

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

I. Sujet et problématique

- Récurrences dans les syntagmes binominaux
 - après un double point
 - dans les titres de documents scientifiques



I. Sujet et problématique

- Récurrences dans les syntagmes binominaux
 - après un double point
 - dans les titres de documents scientifiques

II. Démarche et résultats

- Données et corpus de titres
- Récurrences lexicales
- Trois patrons comme outils
- Récurrences lexico-syntaxiques



I. Sujet et problématique

- Récurrences dans les syntagmes binominaux
 - après un double point
 - dans les titres de documents scientifiques

II. Démarche et résultats

- Données et corpus
- Récurrences lexicales
- Trois patrons comme outils
- Récurrences lexico-syntaxiques

III. Bilan, limites et perspectives



I. Sujet et problématique



Les titres de documents scientifiques

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*»

- L'expression du sujet est beaucoup moins explicite
- Difficilement recherchable



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)



Exemples de Swales & Feak 1 / 2

10

général : spécifique

La foule : un nouvel acteur dans l'accompagnement à la création d'entreprises

(Onnée, Zoukhou & Calme, 2018, Article)

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)



Exemples de Swales & Feak 2/2

11

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)

sujet : méthode

Regard sur l'histoire de quelques prépositions de l'anglais contemporain : Apport de la diachronie

(Mathieu, 2018, Communication)



Notre problématique

12

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

- Étudier les couples de noms communs immédiatement après le double point,
- Analyser leurs récurrences lexicales et syntaxiques,
- Pour en tirer une information sémantique.



II. Démarche et résultats



Des données au corpus

14

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

Type	Pourcentage
Articles	47 %
Communications	27 %
Chapitres d'ouvrage	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

15

- 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

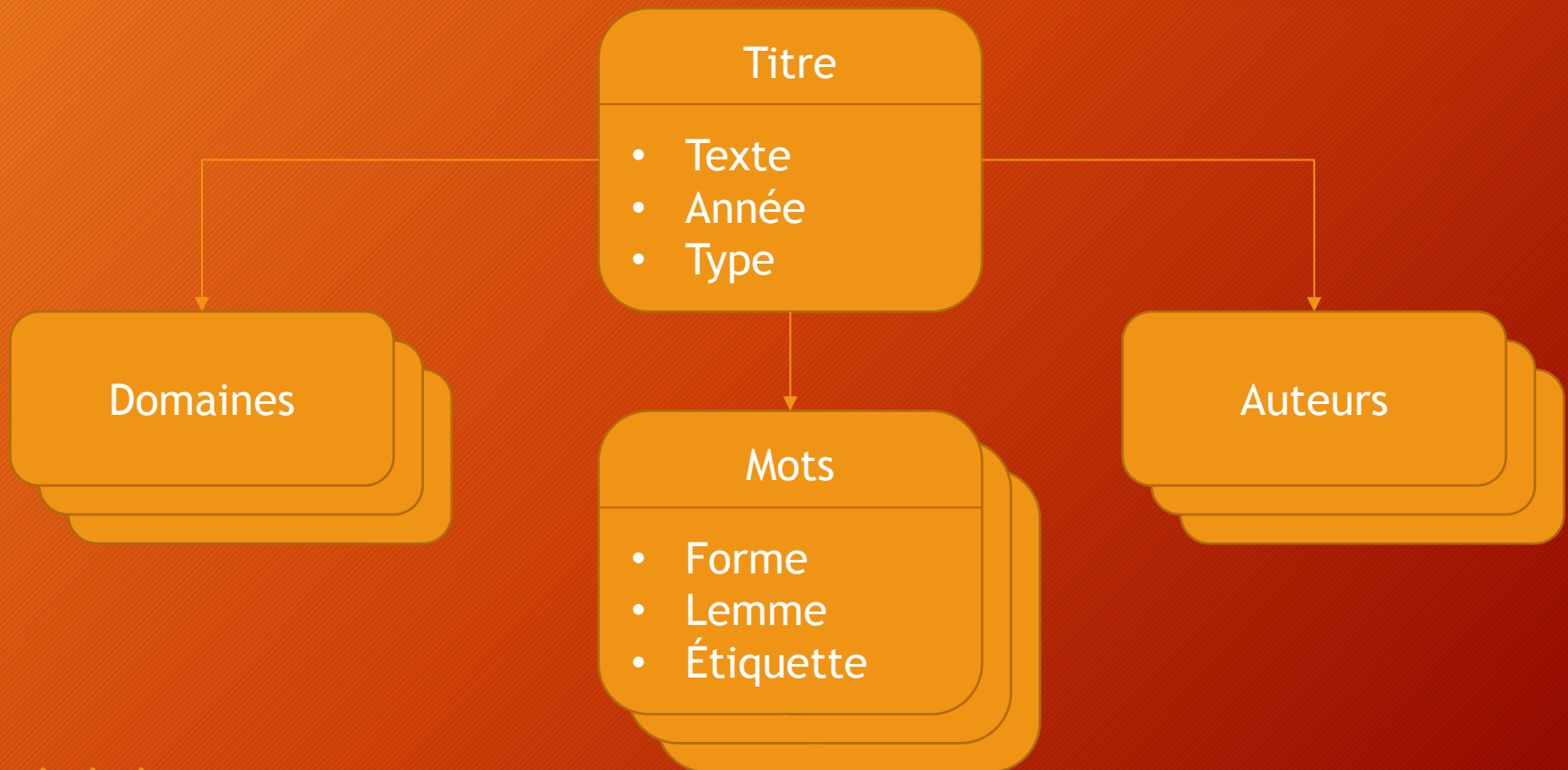
16

- Extraction des titres FR depuis Hyper Article en Ligne (HAL)
- Segmentation et lemmatisation avec Talismane
- Filtrage des titres « aberrants »
 - Corpus général : 278 806 titres de différents types
- Filtrage sur notre problématique :
 - 1 seul double point et entre 1 et 29 mots après
 - Corpus de travail : 85 531 titres (31 %)
 - dont 72 % Sciences de l'Homme et Société



Anatomie d'un titre

17



Réurrences lexicales : noms communs les plus fréquents

18

- Vocabulaire scientifique ; surtout des noms généraux

Lemme	Occurrences	% du total des noms
étude	6 089	1,25 %
cas	4 631	0,95 %
approche	3 036	0,62 %
analyse	3 001	0,62 %
application	2 982	0,61 %
siècle	2 766	0,57 %
pratique	2 609	0,54 %
exemple	2 291	0,47 %



Position des noms communs par rapport au double point

19

- 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



Extension au niveau syntaxique

20

- Dans quel contexte s'inscrivent ces lemmes ?
- En particulier, dans quel syntagme ?
- Matériel de base : les étiquettes POS
- 42 942 séquences d'étiquettes POS après un double point



Regrouper les séquences selon les syntagmes

21

La rue et l'écran

:	la	négociation	de	l'	intimité
PONCT	DET	NC	P	DET	NC

Modèles de régression

:	Comparaison	des	approches	statistiques	et	ensembliste
PONCT	NC	P+D	NC	ADJ	CC	ADJ

Analyses linguistiques de données textuelles

:	Exploitation	quantitative	d'	entretiens	semi-directifs
PONCT	NC	ADJ	P	NC	ADJ



Regrouper les séquences selon les syntagmes

22

La rue et l'écran

:	la	négociation	de	l'	intimité
PONCT	DET	NC	P	DET	NC

Modèles de régression

:	Comparaison	des	approches	statistiques	et	ensembliste
PONCT	NC	P+D	NC	ADJ	CC	ADJ

Analyses linguistiques de données textuelles

:	Exploitation	quantitative	d'	entretiens	semi-directifs
PONCT	NC	ADJ	P	NC	ADJ



Écriture de trois patrons

23

- Séquences d'étiquettes sont associées à un syntagme :
 - Nominal →
 - Prépositionnel →
 - Nominal avec une conjonction de coordination
- Écrire des règles pour capturer ces séquences : un patron
- Définition de trois patrons selon le syntagme capturé
- Une partie du titre après le 2^e nom est écartée



Trois patrons : patron SN pour syntagme nominal

24

- 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)



Trois patrons : patron SP pour syntagme prépositionnel

25

- 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)



Trois patrons : patron SNC pour syntagme nominal avec CC

26

- 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)



Trois patrons : SN, SP, SNC

27

- 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 1 patron ou aucun



Trois patrons : SN, SP, SNC

28

- SN : pour capturer les syntagmes NC P NC 50 %
- SP : pour capturer les syntagmes P NC P NC 5 %
- SNC : pour capturer les syntagmes NC CC NC 10 %
- Une séquence correspond à 1 patron ou aucun
- Couverture de 65% des titres



Réurrences lexico-syntaxiques dans les résultats du patron SN

29

1. Les quatre couples (nom 1, préposition) les plus fréquents :

Nom 1	Suivi par la préposition	Avec une fréquence de	Occurrences
cas	de	99,7 %	2 885
application	à	88,0 %	1 692
exemple	de	96,0 %	1 590
étude	de	72,5 %	1 438



Réurrences lexico-syntaxiques dans les résultats du patron SN

30

1. Les quatre couples (nom 1, préposition) les plus fréquents :

Nom 1	Suivi par la préposition	Avec une fréquence de	Occurrences
cas	de	99,7 %	2 885
application	à	88,0 %	1 692
exemple	de	96,0 %	1 590
étude	de	72,5 %	1 438

2. Comment exprimer le concept d'état des lieux ?



Réurrences lexico-syntaxiques dans les résultats du patron SN

31

1. Les quatre couples (nom 1, préposition) les plus fréquents :

Nom 1	Suivi par la préposition	Avec une fréquence de	Occurrences
cas	de	99,7 %	2 885
application	à	88,0 %	1 692
exemple	de	96,0 %	1 590
étude	de	72,5 %	1 438

2. Comment exprimer le concept d'état des lieux ?

3. Interprétation sémantique de constructions



Résultats du patron SN : exprimer « l'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



Résultats du patron SN : exprimer « l'état des lieux »

33

Pour les titres exprimant cette notion :

- En Informatique
 - ➔ 81 % état de l'art
- En Sciences pour l'ingénieur
 - ➔ 67 % état de l'art
- En Sciences humaines et société :
 - ➔ 56 % état des lieux
- En Sciences de la vie :
 - ➔ 45% état des lieux et 41 % revue de littérature



Résultats du patron SN : interprétations sémantiques

34

- OBJET + CADRE : un outil [de pour] BUT
- OBJET + MÉTHODE : application [à dans en chez] CADRE
- 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)



Réurrences lexico-syntaxiques dans les résultats du patron SP

35

- 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)



Résultats du patron SP : de A à B

36

- De ORIGINE à ARRIVÉE :
 - 1 495 occurrences (37 % des résultats)
 - 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 (4 %) où origine et arrivée sont égales

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

(Mathieu, 2016, Article)



Réurrences lexico-syntaxiques dans les résultats du patron SNC

37

- 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 %



- Contraintes d'ordres des noms dans les syntagmes
 - Sémantico-logique : reprise de la logique temporelle
 - ➔ SNC : bilan et perspective
 - ➔ SP : de origine à ... / de ... à nos jours
 - Rhétorique : d'abord le positif avant le négatif
 - ➔ SNC : intérêt et limites
 - Didactique : la théorie puis la pratique
 - ➔ SNC : théorie et application
 - ➔ SP : de principe à ... / de ... à application

III. Bilan, limites et perspectives



- Récurrences lexicales :
 - Vocabulaire scientifique ; noms généraux
 - Préférence de noms pour la position après
- Récurrences lexico-syntaxiques :
 - Contraintes d'ordres des noms
 - Figement de certains syntagmes
 - Choix de certains syntagmes lié à la discipline
 - Syntagmes possédant une sémantique récurrente



Limites de notre étude

41

- La faiblesse du lien séquence d'étiquettes - syntagme
→ Avoir recours à l'analyse en dépendance
- Une partie des titres du corpus n'est pas couverte : 35 %
- Une partie du titre n'est pas couverte :
 - Avant le double point
 - Après le double point : 39 % (moyenne des trois patrons)



- Création de nouveaux sous-corpus : par type, par domaine
- Patrons plus flexibles → structures plutôt que séquences
- Étude avancée des noms propres
 - 12 toponymes sur les 16 plus fréquents :
1. France 2. Paris 3. 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.



Écriture de patron

45

- 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




Exemple de capture

46

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

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



Résultat global des captures

47

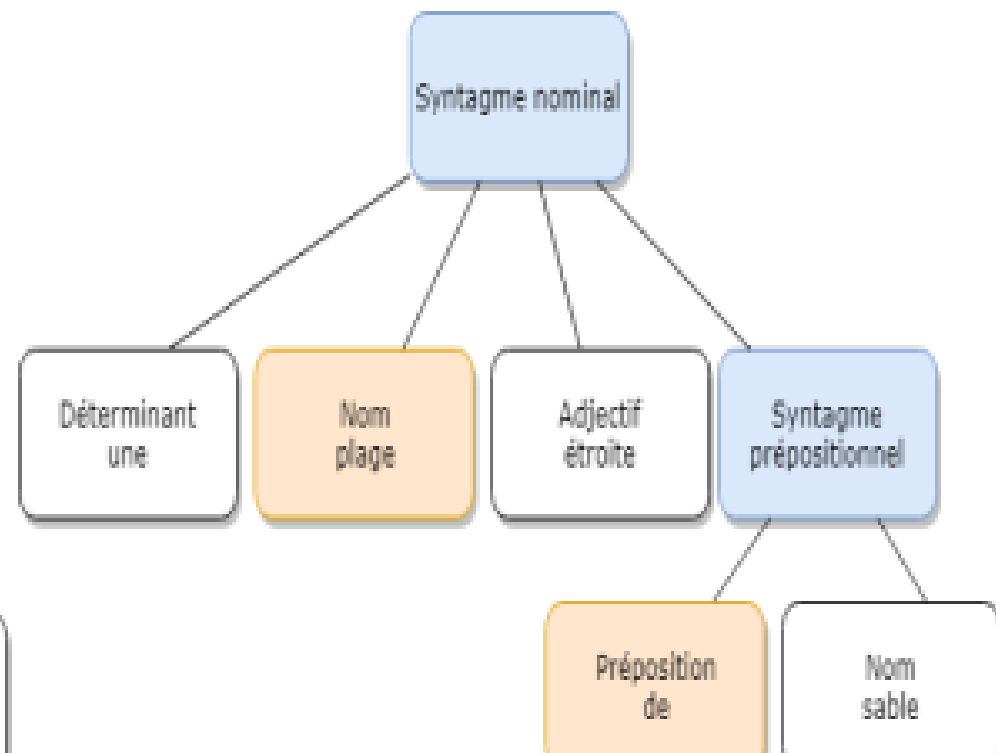
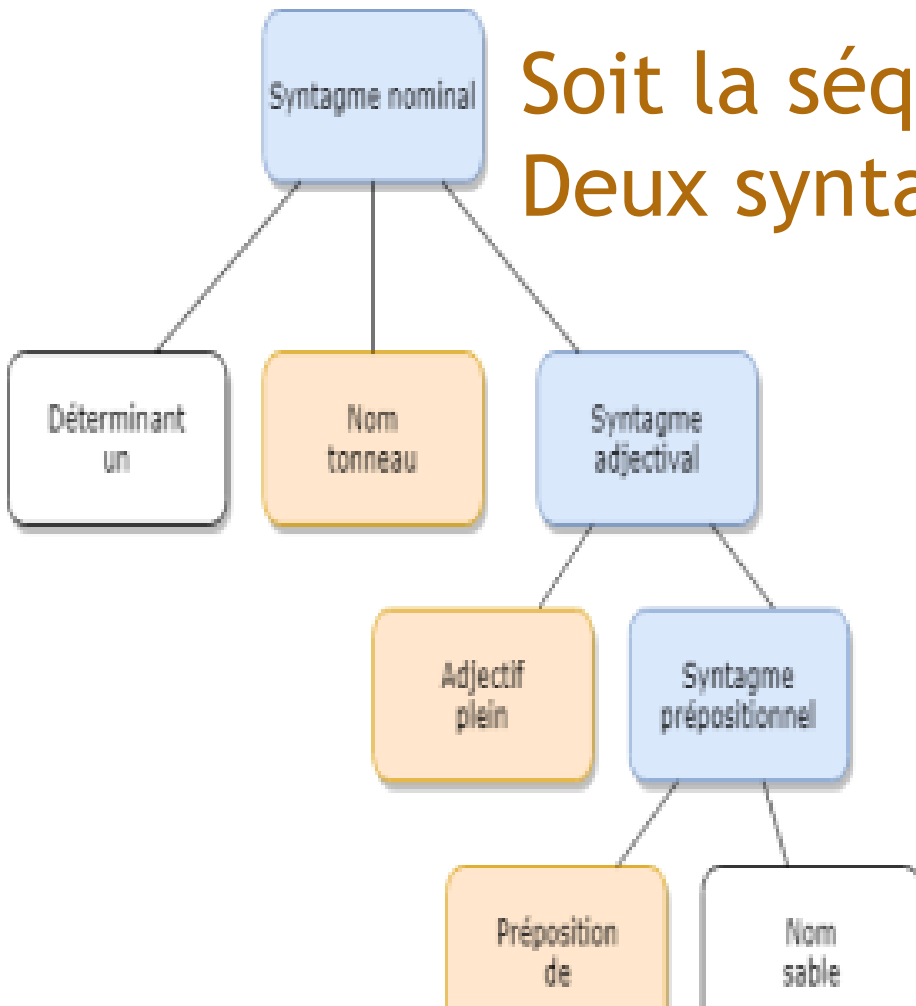
Patrons	Couvertures des séquences	Couvertures sur l'ensemble des titres	Couverture interne des titres
Patron SN	49 %	50 %	49 %
Patron SP	5 %	5 %	71 %
Patron SNC	7 %	10 %	64 %
Total	61 % des 42 942 séquences possibles dans notre corpus	65 % des 85 531 titres de notre corpus	



Faiblesse du lien entre séquences d'étiquettes et syntagmes

48

Soit la séquence : DET NC ADJ P NC
Deux syntagmes possibles



Résultats du patron SN : constructions remarquables

49

- 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)



Résultats du patron SN : constructions remarquables

50

- 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)

