

RETOUR D'EXPÉRIENCE SUR LA DÉTECTION AUTOMATIQUE DE MÉTAPHORES DANS DES TEXTES DE GÉOGRAPHIE

Max Beligné, Isabelle Lefort
Laboratoire EVS, Université de Lyon

Aleksandra Campar, Jean-Hugues Chauchat,
Mélanie Lefeuvre, Sabine Loudcher, Julien Velcin
Laboratoire ERIC, Université de Lyon

Premiers jalons : géographie et métaphores

Contexte

Pas d'étude générale → Des études spécifiques.

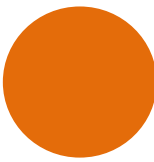
- La métaphore organiciste (P. Bachimon, 1979)
- La métaphore du spectacle (D. Stephen, 1994)
- La métaphore de l'ancrage (B. Debardieux, 2014)
- ...

synthèse

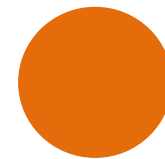
- Importance des métaphores : rôle dans la logique argumentative
- Variation en fonction des époques
- Porte d'entrée dans les systèmes référentiels utilisés plus ou moins implicitement

Objectif

Une étude générale montrant les variations dans le temps de différents systèmes référentiels utilisés en Géographie

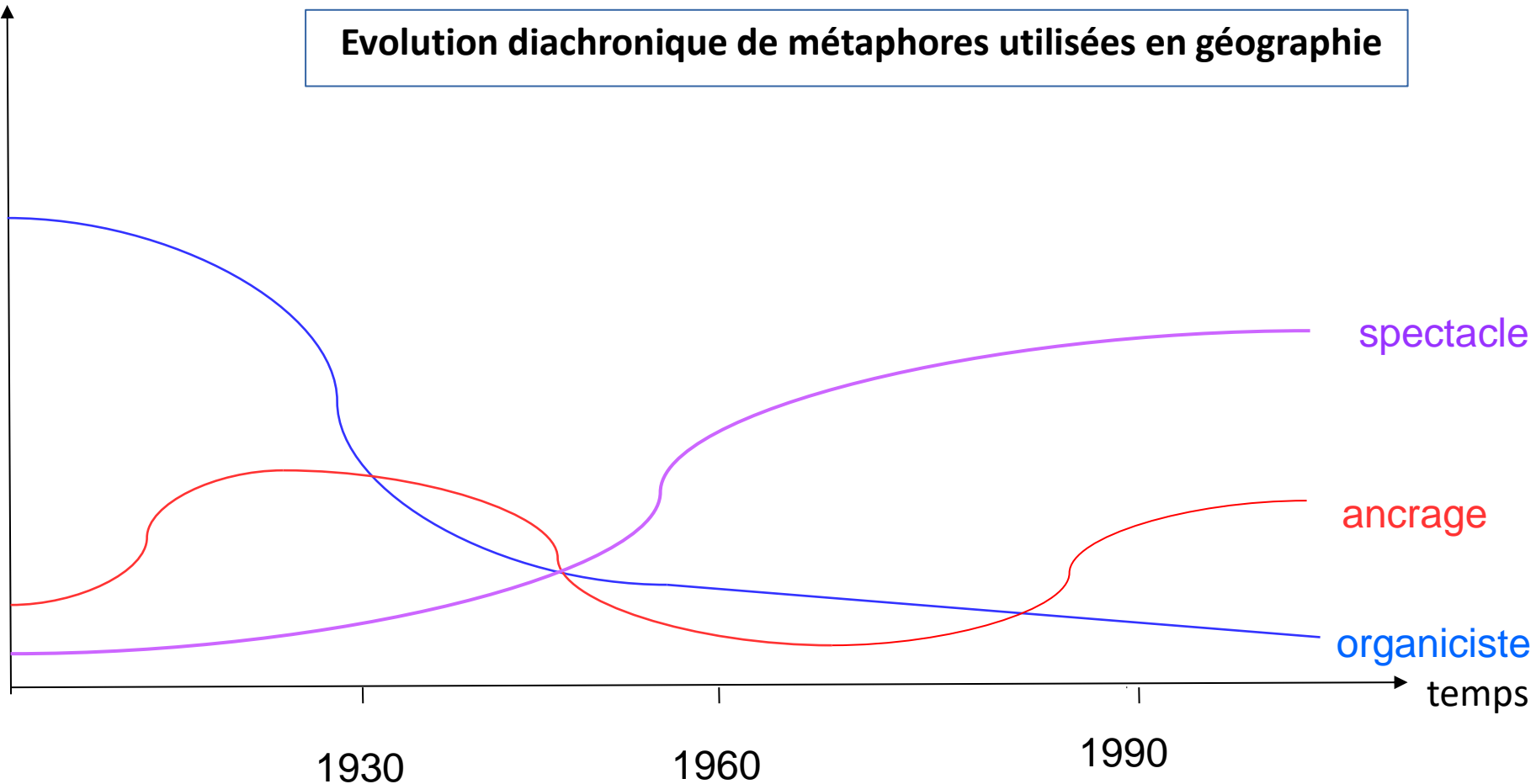


Objectif et hypothèses de recherche



Métaphores
utilisées

Evolution diachronique de métaphores utilisées en géographie



Une définition des métaphores

G. Lakoff et M. Johnson (1980) :

« la métaphore établit une projection partielle entre un domaine source et un domaine cible »

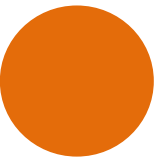
Ex : Des pays ont joué un rôle majeur sur la scène internationale.



Domaine source : spectacle

Domaine cible : géopolitique

➔ **Recherche méthode de détection des domaines sources**



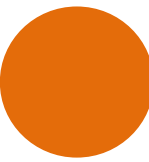
État des lieux : TAL et métaphores

Recherche d'une méthode de détection automatique :

- avec détection des domaines sources
- sans apprentissage supervisé
- sur tous les types de métaphores
- utilisable en français
- avec une bonne efficacité

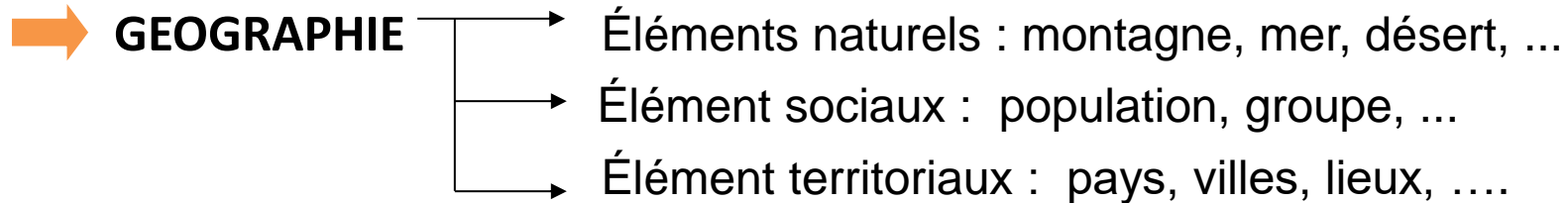


Méthode de Heintz *et al.* , 2013



Contraintes de la méthode de Heintz

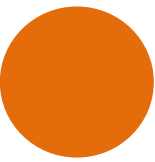
- Choix d'un concept cible obligatoire :



- Choix des concepts sources :

- liste des géographes à partir de leurs connaissances complétée avec la construction du corpus d'évaluation

➔ 42 concepts sources : **ORGANISME, SPECTACLE, GUERRE, MUSIQUE, ANCRAGE, LIVRE, NOURRITURE,**



Constitution du corpus d'évaluation

- Objectif : 1500 phrases annotées
 - En anglais
 - Sélection de 17 articles avec une diversité de thèmes
 - MIP : Metaphor Identification Procedure (Pragglejaz, 2007)
Si un mot dans son contexte n'est pas utilisé dans son sens basique, alors il y a métaphore
 - Pas de méthodologie précise pour trouver les concepts
 - 2 annotateurs indépendants
- ➔ - **1527 phrases dont 365 métaphoriques**
- **Test de Kappa = 0,63**



1ère étape de la méthode de Heintz :

- Sélection aléatoire de la moitié des articles de Wikipédia
- 100 thèmes déterminés sur cet échantillon par topic modeling : LDA (Blei, 2003)

Topic 1	company: 0.0083 / bank: 0.0081 / tax: 0.0066 / market: 0.006 / financial: 0.0056 / ...
Topic 2	music: 0.0212 / opera: 0.0084 / orchestra: 0.0083 / piano: 0.0057 / symphony: 0.0049 / ...
Topic 3	city: 0.007 / population: 0.0065 / century: 0.0054 / world: 0.0051 / government: 0.005 / ...

Ex de 3 thématiques avec les 5 premiers mots les plus probables

2ème étape de la méthode de Heintz :

- Détermination de « 4 mots graines » pour chaque concept source :

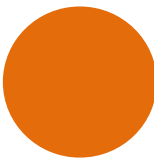
Ex : « music , rhythm, melody, song » pour le concept « music »

- Alignement des concepts avec les thématiques :

Topic 1	company: 0.0083 / bank: 0.0081 / tax: 0.0066 / market: 0.006 /...
Topic 2	music: 0.0212 / opera: 0.0084 / orchestra: 0.0083 / piano: 0.0057/...

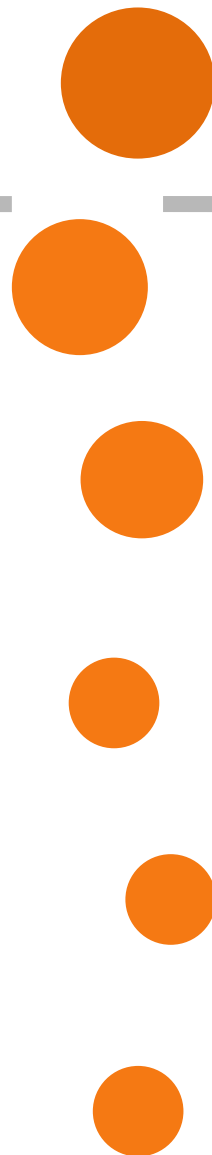
- Au final, le concept « music » est aligné avec 2 thèmes :

Topic 2	Music, opera, orchestra, piano,.....
Topic 78	album, band, song, released,...



Cas du concept cible de « géographie »

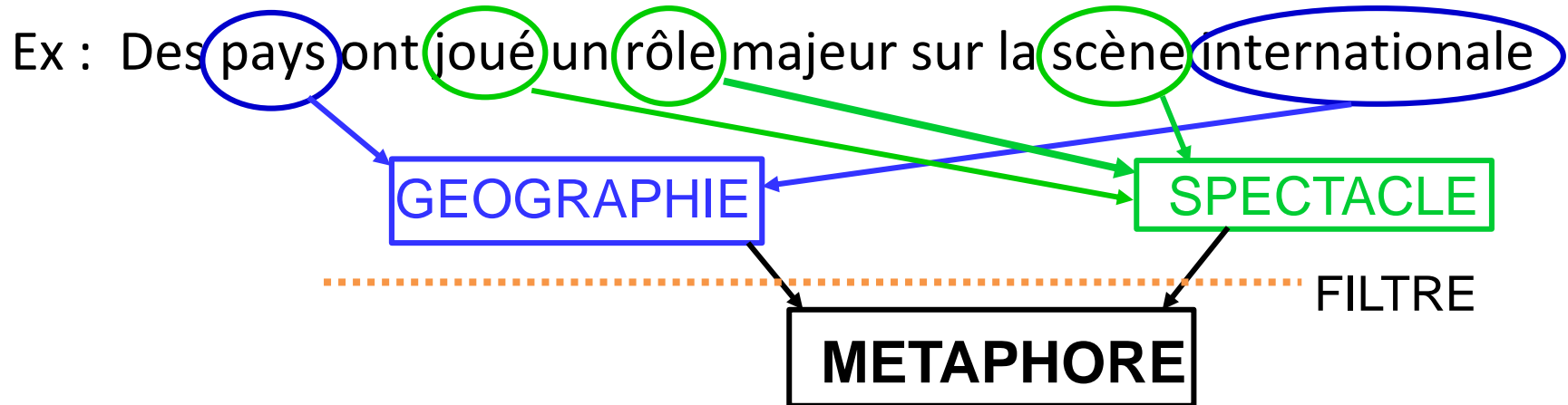
Mots graines pour « geography »	Thématique	Somme proba des mots graines dans thématique
mountain, river, desert, climate, population, society, territory, environment, land	river, lake, estonia, country,....	0.0459
	city, population, century, world...	0.0397
	population, municipality, people,....	0.0361
	district, rural, province, country,...	0.0259
	town, village, area, river, ...	0.0194



3ème étape de la méthode de Heintz :

Dans chaque phrase, calcul de la présence des concepts à partir des thématiques alignées.

Si dans une phrase, le concept cible de « géographie » et un concept source prédéterminé sont présents, alors la phrase est métaphorique. ➡ SEUIL(S)



Jeu sur les seuils

Seuils utilisés	Précision	Rappel
Z_cible =0.06 et Z_source =0.04	0.65	0.04
Z_cible =0.06 et Z_source =0.01	0.625	0.07
Z_cible =0.03 et Z_source =0.01	0.60	0.19
Z_cible =0.01 et Z_source =0.01	0.48	0.23



Approche détaillée des résultats

Une seule métaphore trouvée avec le bon concept source

Ex faux positif :

The urbanisation density threshold is 1,019 people/sq

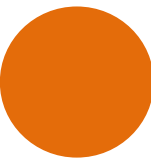
Une métaphore
musicale ????



« Density » et
« threshold » :
termes parfois
utilisés en musique



Effet des
seuils bas



Causes des erreurs

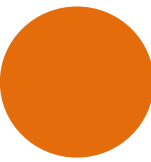
- Maintien de la précision → Les erreurs sont compensées par des métaphores trouvées pour de mauvaises raisons.



Raison : une phrase « géographique » est souvent « métaphorique ».



Raison : MIP est une méthode maximale de détection manuelle.

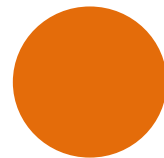


Comparaison par rapport à la méthode de Heintz

Pourquoi des seuils hauts dans la méthode de Heintz ?

Différence des bases évaluatives :

« Toutes les métaphores » VS « un échantillon des phrases les mieux notées mélangées à moitié avec des phrase non sélectionnées comme métaphorique »



Quelques facteurs d'efficacité

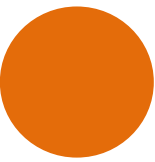
Alignement

26 concepts alignés sur 42 → comment gérer les concepts non alignés ?

Spécificité du concept cible

Pb des textes dont le sujet ne fait pas partie d'une thématique alignée →

- soit réduire à un concept cible avec des textes plus spécifiques.
- soit choisir plusieurs concepts cibles et faire tourner l'algorithme plusieurs fois.

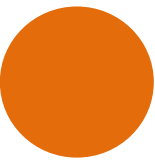


Efficacité spécifique de la méthode

Si on remonte les seuils, recherche métaphore « filée sur une même phrase ».

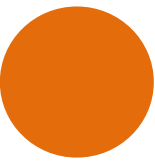
Ex : « La recherche de modèle en géographie consiste à élaguer ce que les faits ont de touffu et de contingent pour en saisir l'essentiel ou, mieux encore, la structure »

Répétition → intégration conceptuelle (P Gréa, 2002)

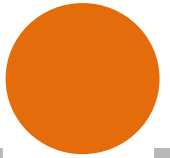


Conclusion

- Articulation thématiques automatiques / concepts choisis.
- Nécessité d'une analyse détaillée des résultats et des bases d'évaluation
- Résultats mitigés mais c'était un premier test !
A permis de construire un corpus, d'esquisser des pistes pour les travaux à venir. A suivre...



Bibliographie



Bachimon, Philippe, « Physiologie d'un langage. L'organicisme aux débuts de la géographie humaine », *Espaces Temps* 13, 1979, p. 75-103.

Daniels, Stephen et Denis Cosgrove, « Spectacle and Text : Landscape Metaphors in Cultural Geography », dans Peter Jackson, *et al.* (dir), *Place/Culture/Representation*, 1994, p. 57-78.

Debarbieux, Bernard, « Enracinement – Ancrage – Amarrage : raviver les métaphores », *L'Espace géographique* 2014/1, t. 43, 2014, p. 68-80.

Group, Pragglejazz. "MIP: A method for identifying metaphorically used words in discourse." *Metaphor and symbol* 22.1 (2007): 1-39.

Gréa, Philippe. "Intégration conceptuelle et métaphore filée." *Langue française* (2002): 109-123.

Heintz, Ilana, et al. "Automatic extraction of linguistic metaphor with Ida topic modeling." *Proceedings of the First Workshop on Metaphor in NLP*. 2013.