

Analogia et Intersector - des jeux comme fondements de l'analyse automatique d'analogies

Analogia and Intersector - games as a basis for automatic analogy analysis

Mathieu Lafourcade¹, Nathalie Le Brun², Jérémie Roux¹

¹ LIRMM (- Laboratoire d'informatique, de robotique et de microélectronique de Montpellier), Université Montpellier, CNRS, Montpellier, France,

² Im@ginat

RÉSUMÉ. La capacité d'un système informatique à analyser sémantiquement des textes passe notamment par l'interprétation de contenus de sens figuré. L'analogie est souvent utilisée afin de transmettre des idées nouvelles par similitude avec des idées connues. Il est possible de modéliser la comparaison et la métaphore sous la forme d'un carré analogique $A : B :: C : D$ (signifiant A est à B ce que C est à D) avec une ou plusieurs variables dont il faut trouver les valeurs les plus pertinentes. Cet article présente deux jeux permettant de collecter des données lexicales en rapport avec les analogies. Le premier jeu, Analogia, est une adaptation de JeuxDeMots aux analogies, où le joueur doit fournir les réponses les plus pertinentes pour une analogie à une variable. Par exemple, trouver x pour "charbon est à noir ce que neige est à x". Le second jeu, Intersector, permet d'indiquer, en répondant à des questions, pour les 4 termes donnés d'une analogie, quels sont leurs points communs. Les termes proposés sont ceux issus des analogies du jeu Analogia. L'ensemble des données récoltées par ces jeux permet d'envisager une résolution automatique d'analogies, et donc l'interprétation de contenus figurés.

ABSTRACT. A computer system's ability to analyze texts semantically involves interpreting figurative content. Analogy is often used to convey new ideas through similarity with known ideas. Comparison and metaphor can be modeled as an analogical square $A : B :: C : D$ (meaning A is to B what C is to D) with one or more variables whose most relevant values must be found. This article presents two games for collecting lexical data related to analogies. The first game, Analogia, is an adaptation of JeuxDeMots to analogies, where the player has to provide the most relevant answers for an analogy to a variable. For example, find x for "coal is to black what snow is to x". The second game, Intersector, involves answering questions to find out what the 4 terms in an analogy have in common. The terms proposed are those taken from the analogies in the Analogia game. All the data collected by these games makes it possible to automatically resolve analogies, and thus interpret figurative content.

MOTS-CLÉS. Jeux à but, Gwap, jeu sérieux, intelligence artificielle, représentation des connaissances, analyse sémantique, relation lexicale, analogie, métaphore, collecte de données

KEYWORDS. Game with a purpose, Gwap, serious game, artificial intelligence, knowledge representation, semantic analysis, lexical relationship, analogy, metaphor, data collection

Introduction

L'analogie est à l'explication ce que la création est à l'innovation.

Les analogies et les métaphores sont des outils linguistiques et cognitifs puissants qui jouent un rôle significatif dans la transmission des connaissances scientifiques ainsi que dans le processus d'innovation. Leur capacité à établir des connexions entre des concepts complexes et des phénomènes familiers les rend particulièrement précieux dans le domaine de l'enseignement et de l'apprentissage, où ils facilitent la compréhension et la mémorisation des idées abstraites. De plus, les analogies et les métaphores ne se limitent pas simplement à la clarification des concepts existants, mais elles ont également le potentiel de catalyser l'innovation en ouvrant de nouvelles perspectives à l'origine de nouvelles interprétations.

Le constat ci-dessus nous amène à développer des systèmes d'analyse de texte capables de détecter et d'interpréter correctement les analogies et leurs formes particulières que sont les métaphores et les comparaisons. Toutefois, pour ce faire, il est nécessaire de disposer d'une base de connaissances contenant à la fois des informations de sens commun mais également des faits relevant de domaines de spécialités. L'analyse automatique de textes peut avoir plusieurs objectifs, dont l'évaluation de la cohérence sémantique d'un document, son indexation afin de le retrouver ultérieurement dans une base documentaire, la détection de plagiat, le croisement de faits à des fins de recherche scientifique, l'évaluation de la pertinence de dépôt de brevets, pour ne mentionner que quelques applications parmi les plus notables.

La base de connaissances utilisée lors de l'analyse doit contenir les informations pertinentes pour l'interprétation des analogies. Il est difficile de déterminer automatiquement de manière anticipée ce que devront être de telles informations, et il est donc intéressant de mettre en place des stratégies pour les acquérir. Une stratégie que nous exposons ici est de faire jouer des personnes sur deux types de jeux. Le premier est de proposer des analogies incomplètes que les joueurs doivent compléter. Ainsi nous procédons à un échantillonnage de réponses possibles pour chacun des termes d'un grand nombre d'analogies. À partir de cet échantillonnage, il est possible via un autre jeu de demander aux joueurs d'indiquer les traits sémantiques communs ou différents qui existent entre les termes d'une analogie donnée. Afin de réduire la grande variabilité des réponses, l'utilisation d'une base de connaissances communes et de mécanismes d'inférence permet de poser des questions fermées aux utilisateurs dans un contexte ludique. Ludifier de telles activités permet d'implémenter non seulement des facteurs motivationnels (gain de points, classement, etc) mais aussi d'évaluer la compétence globale de chaque joueur afin de pondérer les réponses obtenues.

Dans ce qui suit, nous présentons dans un premier temps une définition précise de ce qu'est une analogie et des rapports entre ses différents termes. Nous évoquons ensuite rapidement le projet JeuxDeMots et l'extension *Analogia*, un mode de jeu permettant de construire et renseigner de nouvelles analogies. Enfin, nous détaillons le jeu *Intersector*, qui rend possible l'enrichissement de la base de connaissances JeuxDeMots à partir des analogies construites précédemment.

Carré analogique

Une analogie (aussi appelée carré analogique) est un ensemble de 4 termes liés par des relations sémantiques traduisant une certaine similarité. 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. Nous visons à « atteindre [...] la formalisation d'une opération que tout le monde reconnaît être à l'œuvre dans la langue » [LEP 03] en vue de la rendre accessible à des procédures automatisées, en particulier d'analyse de textes où les analogies sont souvent présentes, par exemple dans des textes relevant de la transmission de connaissances et/ou de l'innovation. L'analogie peut être vue comme une opération fondamentale de pensée utilisée dans les langues [GEN 97]. Hofstadter [HOF 95][HOF 01] estime que la pensée et l'analogie sont indissociables, et 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.

Analogie et similarité

D'une façon générale, une expression imagée vise à intégrer, dans la description d'une idée, des attributs d'une seconde idée, choisie sur la base de sa proximité sémantique avec l'idée de départ. Nous pouvons évoquer alors la distinction entre les deux notions de similarité émanant des travaux de Gentner [GEN 83], qui avance qu'il en existe au moins deux types :

- La **similarité relationnelle** consiste en la correspondance entre les relations de deux paires de concepts.

– La **similarité attributionnelle** 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 [TUR 05].

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 [TUR 10] pourrait être formulée : « le *charpentier* est au *bois* ce que le *maçon* est à la *pierre* ». Les sens des relations qu'entretiennent 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.

Figure 1. Carré analogique : *électron* : *noyau atomique* :: *planète* : *étoile*, où les domaines (liens verticaux de gauche) sont disjoints. L'analogie ainsi formée vise à donner l'intuition de la ligne supérieure à l'aide de la ligne inférieure (où l'inverse qui est aussi possible). (source : <https://analogie.demo.lirmm.fr>)

Figure 2. Carré analogique : corbeau : noir :: oie : blanche, où les domaines (liens verticaux de gauche) ne sont pas disjoints. L'analogie ainsi formée vise plutôt à comparer les deux lignes, sans pour autant avoir un pouvoir explicatif particulier. (source : <https://analogie.demo.lirmm.fr>)

Chaînes de symboles et équation analogique

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 :

charpentier : bois :: maçon : pierre

L'opérateur « : » indique l'existence de rapports relationnels entre ses opérandes, 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.

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 (sont) manquant(s), 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 terme(s) manquant(s). Dans une volonté de formalisme, de façon similaire à King et Gentner [KIN 23], nous partons du postulat que les métaphores et les comparaisons peuvent être formalisées comme étant des manifestations d'analogies à trou(s) (des équations analogiques à 1 ou 2 inconnues).

Quelques exemples d'analogies et leur pertinence

Voici quelques analogies (plus en annexe) :

- Génome : Organisme :: Code source : Programme
- Acide : pH < 7 :: Base : pH > 7
- Magnétisme : Fer :: Gravité : Masse
- Espace : Astronomie :: Molécule : Chimie
- Mutation : Évolution :: Variation : Sélection naturelle
- Gène : ADN :: Mot : Phrase
- Force : Newton :: Énergie : Joule
- Système solaire : Planète :: Galaxie : Étoile
- Photosynthèse : Plante :: Respiration : Animal
- Liquéfaction : Liquide :: Solidification : Solide

Une question qui apparaît alors est de savoir si une analogie est pertinente, parlante et remplit sa fonction, qui est de donner l'intuition d'un phénomène (ou système) par comparaison avec quelque chose de connu. Il est certain que certaines analogies sont meilleures que d'autres. On en conclut donc que pour une analogie $A : B :: C : D$ visant à expliquer la relation entre A et B, les termes C et D doivent être d'un domaine différent et entretenir une relation connue. C'est par exemple, le cas de Génome : Organisme :: Code source : Programme où l'on tenterait d'expliquer à un informaticien ce que serait la génétique.

Il semble clair que plusieurs analogies peuvent avoir des capacités explicatives inégales, que certaines peuvent être qualifiées de *pertinentes*, ou *mauvaises*. Cependant ce n'est pas notre propos ici, car nous nous focalisons plus sur la capacité d'un processus automatique à "comprendre" et à expliquer une analogie donnée, quand bien même elle ne serait pas particulièrement *bonne*. La condition pour ce faire est de disposer dans une base de connaissances des informations partagées entre les différents termes d'une analogie donnée.

Quelles informations sémantiques ?

L'interprétation d'une analogie demande donc de disposer d'informations sémantiques pour les quatre termes impliqués. Plus précisément, il s'agit de répondre aux questions suivantes :

- Est-ce que s'il existe $A R B$, alors il existe $C R D$? (et réciproquement) - R est ici un type de relation sémantique, par exemple *l'hyponymie, le lieu, la relation de caractéristique*, etc.
- Qu'est-ce qui est commun entre A et C ?
- Qu'est ce qui est commun entre B et D ?

Les deux derniers points ne sont pas strictement nécessaires. En fait, pour qu'une analogie remplisse son rôle explicatif, il faut que A et C ne soient pas du même domaine, et que B et D non plus. Cependant, une analogie sera d'autant plus forte qu'il y a des ressemblances entre A et C d'une part et B et D d'autres part.

Dans l'analogie Gène : ADN :: Mot : Phrase, A et C (gène et mot) appartiennent respectivement au domaine de la génétique et de la linguistique. Il en est de même pour B et D (ADN et phrase). La relation entre A et B (gène et ADN) est celle de composition/ holonymie (r_holo) ; elle est préservée entre C et D (mot et phrase).

JeuxDeMots¹ [LAF 07] est un jeu sérieux créé par le LIRMM, qui s'inscrit dans le cadre du modèle Game with a purpose (GWAP) [LAF 15], visant à construire une base de connaissances sous forme d'un réseau lexical. Cette base de connaissances a pour fonction première d'être interrogée par des systèmes d'analyse automatique de textes, en vue d'effectuer des tâches telles que du raisonnement, de la désambiguïsation sémantique, etc.

Principe de JeuxDeMots

JeuxDeMots, lancé en 2007, est un jeu qui vise à recueillir des données lexicales en demandant aux joueurs de répondre à des consignes concernant un mot cible. Ce processus repose sur la validation des termes fournis comme réponses par au moins deux joueurs, sans possibilité de communication entre eux. Ces termes identiques dans les réponses des joueurs à la même consigne deviennent des relations dans le réseau lexical en construction.

Par exemple, si la consigne est de donner des termes génériques pour le terme *chat*, et que deux joueurs répondent "mammifère" et "animal", alors deux relations seront établies dans le réseau : *chat* est un type de *mammifère* et *chat* est un type d'*animal*.

Le jeu se déroule de manière asynchrone en temps limité. Les joueurs gagnent des points lorsqu'ils proposent des réponses communes, mais sont encouragés à être originaux pour enrichir le réseau lexical. Les associations les plus évidentes et déjà proposées de nombreuses fois sont pénalisées, tandis que des primes sont accordées aux associations moins courantes et à la création de nouvelles relations.

Figure 3. Illustration simplifiée d'une partie de rezoJDM, le réseau lexico-sémantique construit par le projet JeuxDeMots. Les relations entre les termes sont orientées et pondérées. Une relation fausse a un poids de valeur négative.

Le réseau lexical se construit progressivement grâce aux relations validées par les réponses communes des joueurs, et ces relations sont pondérées en fonction du nombre de joueurs ayant proposé

¹ <https://www.jeuxdemots.org>

ces associations. JeuxDeMots exploite ainsi le principe que les données fournies par un grand nombre de non-experts sont souvent de meilleure qualité que celles d'un petit groupe d'experts. Les joueurs expérimentés contribuent à améliorer la qualité du réseau en proposant des réponses plus organisées et pertinentes, tout en cherchant à relever des défis plus difficiles pour améliorer leur classement.

Le projet et les jeux de JDM aident à construire rezoJDM [LAF 20] un réseau lexico-sémantique prenant la forme d'un graphe orienté. Les nœuds du graphe représentent les termes, tandis que les arcs désignent des relations typées, pondérées et potentiellement annotées entre ces termes. Par exemple le type de relation "r_isa", représente l'hyponymie : *chat r_isa félin*. Environ 120 types de relations sont présents dans rezoJDM, qui contient un peu plus de 540 millions de relations.

Le graphe représente la polysémie des mots en explicitant des raffinements sémantiques hiérarchisés, où un sens spécifique est affilié au sens général du terme (via une relation de type r_raff de poids positif). Des relations fausses disposent d'une pondération négative, comme par exemple *femme *r_has_part prostate*, ou encore *serpent r_agent-1 voler*.

2.2. Analogia

Un mode de jeu particulier de JeuxDeMots, nommé *Analogia*, permet de jouer sur des analogies à un trou. Le joueur se voit proposer une analogie avec un des quatre termes qui est absent (le trou). Le joueur doit alors fournir au moins une réponse possible.

Figure 4. Exemple de partie d'*Analogia*. L'analogie à trou est *x : dermatologue :: pierre : peau*. Le joueur n'a pour l'instant proposé que le terme *géologue*. Il peut s'arrêter là ou tenter de proposer d'autres réponses.

Le joueur reçoit à l'issue de la partie des récompenses (points, crédits, etc.) dans les mêmes conditions que celles du mode de jeu par défaut.

Voici quelques exemples de jeu analogique (à trou unique) avec des exemples de réponses apportées par les joueurs (les réponses sont pondérées selon leur fréquence d'apparition) :

- corriger : erreur :: ? : offense $\Rightarrow$ réparer (81) ; demander pardon (57) ; pardonner (56) ; excuser>pardonner (55) ; excuser (55) ;
- panneau : indication :: sirène : ? $\Rightarrow$ alarme (90) ; signal (81) ; signal d'alarme (61) ; avertissement (61) ; alarme>signal (61) ; alerte>signal (61) ; alerte (61) ; mise en garde (51) ; avertissement>mise en garde (51) ;
- ? : mouton :: taureau : boeuf $\Rightarrow$ bélier (60) ; bélier>animal (51) ;

- ? : voiture :: rail : train $\Rightarrow$ route (60) ; route>voie de circulation (51) ; boulevard (51) ; autoroute (51) ;
- melon : ? :: camion : véhicule $\Rightarrow$ fruit (80) ; fruit>botanique (71) ; dessert (61) ; dessert>mets (51) ; entrée>repas (51) ; cucurbitacée (51) ; plante (51) ; entrée (51) ; aliment (51) ; aliment>ce qui se mange (51) ; plante>botanique (51) ;
- géologue : ? :: pierre : peau $\Rightarrow$ dermatologue (60) ;
- tableau : ? :: toile : pinceau $\Rightarrow$ craie (80) ; marqueur (61) ; crayon (51) ; craie>bâtonnet (51) ; marqueur>stylo-feutre (51).

La collecte de ces analogies laisse supposer pour chacune d'elles des liens sémantiques entre les termes.

Un point important d'*Analogia*, est que le joueur peut, avec les crédits collectés lors de parties, proposer lui-même ses propres analogies qui seront soumises aux autres joueurs. Les analogies ainsi jouées par d'autres rapportent des points au joueur les ayant proposées, il s'agit en somme d'une "rente". Le joueur propose une analogie complète qui sera à l'origine de 4 parties possibles (une pour chaque terme) . Par exemple, si le joueur propose l'analogie suivante :

- Start-up : Innovation :: Chêne : Gland

(La start-up est à l'innovation ce que le chêne est au gland)

les parties suivantes seront proposées aux joueurs :

- x : Innovation :: Chêne : Gland
- Start-up : x :: Chêne : Gland
- Start-up : Innovation :: x : Gland
- Start-up : Innovation :: Chêne : x

On remarquera que les 4 parties possibles ne sont ni de même difficulté ni de même intérêt. Voici quelques analogies liées à l'innovation (certaines sont clairement discutables) :

1. Start-up : Entrepreneuriat :: Invention : Création (La start-up est à l'entrepreneuriat ce que l'invention est à la création)
2. Imprimante 3D : Fabrication additive :: Robot : Automatisation (L'imprimante 3D est à la fabrication additive ce que le robot est à l'automatisation)
3. Intelligence artificielle : Analyse de données :: Machine à vapeur : Révolution industrielle (L'intelligence artificielle est à l'analyse de données ce que la machine à vapeur est à la révolution industrielle)
4. Blockchain : Sécurité des transactions :: Internet : Communication mondiale (La blockchain est à la sécurité des transactions ce que l'Internet est à la communication mondiale)
5. Réalité virtuelle : Immersion :: Vidéoconférence : Communication à distance (La réalité virtuelle est à l'immersion ce que la vidéoconférence est à la communication à distance)
6. Voiture électrique : Énergie renouvelable :: Téléphone mobile : Communication sans fil (La voiture électrique est à l'énergie renouvelable ce que le téléphone mobile est à la communication sans fil)
7. Nanotechnologie : Matériaux :: Biotechnologie : Santé (La nanotechnologie est aux matériaux ce que la biotechnologie est à la santé)

8. Énergie solaire : Panneaux photovoltaïques :: Énergie éolienne : Éoliennes (L'énergie solaire est aux panneaux photovoltaïques ce que l'énergie éolienne est aux éoliennes)
9. Cryptomonnaie : Finance décentralisée :: Courrier électronique : Correspondance rapide (La cryptomonnaie est à la finance décentralisée ce que le courrier électronique est à la correspondance rapide)
10. Smart home : Domotique :: Smart city : Urbanisme (La smart home est à la domotique ce que la smart city est à l'urbanisme)
11. Biocarburant : Transport durable :: Agriculture biologique : Alimentation saine (Le biocarburant est au transport durable ce que l'agriculture biologique est à l'alimentation saine)
12. Cloud computing : Stockage distant :: Machine learning : Apprentissage automatique (Le cloud computing est au stockage distant ce que le machine learning est à l'apprentissage automatique)
13. Drone : Surveillance aérienne :: Satellite : Observation spatiale (Le drone est à la surveillance aérienne ce que le satellite est à l'observation spatiale)
14. Impression 3D : Prototypage rapide :: Scanner 3D : Modélisation précise (L'impression 3D est au prototypage rapide ce que le scanner 3D est à la modélisation précise)
15. E-commerce : Commerce traditionnel :: Streaming : Télévision par câble (Le e-commerce est au commerce traditionnel ce que le streaming est à la télévision par câble)
16. Bateau autonome : Navigation :: Véhicule autonome : Conduite (Le bateau autonome est à la navigation ce que le véhicule autonome est à la conduite)
17. Fibre optique : Vitesse de connexion :: Moteur à réaction : Vitesse de vol (La fibre optique est à la vitesse de connexion ce que le moteur à réaction est à la vitesse de vol)
18. Assistant vocal : Interaction homme-machine :: Clavier : Interface utilisateur (L'assistant vocal est à l'interaction homme-machine ce que le clavier est à l'interface utilisateur)
19. Réalité augmentée : Enrichissement visuel :: Big data : Enrichissement d'information (La réalité augmentée est à l'enrichissement visuel ce que le big data est à l'enrichissement d'information)
20. Vêtements intelligents : Capteurs corporels :: Montres intelligentes : Surveillance de la santé (Les vêtements intelligents sont aux capteurs corporels ce que les montres intelligentes sont à la surveillance de la santé)

Ces analogies illustrent des concepts et des innovations technologiques qui transforment divers aspects de notre vie quotidienne et industrielle. Les joueurs seront sollicités avec ces analogies où un des quatre termes sera invisibilisé.

Comme, la plupart du temps, il est difficile de trouver un grand nombre de réponses possibles, les réponses données par un joueur sont davantage valorisées que pour le jeu standard de JeuxDeMots. À l'inverse, des réponses connues comme étant fausses (celles à poids négatifs) sont davantage sanctionnées par une forte perte de points.

En définitive, le mode de jeu *Analogia* est plus exigeant que le mode de jeu standard (pour les idées associées) mais les points gagnés et leur influence sur le classement ont été calibrés pour que ce mode de jeu soit particulièrement rentable.

Mécanique du jeu d'*Analogia*

Le mode de jeu *Analogia* permet au joueur, tout comme le jeu standard, de gagner des points en fonction des réponses issues de la partie d'un autre joueur. Plus les réponses communes sont originales, plus les deux joueurs gagnent des points. L'originalité d'une proposition est jugée selon l'inverse de la force d'association entre l'analogie à trou et la proposition dans le réseau lexical.

Par exemple :

loup : agneau :: lion : x $\rightarrow$ zèbre / poids = 100

loup : agneau :: lion : x $\rightarrow$ gazelle / poids = 200

loup : agneau :: lion : x $\rightarrow$ gnou / poids = 50

La réponse *gazelle* a un poids plus important car proposée plus de fois, elle est donc plus commune (moins originale) que la réponse *zèbre*.

Le joueur est donc invité à être à la fois original mais à penser comme l'autre joueur (qu'il ne connaît pas) et est donc confronté à un dilemme qu'il peut surmonter en fournissant le plus grand nombre de réponses possibles. Cependant comme mentionné précédemment, il est parfois difficile de trouver un grand nombre de réponses valables.

Pour compenser de dernier aspect, les réponses proposées la première fois par un joueur sont mémorisées comme telles et rapporte une rente à ce joueur quand un autre joueur fait la même proposition. Ainsi les propositions récurrentes pour une partie analogique donnée rapporte aux joueurs assidus et originaux.

Intersector

Le jeu *Intersector* a pour objet de capturer des relations sémantiques intéressantes à inclure dans rezoJDM en vue de l'analyse automatique d'analogies. Le principe général de la mécanique du jeu est de sélectionner aléatoirement une analogie et de poser des questions relatives aux termes de l'analogie au joueur. La difficulté principale est de poser des questions intéressantes, peu inférables et permettant de mieux analyser en quoi l'analogie est intéressante.

Du côté du joueur, les questions sont proposées les unes après les autres sous forme d'un lot correspondant à une analogie. La forme ressemble un peu à celle qu'on peut avoir avec un agent conversationnel. Des questions sont posées, et l'utilisateur peut y répondre ou passer.

Plus précisément, le système de sélection des questions à poser (SQP) tire au hasard une analogie A : B :: C : D parmi celles existantes dans la base de connaissances, et qui n'ont pas encore été jouées par le joueur. L'objectif est de poser tout ou partie des questions suivantes :

– s'il y a une similarité suffisante entre A et C, demander à partir de relations A R E existantes et positives, si les relations C R E absentes de la base de connaissances sont correctes. On pourra symétriquement exploiter A R E issus de C.

Par exemple, pour l'analogie suivante : poulain : cheval :: chaton : chat
Étant donné que la similarité entre poulain et chaton semble assez forte, nous pouvons nous demander si les relations issues de *poulain* semblent possibles pour chaton. Par exemple :

poulain r_has_part queue $\Rightarrow$?	chaton r_has_part queue
poulain r_has_part crinière $\Rightarrow$?	chaton r_has_part crinière
poulain r_agent-1 brouter $\Rightarrow$?	chaton r_agent-1 brouter
chaton r_agent-1 miauler $\Rightarrow$?	poulain r_agent-1 miauler
...	

Il est important de vérifier que le type de relation entre A et B est identique à celle entre C et D. Ainsi considérons l'analogie suivante :

requin : surfeur :: loup : promeneur

Nous avons les relations :

requin r_can_eat surfeur (le requin peut manger un surfeur)

requin r_interact_with surfeur (le requin peut interagir avec le surfeur)

requin r_against surfeur (le requin peut s'opposer au surfeur)

Il faut donc poser les questions similaires concernant les relations possibles entre loup et promeneur :

? loup r_can_eat promeneur (le loup peut manger un promeneur)

? loup r_interact_with promeneur

(le loup peut interagir avec le promeneur)

? loup r_against promeneur (le loup peut s'opposer au promeneur)

En particulier, pour une analogie à domaines distincts, il est crucial que le parallélisme des relations soit être conservé. En effet, comme il y a peu de similarité entre les termes A et C et B et D, l'essentiel de l'analogie repose sur ce parallélisme. Par exemple pour expliquer le concept d'électricité, on peut avoir :

fil : électricité :: tuyau : eau

fil r_lieu-1 électricité (le fil est un lieu pour l'électricité)

? tuyau r_lieu-1 eau (le fil est un lieu pour l'électricité)

(ou l'inverse, selon la relation connue avant de poser la question)

Le choix des questions se fait donc via une approche contrastive entre les termes A, B, C et D de l'analogie. Il est important de connaître à la fois ce qui rapproche les paires de termes mais aussi ce qui les différencie. Les différences ne seraient en principe pas strictement nécessaires à la compréhension d'une analogie, mais elles sont récupérées par effet collatéral (quand le joueur répond par la négative le plus souvent), et sont dans certains cas d'analyse particulièrement utiles.

On notera que beaucoup de questions sont possibles si chacun des termes est très renseigné dans le réseau lexico-sémantique. Sans entrer dans les détails, les relations non inférables seront sélectionnées en priorité.

Mécanique du jeu Intersector

L'intelligence artificielle d'Intersector (nommée IAI) a comme objet principal la sélection des questions à poser au joueur afin de renseigner le plus efficacement possible les termes impliqués dans les analogies. La récurrence des réponses apportées par les différents joueurs conditionne d'une part la force de celle-ci et d'autre part les points gagnés par les joueurs.

Exemple de discussion avec IAI (intersector)

Figure 5. Exemple de discussion avec IAI – Partie 1

Lors de l'interaction, le joueur peut répondre aux questions, qui sont soit ouvertes (réponses, oui non, nsp), soit fermées. Il peut aussi donner des informations de son cru, mais qui doivent avoir un rapport avec l'analogie. L'analogie n'est pas dévoilée à l'utilisateur, mais celui-ci la devinera assez rapidement, car pour chaque question, au moins un des termes de l'analogie est impliqué.

L'utilisateur peut arrêter l'interaction sur le lot de questions courant, en écrivant 'stop'. Après demande de confirmation, *Intersector* va indiquer le nombre de points d'honneur et de crédits que l'utilisateur a gagnés. On parle alors de *checkout*.

Figure 6. Exemple de discussion avec IAI – Partie 2

IAI peut volontairement poser des questions pièges afin d'évaluer la fiabilité du joueur. Un joueur évalué comme "faible" va avoir pour effet de différer l'insertion dans la base de connaissances (le graphe) des informations qu'il a fournies. Cette évaluation est effectuée selon une fenêtre temporelle glissante (depuis 1 jour, 1 semaine, 1 mois, etc.), et est donc réalisée continuellement.

Figure 7. Exemple de discussion avec IAI – Partie 3

Dans, l'exemple ci-dessus IAI pose une question piège (pas difficile du tout) à laquelle le joueur répond correctement, en donnant une précision (déjà connue du système, mais cela augmente la confiance liée à ce joueur). Manifestement le joueur n'est pas un troll systématique.

IAI peut effectuer des inférences, à la fois pour estimer de lui-même les réponses des joueurs, mais surtout pour rebondir sur des informations pouvant être en rapport avec l'analogie en cours (qui, nous le rappelons, n'est pas explicitement indiquée au joueur). IAI exploite les informations de RezoJDM afin d'interpréter au mieux les réponses des joueurs et pour poser des questions pertinentes.

Figure 8. Exemple de discussion avec IAI – Partie 4

L'analyse des tours de parole est faite de la façon la plus "tolérante" possible en ce qui concerne l'orthographe ou l'expression de la relation sémantique désignée. Par exemple :

Figure 9. *Exemple de discussion avec IAI – Partie 5*

Ci-dessus, la dernière ligne est une écriture totalement non ambiguë sous la forme du nom du type de la relation telle que définie dans JeuxDeMots. Un joueur expérimenté, avec beaucoup d'ancienneté et une bonne connaissance des types de relations, peut écrire directement les relations sous cette forme (s'il le souhaite). Bien que peu naturelle, une écriture sous cette forme est à la fois non ambiguë et plus compacte.

AIA peut se trouver en difficulté pour interpréter ce qu'a écrit l'utilisateur. Il peut s'agir d'une incapacité à trouver le bon type de relation, ou de se trouver face à une forme ambiguë entre plusieurs types de relations. Selon les cas, IAI demande de reformuler ou indique son interprétation en demandant de la confirmer.

Figure 10. *Exemple de discussion avec IAI – Partie 6*

Toutes les relations sémantiques identifiées dans la conversation sont stockées et visibles à tout moment. En cas de contradiction, la dernière occurrence l'emporte. En effet, il est possible que le joueur, en répondant aux questions, se contredise (il répond non à une question, et plus tard y répond oui). L'évaluation des points ne se fait qu'au checkout. Afin d'éviter toute ambiguïté, la relation exacte, telle que définie dans JeuxDeMots, est indiquée dans une reformulation entre parenthèses. Par exemple : *le lézard a une queue* (lézard `r_has_part queue`).

Figure 11. *Exemple de discussion avec IAI – Partie 7*

Le joueur peut corriger les relations qu'il a indiquées, en les reformulant et il peut aussi en ajouter de nouvelles. Seules celles qui concernent un des termes de l'analogie seront prises en compte dans le comptage des points. Cependant, celles sans rapport avec l'analogie en cours sont tout de même mémorisées.

Figure 12. Exemple de discussion avec IAI – Partie 8

La notation avec * indique une relation devant avoir un poids négatif ; cette notation est similaire à une forme négative.

Conclusion

L'analyse automatique d'analogies, très souvent présentes dans des documents textuels ayant trait à la recherche scientifique, l'innovation technologique ou même la publicité, requiert l'identification des caractéristiques communes entre les termes impliqués dans une analogie donnée.

La capture de ces informations sous la forme de relations lexico-sémantiques plongées dans un graphe de grande taille peut avantageusement se faire via des jeux en ligne. Ainsi, une tâche plutôt rébarbative peut se transformer en une activité ludique enthousiasmante et addictive. C'est le principe du "serious game"...

Nous avons présenté deux jeux en rapport avec la découverte d'analogies et des relations que leurs termes entretiennent : 1) *Analogia*, qui est une extension du projet JeuxDeMots qui permet aux joueurs de compléter des analogies à un trou, mais également de proposer de nouvelles analogies aux autres joueurs. 2) *intersector*, qui est un jeu de type conversation textuelle avec un agent conversationnel.

Ces deux GWAPs sont des jeux portant sur la langue, et en particulier sur le concept d'analogie à travers les figures de style qu'il suscite, que ce soit la métaphore ou la simple comparaison. L'interaction est essentiellement textuelle. L'attractivité de ces deux jeux provient des associations insolites qu'il est possible de rencontrer, en raison des phénomènes d'ambiguïté sémantique. Selon les analogies et questions posées, les parties peuvent être tour à tour étonnantes, philosophiques ou encore... choquantes, non politiquement correctes, dérangeantes, voire "scandaleuses". D'une certaine façon, les jeux *Analogia* et *Intersector* sont à la langue ce que Hot Ones² est à la gastronomie.

Bibliographie

- [BAR 12] BARBARA, J-G. « Métaphores, analogies, et modèles comme pratiques interdisciplinaires dans la constitution des objets scientifiques », In F. Darbellay (dir.) *La circulation des savoirs, Interdisciplinarité, concepts nomades, analogies, métaphores*, 2012. ffhalshs-03090939f
- [BEA 17] A DE BEAUNE, S., L. HILAIRE-PEREZ, ET K. VERMEIR. « L'analogie, les techniques et les savoirs ? : Approches cognitives et historiques », In Sophie A. de Beaune, Liliane Hilaire-Pérez et Koen Vermeir (eds.) *L'analogie dans les techniques*, CNRS Éditions, p. 7-56, 2017. ffhalshs-02782128f
- [GEN 83] GENTNER, D. « Structure-mapping: A theoretical framework for analogy », *Cognitive science*, 7(2), p. 155-170, 1983.
- [GEN 97] GENTNER, D. & HOLYOAK, K. J. « Reasoning and learning by analogy: Introduction », *American psychologist*, n° 52(1), 32, 1997.

² https://fr.wikipedia.org/wiki/Hot_Ones

- [GOT 06] GOTTELAND, D., CH. HAON, DANTAS, D. C. « L'innovation par la pensée analogique », *L'Expansion Management Review*, n° 120(1), p. 79-83, 2006.
- [HIL 13] HILGERT, E. « Analyse sémantico-stylistique de la formule analogique A est à B ce que C est à D », *Actes du XXVIIe Congrès international de linguistique et de philologie romanes*, Jul 2013, Strasbourg, France, 2013. (hal-02478986)
- [HOF 95] HOFSTADTER, D. R. *Fluid concepts and creative analogies: Computer models of the fundamental mechanisms of thought*. Basic books, 1995.
- [HOF 01] HOFSTADTER, D. R. « Analogy as the core of cognition », *The analogical mind: Perspectives from cognitive science*, p. 499-538, 2001.
- [KIN 23] KING, D. C., & GENTNER, D. Verb « Metaphors are Processed as Analogies », In *Proceedings of the Annual Meeting of the Cognitive Science Society*, n° 45(45), 2023.
- [LAF 20] LAFOURCADE, M., LE BRUN, N. « JeuxDeMots : Un réseau lexico-sémantique pour le français, issu de jeux et d'inférences », *Ressources lexicales*, 2020.
- [LAF 15] LAFOURCADE, M., LE BRUN N., JOUBERT, A. *Games with a Purpose (GWAPS)*, Wiley-ISTE, 2015.
- [LAF 07] LAFOURCADE, M. « Making people play for Lexical Acquisition with the JeuxDeMots prototype », In *SNLP'07: 7th international symposium on natural language processing*, p. 7, 2007.
- [LEP 03] LEPAGE, Y. *De l'analogie rendant compte de la commutation en linguistique. Interface homme-machine* [cs.HC]. Mémoire d'Habilitation à Diriger les Recherches. Université Joseph-Fourier - Grenoble I, 2003. (tel-00004372)
- [MAR 15] MARQUIS, J-P. *Analogie et métaphore en sciences. Sciences, technologies et sociétés de A à Z*, Frédéric Bouchard et al., Presses de l'Université de Montréal, 2015. <https://doi.org/10.4000/books.pum.4252>.
- [MUR 17] MURENA, P. A., DESSALLES, J. L., & CORNUÉJOLS, A. « A Complexity Based Approach for Solving Hofstadter's Analogies », In *The International Conference on Case-Based Reasoning (ICCBR), Workshops*, p. 53-62, 2017.
- [SHU 10] SHUTOVA, E. « 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*, p. 1029-1037, 2010.
- [TOU 21] TOUSSAINT, C. « Le rôle des analogies et des métaphores dans un grand challenge. Le cas des débris spatiaux », *Revue française de gestion*, n° 8(301), p. 113-125, 2021.
- [TUR 05] TURNEY, P. D. *Measuring semantic similarity by latent relational analysis*, 2005. arXiv preprint cs/0508053.
- [TUR 10] TURNEY, P. D., & PANTEL, P. « From frequency to meaning: Vector space models of semantics », *Journal of artificial intelligence research*, n° 37, p. 141-188, 2010.

Annexes : quelques analogies représentatives

- Chien : Loup :: Chat : Tigre
- Plume : Oiseau :: Écaille : Poisson
- Livre : Lecture :: Aliment : Nourriture
- Rire : Joie :: Pleurs : Tristesse
- Soleil : Jour :: Lune : Nuit
- Peintre : Peinture :: Auteur : Livre
- École : Éducation :: Hôpital : Santé
- Clef : Porte :: Mot de passe : Ordinateur
- Voiture : Roues :: Avion : Ailes
- Fleur : Bouquet :: Arbre : Forêt
-
- Feu : Chaleur :: Glace : Froid
- Peintre : Tableau :: Auteur : Livre
- Plante : Pousse :: Enfant : Grandit
- Mouton : Laine :: Vache : Lait
- Montre : Temps :: Thermomètre : Température
- Professeur : Enseigner :: Médecin : Soigner
- Architecte : Bâtiment :: Compositeur : Symphonie
- Automne : Feuilles :: Hiver : Neige
- Oiseau : Nid :: Abeille : Ruche
- Musicien : Orchestre :: Acteur : Troupe
-
- Proton : Positif :: Électron : Négatif
- ADN : Gènes :: Livre : Chapitres
- Neurone : Impulsion nerveuse :: Fil électrique : Courant
- Chloroplaste : Photosynthèse :: Mitochondrie : Respiration cellulaire
- Enzyme : Catalyseur :: Clé : Serrure
- Voltage : Électricité :: Pression : Eau
- Isotope : Élément chimique :: Allèle : Gène
- Circulation sanguine : Corps humain :: Circuit électrique : Appareil
- Pluie : Cycle de l'eau :: Transpiration : Cycle du carbone
-
- Processeur : Ordinateur :: Cerveau : Corps
- Clavier : Saisie de texte :: Souris : Navigation
- RAM : Mémoire vive :: Disque dur : Mémoire de stockage

- Application : Smartphone :: Logiciel : Ordinateur
- Modem : Internet :: Émetteur : Radio
- PDF : Document numérique :: Papier : Document physique
- Réseau local : Intranet :: Réseau étendu : Internet
- Serveur : Hébergement de sites web :: Bibliothèque : Stockage de livres
- Batterie : Smartphone :: Réservoir de carburant : Voiture
- Algorithme : Traitement des données :: Recette : Préparation de repas
-
- Poète : Poème :: Romancier : Roman
- Métaphore : Comparaison :: Hyperbole : Exagération
- Tragédie : Dramatique :: Comédie : Humoristique
- Antagoniste : Opposant :: Protagoniste : Héros
- Sonnet : Poésie :: Nouvelle : Prose
- Allitération : Sonorité :: Épithète : Description
- Narrateur : Histoire :: Auteur : Livre
- Scène : Acte :: Chapitre : Livre
- Dialogue : Personnages :: Monologue : Individu
- Thème : Sujet :: Motif : Répétition
-
- Poète : Poème :: Romancier : Roman
- Métaphore : Comparaison :: Hyperbole : Exagération
- Tragédie : Dramatique :: Comédie : Humoristique
- Antagoniste : Opposant :: Protagoniste : Héros
- Sonnet : Poésie :: Nouvelle : Prose
- Allitération : Sonorité :: Épithète : Description
- Narrateur : Histoire :: Auteur : Livre
- Scène : Acte :: Chapitre : Livre
- Dialogue : Personnages :: Monologue : Individu
- Thème : Sujet :: Motif : Répétition
-
- Plume : Poète :: Pinceau : Peintre
- Lune : Nuit :: Soleil : Jour
- Rose : Amour :: Épine : Douleur
- Mer : Infini :: Ciel : Éternité
- Vers : Poésie :: Note : Musique
- Métaphore : Comparaison :: Symbole : Signification
- Feuille : Arbre :: Page : Livre

- Vent : Murmure :: Tonnerre : Clameur
- Étoile : Ciel :: Perle : Océan
- Ombre : Lumière :: Silence : Son
- Fleur : Jardin :: Poème : Recueil
- Cascade : Rivière :: Éclair : Tempête
- Sable : Désert :: Neige : Hiver
- Aurore : Espoir :: Crépuscule : Nostalgie
- Rêve : Sommeil :: Fantaisie : Esprit
- Chanson : Harmonie :: Poésie : Rythme
- Arc-en-ciel : Pluie :: Sourire : Larmes
- Brume : Mystère :: Éclaircie : Révélation
- Écho : Montagne :: Reflet : Miroir
- Papillon : Métamorphose :: Poème : Inspiration
-
- Lire : Comprendre :: Écouter : Entendre
- Manger : Nourrir :: Boire : Hydrater
- Parler : Communiquer :: Écrire : Correspondre
- Travailler : Gagner :: Étudier : Apprendre
- Dormir : Reposer :: Courir : Exercer
- Enseigner : Éduquer :: Soigner : Guérir
- Chercher : Trouver :: Questionner : Répondre
- Acheter : Payer :: Vendre : Recevoir
- Conduire : Voiture :: Piloter : Avion
- Chanter : Musique :: Danser : Rythme
- Cuisiner : Nourriture :: Peindre : Art
- Nager : Eau :: Voler : Air
- Programmer : Code :: Composer : Musique
- Lire : Livre :: Regarder : Film
- Réparer : Objet :: Guérir : Corps
- Rire : Joie :: Pleurer : Tristesse
- Observer : Analyse :: Imaginer : Création
- Construire : Bâtiment :: Cultiver : Plante
- Jouer : Jeu :: Travailler : Projet
- Dessiner : Image :: Écrire : Texte
- Rose : Amour :: Épine : Douleur
- Lune : Nuit :: Soleil : Jour
- Poème : Sentiment :: Mélodie : Chanson

- Étoile : Ciel :: Perle : Océan
- Plume : Poète :: Pinceau : Artiste
- Lettre d'amour : Émotion :: Fleur : Parfum
- Baiser : Passion :: Câlin : Tendresse
- Vent : Murmure :: Mer : Mélodie
- Flamme : Bougie :: Éclat : Diamant
- Songe : Nuit :: Fantaisie : Jour
- Arc-en-ciel : Pluie :: Sourire : Larmes
- Papillon : Métamorphose :: Amour : Évolution
- Aurore : Espoir :: Crépuscule : Nostalgie
- Rêve : Sommeil :: Poème : Inspiration
- Fleur : Jardin :: Étoile : Ciel
- Parfum : Souvenir :: Chanson : Mémoire
- Brise : Matin :: Baiser : Amant
- Sable : Plage :: Neige : Montagne
- Éclair : Orage :: Passion : Amour
- Coeur : Émotion :: me : Vérité
- Chat : Miauler :: Chien : Aboier
- Crayon : Dessin :: Pinceau : Peinture
- Soleil : Jour :: Lune : Nuit
- Oiseau : Nid :: Abeille : Ruche
- Pizza : Fromage :: Burger : Viande
- Bateau : Mer :: Avion : Ciel
- Super-héros : Cape :: Pirate : Trésor
- Neige : Bonhomme de neige :: Sable : Château de sable
- Chocolat : Bonbon :: Lait : Jus
- Dinosaur : Fossile :: Chevalier : Armure
- Lutin : Père Noël :: Lapin : Pâques
- Fée : Baguette magique :: Sorcier : Chaudron
- Train : Gare :: Bus : Arrêt
- Crêpe : Sirop :: Glace : Cornet
- Tigre : Rayures :: Zèbre : Zébrures
- Parc : Jeu :: Bibliothèque : Livre
- Poisson : Aquarium :: Hamster : Cage
- Magicien : Chapeau :: Clown : Nez rouge
- Étoile : Ciel :: Coquillage : Plage
- Souris : Fromage :: Écureuil : Noix

-
- Smartphone : Communication :: Drone : Livraison
- Impression 3D : Fabrication :: Internet : Information
- Voiture électrique : Transport :: Énergie solaire : Électricité
- Blockchain : Sécurité :: Cloud : Stockage
- Intelligence artificielle : Analyse :: Robotique : Exécution
- Télétravail : Bureau :: E-commerce : Magasin
- Réalité virtuelle : Expérience :: Simulateur : Entraînement
- Nanotechnologie : Matériaux :: Biotechnologie : Santé
- Big Data : Données :: Machine learning : Modèles
- Fintech : Finance :: Edtech : Éducation
- Voiture autonome : Conducteur :: Maison intelligente : Habitants
- Impression 3D : Prototypage :: Réalité augmentée : Visualisation
- Capteurs : IoT (Internet des objets) :: Récepteurs : Communication
- Biocarburant : Pétrole :: Alimentation à base de plantes : Viande
- Visioconférence : Réunion :: E-book : Livre
- Crowdfunding : Financement :: Réseaux sociaux : Communication
- Wearable tech : Suivi de la santé :: GPS : Navigation
- Smart grid : Réseau électrique :: Fibre optique : Internet
- Impression 3D : Construction :: Génie génétique : Agriculture
- Cryptomonnaie : Monnaie traditionnelle :: Voiture électrique : Voiture à essence