



1 Cadre de la recherche



2 Objet d'étude



3 Méthodologie de l'extraction



4 Résultat de l'extraction



5 Conclusion et perspective



1. Cadre de la recherche

- Enseignement du FOS (Mangiante & Parpette, 2004)
- Haute fréquence des routines de type **N+À+N**, ex. *tartiflette à la reine*, *magret de canard aux pêches*, *pain aux céréales*.
- Problématique de recherche :

Comment extraire automatiquement les routines **N+À+N du domaine culinaire à l'aide d'un outil numérique ?**

2. Définition des routines langagières



- ✓ D'après Tutin & Kraif (2016),
 - Elles sont linéaires, avec parfois des « trous » dans les séquences ;
 - Elles comportent des suites de mots, de lemmes ou de traits morphosyntaxiques ;
 - Elles ne constituent pas nécessairement des constituants syntaxiques classiques ;
 - Elles sont caractérisées par une fonction discursive, textuelle, ou rhétorique spécifiques.

- ✓ Difficulté de l'extraction de séquences polylexicales (Luka *et al.*, 2006).

3. Méthodologie de l'extraction

3.1 *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

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	Vingt	3-5	2016	Mangiante J.M. (Université d'Artois)
Filmages dans deux cuisines françaises	Vingt	60	2015	Nous même

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.

3.2 Modélisation lexicale des données

- Inspiré par les projets de Tutin (2004) et de Mel'čuk (1981 ; 1984 ; 1995 ; 1999), notre modélisation :
 - dans notre coprus *Cuisitext*
 - dans les dictionnaires (*TLFI* et *le petit robert* en ligne)

a. Premier exigence : définir un groupe nominal

Modélisation des routines langagières

N + À + N (ex. *pommes de terre entières à la vapeur*)

Grammaires locales du GN

= Déterminant + Adjectif ou pas + Nom + Adjectif ou pas

= GN + à/de + Nom

= que + phrase

b. Deuxième exigence : récursivité de la grammaire

Modélisation des routines langagières

N + **À** + N (ex. pommes de terre entières à la vapeur)

Grammaires locales du GN

= Déterminant + Adjectif ou pas + Nom + Adjectif ou pas

= **GN** + à/de + Nom

= que + phrase

Grammaires locales du GN

= Déterminant + Adjectif ou pas + Nom + Adjectif ou pas

= **GN** + à/de + Nom

= que + phrase

Grammaires locales du GN

= Déterminant + Adjectif ou pas + Nom + Adjectif ou pas

= GN + à/de + Nom

= que + phrase

c. Troisième exigence : désambiguïisation lexicale

- Routines langagières **N + À + N** non-pertinentes, par exemple,

La pâte à l'intérieur

Fumée à l'aide

Sandwich à la forme

désambiguïisation lexicale

```
graph LR; A[La pâte à l'intérieur] --> B[désambiguïisation lexicale]; C[Fumée à l'aide] --> B; D[Sandwich à la forme] --> B;
```

3.3 Choix du logiciel approprié

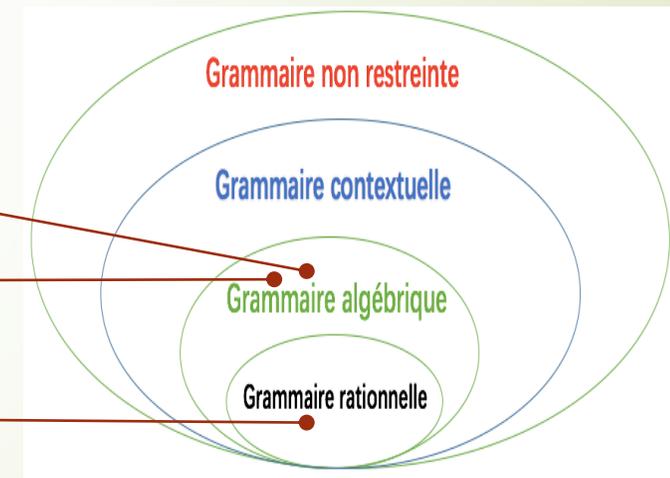
Nos exigences :

Définition d'un groupe nominal

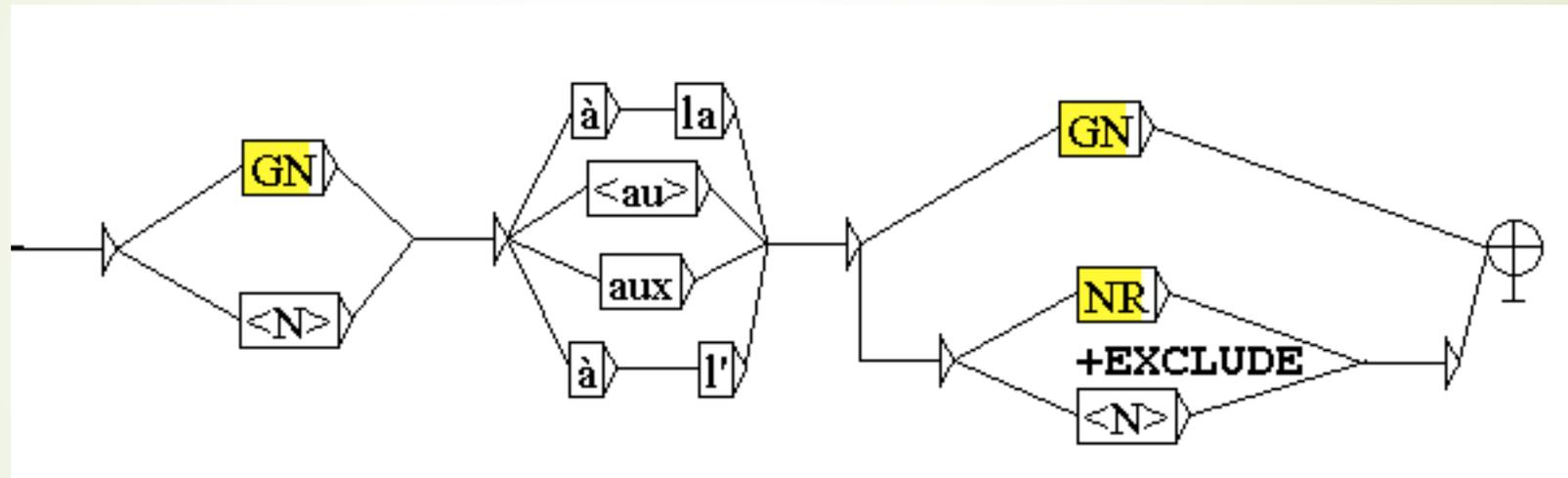
Réversibilité de la grammaire

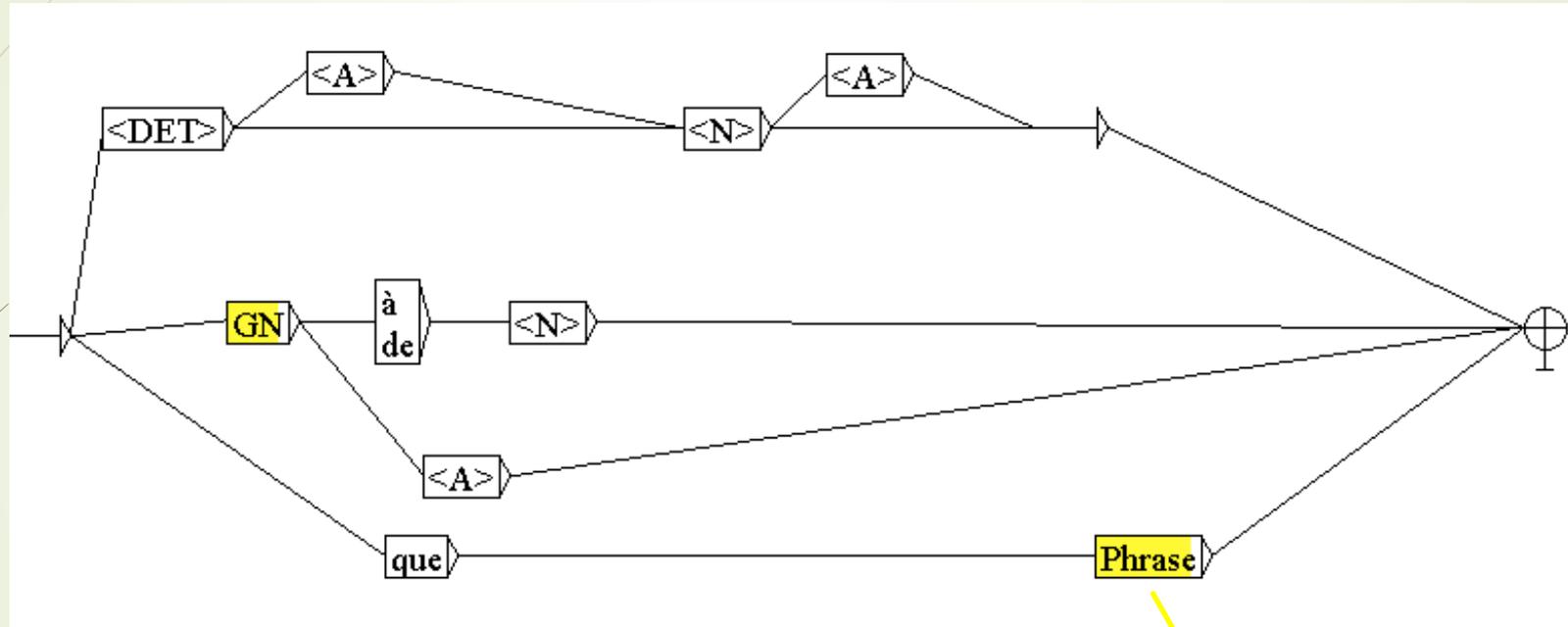
Désambiguïisation lexicale

NooJ :

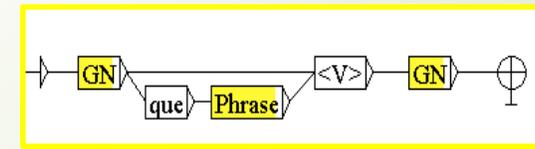


3.4 Implémentation des données lexicales dans *NooJ*





Graphe imbriqué GN



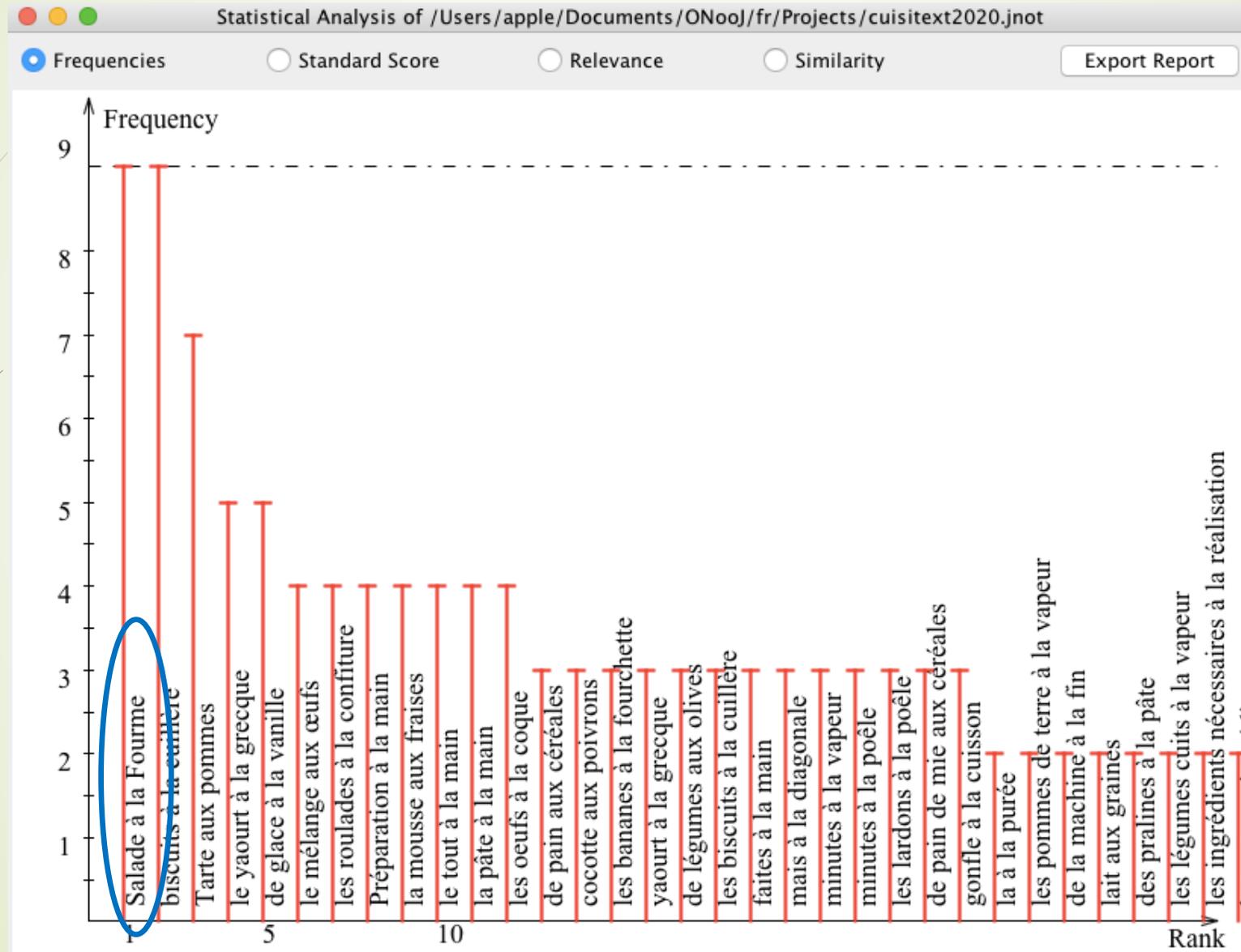
4. Résultat

Concordance for Text cuisitext2020.jnot

Reset Display: characters before, and after. Display: Matches Outputs
 word forms

	Seq.	After
es de saumon fumé. Badigeonnez	le tout de fromage à la ciboulette	. Roulez l'
g de mâche - 1 endive - 2 tranches	de pain de mie aux céréales	Préparati
pointe de sel. Dégustez avec	le pain de mie aux céréales	toasté. T
de mie aux céréales toasté.	Tartiflette à la reine	- 6 bouch
ou vol au vent (pour	bouchée à la reine) - 0,5 rebl
d'olive Préparation: Faites cuire	les pommes de terre entières à la vapeur	, en garda
e semoule - Muscade - Condiment	porto aux fruits	secs - 30
fourchette. Garnissez la pâte avec	le mélange à la viande	hachée. D
la quiche et dégustez-la.	Vol au vent aux courgettes	et pintade
l'oeuf battu et dans	la chapelure aux petites	graines. R
déguster avec une salade verte.	Bricks aux légumes	de soleil e
le persil ciselé et mélangez	cette sauce à la béchamel	. Tartinez
la tomate 1 pâte feuilletée 200 g	de thon à l'huile	3 tomates
rassembler les dés de concombre,	le yaourt à la grecque	, les hachi
finement la sucrine et liez	la à la mayonnaise	. Prélevez
la fraîcheur du foie, moins	la fonte à la cuisson	sera immo

Query 846/846



Conclusion et perspective

- Certaines expressions irrégulières, comme *poivrons grillées à l'huile*, *petits pois à la vapeur*, *olives noires aux pâtes*, *moutarde aux tomates séchées*, *agneau aux légumes et à la bière*, etc.
- Notre sélection des routines langagières N+À+N pour l'enseignement s'arrête au seuil de deux occurrences, car les séquences extraites à une occurrence contiennent des expressions non-pertinentes, comme *mélange aux œufs* et *minute à la vapeur*. Mais nous vérifions les expressions du seuil d'une occurrence pour ne pas rater les expressions pertinentes, par exemple, *rhubarbe à la crème*, *œufs à la framboise* n'apparaissent qu'une fois dans notre corpus, mais elle fait partie des routines langagières N+À+N.

MERCI DE VOTRE ATTENTION

Tong YANG

tong.yang@sorbonne-nouvelle.fr