



LIDILEM

UNIVERSITÉ
Grenoble
Alpes

1

Routines et fonctions rhétoriques dans les écrits scientifiques à partir d'une approche sur corpus

AGNÈS TUTIN, OLIVIER KRAIF (LIDILEM)

Journée d'études

*De la phraséologie aux genres textuels : état des
recherches et perspectives méthodologiques*

Cracovie, 18 octobre 2016

Objectif

2

- Objet d'étude : routines de l'écrit scientifique
- Certaines expressions ou façons de dire sont stéréotypées et relèvent de la phraséologie étendue.
 - Ces routines sont généralement associées à des genres spécifiques, et en particulier à des genres institués (Maingueneau 2004), par exemple les écrits scientifiques ou les écrits professionnels.
 - Elles relèvent plus de l'analyse du discours que de la lexicologie
 - Elles ne relèvent pas de l'anomalie mais plutôt de la régularité
 - Les composantes discursive, rhétorique et énonciative sont essentielles pour ces éléments.

Plan

3

1. Précisions terminologiques et notionnelles
2. Comparaison entre segments répétés et Arbres Lexico-syntaxiques Récurrents (ALR): Etude sur corpus des verbes de communication
3. Extraction et analyse des routines

1. Précisions terminologiques et notionnelles



Un certain flou

5

- Comme pour toute notion linguistique émergente, on observe un certain flou terminologique et notionnel, par exemple :
 - « motifs » (Longrée & Mellet, 2013)
 - « motifs séquentiels » (Quiniou et al., 2012)
 - « routines discursives » (Née, Sitri, Veniard 2014)
 - « segments répétés » (Salem, 1987)
- Quelques précisions terminologiques (Kraif & Tutin, *Corpus*, à paraître)
 - Motifs
 - Routines
 - Arbres Lexico-syntaxiques récurrents

Motifs

6

- Terme utilisé originellement en Tal (Ganascia 2001)
 - **Structures lexico-syntaxiques répétées**, regroupées sous-forme de classes, grâce à un calcul de similarité.
 - Application aux motifs caractéristiques de certains écrits de Madame de Lafayette, comparés à ceux d'autres auteurs
 - Notion développée par D. Longrée et S. Mellet pour la description topologique des textes, pour « l'étude de la structuration interne des textes et à [sa] caractérisation au sein d'un corpus contrastif des expressions récurrentes » (Longrée, Luong & Mellet, 2008, p. 734).
 - Version 1 (2008) : **association récurrente d'éléments dans une structure linéaire** multidimensionnelle, pouvant comprendre des éléments facultatifs.
 - Version 2 (2013) : le « motif textuel » est défini par une **fonction structurante ou discursive**
 - Travaux de Legallois, Charnois, Quiniou et collègues
 - Quiniou *et al.* (2012) : Formalisation de la notion **d'itemset**, combinaison de plusieurs traits issus de dimensions diverses (forme, lemme, classes syntaxique) et intégration de « trous ». Ex : le/la/l' * qui * et *
 - Legallois, Charnois et Poibeau (2016) : Motifs : **séquences récurrentes mêlant lexèmes, catégories syntaxiques, catégories sémantiques**, mais sans « trous ». Etude des clichés typiques des romans sentimentaux , par exemple le patron PRES le NC, il V-PS : « Rouvrant les yeux, elle retint un cri ».

Motifs

7

- Définition de Legallois, Charnois et Poibeau (2016):
 - Quelle que soit la méthode, sont qualifiés de motifs les seuls segments présentant une régularité d'ordre lexical (un même paradigme lexical est employé dans le patron — par exemple, les lexèmes relatifs au corps), et / ou fonctionnel : un motif possède une fonction sémantique, pragmatique, rhétorique discernable, voire une fonction d'organisation des plans textuels (...) (Legallois et al., 2016)
- Définition étendue aux motifs hiérarchiques (configurations lexico-syntaxiques plus que linéaires) :
 - Un motif est une **configuration lexico-syntaxique récurrente** assumant une certaine fonction linguistique, pouvant correspondre à une unité phraséologique (collocations, routines, etc.) ou non, caractéristiques d'un texte ou d'un genre textuel.

Routines sémantico-rhétoriques

8

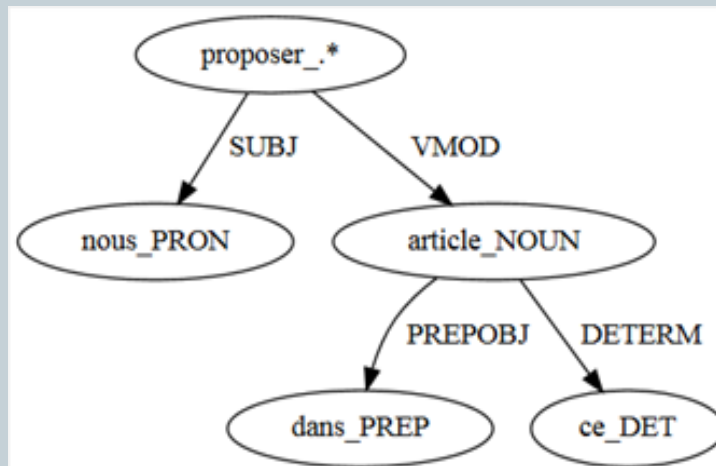
- Un type de motif
 - Construit autour d'un prédicat ou d'une classe de prédicats sémantiques
 - Dont les arguments sont actualisés (associés à une énonciation)
 - Associé à une fonction rhétorique spécifique

- Exemple : Routines impliquant les verbes de constat
 - (*comme*) on/nous le voir/constater/observer LOC
Fonction : Apporter une preuve dans l'argumentation scientifique

Arbres Lexico-syntaxiques récurrents (ALR)

9

- Observable « textométrique » extrait à partir des relations syntaxiques et des mesures d'association (Kraif & Diwersy 2012; Kraif 2016)
- L'ALR peut être un motif ou une suite libre de mots.
 - Ex d'un ALR qui correspond aux exemples :
 - L'étude que **nous proposons dans cet article** met en œuvre un test utilisateur où le sujet , en position debout , est évalué (...) (Psychologie)
 - **Dans cet article, nous proposons** d'analyser la formation endogène d' une ville monocentrique . (Economie)



*2. Comparaison entre segments répétés (SR) et Arbres
Lexico-syntaxiques Récurrents (ALR): Etude sur corpus
des verbes de communication*

2) ALR et segments répétés pour extraire les routines

11

- Point abordé dans Kraif et Tutin (*Corpus*, à paraître)
- Objectif : comparer les méthodes des ALR et des segments répétés (suites de lemmes récurrentes) pour extraire des routines sémantico-rhétoriques dans les écrits scientifiques.
- Hypothèse :
 - Les ALR sont mieux adaptés à notre objectif car :
 - Génèrent moins de bruit (seuls des éléments entretenant des relations syntaxiques seront extraits)
 - Meilleur rappel (moins de dispersion des données)
 - Plus proches des représentations sémantiques

2) ALR et segments répétés pour extraire les routines

12

- **Méthodologie :**
 - Corpus Termith : 500 articles de sciences humaines répartis dans 10 disciplines.
 - Repérage des ALR et segments répétés de 25 verbes de communications des sous-classes sémantiques suivantes : ‘mention’, ‘discussion’, ‘faire savoir’, ‘formulation’, ‘insistance’, ‘proposition’.
 - *citer, débattre, dire, diffuser, discuter, émettre, énoncer, évoquer, exposer, exprimer, formuler, indiquer, insister, intervenir, invoquer, médiatiser, mentionner, préciser, proposer, résumer, révéler, signaler, souligner, suggérer, véhiculer*
- Extraction des segments répétés (SR) lemmatisés de longueur ≥ 3 , apparaissant au moins 8 fois dans 3 disciplines
- Extraction des ALR de longueur ≥ 3 , apparaissant au moins 8 fois dans 3 disciplines (LLR $> 10,83$)
- Exemples de segments extraits :
 - *comme on l'a dit*
 - *nous l'avons souligné*
 - *article se propose de*

Analyse linguistique des segments extraits : repérage du type linguistique

13

● Routines

- *ce résultat suggérer que*
- *comme nous le avoir souligner*
- *il falloir dire que*

● Collocations

- *formuler le hypothèse*
- *insister sur le caractère*
- *proposer un analyse*

● Constructions syntaxiques spécifiques

- *avoir être souligner*
- *être évoquer par*
- *pouvoir se résumer*

● Locutions figées

- *c'est-à-dire*
aller sans dire

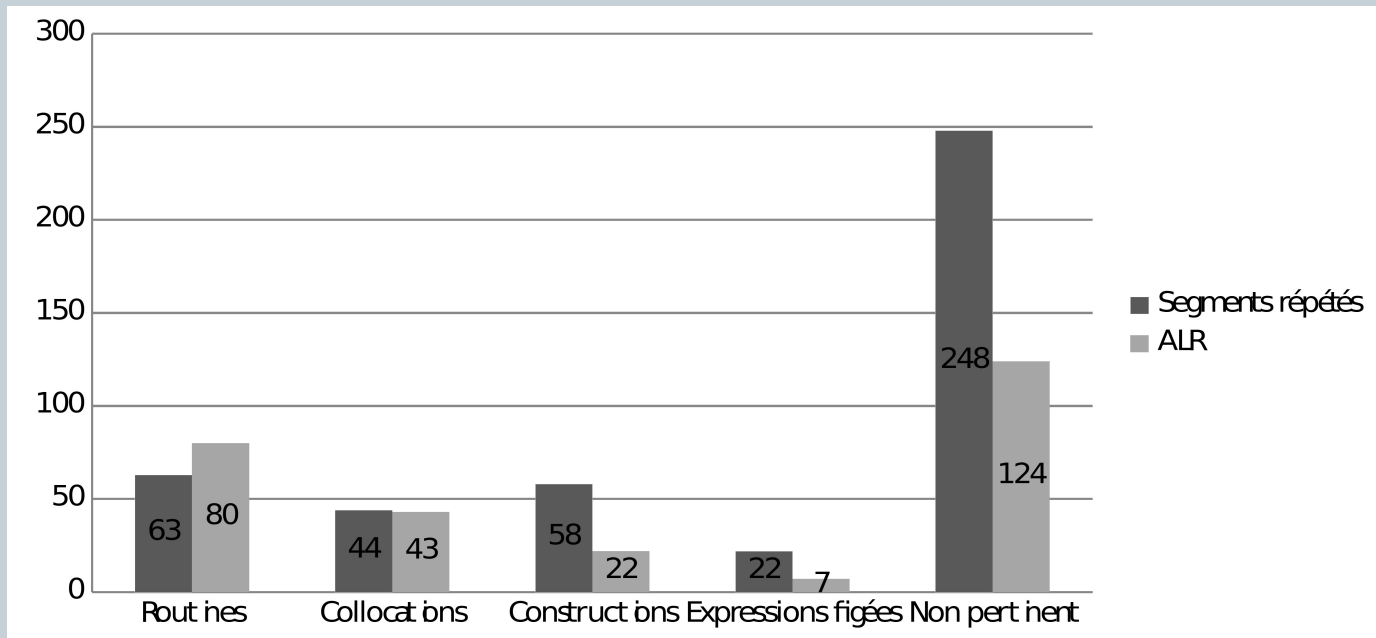
● Non pertinent

- *, il proposer*
- *, proposer par*
- *celui proposer par*

Résultats

14

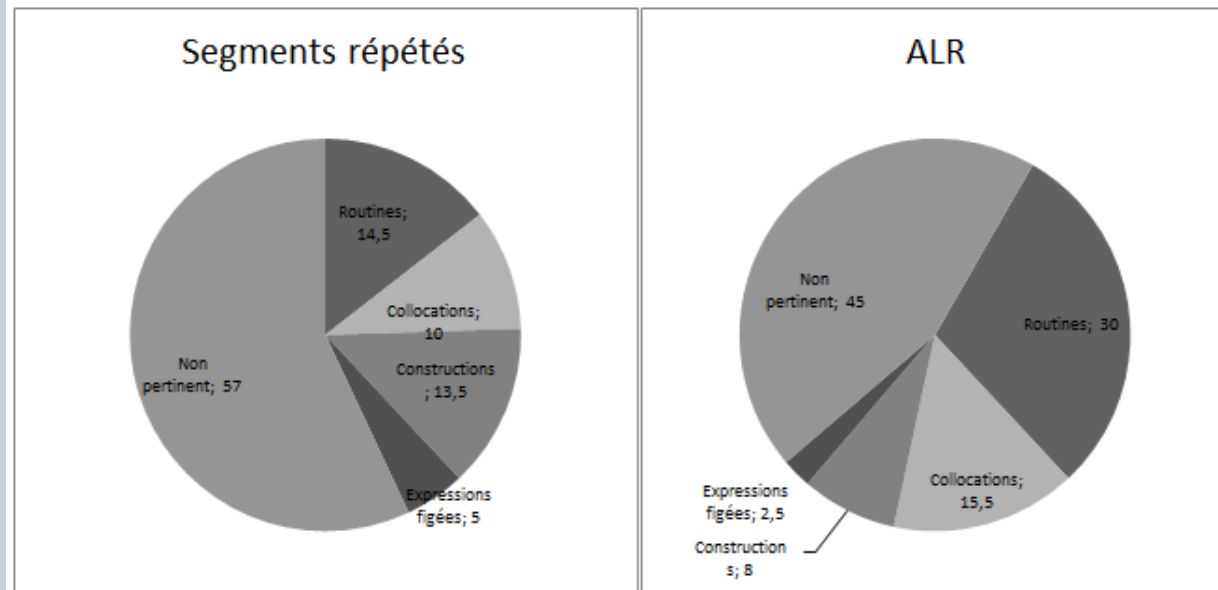
- Davantage de segments extraits par les SR (435) que les ALR (276) mais plus de bruit



Résultats

15

- En proportion, moins de segments non pertinents avec les ALR
 - Points forts des ALR : meilleure précision générale, meilleure extraction des routines, score équivalent pour les collocations
 - Points forts des SR : meilleur repérage de la phraséologie « surfacique » : constructions (alternances) et expressions figées



3. Routines extraites

16

Comparaison pour l'extraction des routines

17

- Les deux techniques extraient des routines, mais d'une manière générale, fréquences plus élevées pour les ALR
 - Insertion possible de modifieurs dans les ALR

Routine	Fréquence d'occurrence	
	SR	ALR
'ça + vouloir + dire'	12	16
'ce + article + proposer'	23	40
'cela + vouloir + dire'	9	30
'comme + le + avoir + souligner'	14	39
'comme + le + dire'	11	15
'comme + le + indiquer'	45	60
'comme + le + souligner'	95	97
'comme + le + suggérer'	31	34
'comme + nous + le + avoir + dire'	9	11

- Quelques problèmes avec les mots grammaticaux pour les ALR
 - Prévoir des post-traitements?

Un bref aperçu des routines extraites

18

- ▣ La plupart des routines extraites peuvent assez facilement se classer dans un petit groupe de fonctions sémantico-rhétoriques
- Fonction interdiscursive
 - Renvoi intertextuel
 - ▣ comme le souligner X
 - ▣ on pouvoir citer X
 - ▣ comme (le) avoir souligner X
 - ▣ comme le suggérer X
 - ▣ comme le dire X
 - ▣ comme le signaler X

- Fonctions métatextuelles (importantes avec les verbes de communication)
 - Cohérence intratextuelle
 - comme nous le avoir souligner
 - , nous le avoir dire,
 - , on le avoir dire,
 - nous le avoir dire ,

La faiblesse des pratiques paysagères en forêt méditerranéenne tient pour partie , on l' a dit , à la définition même qui est donnée de la méditerranéité et de ses caractéristiques visuelles par les sociétés étudiées d'une part , par les gestionnaires forestiers d'autre part .
[géographie]

- Mise en relief de faits
 - nous vouloir souligner
 - on devoir dire
 - il falloir signaler
 - il falloir dire que
 - il falloir préciser que

- Fonctions métatextuelles

- Reformulation

- cela revenir à dire
- pour dire autrement ,

- Cet estompage a l' avantage de mettre l' accent sur un certain nombre de caractéristiques linguistiques et stylistiques du MI : les formes qui relèvent de sa définition doivent être appréhendées en contexte ; **cela revient à dire** que le MI relève d' une analyse contrastive des énonciations , au premier rang desquelles le rapport de la parole du personnage à celle du narrateur . [linguistique]

- Point de vue modalisé

- nous pouvoir dire que
- on pouvoir dire que

- **Présentation de l'objet d'étude**
 - ce article proposer
 - nous proposer ici
 - nous proposer dans article
 - nous proposer de étudier

Cet article propose une revue de littérature sur l'ajustement postural anticipé (APA) et son développement [psychologie]

Pour conclure

22

- Dans l'ensemble, les ALR apparaissent plus satisfaisants pour extraire des motifs ayant fonction de routines que les segments répétés.
 - Moins de bruit
 - Meilleur rappel
 - Ouverture vers des motifs sémantiques
- L'expérimentation sur un sous-ensemble de verbes de communications montre l'intérêt de la technique pour extraire facilement un sous-ensemble de routines pertinentes

Prolongements

23

- Amélioration de la technique avec une meilleure prise en compte des mots grammaticaux
- Exploitation des classes sémantiques du lexique scientifique transdisciplinaire pour proposer des routines plus abstraites
 - Cf expérimentation menée par Sylvain Hatier dans sa thèse
 - Ex : (SUJET Nomdocument) Verbeexamen (OBJET NomLST)
- Inventaire à compléter
- Modélisation des routines à affiner