au cours du temps » (Bottou, 1998, p. 5, notre traduction), ce qui est le cas de la compétence lexicale en L2. Comme Licchelli *et al.* (2004, p. 938) l'ont souligné, puisque les méthodes d'apprentissage automatique classiques requièrent que toutes les données soient disponibles au moment de l'entraînement, ils ne permettent pas de gérer les situations où l'on doit traiter des informations à des moments non déterminés à l'avance. Néanmoins, il semble que la modélisation par apprentissage incrémental ne fasse pas encore l'objet de nombreux travaux en TAL et, moins encore en ce qui concerne la prédiction de la compétence lexicale.

Notre étude vise dès lors à fournir une nouvelle contribution à la prédiction de la compétence lexicale, en analysant la modélisation de cette compétence par apprentissage incrémental. En particulier, notre travail consiste à élaborer un modèle de l'apprenant qui prédit la compétence lexicale en lecture d'un allophone individuel et qui intègre un ensemble de traits de complexité lexicale inspirés de Gala *et al.* (2014) ainsi que les fréquences observées dans FLELex (François *et al.*, 2014). L'hypothèse principale de notre étude est que l'intégration de ces variables dans un modèle adaptatif permet de modéliser de manière plus exacte la compétence lexicale de l'apprenant en lecture.

Dans ce qui suit, nous présentons d'abord comment nous avons obtenu des données de la compétence lexicale en lecture, en exploitant un corpus de textes narratifs et informatifs (Section 2.1) et en faisant annoter ces textes par des apprenants de français langue étrangère (Section 2.2). Ensuite, nous détaillons comment nous avons défini les modèles adaptatifs prédictifs de la compétence lexicale (Section 2.3.1) et comment nous les avons entraînés (Section 2.3.2). Enfin, nous clôturons ce travail par une discussion des résultats de l'évaluation des modèles par rapport au modèle de référence développé par Tack *et al.* (2016) (Section 3.1), ainsi que des principales implications de notre étude pour le domaine de la prédiction automatique de la compétence lexicale en L2 (Section 3.2).

2 Méthodologie

Mesurer la compétence lexicale réceptive d'un apprenant n'est pas chose aisée, en particulier lorsque l'on cherche à dépasser la simple reconnaissance des formes lexicales pour essayer de capturer une connaissance lexicale plus approfondie. Dans ce but, nous avons observé la compétence lexicale réceptive d'apprenants à travers une plateforme de lecture qui leur permettait d'annoter les mots dont ils ne se rappelaient pas le sens en contexte. Pour concevoir une telle expérience de lecture, nous avons exploité un ensemble de textes narratifs et informatifs (Section 2.1) dans lesquels toutes les unités lexicales, y compris les unités polylexicales, ont été balisées et annotées (Section 2.2). Ces annotations ont ensuite été utilisées pour entraîner des modèles de classification adaptatifs (Section 2.3).

2.1 Description du corpus

En vue d'obtenir une annotation de la compétence lexicale en lecture, nous avons exploité deux corpus textuels collectés par Brouwers *et al.* (2014). Les corpus constituent, d'une part, un corpus parallèle de contes authentiques qui ont été simplifiés manuellement pour des apprenants du français (nommé ci-après le corpus *Contes*) et, d'autre part, un corpus comparable de textes informatifs issus de *Wikipédia* et de textes wiki simplifiés pour des enfants francophones issus de *Vikidia* (nommé ci-après le corpus *Wiki*). Toutefois, puisqu'il était impossible de présenter l'intégralité de ces corpus à des sujets allophones pour des raisons matérielles, nous avons procédé à une sélection d'un échantillon plus réduit de textes.

Nous avons d'abord effectué une expérience pilote auprès d'un allophone (ayant le niveau B1 selon le CECR) en lui demandant de lire et d'identifier les mots difficiles dans 11 contes simplifiés correspondant à son niveau. Il devait également évaluer la difficulté des textes et du vocabulaire au moyen d'un questionnaire quantitatif. Il est ressorti de ce questionnaire que le participant avait déjà lu une version de certains contes dans sa langue maternelle et que le vocabulaire des 11 textes était trop répétitif. En conséquence, nous avons décidé d'opter pour un échantillon plus optimal de textes de manière à ce que les unités lexicales à annoter soient les plus diverses possible. Pour ce faire, nous avons d'abord passé au crible l'entièreté des corpus afin d'écarter les textes susceptibles de biaiser l'expérience³, ainsi que les textes qui étaient trop peu informatifs ou trop longs. Ensuite, nous avons conçu un algorithme glouton qui sélectionne, en fonction d'une limite du nombre d'unités lexicales à annoter, un sous-ensemble de textes où le chevauchement lexical entre les vocabulaires V des paires de textes i et j (**Équation 1**) est minimal et où la richesse lexicale (**Équation 2**) est maximale.

$$\text{chevauchement}(V_i, V_j) = \frac{|V_i \cap V_j|}{|V_i \cup V_j|} \quad (1)$$

$$\text{STTR} = \frac{1}{N} \sum_{i=1}^N \left(\frac{\text{Types}_i}{\text{Tokens}_i} \right) \times 100 \quad (2)$$

Après cette étape de sélection, nous avons obtenu une collection de 51 textes comprenant en moyenne 373 mots par texte (**Tableau 1**). Le total des unités lexicales à annoter et à classifier s'élevait à 21 047

3. C'est-à-dire, les textes pour lesquels un lecteur pourrait plus facilement deviner les mots inconnus, grâce à sa connaissance préalable du fil narratif (par ex. les contes déjà lus dans sa langue maternelle).

unités. Signalons que l’algorithme n’a sélectionné que six textes narratifs (corpus *Contes*) et paraît donc largement favoriser les textes informatifs (corpus *Wiki*). Néanmoins, le **Tableau 1** indique aussi que ces textes narratifs, nettement plus longs que les textes informatifs, comprennent à peu près le même nombre de mots que les 45 textes informatifs.

| | Contes | | | Wiki | | | Total |
|-----------|-----------|------------|--------|-----------|---------|--------|--------|
| | originaux | simplifiés | tous | Wikipedia | Vikidia | tous | |
| # textes | 1 | 5 | 6 | 8 | 37 | 45 | 51 |
| # lexèmes | 4 694 | 6 291 | 10 985 | 2 458 | 7604 | 10 062 | 21 047 |

TABLE 1: La sélection des textes utilisés pour l’entraînement des modèles.

2.2 Annotations lexicales

Afin d’identifier automatiquement les unités lexicales à annoter, nous avons tokenisé, lemmatisé et étiqueté en catégories grammaticales la sélection finale de textes au moyen de *TreeTagger* (Schmid, 1994). Cependant, *TreeTagger* ne permet pas d’identifier les expressions polylexicales (EPL), telles que *tandis que* ou encore *maîtrise de soi*. Pour relever automatiquement les expressions multi-mots dans les textes à annoter, nous avons utilisé la version de la ressource FLELex basée sur un étiqueteur par champs conditionnels aléatoires (à savoir FLELex-CRF) et qui comprend des EPL.

Nous avons ensuite converti les textes prétraités dans un format XML où chacune des unités (poly)lexicales identifiées était entourée d’une balise. Ces textes balisés ont ensuite été intégrés au sein d’une interface de lecture web spécialement conçue pour cette étude (**Figure 1**), permettant aux allophones de lire les textes de notre échantillon et de cliquer sur les mots dont le sens leur était inconnu. L’interface choisissait les textes à annoter de manière aléatoire et les présentait phrase après phrase (à l’exception des phrases entretenant un rapport argumentatif direct) afin de limiter l’effet des inférences contextuelles sur l’interprétation du sens de mots inconnus.

Avant de commencer l’expérience, les allophones recevaient des consignes pour savoir comment annoter un mot inconnu. Nous avons limité l’annotation à une simple décision binaire (connu ou inconnu), quoiqu’il existe en réalité plusieurs degrés de la connaissance lexicale (Wesche & Paribakht, 1996). Dès lors, en s’inspirant de l’échelle des connaissances lexicales de ces derniers, nous avons demandé aux participants d’annoter comme étant inconnus les mots qui ne correspondent qu’aux deux premiers degrés de l’échelle, à savoir (1) les mots qu’ils n’ont jamais vus auparavant et (2) les mots qu’ils ont déjà vus, mais dont ils ignorent le sens. Pour les guider dans cette tâche, il leur était demandé d’annoter les mots pour lesquels ils ne pouvaient pas donner de synonyme en français, ni de traduction dans leur langue maternelle, ou pour lesquels ils estimaient avoir besoin d’un dictionnaire pour comprendre le sens.

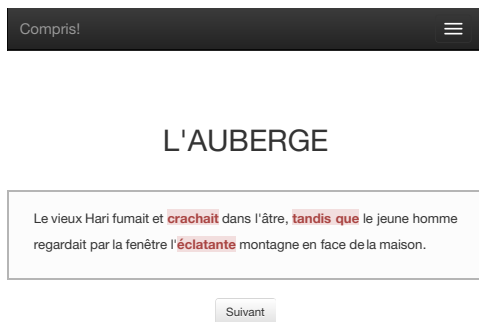


FIGURE 1: Capture d’écran de l’interface utilisée pour collecter les annotations des apprenants de FLE. Lorsqu’un mot est déclaré comme inconnu, il est surligné en rouge.

Les allophones ayant participé aux expériences sont tous des Belges néerlandophones suivant ou ayant suivi la filière générale de l'enseignement libre en Flandre et ayant un niveau en français estimé comme A2 ou B1 sur l'échelle du CECR.⁴ Nous n'avons ciblé que ces deux niveaux de façon à ce que le lexique des textes de notre échantillon ne soit ni trop difficile, ni trop facile pour nos sujets. Nous avons fait appel à quatre néerlandophones, ayant tous fréquenté la même école secondaire. En ce qui concerne le niveau A2, une élève en deuxième année de latin (nommée ci-après l'apprenant A2-2) et un élève en troisième année de sciences (nommé ci-après l'apprenant A2-3) ont participé à l'expérience. Quant au niveau B1, nous avons fait annoter les textes par un élève en quatrième année de latin (nommé ci-après l'apprenant B1-4) et par un étudiant en première année à l'université (nommé ci-après l'apprenant B1-U), n'ayant plus suivi de cours de français depuis la dernière année de l'école secondaire.

2.3 Modélisations lexicales

Nous avons abordé la modélisation de la connaissance lexicale en FLE comme une tâche de classification binaire des unités lexicales d'un texte. Nous avons élaboré trois modèles différents qui classent les unités lexicales des textes en *connu* ou en *inconnu* (Section 2.3.1). Ils se différencient par le public modélisé (groupe d'apprenants d'un niveau donné du CECR ou individu unique) et par la méthode d'entraînement (à base de connaissances symboliques ou de manière supervisée et incrémentale) (Section 2.3.2).

2.3.1 Définition des trois modèles de la connaissance lexicale

Modèle expert (M_E) Notre premier modèle adopte une approche symbolique et estime la compétence lexicale moyenne des apprenants d'un niveau du CECR donné. Ce modèle, nommé ci-après le *modèle expert* (M_E), se repose sur la connaissance lexicale capturée par la ressource FLELex afin de prédire les mots connus/inconnus des apprenants d'un niveau donné. Dans FLELex, chaque entrée lexicale est associée à sa fréquence d'observation dans chacun des six niveaux du CECR, fréquence qui a été estimée sur un corpus de manuels de FLE. En partant de cette distribution, notre modèle expert détermine d'abord quel est le niveau CECR du mot à classer en choisissant le niveau du CECR où le lemme apparaît pour la première fois. Ensuite, le modèle, qui est informé du niveau de l'apprenant, le compare au niveau qu'il vient d'attribuer au mot. Si le niveau du mot est inférieur ou égal au niveau de l'apprenant, le mot est considéré comme étant connu par l'apprenant. Dans le cas inverse, le modèle expert estime que le mot est inconnu. Enfin, tout mot absent de FLELex sera également supposé inconnu par l'apprenant, puisqu'il n'apparaît pas dans les manuels de FLE à partir desquels FLELex a été établi.

Modèle du niveau (M_L) Nous utiliserons le modèle symbolique présenté ci-dessus comme référence (baseline) par rapport à laquelle nous évaluerons les prédictions de deux modèles statistiques. Le premier de ces deux modèles constitue lui aussi une modélisation de la connaissance des apprenants d'un même niveau de maîtrise. Cependant, contrairement au modèle précédent, celui-ci se fonde sur des données a posteriori de la compétence lexicale. Nous l'appellerons le *modèle du niveau* (M_L)

4. Le niveau de maîtrise de chaque participant a été évalué sur la base des objectifs finaux du programme d'études de l'enseignement secondaire en Flandre. Les élèves des 2e et 3e années sont classés A2 et ceux de 4e, 5e et 6e années ont été classés comme B1.

parce qu'il se fonde sur la moyenne de toutes les annotations faites par les apprenants du même niveau, qui a été arrondie afin d'obtenir une valeur binaire, où le 0 correspond aux mots connus et le 1 aux mots inconnus.

Modèle personnalisé (M_P) Le deuxième des modèles statistiques de la connaissance lexicale, nommé ci-après le *modèle personnalisé* (M_P) ou encore le *modèle de l'apprenant*, repose uniquement sur l'annotation de la connaissance lexicale faite par un apprenant individuel. Contrairement au modèle précédent, ce modèle-ci vise dès lors à prédire la compétence lexicale d'un apprenant à partir de ces connaissances personnelles et non à partir de la moyenne du niveau de maîtrise.

Par ailleurs, ces deux derniers modèles statistiques utilisent le même ensemble de variables linguistiques. L'ensemble de ces prédicteurs a été établi en s'inspirant des travaux de Gala *et al.* (2014). Chacune des observations des modèles M_L et M_P est dès lors caractérisée par un ensemble de traits lexicaux qui comprend la catégorie du discours (une variable muette par catégorie), les fréquences de la forme attestées au sein de chacun des six niveaux du CECR (tirées de FLELex), son premier niveau d'attestation dans FLELex, le nombre de lettres, le nombre de voisins orthographiques dans *Lexique3* (New *et al.*, 2007), le nombre de patterns orthographiques complexes et le nombre d'acceptations différentes listées dans *BabelNet* (Navigli & Ponzetto, 2010). Ces indices ont été calculés pour le mot w à prédire, de même que pour son contexte immédiat, à savoir le mot précédent w_{-1} et le mot suivant w_{+1} , ce qui donne un total de 28 prédicteurs par observation.

2.3.2 Apprentissage incrémental des modèles statistiques

Afin de créer des modèles capables d'adapter leurs prédictions de la compétence lexicale d'un apprenant au fil du temps, nous avons employé une méthode par apprentissage automatique incrémental. Contrairement à l'apprentissage par lot (i.e. *batch learning*), nos modèles adaptatifs traitent les données d'entraînement en plusieurs passes (c'est-à-dire par mini-lots), en mettant à jour les paramètres des modèles via une descente de gradient stochastique (SGD). Nous avons opté pour deux types de modèles : (a) un modèle linéaire qui a pris la forme d'une machine à vecteurs de support (SVM) à noyau linéaire et (b) un modèle non-linéaire sous la forme d'un réseau de neurones artificiel (ANN).⁵

Pour l'entraînement des SVM et des ANN, nous disposions d'un jeu de données de 21 047 observations chacune caractérisée par 28 variables explicatives, ainsi qu'une variable à expliquer (respectivement Y_P ou Y_L). Ces deux variables dépendantes correspondent respectivement à l'annotation faite par un apprenant (Y_P) et à l'annotation moyenne faite par tous les apprenants du même niveau (Y_L). De plus, nous disposions également des prédictions \hat{Y}_E déjà effectuées par le modèle baseline M_E . Nous avons partitionné ce jeu de données en deux parties : une partition de développement (50%) et une partition d'évaluation (50%), qui ont été respectivement utilisées pour la sélection des meilleures configurations des modèles et pour l'évaluation des modèles finaux (Section 3).

Sur la première moitié des données, nous avons expérimenté différentes configurations de variables pour les modèles statistiques M_L et M_P en vue de sélectionner les meilleures. Plus précisément, nous avons écarté les caractéristiques du mot qui n'avaient pas d'impact significatif sur la prédiction de la compétence lexicale de l'apprenant concerné (notamment le niveau des mots w_{-1} et w_{+1}). Les

5. Nous avons utilisé deux bibliothèques d'apprentissage automatique disponibles en Python, à savoir *Scikit-learn* (Pedregosa *et al.*, 2011) et *NeuroLab* (Evgenij, 2011).

variables sélectionnées par chacun des modèles sont rapportées dans le **Tableau 2**. Ensuite, nous avons effectué une validation croisée à 10 plis afin de sélectionner la meilleure configuration pour le SVM et pour l'ANN, c'est-à-dire celle qui obtient la plus grande exactitude après avoir traité le dernier mini-lot.

| | A ⁱ | Adv ⁱⁱ | N ⁱⁱⁱ | V ^{iv} | G ^v | n ^{vi} | r ^{vii} | l ^{viii} | v ^{ix} | p ^x | s ^{xi} | w_{-1} | | | | w_{+1} | | |
|------|----------------|-------------------|------------------|-----------------|----------------|-----------------|------------------|-------------------|-----------------|----------------|-----------------|----------|---|---|---|----------|---|---|
| | | | | | | | | | | | | l | v | p | s | l | v | s |
| A2-2 | ✓ | | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| A2-3 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| B1-4 | | | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | | | | ✓ | ✓ |
| B1-U | ✓ | | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | | | | ✓ | ✓ |

ⁱ catégorie adjectivale ⁱⁱ catégorie adverbiale ⁱⁱⁱ catégorie nominale ^{iv} catégorie verbale ^v mot grammatical ^{vi} premier niveau d'attestation dans FLELex ^{vii} fréquences par niveau dans FLELex ^{viii} nombre de lettres orthographiques ^{ix} nombre de voisins orthographiques ^x nombre de patterns orthographiques ^{xi} nombre de polysémies

TABLE 2: Sélection des variables explicatives les plus significatives pour chaque modèle M_P ($p < 0,0001$), après sous-échantillonnage aléatoire des données (voir section 3.1).

3 Évaluation

Afin de vérifier notre hypothèse de recherche qu'un modèle adaptatif est plus prédictif de la compétence lexicale de FLE qu'un modèle expert, nous avons comparé les performances des deux modèles adaptatifs M_P et M_L par rapport à la baseline M_E . À cette fin, nous avons effectué une deuxième validation croisée à 10 plis sur la partition d'évaluation. À chacun des 10 plis, celle-ci a été divisée en un jeu d'entraînement des modèles retenus (90% de la partition d'évaluation) et un jeu de test (10%). En outre, pour l'entraînement, nous avons également ajouté aux 9 plis, l'ensemble des données utilisées lors du développement (Section 2.3.2). Cela nous a permis de disposer de davantage de données d'entraînement, tout en nous assurant que les modèles finaux n'aient jamais rencontré les données de test précédemment.

3.1 Performance et évolution incrémentale des modèles adaptatifs M_P et M_L

Après avoir calculé l'exactitude moyenne pour les prédictions \hat{Y}_E , \hat{Y}_L et \hat{Y}_P en référence aux connaissances lexicales réelles Y_P d'un apprenant donné, nous avons d'abord voulu déterminer l'algorithme adaptatif le plus efficace parmi les deux techniques testées. Il s'est avéré que les réseaux de neurones présentaient des performances moyennes systématiquement supérieures à celles des machines à vecteurs de support et ceci pour tous les modèles entraînés (1,16% de gain en moyenne). Il semblerait dès lors que le modèle non-linéaire soit plus favorable au problème de la modélisation de la compétence lexicale d'un apprenant. Par conséquent, nous rapporterons uniquement les scores d'exactitude des modèles entraînés au moyen des réseaux de neurones dans la suite de cette section.

De surcroît, l'examen de l'exactitude des modèles adaptatifs nous a également amenés à appliquer un meilleur échantillonnage des données. En effet, nous avons observé que les apprenants n'ont annoté qu'un pourcentage relativement restreint de mots inconnus (à savoir entre 3% et 12%). Cette observation est logique dans le contexte de la lecture en L2, notamment en regard des travaux de Hu

& Nation (2000) ou de Laufer & Ravenhorst-Kalovski (2010) cités précédemment. Or, cette réalité entraîne un problème important et bien connu dans les domaines de l'apprentissage automatique, à savoir celui du jeu de données déséquilibré (Chawla, 2005). En conséquence, les scores obtenus par les modèles adaptatifs (**Tableau 3**) ne correspondent pas à des prédictions de qualité, comme le révèlent les scores faibles pour la classe positive (inconnu) (F_1 -mesure entre 0,01 et 0,41).

| | A2-2 | | A2-3 | | B1-4 | | B1-U | |
|-------|------|-------|------|-------|------|-------|------|-------|
| | % | F_1 | % | F_1 | % | F_1 | % | F_1 |
| M_P | 95,8 | 0,01 | 87,6 | 0,37 | 97,0 | 0,08 | 96,7 | 0,04 |
| M_L | 89,5 | 0,25 | 86,8 | 0,41 | 96,8 | 0,11 | 96,6 | 0,08 |
| M_E | 89,9 | 0,25 | 87,2 | 0,40 | 92,2 | 0,22 | 91,9 | 0,21 |

TABLE 3: Exactitude et F_1 -mesure (mots inconnus) sur la partition d'évaluation (données non équilibrées).

Afin de contourner ce problème, nous avons procédé à un ré-échantillonnage de chaque jeu de données pour équilibrer les effectifs des classes à prédire. Dans le domaine de la fouille de données, il est commun d'opérer soit un sur-échantillonnage de la classe minoritaire, soit un sous-échantillonnage de la classe majoritaire. Étant donné que le sur-échantillonnage induirait artificiellement de nouvelles observations de mots inconnus, nous estimions que cette technique n'était pas adaptée à un contexte où l'on voudrait entraîner des modèles à partir d'observations authentiques. Nous avons dès lors opté pour un sous-échantillonnage aléatoire des mots connus. Les résultats obtenus par nos modèles sur ce jeu de données équilibré sont rapportés au **Tableau 4**.

| | A2-2 | | | A2-3 | | | B1-4 | | | B1-U | | |
|-------|------|-------|----------|------|-------|----------|------|-------|----------|------|-------|----------|
| | % | F_1 | κ | % | F_1 | κ | % | F_1 | κ | % | F_1 | κ |
| M_P | 74,6 | 0,75 | 0,49 | 74,9 | 0,72 | 0,50 | 83,0 | 0,83 | 0,66 | 83,1 | 0,82 | 0,66 |
| M_L | 72,4 | 0,74 | 0,45 | 72,2 | 0,71 | 0,44 | 83,8 | 0,84 | 0,67 | 81,7 | 0,80 | 0,63 |
| M_E | 65,7 | 0,54 | 0,31 | 65,3 | 0,52 | 0,31 | 67,6 | 0,55 | 0,34 | 66,3 | 0,51 | 0,32 |

TABLE 4: Exactitude, F_1 -mesure (mots inconnus), et κ de Cohen après sous-échantillonnage.

On note tout d'abord que les modèles statistiques sont devenus nettement plus performants sur la classe minoritaire puisque la F_1 -mesure est cette fois comprise entre 0,71 et 0,84. De surcroît, l'application de tests signés des rangs de Wilcoxon avec correction de continuité indique que les modèles personnalisés M_P présentent un gain d'exactitude considérable par rapport à la baseline M_E et ceci pour tous les sujets (A2-2 : $V = 44$, A2-3 : $V = 55$, $p < 0,05$; B1-4 : $V = 45$, B1-U : $V = 55$, $p < 0,01$)⁶. Ces résultats confirment l'idée qu'une approche basée sur des modèles adaptatifs composés de variables de la compétence lexicale (dont les fréquences issues de FLELex) aboutit à une meilleure prédiction de la compétence lexicale qu'une simple approche symbolique basée sur la ressource FLELex. Quant aux modèles du niveau M_L , par contre, seuls les modèles du niveau B1 s'avèrent significativement plus exacts que la baseline (A2-2 : $V = 44$, $p = 0,10$; A2-3 : $V = 45$, $p = 0,08$; B1-4 : $V = 55$, $p < 0,01$; B1-U : $V = 55$, $p < 0,01$). Cette observation s'explique par le fait que les apprenants du niveau A2 avaient moins de connaissances en commun (coefficient d'accord de Cohen, $\kappa = 0,297$) que les apprenants du niveau B1 ($\kappa = 0,529$) et que les prédictions moyennes

6. Les scores d'exactitude ont été appariés par pli de validation croisée.

par niveau (\hat{Y}_L) s'écartaient dès lors plus de leurs connaissances réelles (Y_P). Il semble donc qu'une modélisation personnalisée soit mieux appropriée pour prédire la compétence lexicale lorsqu'il y a plus de différences entre apprenants d'un même niveau. Notons qu'une étude à plus large échelle devra être menée afin d'en fournir des preuves mieux étayées.

Par contre, deux observations nous empêchent pour l'instant de prouver l'efficacité générale d'un modèle personnalisé par rapport à un modèle stochastique du niveau. Premièrement, sur la base du kappa de Cohen, il s'avère que – contrairement à la baseline M_E qui s'accorde pauvrement avec la compétence de tous les sujets (κ entre 0,31 et 0,34) – les modèles statistiques M_P et M_L indiquent un accord différent selon le niveau de maîtrise (**Tableau 4**). En effet, alors qu'ils obtiennent un accord substantiel pour le niveau B1 (κ entre 0,63 et 0,67), ils n'obtiennent qu'un accord modéré pour le niveau A2 (κ entre 0,44 et 0,50). Cette observation peut s'expliquer par le fait que les annotations faites par les apprenants A2 étaient plus inconsistantes. De fait, ils ont annoté entre 4% (A2-2) et 7% (A2-3) d'unités lexicales tantôt comme connues, tantôt comme inconnues, contrairement aux apprenants de B1 qui en annotent seulement de 1% (B1-4) à 2% (B1-U). Bien que ce problème soit problématique pour la modélisation, il s'agit d'un phénomène bien connu en acquisition des langues. Comme Laufer (1997) l'affirme, l'apprenant peut toujours se tromper et peut croire savoir le sens de certains mots « trompeusement évidents » ou peut avoir inféré le sens d'un mot auparavant inconnu à partir du contexte. Il s'agit par conséquent d'un phénomène qu'il faudrait pouvoir gérer lorsqu'on modélise la compétence lexicale à partir de données authentiques.

On observe encore que le modèle de l'apprenant M_P n'est pas significativement plus exact que le modèle moyen du niveau M_L (A2-2 : $V = 32$, $p = 0,68$; A2-3 : $V = 32$, $p = 0,68$; B1-4 : $V = 11,5$, $p = 0,40$; B1-U : $V = 35$, $p = 0,15$), ce qui s'explique vraisemblablement par la nature des variables qui composent les modèles. De même que le modèle M_L , le modèle M_P effectue ses prédictions sur la base d'indices objectifs de la complexité lexicale (à savoir la fréquence du mot, etc.). Or, afin d'améliorer les prédictions du modèle personnalisé par rapport au modèle du niveau, il faudrait intégrer davantage de variables lexicales subjectives (ex. la familiarité du mot) et de variables d'apprenant (ex. âge et langue maternelle) qui nous permettraient d'enrichir le modèle personnalisé. Par ailleurs, la saturation de l'information véhiculée par les indices lexicaux objectifs se manifeste également dans la stagnation de la performance des modèles M_P au cours du temps (**Figure 2**).

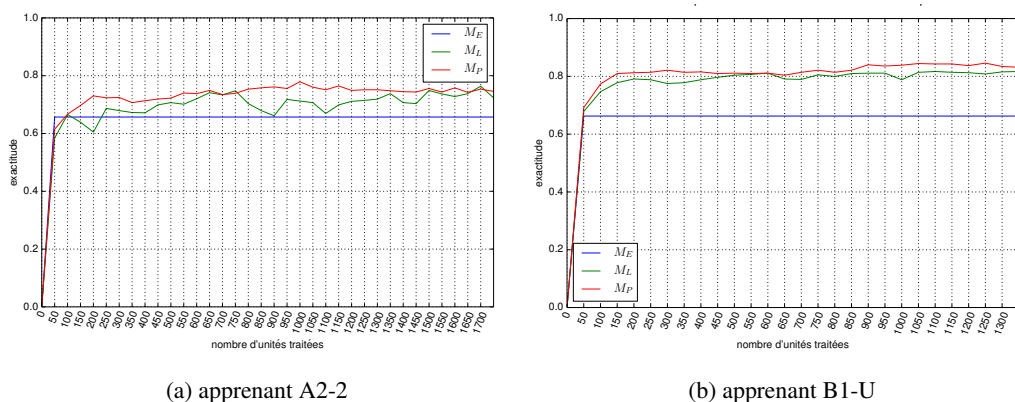


FIGURE 2: L'évolution incrémentale de la performance de M_P par rapport à M_L et M_E (50 unités lexicales/mini-lot). Nous présentons les limites inférieure (A2-2) et supérieure (B1-U) de performance.

3.2 Implications linguistiques

Dans ce qui précède, nous avons constaté que l'intégration de prédicteurs de la compétence lexicale dans un modèle adaptatif permettait d'atteindre une bonne exactitude. Néanmoins, nous avons également relevé plusieurs possibilités d'amélioration des modèles, notamment en ce qui concerne la prise en compte de caractéristiques individuelles des apprenants. Ci-après, nous illustrerons les principales implications de notre étude, ainsi que les principaux points d'amélioration que nous nous proposons d'aborder dans le futur. Ces exemples correspondent aux sorties réelles de nos modèles de l'apprenant M_P sous-échantillonnés et évalués sur le jeu de test. Dans chaque exemple, nous indiquerons en police normale les mots annotés comme connus par l'apprenant et **en gras** les mots annotés comme inconnus par l'apprenant. Ensuite, nous mettrons en couleur les prédictions de chaque modèle sur le jeu de test : en vert les classifications correctes (i.e. **vrais inconnus** et **vrais connus**) et en rouge les erreurs (i.e. **faux inconnus** et **faux connus**).

Dans un premier temps, nous avons observé que la classification correcte des mots réellement connus par l'apprenant concernait surtout les mots de base (1) et les mots grammaticaux (2).

- (1) Maître Cornille ! Maître Cornille ! Ouvrez-moi la **porte**, s'il vous plaît. (*Contes simplifiés, Le secret de maître Cornille*, apprenant B1-U)
- (2) Autrefois, **il** était amoureux de sa fille, mais le temps a passé et, maintenant, il l'aime comme un père. (*Contes simplifiés, Peau d'âne*, apprenant B1-U)

Quant aux mots faussement classifiés comme inconnus, une part importante de ceux-ci s'expliquent parce que nos modèles ne permettent pas encore de prendre en compte l'effet du transfert des connaissances à partir de la langue maternelle (L1) (en l'occurrence, le néerlandais) sur la compétence lexicale d'un apprenant. Ainsi, les modèles ont tendance à classer des mots tels que *ingénieur* (nl. *ingenieur*) ou *insectes* (nl. *insecten*) comme inconnus sur la base de nos indices de complexité, car ils ne sont pas informés qu'un transfert lexical peut s'opérer chez nos apprenants depuis la L1 (3) (4). Dès lors, l'intégration d'une variable « langue maternelle » et d'un dictionnaire de congénères dans M_P nous permettra certainement d'en améliorer les performances.

- (3) Il était à la fois physicien, mathématicien et **ingénieur**. (*Vikidia, Archimède*, apprenant A2-2)
- (4) Les alligators se **nourrissent** d'**insectes**, de poissons, de **grenouilles**, de **crustacés** ou d'escargots. (*Vikidia, Alligator*, apprenant A2-3)

Parmi les mots correctement classifiés comme inconnus, l'on remarque notamment la présence des mots techniques comme *ogives*, *arcs-boutants* (5) ou encore *aéronefs* (6). L'identification de ces lieux de difficulté s'explique vraisemblablement par l'utilisation des variables fréquentielles à partir d'une ressource lexicale spécialisée pour le FLE telle que FLELex. En effet, les mots absents de la ressource, comme (5) et (6), se voient attribuer, dans nos modèles, un niveau CECR élevé (« absent ») et une fréquence nulle. Les variables telles que le « premier niveau d'attestation » ont permis de bien identifier l'absence de connaissance de ces mots-là.

- (5) L'architecture gothique dispose de nouvelles techniques : des arcs **brisés** et des **arcs-boutants**. (*Vikidia, Architecture gothique*, apprenant B1-4)

- (6) On distingue deux catégories d'**aéronefs** : l'**aérostat** et l'**aérodyn**e. (*Wikipédia, Aéronef*, apprenant B1-4)

Toutefois, il importe de noter que le fait que les indices de complexité lexicale se calculent sur la base du lemme et non sur la base de la forme fléchie des mots constitue une autre source importante d'erreurs. De ce fait, les modèles ne sont pas capables de distinguer la difficulté de certaines formes fléchies d'un même lemme verbal. En effet, il s'est avéré que certaines formes fléchies non connues d'un verbe, comme le passé simple du verbe *être* (7), n'ont pas été reconnues par certains apprenants, alors qu'ils ont annoté le présent du même verbe comme connu (8). Or, sur la simple base des indices de complexité, les modèles n'étaient pas capables d'identifier les difficultés liées à l'âge d'acquisition du passé simple par les apprenants (à savoir la quatrième année du secondaire).

- (7) L'empereur Charlemagne **fut** enterré **dans** la **chapelle palatine** en 824. (*Vikidia, Aix-la-Chapelle*, apprenant A2-3)
- (8) Son nom **savant est** l'"**Aquila chrysaetos**". (*Vikidia, Aigle royal*, apprenant A2-3)

4 Conclusion et perspectives

Cette étude nous a permis d'évaluer l'efficacité d'une modélisation adaptative de la compétence lexicale d'apprenants de français langue étrangère (FLE). La compétence lexicale modélisée portait sur le rappel des mots en lecture par quatre apprenants néerlandophones ayant un niveau A2/B1 selon le Cadre européen commun de référence (CECR). Nous avons décrit deux modèles qui reposaient soit sur la compétence lexicale en moyenne de tous les apprenants du même niveau de maîtrise (modèle du niveau), soit sur la compétence lexicale d'un apprenant individuel (modèle personnalisé). Lors de l'entraînement incrémental de ces modèles, il est apparu que les réseaux de neurones artificiels étaient plus robustes qu'un classifieur linéaire par SVM. En outre, nous avons également observé l'intérêt de sous-échantillonner la classe majoritaire de mots connus afin de mieux prédire la connaissance lexicale. Enfin, nous avons observé un gain d'exactitude significatif des modèles personnalisés par rapport à un modèle de référence, ce qui a permis de confirmer notre hypothèse centrale. Toutefois, une étude à plus large échelle devra être menée afin de préciser de mieux les comparer aux modèles du niveau.

Nous envisageons plusieurs perspectives pour rendre plus précise la prédiction de la compétence lexicale d'un apprenant de FLE. Une première démarche concernera l'intégration de variables lexicales subjectives (telles que la familiarité du mot) et de variables d'apprenant (telles que sa langue maternelle) dans le modèle de l'apprenant, s'appuyant sur les travaux sur l'identification automatique de congénères (Mitkov *et al.*, 2008). De cette manière, nous serons capables de personnaliser davantage le modèle de l'apprenant et de mieux en comparer l'efficacité par rapport à un modèle moyen par niveau de maîtrise. Une autre perspective consistera à différencier les différents sens des unités lexicales dans le texte. En effet, les modèles élaborés dans cet article ne se basent que sur la forme et se limitent donc à la modélisation des aspects formels de la compétence lexicale. Dès lors, en désambiguïsant sémantiquement les lemmes, nous serons mieux en mesure de prendre en compte la compétence lexicale d'un apprenant de FLE.

Remerciements

Nous remercions les participants à l'expérience pour le temps qu'ils ont consacré à donner un retour essentiel sur leur compétence lexicale, ainsi que la professeure qui nous a mis en contact avec eux.

Références

- BOTTOU L. (1998). Online Learning and Stochastic Approximations. In D. SAAD, Ed., *Online Learning and Neural Networks*. Cambridge, UK : Cambridge University Press. révisé, octobre 2012.
- BROUWERS L., BERNHARD D., LIGOZAT A.-L. & FRANÇOIS T. (2014). Syntactic Sentence Simplification for French. In *Proceedings of the Third International Workshop on Predicting and Improving Text Readability for Target Reader Populations (PITR'14)*, p. 47–56 : Association for Computational Linguistics.
- CHAWLA N. V. (2005). Data Mining for Imbalanced Datasets : An Overview. In O. MAIMON & L. ROKACH, Eds., *Data Mining and Knowledge Discovery Handbook*, p. 853–867. New York : Springer.
- CHEN C.-M. & CHUNG C.-J. (2008). Personalized mobile English vocabulary learning system based on item response theory and learning memory cycle. *Computers & Education*, **51**(2), 624–645.
- EVGENIJ Z. (2011). NeuroLab : a simple and powerful Neural Network Library for Python. <http://pythonhosted.org/neurolab/>.
- FERRAND L. (2007). *Psychologie cognitive de la lecture*. Bruxelles : De Boeck.
- FRANÇOIS T. & FAIRON C. (2013). Les apports du TAL à la lisibilité du français langue étrangère. *Traitement Automatique des Langues (TAL)*, **54**(1), 171–202.
- FRANÇOIS T., GALA N., WATRIN P. & FAIRON C. (2014). FLELex : a graded Lexical Resource for French Foreign Learners. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, p. 3766–3773, Reykjavik, Iceland : European Language Resources Association (ELRA).
- GALA N., FRANÇOIS T., BERNHARD D. & FAIRON C. (2014). Un modèle pour prédire la complexité lexicale et graduer les mots. In *Actes de la 21e Conférence sur le Traitement Automatique des Langues Naturelles (TALN 2014)*, p. 91–102, Marseille.
- GYLLSTAD H. (2013). Looking at L2 Vocabulary Knowledge Dimensions from an Assessment Perspective – Challenges and Potential Solutions. In C. BARDEL, B. LAUFER & C. LINDQVIST, Eds., *L2 vocabulary acquisition, knowledge and use : New perspectives on assessment and corpus analysis (Eurosla Monographs Series, 2)*, p. 11–28. European Second Language Association.
- HEILMAN M., COLLINS-THOMPSON K., CALLAN J. & ESKENAZI M. (2007). Combining lexical and grammatical features to improve readability measures for first and second language texts. In *Proceedings of NAACL HLT*, p. 460–467.
- HU M. & NATION P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a foreign language*, **13**(1), 403–30.
- KRASHEN S. (1989). We acquire vocabulary and spelling by reading : Additional evidence for the input hypothesis. *The Modern Language Journal*, **73**(4), 440–464.

- LAUFER B. (1997). The lexical plight in second language reading : Words you don't know, words you think you know, and words you can't guess. In J. COADY & T. HUCKIN, Eds., *Second Language Vocabulary Acquisition : A Rationale for Pedagogy*, p. 20–34. Cambridge, UK : Cambridge University Press.
- LAUFER B. & RAVENHORST-KALOVSKI G. (2010). Lexical Threshold Revisited : Lexical Text Coverage, Learners' Vocabulary Size and Reading Comprehension. *Reading in a foreign language*, **22**(1), 15–30.
- LÉTÉ B., SPRENGER-CHAROLLES L. & COLÉ P. (2004). Manulex : A grade-level lexical database from French elementary-school readers. *Behavior Research Methods, Instruments and Computers*, **36**, 156–166.
- LICHELLI O., BASILE T., DI MAURO N., ESPOSITO F., SEMERARO G. & FERILLI S. (2004). Machine Learning Approaches for Inducing Student Models. In B. ORCHARD, C. YANG & M. ALI, Eds., *Innovations in Applied Artificial Intelligence (Lecture Notes in Computer Science, 3029)*, p. 935–944. Berlin : Springer.
- MITKOV R., PEKAR V., BLAGOEV D. & MULLONI A. (2008). Methods for extracting and classifying pairs of cognates and false friends. *Machine Translation*, **21**(1), 29–53.
- NATION I. S. P. (2001). *Learning Vocabulary in Another Language*. Cambridge : Cambridge University Press.
- NAVIGLI R. & PONZETTO S. P. (2010). BabelNet : Building a very large multilingual semantic network. In *Proceedings of the 48th annual meeting of the association for computational linguistics*, p. 216–225.
- NEW B., BRYLSBAERT M., VERONIS J. & PALLIER C. (2007). The use of film subtitles to estimate word frequencies. *Applied Psycholinguistics*, **28**(04), 661–677.
- PEDREGOSA F., VAROQUAUX G., GRAMFORT A., MICHEL V., THIRION B., GRISEL O., BLONDEL M., PRETTENHOFER P., WEISS R., DUBOURG V., VANDERPLAS J., PASSOS A., COURNAPEAU D., BRUCHER M., PERROT M. & DUCHESNAY E. (2011). Scikit-learn : Machine Learning in Python. *Journal of Machine Learning Research*, **12**, 2825–2830.
- PILÁN I., VOLODINA E. & JOHANSSON R. (2014). Rule-based and machine learning approaches for second language sentence-level readability. In *Proceedings of the Ninth Workshop on Innovative Use of NLP for Building Educational Applications*, p. 174–184.
- SCHMID H. (1994). Probabilistic Part-of-Speech Tagging Using Decision Trees. In *Proceedings of International Conference on New Methods in Language Processing*, volume 12, p. 44–49.
- SCHMITT N. (1998). Tracking the Incremental Acquisition of Second Language Vocabulary : A Longitudinal Study. *Language Learning*, **48**(2), 281–317.
- SHARDLOW M. (2013). A Comparison of Techniques to Automatically Identify Complex Words. In *ACL Student Research Workshop*, p. 103–109.
- STOCKWELL G. (2007). Vocabulary on the move : Investigating an intelligent mobile phone-based vocabulary tutor. *Computer Assisted Language Learning*, **20**(4), 365–383.
- TACK A., FRANÇOIS T., LIGOZAT A.-L. & FAIRON C. (2016). Evaluating Lexical Simplification and Vocabulary Knowledge for Learners of French : Possibilities of Using the FLELex Resource. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, p. 230–236 : European Language Resources Association (ELRA).
- WESCHE M. B. & PARIBAKHT T. S. (1996). Assessing second language vocabulary knowledge : Depth versus breadth. *The Canadian Modern Language Review*, **53**(1), 13–40.