

## Nouvelles perspectives en sciences sociales



# Des espaces dans la détection de métaphores Some Spaces in Detecting Metaphors

Max Beligné

Volume 12, numéro 2, mai 2017

URI : <https://id.erudit.org/iderudit/1040905ar>

DOI : <https://doi.org/10.7202/1040905ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

Prise de parole

ISSN

1712-8307 (imprimé)

1918-7475 (numérique)

[Découvrir la revue](#)

Citer cet article

Beligné, M. (2017). Des espaces dans la détection de métaphores. *Nouvelles perspectives en sciences sociales*, 12(2), 69–99. <https://doi.org/10.7202/1040905ar>

Résumé de l'article

Cet article propose un état des lieux des méthodes de détection manuelle et automatique de métaphores en vue par la suite d'une meilleure compréhension d'une science sociale. L'objectif est tout d'abord d'avoir une vue synthétique et organisée des différentes théories et méthodologies (formant à chaque fois des espaces possibles différents de détection) pour mieux orienter la recherche sur cette question difficile. Ensuite, chaque espace a été approfondi pour mieux comprendre ses potentialités, ses limites et ses articulations avec les autres espaces. L'aspect protéiforme des métaphores oblige à cet exercice important à la fois de synthèse et de détail, car c'est en combinant les spécialisations des différents espaces qu'on pourra améliorer la détection globale de ce trope aux définitions multiples. Enfin, il s'agit, à travers cet article, de reconstruire un cadre intégrateur pour appréhender simultanément les détections manuelles et automatiques, car elles sont souvent traitées de manière indépendante du fait des spécialisations alors qu'elles traitent d'un problème commun.

## Des espaces dans la détection de métaphores

**MAX BELIGNÉ**  
Université de Lyon

**D**ans son article « Quelques orientations actuelles de la réflexion épistémologique en géographie », publié en 1985, Paul Claval affirme :

Ce que les géographes attendent désormais de l'épistémologie, ce ne sont plus des recettes infaillibles pour arriver à la vérité, ce sont des lumières sur la façon dont les démarches originales et les idées stimulantes naissent. C'est cela qui fera peut-être des années 1980 la décennie des métaphores comme les années 1960 ont été la décennie des paradigmes et les années 1970 celles des démarches critiques<sup>1</sup>.

S'intéresser aux métaphores en géographie n'est donc pas nouveau mais la dynamique du début des années 1980 est bien vite retombée. En effet, il n'existe pas de travaux visant à une exploration systématique des métaphores (pas seulement un type de métaphore) avec un spectre disciplinaire étendu (dans plusieurs champs de la géographie). Si on se tourne vers les autres sciences sociales, il est possible de citer les travaux de Giovanni Busino<sup>2</sup>

---

<sup>1</sup> Paul Claval, « Quelques orientations actuelles de la réflexion épistémologique en géographie : systèmes structures et métaphores », *Paralelo*, vol. 37, n<sup>os</sup> 8-9, 1985, p. 171-179.

<sup>2</sup> Giovanni Busino, « La place de la métaphore en sociologie », *Revue européenne des sciences sociales*, vol. 42, n<sup>o</sup> 126, 2003, p. 91-101, <https://ress.revues.org/539>, consulté le 2 février 2017.

en sociologie qui ont une visée plus large que les travaux existants en géographie. Cependant, il ne semble pas exister sur la question des métaphores de travaux vraiment détaillés et chiffrés d'évolution globale dans une science sociale.

Pourtant, la reprise de la réflexion de Jean-Claude Passeron<sup>3</sup> par Giovanni Busino place ce trope au cœur de la réflexion sur l'écriture et le raisonnement dans ce domaine : « les métaphores jouent un rôle stratégique dans les sciences sociales car elles font converger des preuves en forme logique différente dans un argumentaire d'ensemble, leurs arguments dans le langage de l'interprétation, leurs interprétations dans une théorie plausible<sup>4</sup> ».

La géographie est, bien sûr, concernée par cette affirmation. Il est possible d'aller même plus loin en suivant la réflexion de Jocelyn Benoist et Fabio Merlini qui affirment que « traditionnellement, la métaphore est associée à la spatialité en vertu même de cette image du déplacement qui serait censée la définir<sup>5</sup> ». La géographie, ayant pour objet principal « la spatialité », pourrait alors être particulièrement concernée par l'utilisation des métaphores. Cependant, comme il s'agit dans cette définition d'un usage lui-même métaphorique de « la spatialité », et en l'absence d'étude détaillée comparative sur l'utilisation des métaphores entre plusieurs sciences sociales, nous resterons ici prudent quant à cette idée d'une spécificité géographique.

Une hypothèse pour expliquer la fragmentation des travaux sur les métaphores en géographie (qui restent limités à un type de métaphore ou à un champ de la discipline) repose sur les difficultés de détection de ce trope de manière systématique autant manuellement qu'automatiquement. L'objectif de cet article est de s'attaquer à cette problématique globale de la reconnaissance des métaphores en se tournant vers la linguistique et le Traitement Automatique des Langues (TAL).

<sup>3</sup> Jean-Claude Passeron, « La forme des preuves dans les sciences historiques », *Revue européenne de sciences sociales*, vol. 39, n° 120, p. 31-76, 2001.

<sup>4</sup> *Ibid.*, p. 101.

<sup>5</sup> Jocelyn Benoist et Fabio Merlini, *Historicité et spatialité : recherches sur le problème de l'espace dans la pensée contemporaine*, Paris, Vrin, 2001, p. 110.

Plus qu'un simple état des lieux de la question dans ces autres disciplines, il s'agit d'une synthèse avec une lecture assez géographique de cette problématique. Le cadre théorique a pour base la considération d'un texte comme un espace dans lequel il est possible de calculer des distances entre les mots. Si on se place à l'échelle d'un article de géographie, il est difficile, voire quasi impossible, dans cet espace seul et sans aucune ressource extérieure (connaissance d'une personne ou d'une base de données), de réaliser un processus de détection des métaphores. Il est nécessaire de se déplacer à partir de ce texte dans un autre espace où il sera plus facile de détecter des distances indicatrices de métaphores. Cependant les définitions de la métaphore étant multiples depuis Aristote jusqu'aux penseurs modernes, il n'existe pas un espace de référence dans lequel se placer mais une multiplicité d'espaces possibles.

L'objectif de cet article est tout d'abord de mieux caractériser ces espaces potentiels de détection. En effet, chaque espace possible est défini par une base référentielle impliquant une méthode et des résultats spécifiques. Comment mieux définir dans chaque espace la distance qu'on calcule? Quelles sont les forces et les faiblesses de chaque positionnement? Il s'agit d'essayer de pousser en quelque sorte chaque méthode dans ses retranchements pour voir ce que chacune peut réellement apporter ou ne peut pas apporter. Contrairement à beaucoup d'approches qui ne traitent que de la détection manuelle ou que de la détection automatique, l'angle choisi a été d'essayer d'articuler ces deux approches : peut-il y avoir une continuité ou y a-t-il des problèmes que l'approche automatique ne peut pas surmonter de la même manière que la détection manuelle, obligeant à essayer de trouver des solutions dans d'autres espaces?

## 1. Les espaces liés à des déviances

### 1.1. Les déviances par rapport à l'usage

Aristote est un des fondateurs de ces espaces liés à des déviances. En effet, pour lui, la principale fonction des métaphores est de créer un « écart par rapport à l'usage<sup>6</sup> ». Il est intéressant de noter que cet auteur précise qu'en rhétorique, à la différence de la poétique, cet écart doit passer inaperçu pour s'exprimer avec un style qui « ne soit ni bas ni d'une dignité excessive, mais convenable<sup>7</sup> » tout en maintenant l'intérêt du récepteur. Un des instruments privilégiés pour obtenir cet objectif rhétorique est la métaphore. L'enjeu de la détection de cette figure de style réside donc dans la reconnaissance de cet écart dont l'objectif initial était de passer inaperçu. Même si la culture de la rhétorique peut paraître un cadre désuet, il est intéressant de rappeler cet aspect originel caché qui peut faire écho à quelques difficultés bien actuelles de la reconnaissance des métaphores. Il est possible dans cette optique de l'« écart à l'usage » de faire appel tout d'abord à la culture, au bagage cognitif de celui à qui a été donné la tâche de détection des métaphores. Plus précisément, il est alors fait appel à une dissonance cognitive de l'analyste entre l'usage habituel qu'il connaît du mot (ou de l'expression) et son usage dans le texte. Le problème d'une telle méthode est qu'elle donne des résultats fort variables suivant les annotateurs.

Une des variables importantes est la culture de l'annotateur, et plus particulièrement son bagage cognitif qui peut être fort différent d'un individu à l'autre. Il faut toutefois souligner que même en faisant l'effort de prendre des annotateurs de même culture avec un niveau d'éducation similaire, les résultats d'annotations peuvent toujours être fort variables. En effet, qu'est-ce qu'on identifie comme « usage », terme qui est une norme assez floue de référence? à partir de quel seuil d'écart considère-t-on qu'il y a une métaphore? Tout écart est-il une métaphore et, sinon, quels critères supplémentaires est-il nécessaire de préciser?

<sup>6</sup> Aristote, *Rhétorique*, Paris, Flammarion, 1404 b.

<sup>7</sup> *Ibid.*, 1404 b.

Ces questions renvoient à des définitions de la métaphore pouvant être fort différentes et qu'il est difficile d'unifier. On peut chercher à essayer de donner plus d'informations par rapport à ce qu'on appelle « usage », « écart » et « métaphores », mais il reste toujours des interrogations face à la multiplicité des cas possibles. Certaines métaphores peuvent être surévaluées du fait de filtres préconçus et non conscientisés de la part de l'annotateur. Pour limiter ce problème, le recours à plusieurs annotateurs et à la discussion sur des cas jugés différemment s'avère nécessaire.

Le recours aux statistiques amène quelques éléments intéressants face à la problématique de l'écart à l'usage. En effet, il est possible de raisonner sur des grands corpus, par exemple des millions de page Web, censés représenter l'usage. Malgré sa taille importante, le choix d'un tel corpus reste problématique et peut passer à côté de certains usages pas forcément présents dans la base choisie. Ensuite, les scores d'association permettent de quantifier la force du lien entre deux unités. Plus le score est élevé, plus l'affinité sémantique entre les unités en question est grande dans le corpus textuel étudié. Cependant, il existe un grand nombre de calculs de scores d'association et rentrer dans leur détail dépasserait le cadre de cet article. Il faut garder à l'esprit que l'on obtient des valeurs toujours dépendantes de la construction choisie. À partir de ces résultats, il est possible de détecter dans un nouveau corpus des associations qui s'écartent de l'usage. Par exemple, quand Isabelle Lefort intitule son article « Le terrain : l'Arlésienne des géographes?<sup>8</sup> », les trois mots principaux présentent une combinaison d'un usage inhabituel. Le choix du seuil statistique permettant de discriminer l'usage habituel de l'usage inhabituel est important. Cette méthode ne peut de toute façon s'appliquer qu'aux nouvelles métaphores ou à celles qui sont très peu utilisées. Son avantage est de pouvoir éclairer des métaphores novatrices, créatrices de sens, vives qui n'ont pas encore été répertoriées. Un des problèmes majeurs est de distinguer les associations inhabituelles qui sont métapho-

<sup>8</sup> Isabelle Lefort, « Le terrain : l'Arlésienne des géographes? », *Annales de géographie*, n° 687-688, 2012, p. 468-486.

riques de celles qui ne le sont pas. Il ne peut s'agir que d'un premier tri qui devra être complété par d'autres méthodes.

Une variante intéressante permettant d'élargir le spectre de recherche consiste à s'intéresser aux associations d'unités lexicales et à les relier à des catégories pour les analyser. Si on prend par exemple l'adjectif « noir », on va trouver comme nom associé « cheveux », « yeux », « individu », « café », « table ». Grâce à l'utilisation d'une ressource lexicale, on remonte ensuite aux catégories. C'est surtout des « parties du corps », des « personnes », des « objets » ou de la « nourriture ». Des expressions comme « pensée noire » ne font pas partie de ces catégories et sont donc plus susceptibles d'être métaphoriques. Cette méthode utilisée par Dan Assaf<sup>9</sup> en complément d'un autre espace que nous verrons par la suite pose plusieurs problèmes : tout d'abord en amont, le choix de la base de données lexicales avec des informations ontologiques. Ensuite, à l'intérieur de l'algorithme, le choix du seuil statistique qui doit sélectionner ce qui est considéré comme écart à l'usage. L'hypothèse sous-jacente qui peut être discutable est que l'usage courant est moins métaphorique que les usages exceptionnels. Les métaphores les plus conventionnelles sont ainsi exclues de la recherche. Enfin, à l'aval, il y a nécessairement une analyse à faire pour déterminer les erreurs de détection. On est donc encore dans cet espace dans une méthode semi-automatique de détection. Cependant, le recours à des catégories permet d'élargir la détection et de ne pas se limiter seulement aux métaphores novatrices.

Pour finir avec cet espace des écarts à l'usage, il est possible de changer d'échelle et de raisonner simultanément sur des catégories plus larges et sur l'ensemble du texte étudié. Pour l'illustrer, citons cette étude sur des textes de vulgarisation de science de Marc Alexander<sup>10</sup>. Grâce à l'utilisation d'un thésaurus, la détection

<sup>9</sup> Dan Assaf *et al.*, « Why “Dark Thoughts” Aren't Really Dark: A Novel Algorithm for Metaphor Identification », *Proceedings of the IEEE Symposium Series on Computational Intelligence*, Singapore, 2013, p. 60-65.

<sup>10</sup> Marc Alexander *et al.*, « Metaphor, Popular Science, and Semantic Tagging: Distant Reading with the Historical Thesaurus of English », *Digital Scholarship in the Humanities*, n° 30, 2015, p. 16-27.

des répétitions autour d'un thème éloigné du thème principal, par exemple des termes évoquant le domaine du textile dans un texte sur le cosmos, correspondent à une forte probabilité qu'il s'agisse d'une métaphore filée. Il est possible sur des textes dont on ne connaîtrait pas le thème initial d'utiliser des outils statistiques (à partir de bases de données comme ProxyDocs ou à partir de méthodes autonomes comme les modèles de thématiques) pour le déterminer et ensuite trouver les thèmes secondaires. Là encore, tout thème secondaire n'est pas forcément métaphorique (la méthode marche bien dans la recherche de Marc Alexander car le thème est très spécialisé et l'aspect de vulgarisation favorise explicitement l'emploi de métaphores éloignées du thème principal), mais cette méthode associée à une expertise humaine est particulièrement intéressante car elle peut rapidement donner des pistes sur des métaphores filées. Par exemple, en appliquant le logiciel Tropes à un texte de Vidal de la Blache sur les genres de vie, la catégorie « conflit » révèle toute une série de métaphores (mais pas seulement) sur une vision de la nature comme lutte pour la survie avant tout. Ces métaphores filées sont particulièrement importantes car elles jouent un rôle d'intégration conceptuelle au sein des sciences humaines et sociales. Avoir un outil qui peut servir pour une pré-détection rapide de quelques-unes de ces métaphores peut être intéressant, mais on est loin de l'objectif initial de détection systématique.

## 1.2. Les déviations par rapport à des règles syntaxiques

L'utilisation de l'espace syntaxique pour détecter des métaphores est basée sur l'hypothèse que cette figure de style s'accompagne souvent d'une « violation des sélections préférentielles<sup>11</sup> ». Pour comprendre cette notion, le plus simple est de partir d'un exemple comme « ma voiture boit de l'essence ». On fait appel à une première base de données qui nous indique que le verbe « boire » a habituellement pour sujet « une personne » et pour objet « un liquide ». On fait appel à une deuxième base de

<sup>11</sup> Cette notion provient des travaux de Yorick Wilks (« Making Preferences more Active », *Artificial Intelligence*, vol. 11, n° 3, 1978, p. 197-223).



données qui nous indique que « voiture » n'est pas une personne. Il y a manifestement une « violation de sélection préférentielle » qui révèle une utilisation métaphorique du verbe « boire ». La distance mise en évidence est ici une déviation sous forme booléenne (oui ou non, il y a déviation) par rapport à des règles syntaxiques précisant l'utilisation de catégories prédéfinies dans des cadres précis (patrons).

Cet espace se prête bien à une automatisation réalisée dès 1991 par Dan Fass<sup>12</sup> mais repose entièrement sur la qualité des bases de données. Par exemple, si « essence » est catégorisée en « source d'énergie », le système va détecter une « violation de sélection préférentielle » avec l'association précédente au verbe « boire ». Alors que si « essence » est catégorisée comme « liquide », le système ne va pas détecter de métaphore sur cette partie. Cet espace tend donc à repousser les problèmes sur le choix d'une base de données lexicales avec des informations ontologiques de qualité, ce qui n'a rien d'évident. Quelle base choisir (quand il en existe... car les situations peuvent être très différentes suivant les langues)? Pourquoi ce choix et avec quelles limites? Comment actualiser la base de données? Dans notre exemple, si de nouvelles boissons apparaissent, il faut qu'elles soient référencées pour être reconnues. De plus, cette méthode ne détecte que quelques métaphores (essentiellement celles liées à certains verbes) et peut sur-générer des cas non métaphoriques : une utilisation incongrue, qu'elle soit voulue ou fruit d'une erreur, va être marquée comme métaphorique. Si la prise en compte des métonymies comme « boire un verre » a été réalisée, bien des cas restent problématiques : comment le système peut-il faire face à des formules comme « boire n'importe quoi » où il n'y a pas de référent explicite et où il est nécessaire de prendre en compte le contexte?

Ces remarques expliquent plusieurs pistes actuelles de travail. Tout d'abord, tout un ensemble de recherches vise à auto-construire ces bases de données à partir d'internet : soit en utilisant des recherches de patrons (par exemple, « des boissons telles

<sup>12</sup> Dan Fass, « Met\*: A Method for Discriminating Metonymy and Metaphor by Computer », *Computational Linguistics*, vol. 17, n° 1, 1991, p. 49-90.

que X ». Le système va alors en déduire qu'il faut classer X dans la catégorie « boisson »), soit par des méthodes de rapprochement de mots utilisés dans le même contexte. Cette voie de recherche nous semble particulièrement intéressante pour des recherches dans des langues où il n'existe pas d'ontologie de référence et pour faire des actualisations de bases existantes qui sont toujours coûteuses en temps et en énergie. Une autre piste de travail lancée par Ekaterina Shutova et Sun Lin est l'utilisation de cet espace par défaut, plus pour éliminer ce qui n'est pas métaphore que pour détecter celles-ci<sup>13</sup>. L'idée est de filtrer les verbes qui ont une faible sélection de préférence, c'est-à-dire qui s'associent à beaucoup de classes d'arguments, comme ayant un petit potentiel métaphorique. Par exemple, le verbe « se rappeler de » : on peut potentiellement se rappeler de tout, donc c'est un verbe qui a un faible potentiel métaphorique. Enfin, les recherches de Dirk Hovy ouvrent une nouvelle voie intéressante dans cet espace<sup>14</sup>. À partir d'un corpus dans lequel des phrases littérales et métaphoriques ont déjà été identifiées, une machine à vecteurs de support (en anglais, *Support Vector Machine*) a été entraînée en prenant en entrée une décomposition des phrases littérales et métaphoriques en arbres lexicaux, grammaticaux et sémantiques (à partir de l'ontologie de Wordnet). Ayant gardé 10 % du corpus pour l'évaluation, Dirk Hovy a ensuite testé les résultats de cet apprentissage. L'obtention d'un F-Score de 0,75 nous montre que les propriétés syntaxiques couplées à une compréhension des constructions des phrases et des catégories ontologiques utilisées restent une piste de recherche intéressante pour la détection automatique des métaphores.

---

<sup>13</sup> Ekaterina Shutova et Sun Lin, « Unsupervised Metaphor Identification Using Hierarchical Graph Factorization Clustering », *Proceedings of NAACL 2013*, Atlanta (GA), 2013, p. 978-988.

<sup>14</sup> Dirk Hovy *et al.*, « Identifying Metaphorical Word Use with Tree Kernels », *Proceedings of the First Workshop on Metaphor in NLP*, Atlanta (GA), 2013, p. 52-57.

### 1.3. Les déviances par rapport au sens propre

L'utilisation de cet espace est bien illustrée par une méthode historique de détection manuelle des métaphores développée par le Pragglejaz Group et appelée *Metaphor Identification Procedure*. Après une lecture du texte en entier pour avoir une compréhension globale du sens, il s'agit de déterminer pour chaque unité lexicale si elle possède un sens plus basique que son utilisation dans le texte donnée. La procédure détaille ce qui est entendu par « sens plus basique » : soit plus concret (ce qu'il évoque est plus facile à imaginer, voir, sentir ou goûter), soit plus relié à l'action corporelle, soit plus précis (par opposition à vague), soit historiquement plus ancien. Les sens basiques ne sont pas nécessairement les sens les plus fréquents de l'unité lexicale. Cette dernière remarque nous indique clairement que l'on n'est pas dans le même référentiel que l'espace des écarts à l'usage. En cas de doute d'un examinateur (ou entre les examinateurs), la référence pour déterminer les sens d'un mot et pour trancher *in fine* sur le fait de savoir si, oui ou non, son utilisation dans le texte dévie d'un sens plus basique est l'utilisation de dictionnaire. Dans cet espace, la distance appréciée se fait à l'échelle de l'unité lexicale (c'est-à-dire le plus souvent le mot) et se réduit à un modèle booléen (existence d'une déviation par rapport à un sens plus basique ou non). L'objectif en adoptant cette procédure systématique avec une ressource externe stabilisée (un dictionnaire) et un choix binaire (métaphorique ou non métaphorique) est d'avoir la méthode la plus fiable et reproductible possible.

La problématique centrale est de savoir si, dans cet espace et avec cette technique, on n'identifie pas de manière générale le sens figuré des unités lexicales, et non spécifiquement le trope de la métaphore. Les débats que pourrait soulever cette remarque renvoient en toile de fond à une définition plus ou moins stricte ou large de la métaphore (schématiquement toute figure de style créant un rapport imagé ou bien une figure de style basée plus spécifiquement sur une projection conceptuelle). Cet espace a toutefois pour avantage d'offrir la plus grande objectivité, ce qui explique son utilisation et sa reconnaissance. L'Amsterdam Vu

Corpus issu de cette méthode est un corpus annoté de références et il existe une communauté active sur ce sujet<sup>15</sup>. Mais il faut souligner que dans le résultat obtenu, la compréhension des métaphores n'est pas directement lisible dans le texte annoté. Si on considère l'étude des métaphores comme une porte d'entrée pour s'intéresser aux sens implicites d'un texte, un travail de réinterprétation est nécessaire.

Il est difficile d'automatiser cette méthode car s'il existe bien des algorithmes de désambiguïsation de mots permettant de trouver le sens d'un mot grâce à ses co-occurents, ceux-ci ne sont pas assez performants pour s'appliquer à chaque mot. En imaginant qu'on sorte du cadre mot-à-mot pour essayer de n'appliquer cette méthode qu'aux mots les plus spécifiques d'un texte (ce qui est d'un point de vue méthodologique facilement automatisable et ce qui est plutôt légitime car ce sont les unités les plus saillantes de l'espace textuel étudié), rien ne nous dit que l'on va avoir assez d'indications pour préciser le sens de ces mots spécifiques. En effet, l'analyse des co-occurents peut être dans bien des cas trop vague pour pouvoir inférer que le mot est utilisé dans une certaine définition prédéfinie. Ensuite, comment savoir si la définition retenue est la plus basique? Prendre comme critère la première trouvée ne peut pas être toujours satisfaisant car l'ordre du dictionnaire ne suit pas forcément cette logique. Enfin, les métaphores explorent parfois des nouveautés qui n'ont pas encore été consignées dans le dictionnaire.

#### 1.4. Les déviances par rapport à un sens connu

Une direction de recherche intéressante est celle de la néosémie car elle permet de sortir du caractère fermé des définitions d'un dictionnaire et de s'ouvrir à l'intégration de nouveaux sens. Ainsi, de nouvelles métaphores intéressantes peuvent être potentiellement détectées. Le principe de base réside dans la transformation des définitions d'un dictionnaire classique lexicographique en traits de contenu. Ce passage nécessite une réduction des données qui n'est pas évidente mais a été réalisé par exemple pour le *Trésor de*

<sup>15</sup> Site du Metaphor Lab Amsterdam : <http://metaphorlab.org/>.

*la langue française informatisé*. À partir de ce travail, Coralie Reutenauer s'appuie sur le co-texte d'un terme pour déterminer si son sens existe déjà ou pas. Par exemple, le terme « tsunami » est réduit à :

- Phénomène et éléments naturels : mer / vague / terre / côte / volcan / falaise / glacier
- Mouvement : vague/ onde / amplification puis effondrement / déferlement / propagation
- Négatif : catastrophe / provoquer des dégâts / danger
- Ampleur : étendue géographique / démesure / fort impact<sup>16</sup>

L'apparition du terme « tsunami » dans le domaine financier pour qualifier certaines crises importantes peut être identifiée comme un trait de contenu non existant du dictionnaire. Elle active des traits de contenus existants (Mouvement, Négatif, Ampleur) mais son co-texte est suffisamment éloigné du thème principal (phénomène et éléments naturels) pour être repéré comme une déviance du sens existant. La projection métaphorique pouvant servir à éclairer de l'inconnu à partir du connu, cette méthode permet de mettre à jour de nouvelles métaphores intéressantes. Toutefois, dans cet espace de la néosémie, on ne repère pas forcément des métaphores (cela peut être juste des évolutions de l'utilisation sémantique d'un mot) et la méthode ne marche que pour des évolutions sémantiques marquées.

Une autre démarche particulièrement intéressante consiste à se passer d'une base de données lexicale de référence en se servant directement du co-texte pour réaliser des agrégations par différentes méthodes statistiques. Dans cette optique, on peut citer les travaux d'Armelle Boussidan qui reposent sur des analyses factorielles associées à du *clustering*<sup>17</sup> ou encore les travaux de Jey Han Lau et de ses collaborateurs qui reposent sur la détection

<sup>16</sup> Coralie Reutenauer, *Vers un traitement automatique de la néosémie : approche textuelle et statistique*, thèse de doctorat, Université de Lorraine, 2012, p. 114.

<sup>17</sup> Armelle Boussidan, *Dynamics of Semantic Change: Detecting, Analyzing and Modeling Semantic Change in Corpus in Short Diachrony*, thèse de doctorat, Lyon 2, 2013.

de thématiques<sup>18</sup> pour créer des regroupements que l'on espère porteurs de sens. Si on utilise ces travaux dans une perspective de détection automatique des métaphores, l'hypothèse sous-jacente sur laquelle on s'appuie, mais qui reste problématique, est que les nouveaux sens détectés sont probablement une projection métaphorique des anciens. Il est de plus très difficile de déterminer à partir de quand un nouveau sens perd son sens métaphorique pour devenir un sens second. D'une manière plus générale, on touche là au processus dynamique de la langue et à des appréciations subjectives de passages entre différents états, problème que l'on retrouve plus spécifiquement pour les métaphores avec la gradation : créative - vive - courante - codée - cliché - usée - morte<sup>19</sup>. Néanmoins, si le passage d'un seuil à l'autre restera toujours une appréciation subjective, une analyse quantitative diachronique de ces phénomènes serait particulièrement intéressante.

L'enjeu est de trouver la meilleure méthode pour réaliser des agrégats qui vont être les révélateurs de différentes utilisations sémantiques d'un même terme. Une fois choisie la technique utilisée dans la kyrielle de méthodes possibles, les paramètres d'entrée et de sortie sont importants : quelle taille du co-texte choisit-on en entrée? Quel nombre de regroupements fait-on en sortie? Ces choix réalisés déterminent une partie du travail final et décisif qu'est l'interprétation des regroupements. Le choix du co-texte comme seule référence peut faire passer à côté d'évolutions sémantiques mais le caractère totalement inductif de la démarche est séduisant. On ne projette pas une grille de lecture *a priori* comme dans l'utilisation d'une base de données lexicales dont on ne maîtrise pas la construction. Le travail fastidieux de construction et d'actualisation d'une base disparaît mais il est remplacé par un travail important d'interprétation. Cette méthode s'applique bien sur des exemples précis pour

<sup>18</sup> Jey Han Lau *et al.*, « Learning Word Sense Distributions, Detecting Unattested Senses and Identifying Novel Senses Using Topic Models », *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, Baltimore (MD), 2014, p. 259-270.

<sup>19</sup> Geoffrey Leech, *Semantics*, Harmondsworth, Penguin, 1974.

confirmer une évolution métaphorique pressentie d'un mot mais il semble difficile de l'appliquer systématiquement dans un texte pour détecter des métaphores.

### 1.5. Les déviances par rapport à un sens concret

Cet espace reprend les méthodes vues précédemment (analyse des co-occurrences, analyse factorielle des correspondances, *clustering* ou *topic modeling*) mais les applique sur un corpus sans la dimension diachronique. Il est délicat dans ces analyses synchroniques de déterminer l'agrégat qui correspond au sens le plus concret des autres agrégats. Une des méthodes possibles est d'utiliser des bases de données psycholinguistiques pour récupérer les valeurs de « concrétude » des termes. Notons que ces bases sont établies en demandant à des personnes leur estimation du caractère concret d'un terme et qu'à la suite de ces expériences une moyenne est établie pour déterminer une valeur de « concrétude » associée au terme. Le travail de Peter Turney et de ses collaborateurs sur ce sujet nous montre que ces bases ne sont pas assez complètes pour être utilisées directement et qu'un travail statistique est nécessaire pour obtenir des estimations pour les données manquantes<sup>20</sup>. Le sens concret repose donc avec cette méthode sur une moyenne de jugements subjectifs ou sur des estimations statistiques.

Une variante intéressante introduite par Tomek Strzalkowski et ses collaborateurs est de se servir dans les mêmes bases de données d'une autre variable : les degrés d'« imageabilité »<sup>21</sup>. Ils se dégagent d'une référence problématique au « concret » et peuvent ainsi déterminer l'agrégat qui évoque le plus facilement une image mentale. Cette variante par rapport au critère de la « concrétude » se justifie par le fonctionnement d'une bonne partie des métaphores qui repose sur une représentation imagée.

<sup>20</sup> Peter Turney *et al.*, « Literal and Metaphorical Sense Identification through Concrete and Abstract Context », *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 2011, p. 680-690.

<sup>21</sup> Tomek Strzalkowski *et al.*, « Robust Extraction of Metaphor from Novel Data », *Proceedings of the First Workshop on Metaphor in NLP*, Atlanta (GA), 2013, p. 67-76.

Une méthode originale présentée par Ekaterina Shutova et Sun Lin<sup>22</sup> utilise par exemple l'hypothèse que les concepts concrets et abstraits s'organisent différemment pour faire leurs détections. Les concepts concrets vont s'organiser en arbre (avec une structure hiérarchique du plus spécifique au plus général) alors que les concepts abstraits vont avoir des organisations plus complexes. L'utilisation d'un *clustering* avec différents niveaux de hiérarchie leur permet de repérer ces différences. Même si les auteurs renvoient à des travaux de sciences cognitives pour justifier ces différences de structuration, ces recherches montrent que la détermination du sens concret n'a rien d'évident.

Un des problèmes principaux non résolus est de déterminer ensuite dans les agrégats secondaires ceux qui correspondent à un sens métaphorique ou pas. Ces méthodes nous placent dans l'espace de la polysémie, des distinctions sémantiques dans lequel les métaphores ne sont qu'un sous-groupe qu'il faut réussir à distinguer. Une solution adoptée par Tomek Strzalkowski et ses collègues est de ne pas seulement déterminer le *cluster* le plus « imageable » mais d'éliminer tous ceux qui sont au-dessus d'un certain seuil statistique<sup>23</sup>. Le choix de ce seuil ne va pas sans poser de nombreuses questions. De plus, à l'opposé, pour qu'un sens soit identifié comme nouvel agrégat représentatif, il faut un co-texte suffisamment important et différent des co-textes connus. Il est fort probable que certaines métaphores, notamment les nouvelles et celles avec un co-texte insuffisamment distinctif, soient noyées dans certains agrégats. Une des solutions est peut-être, pour détecter ces cas spécifiques, d'adopter des méthodologies spécifiques, plus proches par exemple de l'espace de l'écart à l'usage.

## 2. Les espaces liés à des relations

À la différence des espaces des déviances qui se définissaient par rapport à un référentiel (usage, règle, sens propre, connu ou concret), les espaces des relations abordent la détection de métaphore sans que la définition d'un sol de référence soit nécessaire.

<sup>22</sup> Ekaterina Shutova et Sun Lin, *op. cit.*

<sup>23</sup> Tomek Strzalkowski *et al.*, *op. cit.*



## 2.1. Les relations à partir des polarités

Les métaphores relient fréquemment des éléments avec des polarités différentes : concret/abstrait, animé/inanimé, statique/dynamique... Il est possible d'imaginer des polarités avec plus de deux éléments en prenant par exemple appui sur les cinq sens pour détecter toute une gamme de métaphore comme « la musique qui nous touche ». Toutefois, la polarité qui semble la plus explorée (et on peut penser que son lien avec les espaces des déviances n'y est pas pour rien) est la polarité concret/abstrait. Certes, nous sommes ici à la transition des deux espaces mais il faut bien comprendre que dans cette nouvelle approche la focale est plus portée sur la relation entre deux termes. Pour reprendre l'expression déjà étudiée précédemment « des pensées noires », il y a ici métaphore car on peut identifier l'association du terme abstrait « pensée » avec le terme concret « noire » (et non pas comme vu précédemment parce qu'on s'éloigne d'un écart à l'usage). Le travail de Peter Turney et de ses collaborateurs, déjà cité<sup>24</sup>, utilise en fait ces deux espaces de détection, polarité et usage, montrant ainsi que les espaces s'appuyant sur les déviances et les relations ne sont donc pas incompatibles et peuvent être utilisés pour certains de manière complémentaire en vue d'une meilleure détection des métaphores.

Une autre polarité intéressante est statique/dynamique car elle a fait l'objet de développements théoriques et pratiques, notamment suite aux travaux de Léonard Talmy. Il est intéressant de constater que son exemple phare « la clôture va du plateau à la vallée » pour illustrer la dissonance existant entre une représentation factive (qui correspond à la clôture immobile) et la représentation fictive (qui correspond à la représentation mentale de la clôture se déplaçant) est très géographique<sup>25</sup>. On peut faire ici l'hypothèse que l'expression de cette polarité peut être fortement intéressante pour les géographes et cela d'autant plus qu'elle semble ne pas avoir été étudiée par l'épistémologie de la discipline. Il est intéressant

<sup>24</sup> Peter Turney *et al.*, *op. cit.*

<sup>25</sup> Léonard Talmy, *Toward a Cognitive Semantics*, Cambridge (MA), 2000, MIT Press, chap 2, p. 1.

de constater que d'autres sciences, comme la science médicale, se sont penchées sur ces questions avec les travaux de Sylvie Vandaele et de ses collaboratrices<sup>26</sup>. Ces métaphores permettent de mettre en avant des réseaux lexicaux et des modes de conceptualisation comme par exemple « celui qui attribue aux molécules biologiques et aux éléments cellulaires une volonté, comme des personnes<sup>27</sup>[sic] ». Toutefois, l'approche reste manuelle et aucune tentative d'automatisation ne semble avoir été réalisée.

Un des principaux problèmes liés à une telle voie de recherche est que le caractère statique ou dynamique d'un terme peut changer suivant les contextes, ce qui complexifie la tâche de l'automatisation. Si un géographe décrit les montagnes devant lui en disant qu'« elles s'élèvent », on est dans un registre métaphorique mais s'il parle du processus de formation des massifs liés à la tectonique des plaques sur le long terme, l'élévation des montagnes perd son sens métaphorique. D'un point de vue plus global, aborder la détection automatique de métaphores par le prisme des polarités fait courir le risque de n'attraper qu'un type de métaphores. On peut imaginer des recherches impliquant des combinaisons de polarités pour élargir le spectre de recherche mais certaines métaphores reposant sur des polarités très spécifiques passeront toujours à travers les cribles mis en place. Il n'en reste pas moins que cette méthode peut mettre à jour des métaphores, en particulier des très conventionnelles et lexicalisées qui peuvent poser des problèmes de détection pour d'autres méthodes.

## 2.2. Les relations à partir des analogies strictes

Cet espace a aussi comme fondateur Aristote qui transpose au langage le principe de l'analogie mathématique : un premier élément est à un deuxième ce qu'un troisième est à un quatrième. On peut citer pour mieux comprendre cet exemple de Roger Brunet, Rober Ferras et Hervé Théry : « Le territoire est à

<sup>26</sup> Sylvie Vandaele *et al.*, « La conceptualisation métaphorique en biomédecine : indices de conceptualisation et réseaux lexicaux », *Glottopol*, vol. 8, 2006, p. 73-94.

<sup>27</sup> *Ibid.*, p. 82.

l'espace ce que la conscience de classe est à la classe : quelque chose que l'on intègre comme partie de soi, et que l'on est donc prêt à défendre<sup>28</sup> ». Au-delà de ces formulations très spécifiques, il est possible de trouver dans des métaphores cette combinaison d'un quadruplet de manière sous-jacente. De manière assez littéraire, on peut citer l'exemple de Yann Desalle qui décompose par exemple la formule métaphorique « déshabiller une pomme » par le quadruplet « peler:pomme :: déshabiller:personne »<sup>29</sup>. De manière plus conceptuelle, on peut citer l'approche de Roger Tourangeau et Robert J. Sternberg qui considèrent une métaphore comme le rapprochement de termes de deux sous-espaces distincts avec une position équivalente de ces termes dans leur espace respectif (l'idée du quadruplet se retrouve en toile de fond)<sup>30</sup>. Une difficulté vient du fait que le terme « analogie » est aussi utilisé de manière étendue pour désigner plus généralement des figures fondées sur des rapports de ressemblance. On sort alors de l'analogie mathématique et la métaphore devient un cas particulier d'analogie (qui comprend aussi les comparaisons, les personnifications, les allégories entre autres). À l'inverse, dans l'approche issue des mathématiques et d'Aristote, l'analogie est un cas particulier de métaphore.

Au niveau des bases de données, même en s'orientant vers les lexicographies les plus avancées avec des systèmes de graphes représentant les différents liens entre les lexies, il est difficile d'extraire des analogies strictes comme le montre le travail de Sandrine Ollinger sur le Réseau Lexical du Français<sup>31</sup>. Au niveau des méthodologies, une piste intéressante est le « plongement de

<sup>28</sup> Roger Brunet, Robert Ferras et Hervé Théry, *Les mots de la géographie. Dictionnaire critique*, 3<sup>e</sup> édition, Paris, Reclus, La Documentation française, 1993, p. 480.

<sup>29</sup> Yann Desalle, *Réseaux lexicaux, métaphore, acquisition : une approche interdisciplinaire et inter-linguistique du lexique verbal*, thèse de doctorat, Toulouse 2, 2012.

<sup>30</sup> Roger Tourangeau et Robert J Sternberg, « Understanding and Appreciating Metaphors », *Cognition*, vol. 11, n° 3, 1982, p. 203-244.

<sup>31</sup> Sandrine Ollinger, *Le raisonnement analogique en lexicographie, son informatisation et son application au Réseau Lexical du Français*, thèse de doctorat, Université de Lorraine, 2014.

mot » (en anglais *word embedding*) qui, à partir de corpus volumineux et de raisonnements sur les vecteurs, arrive à trouver des analogies. Si les tests de cette technique montrent qu'elle marche assez bien pour des analogies simples, des recherches sur des analogies plus complexes restent à faire. Quoiqu'il en soit, les métaphores reposant sur des analogies strictes de manière directe ou sous-jacente (comme dans l'exemple de Yann Desalle) représentent une faible partie des métaphores. Si on étend la définition, on sort de cet espace spécifique de recherche. Comme le souligne Christian Plantin, la recherche par analogie peut alors perdre de sa pertinence car « tout est analogue à tout sous l'un ou l'autre aspect<sup>32</sup> ». Il semble alors nécessaire de préciser à quel niveau se joue la ressemblance dans l'analogie pour essayer d'améliorer la détection des métaphores : on peut se situer au niveau soit des concepts, soit des attributs, soit des sèmes.

### 2.3. Les relations à partir des concepts

Dans l'espace des relations, l'espace conceptuel est le plus développé pour la détection des métaphores suite aux travaux de George Lakoff et Mark Johnson<sup>33</sup>. Il s'agit de déterminer le domaine source et le domaine cible de chaque expression métaphorique définie comme une projection de l'un sur l'autre. Cela peut conduire à une meilleure compréhension de ce qui se joue derrière l'utilisation de cette figure de style. Par exemple, pour reprendre leur terminologie, si on a une métaphore du type « Le processus d'intégration à l'Union Européenne est un train », alors ce train peut rouler mais il peut aussi s'arrêter, avoir plusieurs wagons qui déraillent... En se plaçant ainsi à un niveau conceptuel, on saisit mieux les enjeux possibles de l'utilisation d'une telle image. On peut ainsi faire le repérage des métaphores de manière inductive et manuelle. Dans le cas où aucune autre consigne n'est donnée, il y a fort à parier que le repérage dépende beaucoup du

<sup>32</sup> Christian Plantin, « Analogie et métaphore argumentatives », *A contrario*, n° 16, 2011, p. 110-130, paragraphe 93, <http://www.cairn.info/revue-a-contrario-2011-2-page-110.htm>, consulté le 28 mars 2017.

<sup>33</sup> George Lakoff and Johnson Mark, *Metaphors We Live by*, Chicago, University of Chicago Press, 2008 [1980].

bagage cognitif de l'annotateur. En effet, avant même d'évaluer la distance entre deux concepts, il est nécessaire d'identifier ces concepts qui sont souvent sous-entendus dans une métaphore (à la différence d'une comparaison). Gerard Steen<sup>34</sup> essaye de faire le pont avec la procédure *Metaphor Identification Procedure* précédemment vue et essaye de clarifier une méthode d'identification des domaines source et cible. Si cet effort est louable, il n'en reste pas moins que, dans la pratique, il n'est pas évident d'identifier ces domaines source et cible comme le montre la rareté des corpus annotés dans cette direction pourtant fort porteuse de sens.

Si on continue de se placer dans la théorie de George Lakoff et Mark Johnson, toutes les métaphores sont conceptuelles et elles mettent toutes en jeu des champs conceptuels préétablis. Il découle de cette approche qu'on peut appliquer une démarche déductive en recherchant directement dans les textes ces champs conceptuels préétablis. Ainsi, Alice Deignan<sup>35</sup> et Anatol Stefanowitsch<sup>36</sup> cherchent respectivement à partir de l'établissement de lexiques clés de ces champs préétablis leurs présences dans des textes. Alice Deignan utilise ensuite un concordancier pour voir si ces termes sont utilisés métaphoriquement ou pas. On est par conséquent dans une méthode de détection semi-automatique. Par ailleurs, Éric Baumer *et al.*<sup>37</sup> essayent également de descendre du niveau conceptuel dans les textes en s'appuyant sur le repérage de structures communes (comme « divisé en deux arches » dans le domaine de l'architecture qui se retrouve dans les textes de biologie sous la forme « divisé en deux

<sup>34</sup> Gerard J. Steen, « From Linguistic to Conceptual Metaphor in Five Steps », *Amsterdam Studies in the Theory and History of Linguistic Science*, series 4, 1999, p. 57-78.

<sup>35</sup> Alice Deignan, *Metaphor and Corpus Linguistics*, Amsterdam, John Benjamins Publishing Company, 2005.

<sup>36</sup> Anatol Stefanowitsch, « Words and Their Metaphors: A Corpus-Based Approach », *Corpus-Based Approaches to Metaphor and Metonymy*, Berlin, Mouton de Gruyter, 2006, p. 63-105.

<sup>37</sup> Eric Baumer, Bill Tomlinson et Lindsey Richland, « Computational Metaphor Identification: A Method for Identifying Conceptual Metaphors in Written Text », *Proceedings of Analogy '09*, Sofia, 2009, p. 20-29.

cellules ») pour identifier des correspondances métaphoriques. Comme le montre cet exemple, on est plus alors dans la correspondance métaphorique que dans la métaphore proprement dite.

La voie de recherche la plus porteuse semble être de remonter des textes vers les concepts. Deux directions sont possibles : soit en utilisant d'autres méthodes de détections et en essayant ensuite, avec l'aide de ressources lexicales ou/et l'aide de méthodes statistiques, de remonter des métaphores déjà identifiées vers les concepts. Mais cela ne résout pas en amont le problème de l'identification automatisée des métaphores dans le texte. Soit en partant directement des textes : par exemple, Ekaterina Shutova et Sun Lin<sup>38</sup> utilisent des méthodes de *clustering* sur les associations verbes-noms comme « entretenir une relation », « rompre une relation »... pour essayer de trouver des métaphores comme une « relation de couple est un mécanisme ». Ilana Heintz *et al.* utilisent, quant à eux, les modèles de thématiques pour essayer de trouver les métaphores à partir du concept cible de « gouvernance » et de différents concepts sources prédéterminés<sup>39</sup>. Ces exemples nous montrent que les auteurs essaient de décomposer le problème soit en se concentrant sous une forme syntaxique, soit en fixant les concepts *a priori*. Le champ de recherche est ouvert pour trouver la ou les meilleures méthodes statistiques de détection.

#### 2.4. Les relations à partir des attributs

Les champs conceptuels préétablis de la théorie de George Lakoff et Mark Johnson, recensés et organisés au sein de la *Metaphor Master List*, ont été assez critiqués, par exemple par Birte Lönneker-Rodman<sup>40</sup>. Ils permettent à ces auteurs de montrer que tout notre système conceptuel est sous-tendu et organisé par des

<sup>38</sup> Ekaterina Shutova et Sun Lin, *op. cit.*, p. 978-988.

<sup>39</sup> Ilana Heintz *et al.*, « Automatic Extraction of Linguistic Metaphors with LDA Topic Modeling », *Proceedings of the First Workshop on Metaphor in NLP*, Atlanta (GA), 2013, p. 58-66.

<sup>40</sup> Birte Lönneker-Rodman, « The Hamburg Metaphor Database Project: Issues in Resource Creation », *Language Resources and Evaluation*, vol. 42, n° 3, 2008, p. 293-318.

métaphores. Toutefois, même si les champs sont assez vastes pour contenir toutes les métaphores qu'ils ont répertoriées, on peut considérer qu'une telle approche rend mal compte de l'inventivité de ce trope et de son processus dynamique de renouvellement. Des approches psycholinguistiques sont venues combler cette lacune sortant les métaphores de ces champs préétablis en les envisageant comme des processus cognitifs dynamiques. Cependant, comme le montre Yann Desalle<sup>41</sup>, ces théories ont eu tendance à se concentrer sur les métaphores nominales prédictives du type « A est un B ». Dans ces théories (*Salience Imbalance* d'Andrew Ortony, *Structure Mapping* de Dedre Gentner et Arthur B. Markman, *Class-inclusion* de Sam Glucksberg et *Approximate Semantic Transference* chez Bipin Indurkha), celle qui a été la plus développée dans le cadre de la détection automatique de métaphores est celle de Dedre Gentner<sup>42</sup> à travers la mise en place d'une modélisation informatique : le « Structure Mapping Engine (SME) ».

L'objectif est de rechercher des parallélismes de structure entre des attributs de composition, d'organisation ou de fonction. Sans rentrer dans les détails, les « racines des arbres » peuvent être détectées comme ayant une connexion métaphorique avec des « pailles » car les deux partagent des attributs d'organisations (elles sont cylindriques et allongées) et de fonction (elles transportent du liquide). Ce système demande de rentrer en amont les attributs, ce qui nécessite un temps long de constitution d'une base de données. En aval, ce sont seulement des rapprochements préexistants de structure qui sont trouvés. De plus, tous les rapprochements ne sont pas des métaphores intéressantes. Enfin, comme le souligne Giovanni Busino dans son article « La place de la métaphore en sociologie » : « L'analogie fonctionnelle et l'homologie sont rares car la similarité caractérisant des objets semblables est normalement indéfectible. L'isomorphisme de forme ou de structure entre deux systèmes à contenus différents

<sup>41</sup> Yann Desalle, *op. cit.*

<sup>42</sup> Dedre Gentner et Arthur B. Markman, « Structure Mapping in Analogy and Similarity », *American Psychologist*, vol. 52, n° 1, 1997, p. 45-56.

n'est presque jamais réalisé<sup>43</sup> ». La recherche dans cet espace doit donc s'orienter vers des rapprochements incomplets d'attributs posant la problématique du seuil à partir duquel un rapprochement est pertinent.

Pour toutes les raisons citées ci-dessus, cet espace de recherche n'est plus utilisé dans les recherches actuelles mais pourrait encore être adapté pour des usages très précis comme par exemple la décomposition des métaphores traditionnelles pour améliorer la détection globale de l'ensemble des métaphores d'un texte. Par exemple, il est possible de donner l'attribut « négatif » pour le terme « noir » quand il est activé avec une combinaison d'un terme abstrait et d'un terme concret comme on l'a vu précédemment à partir de l'exemple de « pensée noire ». Ces attributs pourraient être mis en relation ensuite avec des traits de contenus (exemple précédent de tsunami) pour améliorer la détection des métaphores filées. Il s'agit d'une piste de recherche qui demande en amont un travail important d'identification des métaphores traditionnelles et de leurs attributs. Il est possible de s'appuyer pour cela soit sur des listes existantes de métaphores comme la *Metaphor Master List* (on transforme alors une approche à la base conceptuelle en une approche par attributs), soit sur l'extraction de données dans des dictionnaires électroniques qui indiquent les usages métaphoriques connus de certains mots.

## 2.5. Les relations à partir des sèmes

Cet espace se base sur la théorie de la sémantique interprétative développée notamment par François Rastier. L'identification des métaphores repose sur la notion de sème, ou trait de contenu. Par exemple dans la phrase « cette femme est une fleur », il y a connexion métaphorique car on peut identifier dans les deux termes en jeu (femme et fleur) les sèmes incompatibles /humain/ et /végétal/ alors qu'ils partagent le sème commun /beauté/. Le contexte est central pour définir les sèmes en jeu. Par exemple, dans un contexte médiéval ou contemporain, l'expression « monter au créneau » sera composée de sèmes différents et ne sera pas

<sup>43</sup> Giovanni Busino, *op. cit.*, paragraphe 34.



porteuse du même potentiel métaphorique. Dans cette théorie, les sèmes n'ont pas de référence absolue mais toujours un caractère relatif se caractérisant par une souplesse et une ouverture possible à la nouveauté. L'importance de la prise en compte du contexte va de pair avec une détection qui peut se faire aussi à un niveau supérieur par l'analyse des récurrences des connexions métaphoriques pouvant aboutir dans certains cas à une métaphore filée. Il s'agit ici d'un survol théorique qui demanderait un cadre beaucoup plus grand pour être précisé.

Un premier apport important de cet espace est qu'il permet indirectement de situer les autres espaces en partant de la distinction développée par François Rastier entre tradition logico-grammaticale et tradition rhétorique-herméneutique<sup>44</sup>. Dans la première, la compréhension du sens passe par l'examen d'un ensemble fixe, prédéterminé et composé des valeurs possibles associées au mot qu'on étudie. Le contexte agit surtout par le biais des contraintes liées à la syntaxe pour préciser quelle est la valeur du mot dans la phrase en question. Dans la seconde, la tradition rhétorique-herméneutique, la compréhension du sens passe par l'examen du contexte qui n'est plus réduit à la syntaxe d'une phrase. L'ensemble du texte devient potentiellement porteur d'un sens qui n'est plus donné ontologiquement. Cet axe organise clairement l'espace lié aux déviations par rapport à des règles syntaxiques, ensuite par rapport au sens propre (dans la démarche du *Metaphor Identification Procedure*, il y a une lecture préalable du texte pour prendre en compte le contexte mais l'utilisation des dictionnaires comme ressource externe de référence les rapproche de ce pôle), puis par rapport au sens connu (où il y a prise en compte du contexte avec ouverture possible à la nouveauté). Enfin, l'espace lié aux relations à partir des sèmes se place à l'autre extrémité de cet axe organisateur.

Si on se place dans le cadre théorique strict de la sémantique interprétative, l'automatisation est très difficile car les sèmes ne

---

<sup>44</sup> François Rastier « De la signification au sens. Pour une sémiotique sans ontologie », *Texto!*, juin-sept 2003, [http://www.revue-texto.net/Inedits/Rastier/Rastier\\_Semiotique-ontologie.html](http://www.revue-texto.net/Inedits/Rastier/Rastier_Semiotique-ontologie.html), consulté le 2 février 2017.

sont jamais définis au préalable et dépendent du contexte. Nous avons déjà décrit, dans le cadre de l'espace des déviations, le travail de Coralie Reutanaeur qui permet d'ajouter des sèmes à une base de données existante si le co-texte est suffisamment distinctif des traits de contenus déjà connus. En se plaçant dans l'espace relationnel, le travail est plus délicat car il s'agit de mettre en parallèle plusieurs sèmes qui peuvent être définis par le contexte. Bénédicte Pincemin, dans son article « Sémantique interprétative et analyses automatiques de textes : que deviennent les sèmes? », publié en 1999, montre l'importance de la notion d'isotopie en la matière, c'est-à-dire de récurrence d'un sème dans un même texte<sup>45</sup>. Les différentes méthodes statistiques de recherche de co-occurrences permettent de rechercher ces isotopies. Dans une optique relationnelle, elle cite l'expérience de Ronan Pichon et Pascale Sébillot<sup>46</sup> menée sur un corpus d'articles du *Monde Diplomatique*. Ces auteurs recherchent des groupements de mots au voisinage du terme « militaire » dans différents contextes :

- dans le contexte de « territoire » (à l'exclusion de « négociation »), les résultats sont : moyen, opération, régime, russe, victoire, base, massif, occupation. L'isotopie est plutôt centrée sur le sème /guerrier/ de « militaire ».
- dans le contexte de « négociations » (à l'exclusion de « territoire ») : action, ordre, effort, responsable, pays, dépense, atlantique, Europe, OTAN, organisation. L'isotopie est plutôt centrée sur les sèmes /organisé/ et /structuré/ de « militaire ».

Une voie de recherche possible serait de chercher sur de très gros corpus par exemple les groupements de mot proche de « femme », de « fleur », puis de « femme » dans le contexte de « fleur ». Une des difficultés est d'arriver à déduire des différences trouvées la présence ou non de métaphore. En effet, comme nous

<sup>45</sup> Bénédicte Pincemin, « Sémantique interprétative et analyses automatiques de textes : que deviennent les sèmes? », *Sémiotiques*, n° 17, 1999, p. 71-120.

<sup>46</sup> Ronan Pichon et Pascale Sébillot, « Différencier les sens des mots à l'aide du thème et du contexte de leurs occurrences : une expérience », *6<sup>e</sup> Conférence francophone internationale sur le Traitement Automatique des Langues Naturelles (TALN 99)*, 1999.

le montre l'exemple de « militaire », il peut y avoir des activations de sèmes différents suivant les contextes sans qu'il y ait de métaphore. Il s'agit là d'une difficulté majeure qui explique l'absence de recherche en détection automatique de métaphores (à notre connaissance) qui utilise cet espace spécifique de la sémantique interprétative.

## 2.6. Les relations à partir d'intentions

Cet espace fait rentrer dans la dimension communicative des métaphores. Elle a surtout été développée par Gerard Steen qui met en avant le concept de « métaphore délibérée ». Il nomme sa théorie DMT (*Deliberate Metaphor Theory*)<sup>47</sup> pour faire le pendant à la théorie de George Lakoff et Mark Johnson (*CMT : Conceptual Metaphor Theory*). En effet, Steen met en avant que toutes les métaphores n'activent pas un transfert conceptuel chez les récepteurs contrairement à ce que pourrait laisser penser la CMT. Certaines métaphores, objets de la DMT, au contraire activent fortement un transfert conceptuel, en particulier, selon son analyse, les nouvelles métaphores, celles qui sont filées et celles introduites par des formules marquant leurs caractères imagés. Mis à part ces trois critères, il n'y a pas de définition claire pour attester le fait qu'une métaphore soit délibérée. Cette notion nous fait rentrer dans le domaine de l'intention du locuteur. Ainsi, le langage n'est plus perçu comme un système indépendant comme dans les précédents espaces mais comme le produit d'interactions sociales entre individus. Les disciplines potentiellement mobilisables pour ce type d'étude sont alors la psychologie, la sociologie, la pragmatique... Les approches peuvent être jugées plus subjectives mais aussi ouvrant à une compréhension plus fine de ce qui joue derrière les métaphores.

Du point de vue de la détection manuelle, on peut citer l'étude d'Anke Beger sur des cours de biologie, de chimie, de psychologie et de philosophie<sup>48</sup>. Celle-ci montre que les métaphores délibérées

<sup>47</sup> Gerard J. Steen, « When Is Metaphor Deliberate », in *Selected Papers from the Stockholm*, 2008, p. 43-63.

<sup>48</sup> Anke Beger, « Deliberate Metaphors? An Exploration of the Choice and Functions of Metaphors in US-American College Lectures », *Metaphorik.de*,

sont utilisées pour expliquer des nouveaux concepts, pour résumer des abstractions mais aussi pour convaincre les étudiants. Un des apports de l'étude est également de relier les métaphores à la structure des exposés pour mettre en lumière que les métaphores interviennent plus à certains moments (explication, résultat, théorie et conclusion) que les autres. Il serait intéressant de rentrer plus finement dans la structure argumentative pour essayer d'analyser plus en détail le rôle de ces métaphores volontairement employées. D'un point de vue de la détection automatique, on peut imaginer une approche essayant de travailler sur les trois critères (nouveau, filée, introduite) des métaphores délibérées. Une métaphore serait considérée comme d'autant plus délibérée que la combinaison de ces trois critères serait importante. Cette idée rejoint celle de Jonathan Dunn qui distingue dans sa recherche des niveaux de métaphoricité différents<sup>49</sup>. À la différence des travaux précédents qui envisageaient le problème sous une forme binaire (métaphorique ou littéral), cette nouvelle perspective sous forme d'un gradient peut apporter des résultats intéressants.

## Conclusion

Cet état des lieux nous montre qu'il n'y a pas de solution simple et unique en termes de détection de métaphore. Toutes les pistes évoquées offrent des possibilités mais apportent à chaque fois un lot important de difficultés. Quelques spécialisations ont été mises en lumière : l'espace lié à des déviations par rapport à l'usage pour rechercher les métaphores novatrices, l'espace lié à des relations à partir des attributs pour travailler les métaphores traditionnelles... Nous rejoignons à ce titre le constat de Jonathan Dunn qui réalise une recherche pratique en testant quatre méthodes d'identification sur un même corpus : « *This suggests that there are multiple types of metaphor and that each system is*

---

n° 20, 2011, p. 39-60.

<sup>49</sup> Jonathan Dunn, « Measuring Metaphoricity », *Proceedings of the 52<sup>nd</sup> Annual Meeting of the Association for Computational Linguistics (Short Papers)*, 2014, p. 745-751.

*stronger at identifying one type over another type*<sup>50</sup> ». L'objectif est d'associer les forces de chaque espace d'identification pour améliorer globalement la détection automatique. Cette segmentation du problème renvoie à la difficulté générale de définir avec des critères précis un phénomène aussi complexe et multiforme que les métaphores (allant des clichés redondants à l'imagination verbale, en passant par les jeux contextuels). Dans ce cadre, soumettre la détection d'une procédure de langage figuré à une procédure logico-informatique s'avère particulièrement compliqué et nécessite d'appliquer la méthode cartésienne de division du problème. Cependant si on veut par la suite refaire la synthèse en combinant les différents sous-espaces de détection, une vision globale et d'ensemble du problème reste nécessaire. Un axe organisateur (logico-grammaticale *versus* rhétorique-herméneutique) de plusieurs espaces a été mis à jour afin de permettre une meilleure compréhension des articulations possibles.

Du point de vue de la détection manuelle, trois pôles apparaissent comme dominants : la *Metaphore Identification Procedure* du Pragglejaz Group, l'approche conceptuelle de George Lakoff et Mark Johnson et l'approche « délibérée » de Gerard Steen. Les autres pôles qu'on peut qualifier de secondaires car moins développés théoriquement et pratiquement ne doivent pas être oubliés. En effet, il n'y a pas de continuité simple entre ces trois approches manuelles et des approches automatisées. Si la définition des figures de styles comme déviance est fortement remise en cause par la linguistique moderne<sup>51</sup>, il n'en reste pas moins qu'elle constitue un socle important pour les détections des métaphores comme nous le montre l'importance de la première partie dans cette présentation. À l'inverse, les espaces liés aux

<sup>50</sup> Traduction : « Ceci suggère qu'il existe plusieurs types de métaphores et que chaque système est plus à même d'identifier un type plutôt qu'un autre type », Jonathan Dunn, « Evaluating the Premises and Results of Four Metaphor Identification Systems », *Computational Linguistics and Intelligent Text Processing, Conference Proceedings of CICLing'13*, Samos, 2013, p. 484.

<sup>51</sup> Par exemple, lire Laurent Jenny, « Les figures de rhétorique », *Méthodes et problème*, Genève, Département de français moderne, 2003, <http://www.unige.ch/lettres/framo/enseignements/methodes/frhetorique/>, consulté le 2 février 2017.

relations privilégiés par la linguistique moderne<sup>52</sup> semblent plus difficiles à mettre en œuvre de façon systématique comme nous le montre l'aspect plus réduit de cette seconde partie. Il y a là une discontinuité, une tension intéressante qui justifie pleinement un travail interdisciplinaire de fond entre linguistes, informaticiens et géographes.

## Bibliographie

- Alexander, Marc *et al.*, « Metaphor, Popular Science, and Semantic Tagging: Distant Reading with the Historical Thesaurus of English », *Digital Scholarship in the Humanities*, n° 30, 2015, p. 16-27.
- Aristote, *Rhétorique*, Paris, Flammarion, 2007.
- Assaf, Dan *et al.*, « Why “Dark Thoughts” Aren’t Really Dark: A Novel Algorithm for Metaphor Identification », *Proceedings of the IEEE Symposium Series on Computational Intelligence*, Singapore, 2013, p. 60-65.
- Baumer, Eric, Bill Tomlinson et Lindsey Richland, « Computational Metaphor Identification: A Method for Identifying Conceptual Metaphors in Written Text », *Proceedings of Analogy '09*, Sofia, 2009, p. 20-29.
- Beger, Anke, « Deliberate Metaphors? An Exploration of the Choice and Functions of Metaphors in US-American College Lectures », *Metaphorik.de*, n° 20, 2011, p. 39-60.
- Benoist, Jocelyn et Fabio Merlini, *Historicité et spatialité : recherches sur le problème de l'espace dans la pensée contemporaine*, Paris, Vrin, 2001.
- Boussidan, Armelle, *Dynamics of Semantic Change: Detecting, Analyzing and Modeling Semantic Change in Corpus in Short Diachrony*, thèse de doctorat, Lyon 2, 2013.
- Brunet, Roger, Robert Ferras et Hervé Théry, *Les mots de la géographie. Dictionnaire critique*, 3<sup>e</sup> édition, Paris, Reclus, La Documentation française, 1993.

---

<sup>52</sup> *Ibid.*

- Busino, Giovanni, « La place de la métaphore en sociologie », *Revue européenne des sciences sociales*, vol. XLI, n° 126, 2003, p. 91-101, <https://ress.revues.org/539>, consulté le 2 février 2017.
- Claval, Paul, « Quelques orientations actuelles de la réflexion épistémologique en géographie : systèmes, structures et métaphores », *Paralelo*, vol. 37, n°s 8-9, 1985, p. 171-179.
- Deignan, Alice, *Metaphor and Corpus Linguistics*, Amsterdam, John Benjamins Publishing Company, 2005.
- Desalle, Yann, *Réseaux lexicaux, métaphore, acquisition : une approche interdisciplinaire et inter-linguistique du lexique verbal*, thèse de doctorat, Toulouse 2, 2012.
- Dunn, Jonathan, « Evaluating the Premises and Results of Four Metaphor Identification Systems », *Computational Linguistics and Intelligent Text Processing, Conference Proceedings of CICLing'13*, Samos, 2013, p. 471-486.
- Dunn, Jonathan, « Measuring Metaphoricity », *Proceedings of the 52<sup>nd</sup> Annual Meeting of the Association for Computational Linguistics (Short Papers)*, 2014, p. 745-751.
- Fass, Dan, « Met\*: A Method for Discriminating Metonymy and Metaphor by Computer », *Computational Linguistics*, vol. 17, n° 1, 1991, p. 49-90.
- Gentner, Dedre et Arthur B. Markman, « Structure Mapping in Analogy and Similarity », *American psychologist*, vol. 52, n° 1, 1997, p. 45-56.
- Heintz, Ilana *et al.*, « Automatic Extraction of Linguistic Metaphors with LDA Topic Modeling », *Proceedings of the First Workshop on Metaphor in NLP*, Atlanta (GA), 2013, p. 58-66.
- Hovy, Dirk *et al.*, « Identifying Metaphorical Word Use with Tree Kernels », *Proceedings of the First Workshop on Metaphor in NLP*, Atlanta (GA), 2013, p. 52-57.
- Lakoff, George et Mark Johnson, *Metaphors We Live by*, Chicago, University of Chicago Press, 2008 [1980].
- Lau, Jey Han *et al.*, « Learning Word Sense Distributions, Detecting Unattested Senses and Identifying Novel Senses Using Topic Models », *Proceedings of the 52<sup>nd</sup> Annual Meeting of the Association for Computational Linguistics*, Baltimore (MD), 2014, p. 259-270.
- Lönneker-Rodman, Birte, « The Hamburg Metaphor Database Project: Issues in Resource Creation », *Language Resources and Evaluation*, vol. 42, n° 3, 2008, p. 293-318.

- Ollinger, Sandrine, *Le raisonnement analogique en lexicographie, son informatisation et son application au Réseau Lexical du Français*, thèse de doctorat, Université de Lorraine, 2014.
- Pichon, Ronan et Pascale Sébillot, « Différencier les sens des mots à l'aide du thème et du contexte de leurs occurrences : une expérience », *6<sup>e</sup> conférence francophone internationale sur le Traitement Automatique des Langues Naturelles (TALN 99)*, 1999.
- Reutenauer, Coralie, *Vers un traitement automatique de la néosémie : approche textuelle et statistique*, thèse de doctorat, Université de Lorraine, 2012.
- Shutova, Ekaterina et Lin Sun, « Unsupervised Metaphor Identification Using Hierarchical Graph Factorization Clustering », *Proceedings of NAACL 2013*, Atlanta (GA), 2013, p. 978-988.
- Steen, Gerard J., « From Linguistic to Conceptual Metaphor in Five Steps », *Amsterdam Studies in the Theory and History of Linguistic Science*, series 4, 1999, p. 57-78.
- Steen, Gerard J., « When is Metaphor Deliberate », *Selected Papers from the Stockholm, 2008*, p. 43-63.
- Stefanowitsch, Anatol, « Words and their Metaphors: A Corpus-Based Approach », *Corpus-Based Approaches to Metaphor and Metonymy*, Berlin, Mouton de Gruyter, 2006, p. 63-105.
- Strzalkowski, Tomek *et al.*, « Robust Extraction of Metaphor from Novel Data », *Proceedings of the First Workshop on Metaphor in NLP*, Atlanta (GA), 2013, p. 67-76.
- Turney, Peter *et al.*, « Literal and Metaphorical Sense Identification through Concrete and Abstract Context », *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 2011, p. 680-690.
- Vandaele, Sylvie *et al.*, « La conceptualisation métaphorique en biomédecine : indices de conceptualisation et réseaux lexicaux », *Glottopol*, vol. 8, 2006, p. 73-94.
- Wilks, Yorick, « Making Preferences more Active », *Artificial Intelligence*, vol. 11, n° 3, 1978, p. 197-223.