



HAL
open science

Explication de métaphores via la résolution d'analogies à l'aide d'un graphe de connaissances

Jérémie Roux, Hani Guenoune, Mathieu Lafourcade, Richard Moot

► To cite this version:

Jérémie Roux, Hani Guenoune, Mathieu Lafourcade, Richard Moot. Explication de métaphores via la résolution d'analogies à l'aide d'un graphe de connaissances. JADT 2024 - 17es Journées internationales d'Analyse statistique des Données Textuelles, SeSLa (Séminaire des Sciences du Langage de l'UCLouvain – Site Saint-Louis), en collaboration avec le LASLA (Laboratoire d'Analyse statistique des Langues anciennes de l'Université de Liège), Jun 2024, Bruxelles, Belgique. pp.813-822. lirmm-04643469

HAL Id: lirmm-04643469

<https://hal-lirmm.ccsd.cnrs.fr/lirmm-04643469v1>

Submitted on 10 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Public Domain

Explication de métaphores via la résolution d'analogies à l'aide d'un graphe de connaissances

Jérémie Roux¹, Hani Guenoune¹, Mathieu Lafourcade¹, Richard Moot¹

¹LIRMM, Univ Montpellier, CNRS, Montpellier, France – prenom.nom@lirmm.fr

Abstract

An analogy is a relation which operates between two pairs of terms representing two distant domains. It operates by transferring meaning from a concept that is known to another that one would like to clarify or define. In this report, we address analogy both from the aspect of modeling and by automatically explaining it. We will then propose a system of resolution of analogical equations in their notation in symbol chains. The model, based on the common sense knowledge base *JeuxDeMots* (a semantic network), operates by generating a list of potential candidates from which it chooses the most suitable solution. We conclude by evaluating our model on a collection of equations, and reflecting upon future work.

Keywords: analogy, metaphor, figurative language, natural language processing, knowledge base

Résumé

La relation d'analogie opère entre deux paires de termes représentant deux domaines différents. Elle transfère la connaissance d'un concept connu vers un autre qu'on souhaiterait clarifier ou définir. Dans ce document, nous abordons d'abord les problématiques de modélisation de l'analogie puis celles de l'automatisation de son explication. Nous proposerons ensuite un système de résolution d'équations analogiques notées en chaînes de symboles. Le modèle, basé sur la base de connaissances *JeuxDeMots*, procède en générant une liste de candidats potentiels parmi lesquels il tente de choisir la solution la plus pertinente. Nous finissons par tester le modèle sur une collection d'équations et évoquons des perspectives suite au travail réalisé.

Mots clés : analogie, métaphore, contenu figuré, traitement automatique du langage naturel, base de connaissances

1. Introduction et état de l'art

À partir d'idées et de raisonnements plus ou moins complexes, peuvent émerger de façon non triviale des expressions éloquentes et persuasives, portant des significations nettes ou nuancées. Les productions langagières ainsi que leurs compréhensions sont admises en tant que facultés propres aux humains, capables d'interprétation sémantique de haut niveau. Un locuteur semble se référer de manière inconsciente à un appareil complexe et hiérarchique du fait de sa connaissance du monde, sans que cette réflexion ne demande un effort particulier. Nous cherchons ici à « *atteindre [...] la formalisation d'une opération que tout le monde reconnaît être à l'œuvre dans la langue* » (Lepage, 2003) en vue de la rendre accessible à des procédures automatisées. L'enjeu de notre travail est de réaliser automatiquement des raisonnements sémantiques élaborés.

Nous pouvons décrire l'analogie comme une opération fondamentale de pensée utilisée dans les langues à travers le monde (Gentner et Holyoak, 1997). Au-delà de cette universalité,

l'omniprésence des analogies et des métaphores dans le langage écrit justifie l'intérêt de pousser les recherches dans le domaine. L'annotation manuelle des figures métaphoriques du *British National Corpus* a révélé que 241 phrases sur 761 contenaient ce type de langage (Shutova, 2010). L'étude, bien que marquée d'une dimension idiomatique spécifique à la langue anglaise, représente un indicateur de l'occurrence des analogies dans le langage naturel. Ces chiffres ne peuvent cependant pas être généralisés *a priori* à toutes les langues.

Il est important de souligner le caractère central de l'opération d'analogie dans l'appareil cognitif humain, que ce soit dans ses opérations naturelles fondamentales, ou dans des démonstrations et raisonnements plus complexes et méthodiques (système 1/système 2 de Kahneman). Nous sommes habitués à aborder la résolution d'un problème difficile ou inédit en tentant de le réduire à un autre dont la solution serait connue. Hofstadter (1995, 2001) estime que la pensée et l'analogie sont indissociables, il avance que l'analogie est le noyau du fonctionnement cognitif des êtres humains, et que chaque problème que nous rencontrons n'est autre qu'un assemblage d'analogies que nous parcourons avec plus ou moins de fluidité par nos raisonnements. Il fonde ainsi ses théories sur cette hypothèse de liaison étroite entre raisonnement et analogie. À noter également les travaux de Murena et al. (2017) qui choisissent une approche fondée sur la recherche de la solution de complexité minimale.

L'interprétation des analogies pourrait représenter une avancée décisive dans le développement des travaux en Traitement Automatique du Langage Naturel (TALN). Dans la perspective ambitieuse de cerner puis de synthétiser ce mécanisme, il se révèle crucial de scruter les subtilités du langage, et d'analyser leurs fonctionnements des points de vue cognitif et linguistique. Une grande partie des travaux en TALN se concentre toujours sur des tâches linguistiques élémentaires, voire de premier niveau (étiquetage morpho-syntaxique, analyse syntaxique, coréférence, reconnaissance d'entités nommées, ...), tandis qu'une autre partie des recherches vise à améliorer les mécanismes d'inférence automatique et l'extraction de connaissances nouvelles à partir de corpus textuels. Au final, encore moins de travaux s'attachent à rassembler les enseignements et progrès apportés par chacune de ces directions scientifiques, afin de se rapprocher de la capacité linguistique humaine et d'ainsi simuler des raisonnements linguistiques de haut niveau tels que la compréhension des expressions imagées dans toute leur créativité.

Dans ce travail, nous cherchons à mettre en place un prototype d'interprétation des analogies en exploitant une base de connaissances de sens commun sous forme de graphe, le réseau *JeuxDeMots (JDM)* (Lafourcade, 2007). Nous formaliserons des énoncés en carrés analogiques (*cf. Section 2*) et présenterons des méthodes pour évaluer la qualité de ces analogies (*cf. Section 3*). Pour une analogie, il s'agit de faire émerger des correspondances entre termes qui sont fournis au système par l'utilisateur. Le démonstrateur d'une première preuve de concept associé à cet article est disponible à l'adresse suivante : <https://analogie.lirmm.fr>.

2. Carré analogique

Une analogie (ou carré analogique) est un ensemble de 4 termes liés par des relations de similarités (voir *Figure 1*). La force d'une analogie procède de la similarité des termes qui la composent, son explicabilité se fonde sur les multiples relations entre ces termes.

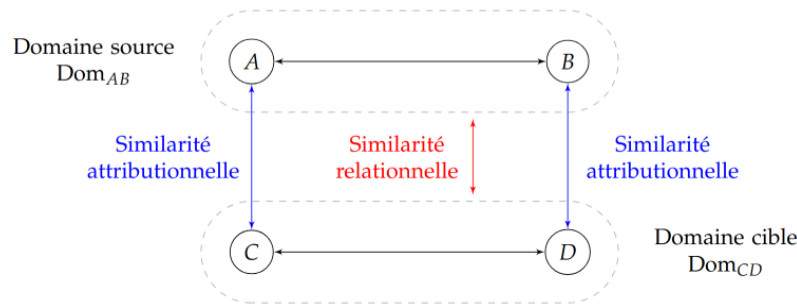


Figure 1 - Schéma simplifié du carré analogique (exemple : A=oeil, B=vue, C=main, D=toucher)

2.1. Analogie et similarité

De façon générale, une expression imagée vise à intégrer, dans la description d'un concept, des attributs d'un deuxième, choisis sur la base de sa similarité sémantique avec le concept décrit. Nous évoquons alors la distinction entre les deux notions de similarité émanant des travaux de Gentner (1983), qui avance qu'il existe au moins deux types de similarités :

- La **similarité relationnelle** (*relational similarity*) qui consiste en la correspondance entre les relations de deux paires de concepts.
- La **similarité attributionnelle** (*attributonal similarity*) qui est la correspondance entre les attributs de deux concepts.

Les notions d'attribut et de relation sont admises dans le sens de la logique du premier ordre, où un attribut est un prédicat à une seule variable, tandis qu'une relation est un prédicat à deux variables. Nous qualifions deux termes désignant chacun un concept, de synonymes, lorsque leur **similarité attributionnelle** est suffisamment élevée tandis que nous désignons deux paires de termes comme analogues si leur **similarité relationnelle** est élevée (Turney, 2005).

Nous pouvons, suite à cette similarité relationnelle, avancer qu'il serait possible de générer une correspondance (analogie) $A \rightarrow B$ transférant la connaissance d'un concept dit *source* A à un concept *cible* B. Le concept *source* est généralement abstrait, incertain ou méconnu, celui sur lequel nous souhaitons asseoir une *cible* facile à comprendre et concrète. Voici des exemples de paires de concepts constituant des analogies :

(*confiance* et *succès*) avec (*soleil* et *fleur*) (1) : abstrait-concret

(*électron* et *plasma*) avec (*personne* et *foule*) (2) : méconnu-connu

(*charpentier* et *bois*) avec (*maçon* et *pierre*) (3)

L'analogie (3) tirée des travaux de Turney et Pantel (2010) pourrait être formulé : « le *charpentier* est au *bois* ce que le *maçon* est à la *pierre* ». Les sens des relations

qu’entretennent respectivement les concepts *maçon* et *charpentier* avec *pierre* et *bois* sont effectivement similaires. D’une part, il s’agit de métiers, et d’autre part, de matériaux en lien très étroit avec ces métiers respectifs. L’introduction d’une caractéristique propre au concept distant (*source*) met donc à disposition de l’orateur toute la connaissance contextuelle du concept et opère, dans le cas d’un discours explicatif ou d’un raisonnement, en tant qu’appui solide et familier, ou alors, dans le cas d’une intention poétique, comme un agent évocateur aidant à teinter le langage et offrant une richesse lyrique complémentaire.

2.2. Chaînes de symboles et équation analogique

#	A	:	B	::	C	:	D	Phrase	Type (nombre d’inconnues)
1	<i>a</i>		<i>b</i>		<i>c</i>		<i>d</i>	<i>a</i> est à <i>b</i> ce que <i>c</i> est à <i>d</i>	Analogie (0)
2	<i>a</i>				<i>c</i>			<i>a</i> est comme <i>c</i>	Comparaison (2)
3			<i>b</i>				<i>d</i>	<i>b</i> est comme <i>d</i>	
4	<i>a</i>						<i>d</i>	<i>a</i> de <i>d</i> <i>d</i> de <i>a</i>	Métaphore (2)
5			<i>b</i>		<i>c</i>			<i>b</i> de <i>c</i> <i>c</i> de <i>b</i>	
6	<i>a</i>		<i>b</i>		<i>c</i>			<i>c</i> de <i>b</i> est <i>a</i> <i>c</i> est comme <i>a</i> : il a son <i>b</i>	Métaphore comparée (1) Comparaison métaphorique (1)
7	<i>a</i>		<i>b</i>				<i>d</i>	<i>d</i> de <i>a</i> est <i>b</i> <i>d</i> est comme <i>b</i> : il a son <i>a</i>	
8	<i>a</i>				<i>c</i>		<i>d</i>	<i>a</i> de <i>d</i> est <i>c</i> <i>a</i> est comme <i>c</i> : il a son <i>d</i>	
9			<i>b</i>		<i>c</i>		<i>d</i>	<i>b</i> de <i>c</i> est <i>d</i> <i>b</i> est comme <i>d</i> : il a son <i>c</i>	

Exemple

a = charpentier, *b* = bois, *c* = maçon, *d* = pierre

1.	Le charpentier est au bois ce que le maçon est à la pierre .	(analogie)
2.	Le charpentier est comme le maçon .	(comparaison)
3.	Le bois est comme la pierre .	(comparaison)
4.1.	Le charpentier de la pierre .	(métaphore)
4.2.	La pierre du charpentier .	(métaphore)
5.1.	Le bois du maçon .	(métaphore)
5.2.	Le maçon du bois .	(métaphore)
6.1.	Le maçon du bois est le charpentier .	(métaphore comparée)
6.2.	Le maçon est comme le charpentier : il a son bois .	(comparaison métaphorique)
7.1.	La pierre du charpentier est le bois .	(métaphore comparée)
7.2.	La pierre est comme le bois : elle a son charpentier .	(comparaison métaphorique)
8.1.	Le charpentier de la pierre est le maçon .	(métaphore comparée)
8.2.	Le charpentier est comme le maçon : il a sa pierre .	(comparaison métaphorique)
9.1.	Le bois du maçon est la pierre .	(métaphore comparée)
9.2.	Le bois est comme la pierre : il a son maçon .	(comparaison métaphorique)

Figure 2 - Formulations observées selon le positionnement des inconnues dans l’équation analogique sous-jacente. Analyse du cas « *maçon : pierre :: charpentier : bois* »

Quand il s’agit de la notion d’égalité des rapports (proportionnalité), entre deux paires de termes (*A*, *B*) et (*C*, *D*), l’énoncé (3) peut être écrit de façon plus concise avec la notation en chaîne de symboles « *A : B :: C : D* ». La notation illustrée par l’exemple de l’analogie proportionnelle entre les paires (*charpentier*, *bois*) et (*maçon*, *pierre*) précédemment évoquée devient alors :

$$\textit{charpentier} : \textit{bois} :: \textit{maçon} : \textit{pierre}$$

L’opérateur « : » indique l’existence de rapports relationnels entre ses opérands, en l’occurrence les termes *charpentier* et *bois*. L’opérateur « :: » transfère ensuite cette relation **de la source à la cible** en affirmant l’existence d’un rapport de même valeur sémantique entre les termes de la deuxième paire (termes *maçon* et *pierre*). Cette notation est adoptée pour sa

séparation claire entre la *source* (*charpentier* et/ou *bois*) et la *cible* (*maçon* et/ou *pierre*) dans une analogie.

2.3. Métaphores et comparaisons, des analogies à trou

Dans le cadre des analogies Aristotéliennes, nous retrouvons le cas où un (ou plusieurs) des quatre symboles est manquant, soulevant ce qui est appelé une **équation analogique** de la forme « $A : B :: C : ?$ ». L'interprétation de l'analogie revient à la résolution de cette équation et consiste à déduire les valeurs possibles du ou des termes manquants. Dans une volonté de formalisme, de façon similaire à King et Gentner (2023), nous partons du postulat que les métaphores et les comparaisons peuvent être formalisées comme étant des manifestations d'analogies à trou (des équations analogiques à 1 ou 2 inconnues). La *Figure 2* illustre les différents cas observés.

3. Force d'une analogie et élection de candidats

Voyons maintenant les méthodes proposées pour l'évaluation des similarités en vue d'élire le ou les meilleurs termes candidats à la résolution d'une analogie à trou.

3.1. Relations entre les mots

Nous notons « $a r_t b$ », la relation r de type t de a vers b et son poids « $p(a, r_t, b)$ ». Dans notre base de connaissances, nous disposons d'informations à propos d'un terme (ou nœud dans le cadre d'un graphe de connaissances) : ses relations avec d'autres mots. Nous exploitons les types de relations présentes dans JDM^1 et nous nous limitons à celles relevant principalement de la sémantique : $r_associated$, r_domain , r_isa , r_anto , r_hypo , r_has_part , r_holo , r_agent , $r_patient$, r_lieu , r_instr , r_carac^2 .

Nous pouvons regrouper les relations entre des termes/nœuds A, B , comme présenté en *Figure 3*, selon si elles sont directes ou indirectes (avec des nœuds iAB) ainsi qu'orientées dans un sens ou l'autre. Pour une relation donnée, il est possible de calculer un poids normalisé p_{norm} : le ratio du poids de la relation $a r_t b$ par le poids max de toutes les relations de type t depuis le nœud a . La valeur p_{norm} permet de classer les relations sortantes d'un nœud et de sélectionner celles considérées comme les plus fortement associées, c'est à dire les n premières³ relations avec les meilleurs poids normalisés eu égard au type t de la relation au départ du nœud a . Cela permet un contrôle de la combinatoire lors du calcul de la force de similarité (cf. *Section 3.2*). La formule détaillée du poids normalisé pour des a, r_t et b fixés est définie dans l'énoncé (4).

$$p_{norm}(a, r, b) = p(a, r_t, b) / \max_val_for_type(a, t) \quad (4)$$

¹ <https://www.jeuxdemots.org/jdm-about-detail-relations.php>

² Certaines relations sont conversives, c'est à dire que $a r_t b \Leftrightarrow b r_{t_i} a$ avec r_{t_i} la relation converse à r_t (exemple : r_isa et r_hypo).

³ Nous avons arbitrairement choisi dans notre démonstrateur les 2 relations les plus pertinentes (s'il y en a) pour des raisons de simplification. À noter qu'il se peut que le nœud intermédiaire soit le même pour ces 2 relations.

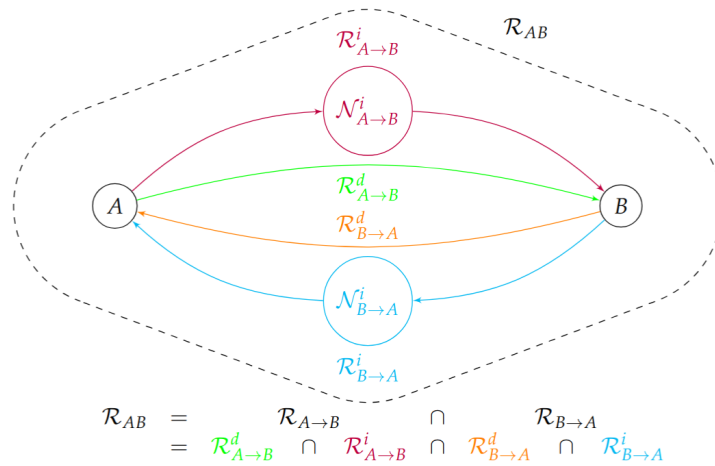


Figure 3 - Notation des ensembles de relations directes et indirectes entre A et B

3.2. Similarité relationnelle

Lors de l'analyse d'une analogie $A : B :: C : D$, nous souhaitons récupérer l'ensemble des relations directes et indirectes entre A et B (dans les 2 sens) ainsi qu'entre C et D (dans les 2 sens). Le but est alors d'effectuer une intersection entre ces types de relations respectivement entre A et B d'une part et C et D d'autre part de manière à évaluer la force de la similarité relationnelle.

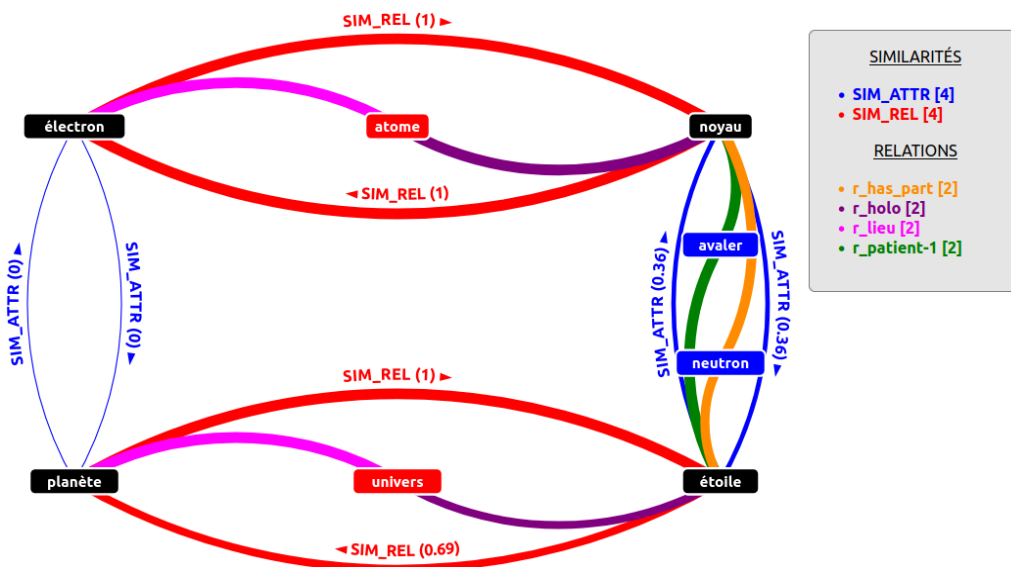


Figure 4 - Exécution de notre démonstrateur (<https://analogie.lirmm.fr>) sur le carré analogique $\text{électron} : \text{noyau} :: \text{planète} : \text{étoile}$

Dans l'exemple présenté en Figure 4, nous constatons que l'on a bien les mêmes types de relations entre *électron* et *noyau* qu'entre *planète* et *étoile*. En l'occurrence, les relations sont indirectes, car aucune de même type et directe n'est trouvée. Nous établissons la force de la similarité relationnelle de A vers B⁴ en faisant la moyenne des poids normalisés des relations

⁴ A noter que la force de similarité relationnelle de A vers B n'est pas forcément la même que celle de B vers A puisqu'il existe des relations orientées dans les 2 sens et de poids différents.

de A vers B combinées à celles de A vers les nœuds intermédiaires iAB . Nous opérons de même pour les relations de B vers A . En pratique, les résultats sont satisfaisants dans le sens où les valeurs obtenues correspondent à l'intuition, les valeurs se rapprochent de 1 pour une similarité jugée forte. La similarité est considérée comme moins forte lorsque la valeur calculée s'approche de 0. Elle est égale à 0 lorsqu'aucune relation (directe ou indirecte) n'est observée. Notons que cette mesure est prise en tant que *baseline* et pourra être affinée afin d'apporter une plus grande précision au résultat.

Nous mettons également en place une légère pondération selon le type de relation, ce qui permet, par exemple, de considérer une relation du type r_{isa} , r_{lieu} ou r_{has_part} comme plus importante au regard de son apport sémantique, relativement à une relation $r_{associated}$ qui est plus vague.⁵ Nous posons l'hypothèse qu'établir un lien précis entre *électron* et *noyau* puis retrouver ce même lien de l'autre côté de l'analogie entre *planète* et *étoile* (avec d'autres nœuds intermédiaires) permet de s'assurer de la force de cette **similarité relationnelle** plus que s'il s'agissait d'une simple relation d'association d'idées ou même s'il n'en existait pas.

3.3. Similarité attributionnelle

Nous analysons maintenant les relations de mêmes types entre chacun des termes A , B et un certain nœud intermédiaire iAB . Le schéma recherché correspond donc à $A r_t iAB$ et $B r_t iAB$. Nous procédons de même et de manière indépendante avec C et D . Dans la *Figure 4*, nous constatons qu'il n'existe pas de tel nœud intermédiaire entre *électron* et *planète* dans *JDM*, la similarité attributionnelle est donc à 0. Cela ne suffit pas à affirmer qu'*électron* et *planète* n'ont rien à voir, ceci signifie plutôt une absence de similarité attributionnelle du point de vue de la base de connaissances dans son état actuel. En revanche, nous constatons entre *noyau* et *étoile* la présence du nœud *avalé* via une relation $r_{patient-1}$, sans doute car une *étoile* peut par exemple « être avalée par un trou noir » et un *noyau* peut « se faire avaler par un enfant ». Un locuteur humain remarquera immédiatement que deux sens différents du mot *noyau* sont impliqués. Il faudra donc procéder à un raffinement sémantique en distinguant les différents sens de chaque mot pour obtenir de meilleurs résultats ; *JDM* permet de faire cela.

3.4. Force globale et élection de candidats dans le cas de la métaphore

Une analogie est considérée comme plus parlante quand les similarités relationnelles et attributionnelles qu'elle contient sont suffisantes, ou d'un point de vue statistique si leur score est élevé. En faisant une moyenne⁶ entre l'ensemble de ces similarités nous obtenons une valeur que l'on va considérer comme une force d'association sous la forme d'une probabilité. Ce score prend son sens lorsqu'il s'agit de classer différents candidats dans le cas d'une analogie à trou⁷. Nous pouvons donc évaluer la pertinence de ces candidats en calculant le

⁵ Dans le futur, nous pourrions utiliser une approche de type *TF-IDF* qui consiste à voir en quoi un $A r B$ est aussi spécifique que possible à A , une telle démarche serait toutefois gourmande en calcul.

⁶ Les moyennes arithmétiques et géométriques produisent des résultats similaires.

⁷ Nous n'avons abordé pour l'instant que les métaphores, les comparaisons étant un sujet plus vaste étant donné que leurs deux inconnues dans le cadre du carré analogique constituent un défi combinatoire.

score pour chacun d’eux avec les nœuds connus de l’analogie avec lesquels ils entretiennent des rapports de similarité suffisamment élevés. Prenons par exemple l’analogie à trou *jambe : genou :: bras : ?*, le candidat proposé est *coude* ce qui correspond à une réponse qui vient naturellement. Dans notre démonstrateur, les relations $a \ r_t \ b$ affichées et retenues actuellement pour le calcul du poids sont les 2 premières par ordre de poids décroissant pour chaque couple (a, b) ; cela permet de rendre les temps de calcul raisonnables et donne des résultats satisfaisants pour une preuve de concept.

3.5. Evaluation de la résolution d’équation analogique

Nous disposons d’une liste d’analogies avec une distribution de réponses fournies par les joueurs de *JDM* (témoin)⁸. Nous vérifions si les 4 meilleurs candidats prédits incluent le meilleur terme témoin (voir *Figure 5*). Ce travail préliminaire aboutit à une précision de l’ordre de 37%. Une perspective d’amélioration consiste, en premier lieu, à prendre en compte un ensemble de relations sémantiques plus couvrant, et ensuite calibrer les méthodes de calcul des similarités et leurs agrégations. Il sera nécessaire de mettre en place des procédures plus approfondies d’exploration de la base de connaissances telles que des mécanismes d’inférence et de raisonnement. Il est aussi possible de procéder à une préparation des données, comme une normalisation morpho-syntaxique⁹ ou encore une désambiguïisation sémantique.

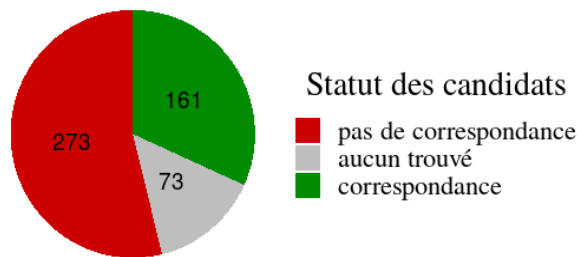


Figure 5 - Evaluation de la pertinence des candidats (4 premiers) vis à vis du témoin le plus joué dans *JDM* pour la même métaphore (sur un corpus de 507 métaphores)

4. Synthèse

Nous considérons l’analogie Aristotélicienne dans sa notation en chaînes de symboles $A:B::C:D$, ce qui signifie “**A est à B ce que C est à D**” et on dit qu’il y a (voir *Figure 6*) :

- **similarité relationnelle** (correspondance entre les relations de 2 paires de concepts) :
 - $R_{AB} = A : B$ (avec A et $B \in$ domaine *source* noté Dom_{AB})
 - $R_{CD} = C : D$ (avec C et $D \in$ domaine *cible* noté Dom_{CD})
 - $SimRel = R_{AB} \cap R_{CD}$

⁸ En simulant les intersections pour une liste d’un peu plus de 2000 métaphores jouées dans *JDM* (<http://jeuxdemots.org/analogies.php> → “exporter données”), nous constatons qu’environ 87% des cas d’échec sont dus à des relations manquantes. Des processus d’inférence déductive peuvent pallier ce problème dans 70 % des cas. Exemple d’inférence : *enfant r_has_part jambe* → *enfant r_isa humain r_has_part jambe*.

⁹ Passage par la lemmatisation en observant l’ensemble des relations de mots proches comme *petit, petite, petits, petites* quand l’un d’eux est concerné.

- Il est possible d'avoir une correspondance $A \rightarrow B$ (respectivement $C \rightarrow D$) transférant la connaissance d'un concept *source* généralement familier et concret A (respectivement C) à un concept *cible* généralement méconnu et abstrait B (respectivement D)
- **similarité attributionnelle** (correspondance entre les attributs de 2 concepts) :
 - SimAttr_{AC} entre A et C (co-P)¹⁰
 - SimAttr_{BD} entre B et D (co-P)
- **analogie** quand il existe une intersection non vide des relations de \mathcal{R}_{AB} et de \mathcal{R}_{CD} , sa compréhension s'en voit améliorée avec la présence de similarités attributionnelles

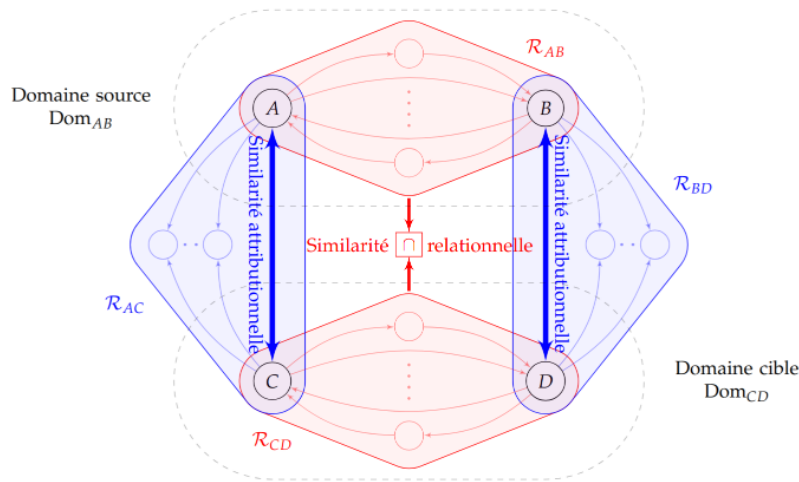


Figure 6 - Schéma détaillé du carré analogique

Nous pensons que la force d'une analogie se définit à travers plusieurs aspects. Tout d'abord, il faut noter que la démarche ne s'inscrit pas, au regard de la base de connaissances, dans un contexte de monde clos, dans le sens où une relation $a r_t b$ manquante ne signifie pas que celle-ci n'est pas possible. Les défauts de connaissance peuvent résulter en des aberrations dans le résultat de l'analyse de l'analogie, défauts qu'il conviendra alors d'identifier puis de rectifier ou de compenser. Les améliorations en cours de réalisation dans cette preuve de concept incluent entre autres, des mécanismes de raisonnement plus efficaces, ou encore des démarches de résolution de la polysémie des termes en entrée.

Dans un carré analogique, les relations entre les mots du domaine *source* sont de même type (on pourrait dire analogues) à celles du domaine *cible* établissant la force de la similarité relationnelle. Les termes équivalents de domaines opposés doivent avoir une relation de même type vers un nœud intermédiaire (attribut). Il s'agit de similarité attributionnelle, clef d'une meilleure explication de l'analogie.

Dans tous les cas, plus le type de relation est précis sémantiquement, plus ce type est utile pour expliquer l'analogie dans son ensemble. La force globale de l'analogie pourrait correspondre à la moyenne de l'ensemble des poids de similarité relationnelle et attributionnelle. Cela permet actuellement d'élire le ou les candidats s'avérant les plus

¹⁰ P étant la relation sémantique vue comme un prédicat unaire de même valeur

adéquats pour une métaphore donnée (une analogie à trou à une seule inconnue) dans 37% des cas avec un algorithme simple (constituant une *baseline*). Néanmoins, une agrégation plus précise des différents scores de similarité permettrait de tenir compte des nuances portées par chaque type de relation ; raison pour laquelle la définition de cette méthode représente un des aspects centraux de notre recherche. Par définition, un langage figuré est soumis à interprétation, une même analogie ou son explication peuvent se révéler plus ou moins parlantes selon l'angle par lequel l'explication est approchée. Le but d'un analyseur automatique est de mettre en évidence ce qui peut le plus être considéré comme une interprétation convaincante ou satisfaisante (des concepts subjectifs) pour une analogie donnée. Dans l'état de l'art, il est à noter que le sujet est souvent traité à travers des points de vue théoriques et linguistiques plutôt que dans des travaux computationnels et appliqués.

Ce travail servira de modèle de résolution mais fera également office d'outil de repérage des anomalies dans la base de connaissances (imparfaite et incomplète par nature). Les manques éventuels seront mis en évidence par les cas d'échec de l'algorithme et permettront de consolider les types de relations appropriés. Une piste pourrait alors être de concevoir un jeu sur les intersections d'associations pour élargir la base de connaissances en ce sens.

Bibliographie

- Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive science*, 7(2), 155-170.
- Gentner, D. & Holyoak, K. J. (1997). Reasoning and learning by analogy: Introduction. *American psychologist*, 52(1), 32.
- Hofstadter, D. R. (1995). *Fluid concepts and creative analogies: Computer models of the fundamental mechanisms of thought*. Basic books.
- Hofstadter, D. R. (2001). Analogy as the core of cognition. *The analogical mind: Perspectives from cognitive science*, 499-538.
- King, D. C., & Gentner, D. (2023). Verb Metaphors are Processed as Analogies. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 45, No. 45).
- Lafourcade, M. (2007, December). Making people play for Lexical Acquisition with the JeuxDeMots prototype. In *SNLP'07: 7th international symposium on natural language processing* (p. 7).
- Lepage, Y. (2003). *De l'analogie rendant compte de la commutation en linguistique* (Doctoral dissertation, Université Joseph-Fourier-Grenoble I).
- Murena, P. A., Dessalles, J. L., & Cornuéjols, A. (2017, June). A Complexity Based Approach for Solving Hofstadter's Analogies. In *ICCB (Workshops)* (pp. 53-62).
- Shutova, E. (2010, June). Automatic metaphor interpretation as a paraphrasing task. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics* (pp. 1029-1037).
- Turney, P. D. (2005). Measuring semantic similarity by latent relational analysis. *arXiv preprint cs/0508053*.
- Turney, P. D., & Pantel, P. (2010). From frequency to meaning: Vector space models of semantics. *Journal of artificial intelligence research*, 37, 141-188.