Extraction des expressions phraséologiques et semi-phraséologiques par NooJ

Tong YANG

Deuxième année de doctorat à l'Université de Sorbonne Nouvelle Paris 3, ED 268

Sous la co-direction de C. Cavalla (Paris 3) et de J.M. Debaisieux (Paris 3)

International NooJ 2017 Conference, Kenitra, les 18, 19, 20 Mai, 2017

Cadre et objectif

- Cadre : FOS
- Domaine : culinaire
- Expressions phraséologiques : N+ADJ
- Outil d'extraction : NooJ

1. Corpus

Cuisitext (Yang, 2016)

Corpus écrits

Sites culinaires français	
Nom du site	URL
Marmiton	http://www.marmiton.org
750g	http://www.750g.com
Cuisine AZ	http://www.cuisineaz.com
Ôdélice	http://www.odelices.com
Cuisine actuelle	http://www.cuisineactuelle.fr

Figure 1 : Partie écrite de *Cuisitext*

Corpus oraux

Corpus oral	Nombre de vidéos	Mins/Vidéo	Année	Source
Clips vidéos culinaires sur internet	Cent	5	Après 2010	Chaines de télévision
Filmages dans une école hôtelière	Cinq	1	2016	Mangiante J.M. (Université d'Artois)
Filmages dans deux cuisines	Vingt	60	2015	Nous même

Figure 2 : Partie orale de *Cuisitext*

Yang, T., (2016). *Cuisitext* : un corpus écrit et oral pour l'enseignement, colloque *LOSP (Langues sur objectifs spécifiques : perspective croisées entre linguistique et didactique)* organisé par les laboratoires LIDILEM et ILCEA4 – GREMUTS) à Grenoble. http://losp2016.u-grenoble3.fr/index.php?pg=4&lg=fr
Mangiante J.M. (2016). Spécialiste du FOS, Université d'Artois.

2. Expressions phraséologiques et semi-phraséologiques

- « La phraséologie d'une langue est l'ensemble de toutes les expressions non libres de cette langue » (Polguère, 2016 : 62).

  Les expressions non libres sont les expressions phraséologiques.

  Exemple: *steak haché*.

- « Dans le segment *AB*, *A* peut être sélectionné de façon libre par le locuteur et *B* de façon contrainte. Dans un tel cas, *AB* sera dit semi-phraséologique » (Polguère 2016 : 64).

  Exemple: *viande hachée*

  la collocation fait partie des expressions semi-phraséologiques.

Polguère, A., (2016). Lexicologie et sémantique lexicale. Notions fondamentales : Troisième édition. Les Presses de l'Université de Montréal.

3. Comparaisons de logiciels d'extraction

	Exemple de logiciels d'extraction de la collocation	Fonctions
Concordancier	*AntConc* (Anthony, 2005)	Index, concordance, n-gram
Textométrie	*Lexico* 3 (Lafon & Salem, 1983)	Segments répétés
Formalisation des langues	*NooJ* (Silberztein, 2004)	Grammaire générative et grammaire transformationnelle

Figure 3 : Comparaisons de logiciels d'extraction

Anthony, L. (2005). AntConc: design and development of a freeware corpus analysis toolkit for the technical writing classroom. In *Professional Communication Conference*, 2005. Proceedings. International, p. 729-737.
Lafon, P. & Salem, A., (1983). L'inventaire des segments répétés d'un texte. *Mots*, 6(1), p.161-177.
Silberztein, M., (2004). NooJ : an oriented object approach. In Royauté, J. & Silberztein, M. (dir.) *INTEX pour la Linguistique et le Traitement Automatique des Langues*. Besançon : Presses universitaires de Franche-Comté.

4. Approche de NooJ

- Système de traitement de corpus qui peut offrir des possibilités pour traiter un corpus pour l'enseignement (Silberztein & Tutin, 2005).
- Quatre grammaires

Figure 4 : Hiérarchie de Chomsky-Schützenberger

Silberztein M. & Tutin, A., (2005). NooJ, un outil TAL pour l'enseignement des langues. Application pour l'étude de la morphologie lexicale en FLEM, *Alsic*, Vol. 8, n° 2.

5. Codage et modélisation

- Propriétés lexicales à la base d'observation :

sur *Cuisitext*
sur les dictionnaires (TLFI et LAROUSSE)

- Fonction lexicale

6. Fonction lexicale

- Outil conceptuel des descriptions des langues en modélisant et encodant les liens paradigmatiques (sémantique) et syntagmatiques (Polguère, 2000).

- Formule mathématique :

$$\mathbf{f(x)=y}$$

x représente l'argument (mot-clé)
y est sa valeur

Par exemple,

PreparFact0(*ail*) = *hacher* [ART ~]
A2PerfPreparFact0(*ail*) = *haché*

Donc, deux structures :

NOM+ADJ (*ail écrasé/coupé/haché/pressé*) ;
NOM+VERBE D'ETAT +ADJ (*ail est écrasé/coupé/haché/pressé*).

Polguère, A., (2000). Towards a theoretically-motivated general public dictionary of semantic derivations and collocations for French. *Proceedings of EURALEX'2000*. Stuttgart. p.517-527.

7. Modélisation lexicale des données

- Pour la première structure, la modélisation est facile : NOM+ <WF> ou <E> + ADJ.
- Pour la deuxième structure, il est impossible de créer tant de modélisations pour toutes les extractions. Par exemple, l'ail est écrasé ; l'ail n'est pas écrasé ; c'est l'ail qui est écrasé ; c'est l'ail qui n'est pas écrasé ; etc.

Constat : ces phrases au-dessus partagent le même matériaux lexical.

Solution : recourir à la grammaire transformationnelle.

- Elle s'intéresse aux relations entre phrases.
- À partir d'une phrase, le générateur de NooJ peut produire toutes ses phrases transformationnelles. Bien entendu, il peut aussi les reconnaître.
- Grâce aux variables globales, la modélisation de la deuxième structure : Sujet + Verbe + ADJ

8. Implémentation de données lexicales dans *NooJ*

- Pour implémenter la modélisation NOM+ <WF> ou <E> + ADJ, nous utilisons la variable N pour que les noms et les adjectifs s'accordent en genre et en nombre.

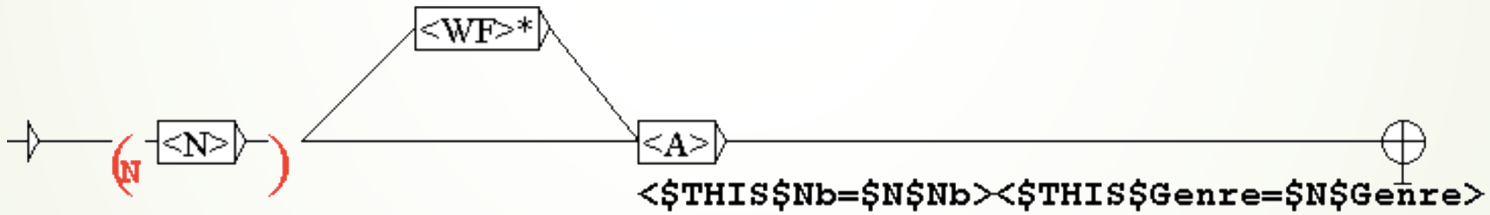

Figure 5 : Implémentation de la première modélisation Nom + <WF> ou <E> + ADJ

<N> : toutes les formes nominales
<WF> (*Word Form*) : une séquence de mots
<A> : les adjectif

8. Implémentation de données lexicales dans NooJ

- Pour implémenter la modélisation : Sujet + Verbe + ADJ, nous recourons aux trois variables globales.

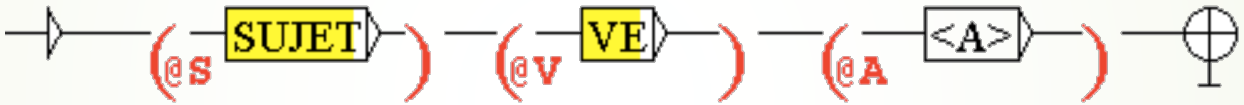

Figure 6 : Implémentation de la modélisation SUJET + VERBE D'ÉTAT + ADJ

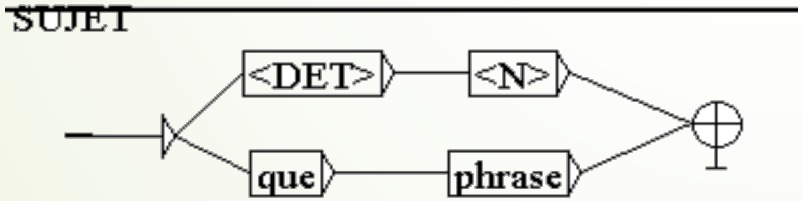

Figure 7 : Graphe imbriqué du sujet

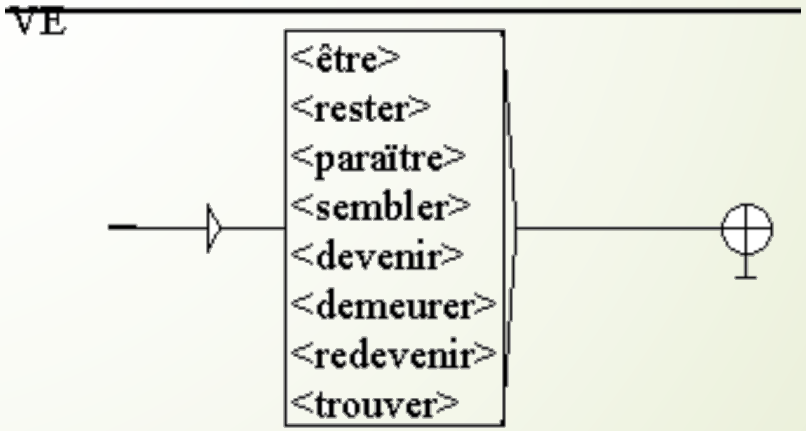

Figure 8 : Graphe imbriqué du verbe d'état

Conclusion

- NooJ peut résoudre notre problème des extractions des expressions phraséologiques et semi-phraséologiques.

- Une fois achevée l'extraction, nous commençons à réfléchir aux aspects didactiques.

Merci !

Tong YANG
tongyangparis3@gmail.com
Université de Sorbonne Nouvelle-Paris 3