

Soutenance du 17 septembre 2018

Les titres de documents scientifiques : réurrences dans les syntagmes binominaux après un double point

Damien Gouteux

Sous la direction de Mme Josette Rebeyrolle et M. Ludovic Tanguy



2017 - 2018

Plan

2

I. Sujet et problématique

- Le titre, le double point et notre problématique



I. Sujet et problématique

- Le titre, le double point et notre problématique

II. Démarche et résultats

- Nos données et notre corpus
- Les récurrences lexicales
- Nos trois patrons : besoin, écriture et exemple
- Les récurrences lexico-syntaxiques



I. Sujet et problématique

- Le titre, le double point et notre problématique

II. Démarche et résultats

- Nos données et notre corpus
- Les récurrences lexicales
- Nos trois patrons : écriture et exemple
- Les récurrences lexico-syntaxiques

III. Limites, bilan et perspectives

- Nos limites, notre bilan et les perspectives



I. Sujet et problématique



Le titre comme objet d'études

6

- Un espace court (15 mots)
 - Plus il y a d'auteurs, plus le titre est long
 - Différences selon les domaines scientifiques
- Le premier contact décisif dans la décision de lire



- Deux fonctions : INFORMER OU ATTIRER



Informer...

7

- Titre de Hartley (2004)

*« Were there any sex differences?
Missing data in psychology journals »*



... ou attirer ?

8

- Titre choisi par l'éditeur pour le même article

«*More sex please, we're psychologists* »



Le double point

9

- Le double point introduit (Grevisse et Goosse, 2011) :
 - énumération, citation, exemple, cause, **synthèse, description, définition, explication**
- Présence dans les titres très étudiée (Dillon, 1981)
 - Si présent, longueur du titre plus grande
 - Un marqueur de qualité et d'impact ?
- Haute fréquence de titres à deux segments avec un double point dans les titres scientifiques (Haggan, 2004)



général : spécifique

10

- Utilisation dans les titres (Swales & Feak, 1994)

général : spécifique

La foule : un nouvel acteur dans
l'accompagnement à la création d'entreprises
(Onnée, Zoukhoua & Calme, 2018, Article)



majeure : mineure

11

- Utilisation dans les titres (Swales & Feak, 1994)

majeure : mineure

Santé mobile pour le suivi de l'insomnie
chronique : design de services et sciences de
l'information et de la communication

(Catoir-Brisson, 2018, Chapitre)



problème : solution

12

- Utilisation dans les titres (Swales & Feak, 1994)

problème : solution

Violences et justice dans les cours de récréation
d'écoles élémentaires classées REP+ : effets des
dispositifs pédagogiques mis en place par les
enseignants

(Boxberger, 2018, Poster)



- Utilisation dans les titres (Swales & Feak, 1994)

sujet : méthode

Regard sur l'histoire de quelques prépositions de
l'anglais contemporain : Apport de la diachronie
(Mathieu, 2018, Communication)



Notre problématique

14

Qu'apporte le segment après le double point
au segment qui le précède ?

- Étudier les couples de noms communs qui suivent le double point
- Analyser leurs récurrences lexicales et syntaxiques
- Pour en tirer une information sémantique



II. Démarche et résultats



Des données au corpus

16

- Extraction des titres FR depuis Hyper Article en Ligne (HAL)

Type	Pourcentage
Articles	47 %
Communications	27 %
Chapitres	7 %
Thèses	5 %
Mémoires d'étudiants	2 %
Autres publications	2 %
Ouvrages, rapports, prépublications, posters, HDR, vidéo, cours, son, brevet, rapport d'activité, note de lecture, note de synthèse	10 %



Des données au corpus

17

- Extraction des titres FR depuis Hyper Article en Ligne (HAL)
- Segmentation et lemmatisation avec Talismane
- Filtrage des titres « aberrants »
 - ➔ Corpus général : 278 806 titres



Des données au corpus

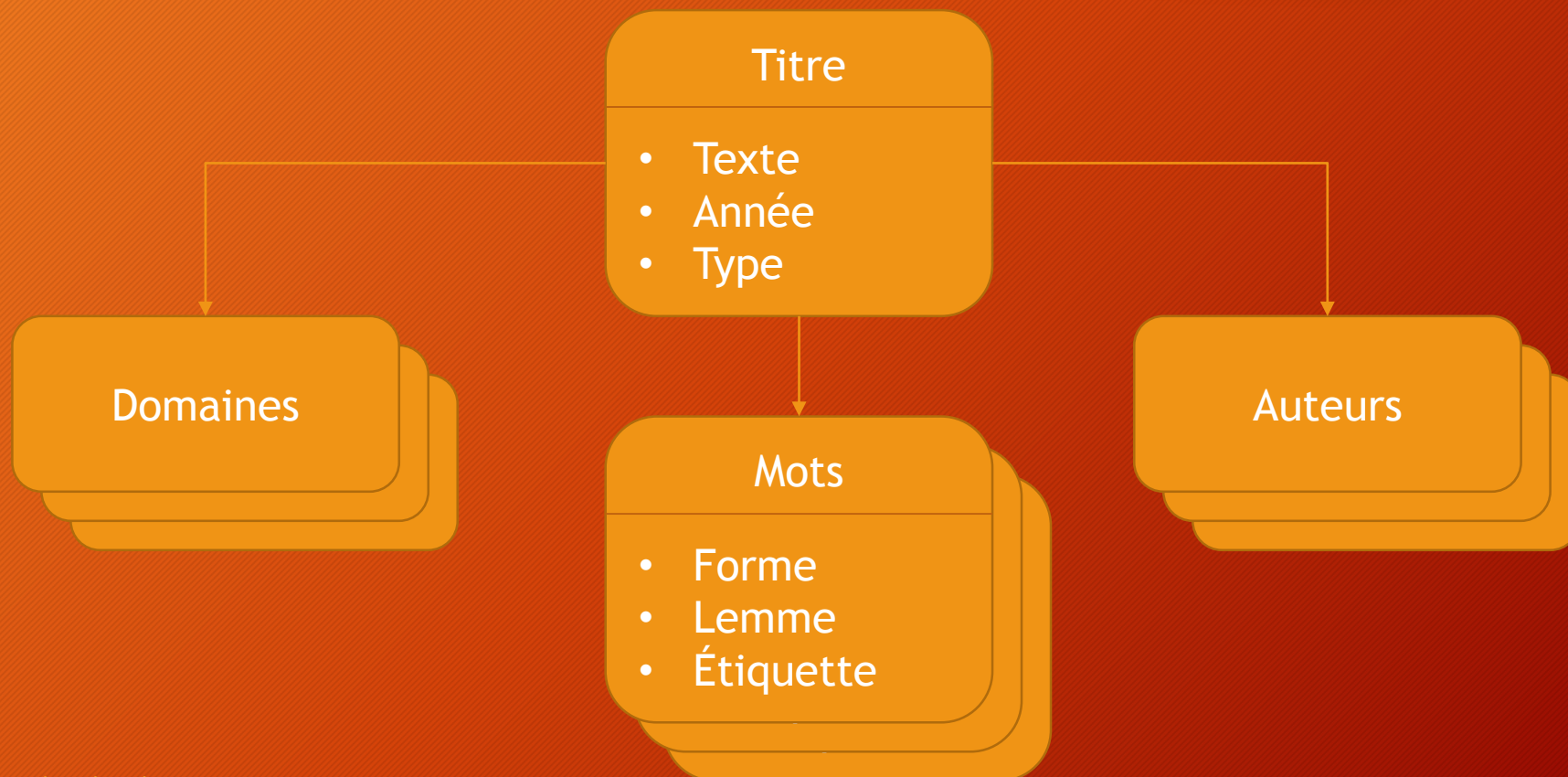
18

- Extraction des titres FR depuis Hyper Article en Ligne (HAL)
- Segmentation et lemmatisation avec Talismane
- Filtrage des titres « aberrants »
 - Corpus général : 278 806 titres
- Filtrage sur notre problématique :
 - 1 seul double point
 - Entre 1 et 29 mots après le double point
 - Corpus de travail : 85 531 titres de différents types



L'anatomie d'un titre

19



Les récurrences lexicales

20

- Les noms les plus fréquents sont des noms généraux :

Lemme	Occurrences	% du total des noms
étude	6089	1,25 %
cas	4631	0,95 %
approche	3036	0,62 %
analyse	3001	0,62 %
application	2982	0,61 %
siècle	2766	0,57 %
pratique	2609	0,54 %
exemple	2291	0,47 %



La position des lemmes

21

- Des noms ont une préférence marquée pour une position :

Lemme	Occurrences	% après double point	Distance (0 = :)
cas	4 631	95 %	2,3
exemple	2 291	94 %	1,8
application	2 982	91 %	1,3
résultat	627	89 %	2,5
perspective	1 317	86 %	2,9
proposition	468	85 %	1,7
enjeu	2 133	83 %	1,7



Les récurrences syntaxiques

22

- Dans quel contexte s'inscrivent ces lemmes ?
- En particulier, dans quel syntagme ?
- Matériel de base : les étiquettes POS

La rue et l'écran	:	la	négociation	de	l'	intimité
	PONCT	DET	NC	PREP	DET	NC

(Trainoir, 2018, Communication)



- 42 942 séquences d'étiquettes POS dans notre corpus
- Séquences associées à un syntagme :
 - Nominal →
 - Prépositionnel →
 - Nominal avec une conjonction de coordination
- Écrire des règles pour capturer ces séquences
- Définition d'un langage pour écrire ces règles
- Définition de trois patrons selon le syntagme capturé



L'écriture de patron

24

- Exemple de séquence :

La rue et l'écran	:	la	négociation	de	l'	intimité
	PONCT	DET	NC	P	DET	NC

- Sera capturée par :

DET? NC [(P DET?) P+D] NC ADJ?

- ? indique une option, [] un choix, () un regroupement



Un exemple de capture

25

La rue et l'écran	:	la	négociation	de	l'	intimité
	PONCT	DET	NC	P	DET	NC

DET? NC [(P DET?) P+D] NC ADJ?



Nos trois patrons : SN

26

- SN : pour capturer les syntagmes NC P NC

Regard sur l'histoire de quelques prépositions de
l'anglais contemporain : Apport de la diachronie
(Mathieu, 2018, Communication)



Nos trois patrons : SP

27

- SN : pour capturer les syntagmes NC P NC
- SP : pour capturer les syntagmes P NC P NC

La co-construction textuelle avec de jeunes enfants :
de la phrase au texte, et vice versa

(Torterat, 2018, Article)



Nos trois patrons : SNC

28

- SN : pour capturer les syntagmes NC P NC
- SP : pour capturer les syntagmes P NC P NC
- SNC : pour capturer les syntagmes NC CC NC

La mise en place du dispositif d'orientation active à l'Université : enjeux et perspectives

(Obajtek, 2018, Article)



Nos trois patrons

29

- SN : pour capturer les syntagmes NC P NC
- SP : pour capturer les syntagmes P NC P NC
- SNC : pour capturer les syntagmes NC CC NC
- Une séquence correspond à un et un seul patron ou aucun (exclusivité)



Le résultat global des captures

30

Patrons	Couvertures des séquences	Couvertures des titres
Patron SN	49 %	50 %
Patron SP	5 %	5 %
Patron SNC	7 %	10 %
Total	61 % des 42 942 séquences possibles dans notre corpus	65 % des 85 531 titres de notre corpus



Les récurrences lexico-syntaxiques dans les résultats du patron SN

31

- Les couples (nom, préposition) les plus fréquents :

Nom	Préposition	Occurrences	Fréquence
cas	de	2 885	99,7 %
application	à	1 692	88,0 %
exemple	de	1 590	96,0 %
étude	de	1 438	72,5 %



Les résultats du patron SN : le cas d'état des lieux

32

Syntagme	Occurrences	
État des lieux	336	
Revue de littérature	171	
État de l'art	86	
État des connaissances	40	
État de la question	40	
État de la recherche	21	
État des savoirs	4	
Revue de connaissance	3	
État du débat	2	
État de la littérature	1	
Revue de la question	1	



Les résultats du patron SN : le cas d'état des lieux

33

Syntaxme	Occurrences	Répartition
État des lieux	336	56 % SHS 45 % SDV
Revue de littérature	171	41 % SDV
État de l'art	86	81 % INFO 67% SPI
État des connaissances	40	
État de la question	40	
État de la recherche	21	
État des savoirs	4	
Revue de connaissance	3	
État du débat	2	
État de la littérature	1	
Revue de la question	1	



Les résultats du patron SN : constructions remarquables

34

- OBJET + MÉTHODE : **application** [à dans en chez] CADRE

Comparaison de différentes méthodes d'interprétation de la prédiction de l'eau corporelle par la méthode de dilution de l'eau lourde : **application chez le chevreau mâle**

(Schmidely, Robelin & Bas, 1989, Article)

- OBJET + CADRE : **application de** MÉTHODE

Échanges thermiques chez le porcelet nouveau-né : **application de la méthode du bilan d'énergie**

(Berbigier, Le Dividich & Kobilinsky, 1978, Article)



Les résultats du patron SN : constructions remarquables

35

- OBJET + CADRE : un outil [de pour] BUT

Micro-impression de BMP-2 et fibronectine sur des matériaux mous : **un outil pour** recréer la niche de cellules souches in vitro

(Fitzpatrick, 2018, Thèse)



Les récurrences lexico-syntaxiques dans les résultats du patron SP

36

- La plus fréquente : à propos de (255 occurrences)

De la salle de cinéma à la caverne : à propos de quelques tentatives artistiques d'ensevelissement
(Honoré, 2018, Communication)



Les résultats du patron SP : de A à B

37

- La plus fréquente : à propos de (255 occurrences)
- De ORIGINE à ARRIVÉE :
 - Extension de sens (spatial, temporel, notionnel)
 - Noms exclusifs à l'origine : diagnostic, observation, origine
 - Noms exclusifs à l'arrivée : application, mise, jour, action
 - 67 occurrences où origine et arrivée sont égales

Perspectives 2007-2008. Royaume-Uni : de B à B...

(Mathieu, 2016, Article)



Les récurrences lexico-syntaxiques dans les résultats du patron SNC

38

- Faible flexibilité de la conjonction de coordination :
→ sauf pour (mythe, réalité), (continuité, rupture)

Ordre :	Premier	Second	Occurrences	% dans cet ordre
	mythe	réalité	56	100 %
	enjeu	perspective	51	98 %
	bilan	perspective	50	100 %
	intérêt	limite	24	100 %
	apport	limite	21	100 %
	théorie	application	13	100 %
	principe	application	13	100 %



III. Limites, bilan et perspectives



Les limites de notre étude

40

- La faiblesse du lien séquence d'étiquettes - syntagme
→ Avoir recours à l'analyse en dépendance

- Une partie du titre n'est pas couverte :

- Avant le double point
- Après :

SN : 49 %

SP : 71 %

SNC : 64 %

- Une partie du corpus n'est pas couverte : 35%



- Récurrences lexicales : vocabulaire scientifique
- Préférence de noms pour la position après le double point
- Récurrences syntaxiques : différents syntagmes binominaux
- Contraintes d'ordres des noms dans les syntagmes SP et SNC
 - Sémantico-logique : reprise de la logique temporelle
 - Rhétorique : d'abord le positif avant le négatif
 - Didactique : la théorie puis la pratique
- Figement de certaines structures



Les perspectives

42

- Création de nouveaux sous-corpus : par type, par domaine
- Patrons plus flexibles → structures plutôt que séquences
- Étude avancée des noms propres
 - toponymes 12 / 16 (France, Paris, Europe)
- Compréhension automatique des titres
 - Extraction d'information
 - Recherche sémantique dans HAL



Remerciements & Questions



- Dillon, J. (1981). The emergence of the colon: an empirical correlate of scholarship. *American Psychologist*, 36, 879-884.
- Grevisse, M. & Goosse, A. (2011). *Le bon usage : grammaire française*. Bruxelles: Duculot.
- Haggan, M. (2004). Research paper titles in literature, linguistics and science: dimensions of attraction. *Journal of Pragmatics*, 36(2), 293-317.
- Huyghe, R. (2015). Les typologies nominales : présentation. *Langue française*, 185, 5-27.
- Legallois, D. & Tutin, A. (2013). Présentation : Vers une extension du domaine de la phraséologie. *Langages*, 189(1), 3-25.
- Swales, J. M. & Feak, C. B. (1994). *Academic Writing for Graduate Students*. Ann Arbor: University of Michigan Press.

