

Apports du TAL à la prédiction automatique de la difficulté lexicale dans le cadre de l'aide à la lecture

Núria Gala, Mokhtar B. Billami, Thomas François

Séminaire TALEP, Aix Marseille Univ, CNRS, LIF

27 juin 2016

Un séminaire à trois voix...

- **Contexte** : ressources lexicales pour l'apprentissage du vocabulaire et l'aide à la lecture
- Construction automatisée de **ReSyf**, un lexique de synonymes désambiguïsés (et triés)
- Méthodes TAL pour la **prédiction automatisée de la difficulté lexicale**

Contexte

- Lire un texte afin de le comprendre > activité cognitive très complexe, spécialement pour certains publics ciblés
 - **Dyslexiques**, faibles lecteurs (5 à 15% des enfants)
 - **Malvoyants** souffrant de dégénérescence maculaire liée à l'âge (1,3 millions de personnes)
 - **Illettrés** (7% de la population adulte)
 - **Sourds** (avec LSF comme langue maternelle)
 - ...
- Les différentes étapes de la lecture ont été largement décrites et validées par les psycholinguistes, i.e. la reconnaissance de mots [Ferrand and Ayora, 2009]
- Moins de travaux : façon de **remédier aux difficultés de lecture...**

Apports du Traitement Automatique des Langues

Puisque les mots sont au cœur de la pratique et de l'apprentissage de la lecture, et que l'un des fondements du TAL est le traitement du lexique, **le TAL peut apporter** :

- des **lexiques spécifiques**
- des méthodes pour l'analyse de la **lisibilité computationnelle**
- des outils de **simplification de textes** (niveaux lexical, syntaxique, discursif)

★ Objectifs : rendre la lecture accessible et plaisante, faciliter l'apprentissage du vocabulaire et de la lecture (publics différents)

Lexiques, vocabulaires, inventaires de mots...

- De nombreuses listes pour maîtriser les "mots difficiles" dans l'apprentissage d'une langue
- **Vocabulaires de base** (listes de termes fondamentaux estimés selon fréquence, familiarité/disponibilité...) vs **lexiques gradués** (indication explicite de la difficulté)
- Deux approches méthodologiques :
 - Vocabulaires "logiques" (sélection de mots par des procédés rationnels c.à.d. **subjectifs**) [Suso Lopez, 2006]
 - Extraction sur **corpus** (productions écrites d'enfants ou textes de lecture (manuels scolaires, littérature, etc.) avec **niveaux identifiés** d'avance), statistiques sur les formes (fréquence, distribution...)

Exemples

Ressource	Auteur(s)	Nbre entrées
<i>Basic English</i>	[Ogden, 1930]	850
<i>The Teacher's Word Book</i>	[Thorndike, 1921]	10 000
<i>Le Français Fondamental</i>	[Gougenheim, 1958]	7 995
<i>Les Listes orthographiques de Base</i>	[Catach, 1984]	2 353
Manulex (L1)	[Lété et al., 2004]	23 812
FLELex (L2)	[François et al., 2014]	17 871
CASSAurus	[Baeza-Yates et al., 2016]	9 928
ReSyf	[Gala et al., 2013]	19 397 (*)

TABLE : Vocabulaires simplifiés (listes de base) et ressources lexicales graduées (intégrant la notion de difficulté lexicale)

(*) En préparation : [Gala et al., 2017]

-  Baeza-Yates, R., Debowski, J., and Rello, L. (2016).
CASSAurus : a resource of simpler Spanish synonyms.
In Language Resources and Evaluation Conference, LREC-2016.
-  Catach, N. (1984).
Les listes orthographiques de base du français.
Nathan. Collection Nathan Recherche, Paris.
-  Ferrand, L. and Ayora, P. (2009).
Psychologie cognitive de la lecture.
De Boeck, Bruxelles.
-  François, T., Gala, N., Watrin, P., and Fairon, C. (2014).
FLELex : a graded lexical resource for French foreign learners.
In Proceedings of LREC 2014, Reykjavik, Islande.
-  Gala, N., François, T., Billami, M. B., and Bernhard, D. (2017).
ReSyf : a lexical resource with difficulty levels for reading assistance and automated text simplification (*en préparation*).
Behavior Research Methods, Instruments, and Computers, Springer Journals, xx.
-  Gala, N., François, T., and Fairon, C. (2013).
Towards a French lexicon with difficulty measures : NLP helping to bridge the gap between traditional dictionaries and specialized lexicons.
In E-lexicography in the 21st century : thinking outside the paper, Tallin, Estonie.
-  Gougenheim, G. (1958).
Dictionnaire fondamental de la langue française.
Didier, Paris.



Lété, B., Sprenger-Charolles, L., and Colé, P. (2004).
Manulex : A grade-level lexical database from French elementary-school readers.
Behavior Research Methods, Instruments and Computers, 36 :156–166.



Ogden, C. K. (1930).
Basic English : A General Introduction with Rules and Grammar.
Paul Treber, London.



Suso Lopez, J. (2006).
Vocabulaires logiques, vocabulaires simplifiés et français élémentaire.
Documents pour l'histoire du français langue étrangère ou seconde [En ligne], 36.



Thorndike, E. (1921).
The Teacher's Word Book.
Teachers College, Columbia University, New York.