

Soutenance du 27 septembre 2019

Sous-spécification dans les titres d'articles scientifiques

Damien Gouteux

Sous la direction de Mme J. Rebeyrolle et M. L. Tanguy

2018 - 2019

Plan

1. Sujet et problématique

Sous-spécification dans les titres scientifiques

2. Démarche

- Construction d'un corpus de travail
- Détermination des têtes spécifiques et transdisciplinaires
- Recherche des constructions spécificationnelles
- Fouille de données pour en détecter de nouvelles
- Schémas des nouvelles constructions

3. Résultats : filtres

4. Bilan, limites et perspectives

1. Sujet et problématique

La sous-spécification
dans les titres scientifiques

Sujet et problématique

Titres

- Deux fonctions : **informer** et attirer
- Un espace restreint

Travail de 1^{ère} année

- Noms privilégient d'être après le double point
- Têtes de segments fréquentes + transdisciplinaires
- Abstraits et similitudes avec la liste des NSS

Ces têtes sont-elles en emploi sous-spécifié ?

Les noms sous-spécifiés

- Classe fonctionnelle
- Néanmoins prédisposition lexicale
- Faible contenu sémantique
- 3 fonctions
 - **Cognitive** : encapsuler dans un concept temporaire
 - **Sémantique** : caractériser
 - **Discursive** : lier

Nom sous-spécifié  Contenu spécifiant
 Construction spécificationnelle 

2. Démarche

Construction d'un corpus

- 340 000 titres tirés de HAL
- Articles > Communications > Chapitres > Posters
- 25 domaines scientifiques **1 titre ⇔ 1 domaine**
- Lemmatisation et analyse par Talismane + Corrections
- Segmentation (Anthony, 2001) : ,
- Repérage des têtes
- From-Corpus-To-Cognition : « *despite the indisputable charm of rare or exotic examples, one should mainly be interested in frequent and therefore systemically and cognitively more important items* ». Schmid (2000, p. 47)

Construction d'un corpus

- 340 000 titres tirés de HAL
- Articles > Communications > Chapitres > Posters
- 25 domaines scientifiques **1 titre ⇔ 1 domaine**
- Lemmatisation et analyse par Talismane + Corrections
- Segmentation **1 ou 2**
- Repérage des têtes **1 segment ⇔ 1 tête**
- 250 998 titres 74 % du matériau initial
- Sciences exactes 72% mono 28% bi
- Sciences humaines et sociales 54% mono 46% bi

Exemples de titres

Le **problème** de la prévision du temps

Problèmes épistémologiques de l'induction

Travailler en équipe : le **choix** social appliqué au problème de la patrouille multi-agents

L'aquaculture : un **problème** pour l'environnement ?

Les **femmes** et les dettes : **problèmes** de responsabilité dans la Mésopotamie du IIe millénaire avant Jésus-Christ

Têtes spécifiques

- Classement des têtes par degré de spécificité à un domaine
- 25 documents virtuels avec tous les titres d'un domaine
- $TF * IDF$

Philosophie 2 800 lemmes / 7 856 occurrences / 6 152 titres
philosophie, Leibniz, Spinoza, Descartes, Bergson, Kant, Habermas, Nietzsche, Poincaré, Henri Poincaré

Archéologie et préhistoire 3 444 / 13 391 / 10 124
céramique, sanctuaire, décor, nécropole, sépulture, occupation, mobilier, archéologie, vaisselle, habitat

- Calcul des distances dans un espace à 30 410 dimensions

Têtes transdisciplinaires

- Seuil de médiane $> 0,1 \%$
= Dans la moitié des domaines, sa fréquence est $> 0,1 \%$
- 94 lemmes sélectionnées / 123 227 lemmes = $0,08 \%$
- 94 739 occurrences des sélectionnées / 354 168 occ. = 27%

Les 20 premières têtes des 94 classés par la médiane sont :
étude, analyse, cas, approche, exemple, enjeu, évolution, apport, rôle, modèle, réflexion, évaluation, outil, question, représentation, application, construction, introduction, histoire et développement.

Têtes transdisciplinaires

- 27 % des titres monosegmentaux
- 17 % premiers segments de titres bisegmentaux
- 36 % seconds segments de titres bisegmentaux
- 79 % présentes dans le LTES (Tutin, 2008)
- 87 % présente dans le LST (Hatier, 2016)
- Dans les seconds segments :
 - 4 % cas, 3 % exemple des occurrences
 - 97 % des occurrences de cas et 93 % de celles d'exemple
 - 4 couples médiane > 0 :
 - Tête 1 : rôle | approche | effet Tête 2 : cas
 - Tête 1 : apport Tête 2 : exemple

Têtes transdisciplinaires

- Modèle général : spécifique (Swales et Feak, 1994)
- Resserrement (Haggan, 2004)
- Exemple :

*L'apport des archives privées à l'histoire politique :
l'**exemple** de Louis Costa, le notaire rouge*

- Ces têtes semblent les plus propices à un emploi sous-spécifié car :
 - Fréquentes
 - Abstraites : dénote entités de type 2 ou 3 (Lyons (1977))
 - Faible contenu sémantique

Constructions spécificationnelles

- NSS [[, ce] + être] + proposition subordonnée avec que
proposition infinitive avec de

Démystification de l'idée que le réseau d'aide informelle se délite

La tentation d'instituer des « Cours constitutionnelles régionales »

- NSS [, ce] + être + NC

Le plus grand danger social, c'est le bandit imberbe.

Constructions spécificationnelles

- NSS de NC

Trop de résultats à analyser...

Faible valeur sélective (Schmid, 2000, p. 26 ; 2018, p. 115)

Fouille de données

- Trouver d'autres constructions spécificationnelles
- Autour des têtes transdisciplinaires
- Séquences de 2 à 4 tokens

Les **problèmes** d'environnement dans une région d'extraction pétrolière

2 - DET TT

2 - TT de

3 - INIT DET TT

3 - DET TT de

3 - TT de NC

4 - INIT DET TT de

4 - DET TT de NC

4 - TT de NC dans

5 - INIT DET TT de NC

5 - DET TT de NC dans

5 - TT de NC dans DET

Fouille de données

- Motifs fréquents, motifs émergents :
 - 1 - SEG TransHead à::P DET
 - 2 - SEG TransHead sur::P DET
 - 3 - INIT TransHead sur::P DET
 - 6 - SEG DET TransHead de::P+D
 - 7 - SEG TransHead de::P+D NC
- On retrouve la construction spécifique NSS de NS
- Mais avec les informations suivantes :
 - Toujours en début de segments
 - Autres prépositions : à / de / sur

Schémas de filtrage

- Deux schémas :
- INIT [DET] TRANSHEAD [ADJ] (((à P | sur P | de P) [DET])) | (à P+D | de P+D)) NC
- SEG [DET] TRANSHEAD [ADJ] (((à P | sur P | de P) [DET])) | (à P+D | de P+D)) NC
- Obligation de prévoir tous les éléments
- Ce n'est pas définition mais une opérationnalisation (Schmid, 2018, p. 113)

Schémas de filtrage

- Toujours trop de résultats
- Restriction sur le nom du contenu spécifiant :
 - Les propositions ont un verbe
 - En anglais, le nom est un gérondif ou un déverbal
 - Restreindre aux noms désignant des actions
 - On utilise pour cela VerbAction (Tanguy et Hathout, 2002)
- Encore des faux positifs :
 - Le problème du **sac** à dos
 - Le problème de l'**économie** de l'eau en pisciculture

3. Résultats

Schémas de filtrage

- Schémas filtrants pour ramener des candidats à la sous-spécification

Effets de la substitution du maïs par du sorgho sur la durabilité de la production de foie gras d'oie

- 24 776 : 9 % de candidates chez les non transdisciplinaires
- 23 497 : 25 % de candidates chez les têtes transdisciplinaires

4. Bilan, limites et perspectives

Bilan et limites

- Nécessité d'un filtrage manuel et donc subjectif
- Schémas pour des noms candidats pour les NSS de NC
- NC désignant une action
- Impossibilité d'avoir une précision ou un rappel global
- Manque d'un corpus de contraste

Perspectives

- Amélioration de la sélectivité
- Utilisation de la typologie des NSS de Schmid (2000)
- Étude de constructions particulières : le problème posé par X
- Interpréter automatique les titres

Remerciements & Questions

Références

- Flowerdew, J. & Forest, R. W. (2015). Signalling nouns in English.
- Haggan, M. (2004). Research paper titles in literature, linguistics and science: dimensions of attraction.
- Hatier, S. (2016). Identification et analyse linguistique du lexique scientifique transdisciplinaire.
- Huyghe, R. (2018). Généralité sémantique et portage propositionnel: le cas de fait.
- Legallois, D. (2008). Sur quelques caractéristiques des noms sous-spécifiés.
- Merrill, E. et Knipps, A. (2014). What's in a Title?
- Nakamura, T. (2017). Extensions transitives de constructions spécificationnelles.
- Schmid, H.-J. (2000). English Abstract Nouns as Conceptual Shells. From Corpus to Cognition.
- Roze, C., Charnois, T., Legallois, D., Ferrari, S. et Salles, M. (2014). Identification des noms sous-spécifiés, signaux de l'organisation discursive.
- Schmid, H. J. (2018). Shell nouns in English-a personal roundup.
- Tanguy, L. et Hathout, N. (2002). Webaffix : un outil d'acquisition morphologique dérivationnelle à partir du Web.
- Tanguy, L., Rebeyrolle, J. (à paraître). Les titres des publications scientifiques en français : fouille de texte pour le repérage de schémas lexico-syntaxiques.

Soutenance du 27 septembre 2019

Sous-spécification dans les titres d'articles scientifiques

Damien Gouteux

Sous la direction de Mme J. Rebeyrolle et M. L. Tanguy

2018 - 2019