

# Vers un système de raisonnement neurosymbolique ancré dans des représentations schématiques

François Olivier<sup>1</sup>, Zied Bouraoui<sup>2</sup>

<sup>1</sup>STIH, Sorbonne Université, Paris, France

<sup>2</sup>CRIL, CNRS & Université d'Artois, Lens, France

francois.olivier@sorbonne-universite.fr    bouraoui@cril.fr

## Résumé

*Cet article présente Embodied-LM, un prototype de système neurosymbolique qui ancre le raisonnement dans des représentations schématiques inspirées de la théorie cognitive des image schemas. Le modèle de langage identifie les structures schématiques pertinentes dans des énoncés et les traduit en programmes logiques exécutés par un système de raisonnement ASP enrichi de prédicats spatiaux. Des expériences sur des tâches de déduction logique montrent que ces représentations permettent un raisonnement interprétable et efficace.*

## Mots-clés

*IA neurosymbolique, représentation des connaissances, linguistique cognitive, raisonnement spatial, modèles de langage*

## Abstract

*This paper presents Embodied-LM, a prototype neurosymbolic system that grounds reasoning in schematic representations inspired by the cognitive theory of image schemas. The language model identifies relevant schematic structures in sentences and translates them into logical programs executed by an ASP-based reasoning system enriched with spatial predicates. Experiments on logical deduction tasks show that these representations enable interpretable and effective reasoning.*

## Keywords

*Neurosymbolic AI, knowledge representation, cognitive linguistics, spatial reasoning, language models*

## 1 Le raisonnement incarné

Le raisonnement à partir d'énoncés en langage naturel pose un problème central en intelligence artificielle : comment représenter de manière structurée les contraintes décrites dans le langage afin de permettre une inférence fiable ? Si les systèmes récents fondés sur des modèles de langage atteignent des performances impressionnantes, leurs processus de raisonnement restent souvent difficiles à inspecter et peuvent produire des inférences erronées, tout en demeurant peu alignés avec les mécanismes cognitifs humains [13, 11, 23]. Une approche consiste alors à compléter

ces systèmes par un module de raisonnement symbolique qui ancre l'inférence dans des représentations structurées proches de celles mobilisées par l'humain [17, 16].

Dans le cadre de la théorie de la cognition incarnée, Johnson a introduit les *image schemas* [7], définis comme des structures récurrentes abstraites issues de nos interactions avec le monde physique, qui organisent aussi bien la pensée concrète qu'abstraite [12, 6]. Par exemple, le schéma CONTAINER, dérivé de nos expériences d'objets placés dans des récipients, peut être mobilisé pour comprendre des phrases comme "Bill est dans la maison", mais aussi des expressions plus abstraites comme "Berlin est en Allemagne", "she is in love" ou encore des expressions mathématiques comme  $2 \in \mathbb{N}$  [9, 10]. Le schéma PATH, issu de nos expériences de déplacement d'un point de départ vers un point d'arrivée, structure de la même manière des expressions comme "ne pas s'arrêter en si bon chemin", mais aussi des représentations plus abstraites comme la ligne mentale des nombres [10] ou des comparaisons temporelles. Le schéma BALANCE, lié à l'expérience corporelle de l'équilibre, peut quant à lui intervenir dans la compréhension d'expressions comme "pencher en faveur de quelqu'un", ou encore dans la représentation des équations mathématiques comme deux termes devant rester en équilibre.

Lorsqu'on donne une réalisation graphique ou diagrammatique à ces structures schématiques, l'information organisée par les *image schemas* prend la forme d'une configuration spatiale manipulable [15]. Certaines relations peuvent alors devenir immédiatement explicites, alors qu'elles n'étaient pas directement données dans le texte. Un exemple simple est celui de l'utilisation du schéma PATH. Considérons l'énoncé suivant : « Alice est plus âgée que Bill, et Charles est plus jeune que Bill ». Comprendre cet énoncé revient à projeter les individus sur une dimension ordonnée représentant l'âge, ce qui conduit à une configuration comme celle de la figure 1.

Une fois cette configuration construite, certaines conséquences deviennent immédiatement visibles : par exemple, Alice apparaît directement comme plus âgée que Charles, bien que cette relation ne soit pas explicitement énoncée. Inférer cette nouvelle information repose alors sur la simple observation de la structure même de la représentation, plutôt que sur une manipulation symbolique [21].

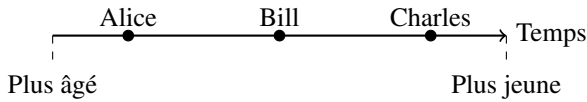


FIGURE 1 – Projection spatiale d’un ordre temporel au moyen du schéma PATH.

C’est précisément ce principe inférentiel, proche du raisonnement diagrammatique, que le système Embodied-LM cherche à formaliser dans un cadre de raisonnement automatique [17]. La section suivante en présente l’implémentation.

## 2 Implémentation

Le prototype Embodied-LM [17] repose sur un pipeline neurosymbolique simple exposé à la figure 2. À partir d’un énoncé en langage naturel, un modèle de langage produit un programme d’entrée où les contraintes logiques de l’énoncé sont interprétées métaphoriquement à l’aide de schémas spatiaux. Ce programme est ensuite résolu par un système en *Answer Set Programming* [8], enrichi de contraintes spatiales exprimées sous forme de contraintes polynomiales [1]. Les détails de la génération des différentes parties du programme d’entrée, ainsi que les aspects techniques du raisonneur symbolique sont décrits dans ce qui suit.

**Module linguistique.** Ce premier module utilise GPT-4 [19] par *prompting* pour interpréter l’énoncé au moyen d’*image schemas* et produire un programme logique décrivant le contexte du problème. Les entités mentionnées dans l’énoncé et les relations qui les relient sont ainsi identifiées puis traduites dans le vocabulaire spatial spécifié dans le prompt.

### Extrait du prompt fourni à GPT-4

You are an expert at writing declarative programs that encode the interpretation of natural language narratives through image schemas - recurring patterns derived from sensorimotor experience that structure human cognition. Declarative programs are written in Answer Set Programming augmented with spatial theory predicates.

#### OBJECT DECLARATION:

Objects in narratives correspond to geometric objects in the program. Specify the geometric sort of one or more objects by "point(obj1;obj2;...)." ...

#### BUILTIN SPATIAL PREDICATES AVAILABLE:

- left(P1,P2): Point P1 is left to point P2.  
- on(P1,L2): Point P1 is on line L2.  
- ...

Generate the Clingo program with the qualitative spatial predicates available to encode the spatial schematic interpretation of this natural language input.

{CONTEXT}

Dans le cas de l’énoncé introduit dans la section précédente, les individus sont ainsi représentés comme des points situés sur un segment abstrait correspondant à une dimension ordonnée. Le programme généré par ce module linguistique pourrait alors être le suivant.

Listing 1 – Programme généré pour l’énoncé "Alice est plus âgée que Bill, et Charles est plus jeune que Bill".

```
1 point(alice; bill; charles).
2 segment(timeline).
3
4 on(X, timeline) :- point(X).
5 left(alice, bill).
6 right(charles, bill).
```

GPT-4 n’a donc pas à inventer la signification des prédicats utilisés : il sélectionne, parmi un vocabulaire spatial fixé à l’avance, les prédicats qui correspondent aux contraintes exprimées dans l’énoncé. Une première utilisation du module symbolique vérifie que le programme généré est conforme aux spécifications données et relance la génération d’un programme si ce n’est pas le cas.

Lorsque l’énoncé est accompagné d’une question, le module linguistique est aussi chargé, par le biais d’un second prompt, d’étendre le programme par des règles de la forme `answer(a) :- condition`, représentant les différentes options de réponse. Ces règles sont ensuite ajoutées au programme initial.

**Module de raisonnement.** Le programme obtenu est ensuite exécuté par *clingo*, dont la syntaxe d’entrée est enrichie par des prédicats spatiaux [8, 3]. La sémantique de ces prédicats est définie analytiquement par des équations polynomiales portant sur les paramètres géométriques des objets, comme décrit dans le *Declarative Spatial Reasoning framework* [1, 24]. Par exemple, une relation comme *left* entre deux points peut être interprétée par la contrainte  $left(a, b) \equiv x_a < x_b$ , où  $x_a$  et  $x_b$  désignent les coordonnées des points  $a$  et  $b$  sur l’axe des abscisses. De même, une relation comme *on* entre un point et un segment peut être définie par des contraintes exprimant que le point appartient à la droite portée par le segment et se situe entre ses deux extrémités.

Pour la recherche de modèles, *clingo* traite la partie logique du programme en déterminant quelles relations doivent être satisfaites à partir des règles déclarées, tandis que les contraintes paramétriques associées à ces relations sont résolues par le solveur SMT Z3 [5]. Un modèle est généré lorsque les contraintes logiques et spatiales du programme sont satisfaites simultanément. Lorsqu’un même objet apparaît dans plusieurs relations, les contraintes correspondantes portent sur les mêmes paramètres géométriques. Par exemple, si le programme contient à la fois  $left(a, b)$  et  $right(c, b)$ , Z3 cherche une même affectation des coordonnées de  $a$ ,  $b$  et  $c$  qui satisfasse simultanément  $x_a < x_b$  et  $x_c > x_b$ .

Une option de réponse, comme par exemple `answer(a)`, est considérée comme correcte lorsqu’elle apparaît dans tous les modèles générés par le programme. Les solutions

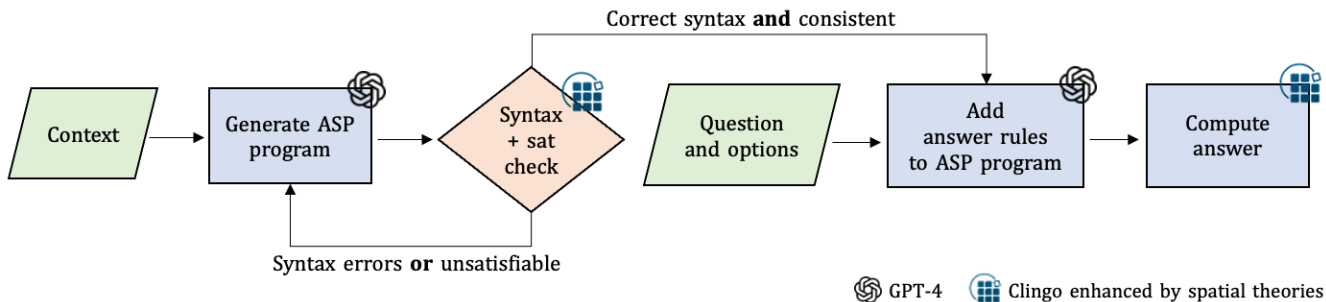


FIGURE 2 – Architecture d’Embodied-LM [17] : GPT-4 traduit l’énoncé en programme ASP enrichi par des prédicats spatiaux, puis le module de raisonnement combine *clingo* et Z3 afin de produire une réponse et, lorsque c’est utile, une représentation schématique du problème.

géométriques satisfaisant les contraintes spatiales constituent alors des *witnesses* qui attestent de la réalisabilité des relations spatiales impliquées dans le raisonnement. Ces *witnesses* peuvent également être visualisés [17] afin de produire des représentations schématiques du problème, analogues à celle de la figure 1, et ainsi rendre explicite la structure spatiale sur laquelle repose l’inférence.

### 3 Résultats et discussion

Nous évaluons Embodied-LM sur le dataset *LogicalDeduction* [22], composé de problèmes proches de l’exemple considéré ici : des entités sont ordonnées ou comparées à partir de relations exprimées en langage naturel, puis une question à choix multiple porte sur les conséquences logiques de ces prémisses.

System	Score
SymbCoT	93.00
<b>Embodied-LM</b>	<b>91.00</b>
VERUS-LM	88.67
Logic-LM	87.63
GPT4-CoT	75.25
GPT4	71.33

TABLE 1 – Performance sur le jeu de données *LogicalDeduction*. Résultats des autres systèmes repris de [4].

Comme le montre la Table 1, Embodied-LM atteint un score de 91%, proche des meilleurs systèmes récents. Nous avons également testé l’approche sur des *zebra puzzles* (aussi appelés *énigmes d’Einstein*), une famille classique de problèmes logiques consistant à déterminer une configuration d’objets satisfaisant un ensemble de contraintes relationnelles. Contrairement au dataset *LogicalDeduction*, ces problèmes sont plus naturellement représentés à l’aide du schéma CONTAINER, et Embodied-LM parvient également à les résoudre.

Ces expériences fournissent une validation conceptuelle de l’approche : elles montrent qu’un raisonnement fondé sur des structures schématiques peut être implémenté sous forme de programmes exécutables et appliqué à des tâches de déduction logique exprimées en langage naturel. Le ré-

sultat principal n’est donc pas seulement la performance obtenue, mais le fait qu’une même représentation puisse à la fois être inspectable, manipulable par un solveur, et reliée à des principes cognitifs de compréhension du langage.

Cette perspective suggère également qu’un langage schématique relativement simple peut servir de base commune pour représenter des situations variées. Les prédicats spatiaux utilisés possèdent en effet une sémantique géométrique fixée indépendamment du modèle de langage, ce qui réduit la charge imposée à celui-ci, puisqu’il n’a plus à apprendre ou à reconstruire le sens des prédicats dans chaque programme généré. Cette propriété contraste avec d’autres approches neurosymboliques reposant sur des prédicats définis *ad hoc* pour chaque problème [20, 4], dont la généralisation à de nouveaux domaines peut alors dépendre plus fortement de la capacité du modèle de langage à produire des prédicats correctement interprétables.

L’approche bénéficie également d’un ancrage formel solide, dans la mesure où la sémantique logique sous-jacente aux programmes ASP, à savoir la logique de *Here-and-There* (HT), a fait l’objet de nombreux travaux et a été étendue avec contraintes dans la logique HTc. L’étude de HTc avec des contraintes spatiales a été proposée dans [18], établissant ainsi le cadre formel sous-jacent aux programmes d’Embodied-LM. Plus récemment, la logique HTc a été enrichie par des opérateurs temporels [2], fournissant ainsi un cadre naturel pour caractériser des représentations schématiques dynamiques, où les objets, leurs paramètres et les relations qui les lient peuvent évoluer au cours du temps.

Les prochains axes de développement consistent, d’une part, à évaluer le système sur d’autres jeux de données où un raisonnement statique suffit, et, d’autre part, à exploiter cette possibilité d’extension dynamique afin de traiter des problèmes de raisonnement tels que la *tour de Hanoï* ou encore le *Blocks World*. Cette même extension viserait également à faire d’Embodied-LM un cadre de psychologie computationnelle pour la modélisation du raisonnement humain [14, 15] : en ajoutant des mécanismes comme un focus attentionnel, le système pourrait servir à simuler non seulement les processus de résolution des participants, mais aussi certaines préférences de construction des représentations ou certains biais observés dans le raisonnement.

## Références

- [1] Mehul Bhatt, Jae Hee Lee, and Carl Schultz. CLP(QS) : A declarative spatial reasoning framework. In Max J. Egenhofer, Nicholas A. Giudice, Reinhard Moratz, and Michael F. Worboys, editors, *Spatial Information Theory - 10th International Conference, COSIT 2011, Belfast, ME, USA, September 12-16, 2011. Proceedings*, volume 6899 of *Lecture Notes in Computer Science*, pages 210–230. Springer, 2011.
- [2] Pedro Cabalar, Martín Diéguez, François Olivier, Torsten Schaub, and Igor Stephan. Towards constraint temporal answer set programming. *Theory and Practice of Logic Programming*, 2025.
- [3] Francesco Calimeri, Wolfgang Faber, Martin Gebser, Giovambattista Ianni, Roland Kaminski, Thomas Krennwallner, Nicola Leone, Marco Maratea, Francesco Ricca, and Torsten Schaub. Asp-core-2 input language format. *Theory and Practice of Logic Programming*, 20(2) :294–309, 2020.
- [4] Benjamin Callewaert, Simon Vandeveld, and Joost Vennekens. Verus-lm : a versatile framework for combining llms with symbolic reasoning, 2025.
- [5] Leonardo De Moura and Nikolaj Bjørner. Z3 : An efficient smt solver. In *Proceedings of the 14th International Conference on Tools and Algorithms for the Construction and Analysis of Systems (TACAS)*, pages 337–340. Springer, 2008.
- [6] Maria M Hedblom. *Image schemas and concept invention : cognitive, logical, and linguistic investigations*. Springer Nature, 2020.
- [7] Mark Johnson. *The body in the mind : The bodily basis of reason and imagination*. Chicago : University of Chicago Press, 1987.
- [8] Roland Kaminski, Javier Romero, Torsten Schaub, and Philipp Wanko. How to build your own asp-based system ?! *Theory and Practice of Logic Programming*, 23(1) :299–361, 2023.
- [9] George Lakoff and Mark Johnson. *Metaphors We Live By*. University of Chicago Press, 1980.
- [10] George Lakoff and Rafael Núñez. *Where mathematics comes from*, volume 6. New York : Basic Books, 2000.
- [11] Kyle Mahowald, Anna A. Ivanova, Idan A. Blank, Nancy Kanwisher, Joshua B. Tenenbaum, and Evelina Fedorenko. Dissociating language and thought in large language models. *Trends in Cognitive Sciences*, 28(6) :517–540, 2024.
- [12] Jean M Mandler and Cristóbal Pagán Cánovas. On defining image schemas. *Language and Cognition*, 6(4) :510–532, 2014.
- [13] R. Thomas McCoy, Shunyu Yao, Dan Friedman, Matthew D. Hardy, and Thomas L. Griffiths. Embers of autoregression show how large language models are shaped by the problem they are trained to solve. *Proceedings of the National Academy of Sciences of the United States of America*, 121(41) :e2322420121, 2024.
- [14] François Olivier. A logical framework for spatial mental models. In J. Škilters, N. Newcombe, and D. Uttal, editors, *German Conference on Spatial Cognition XII*, pages 268–280. Springer, 2020.
- [15] François Olivier. *Spatial relations in reasoning : a computational model*. PhD thesis, Université Paris sciences et lettres, 2022.
- [16] François Olivier and Zied Bouraoui. Grounding agent reasoning in image schemas : A neurosymbolic approach to embodied cognition. In *Proceedings of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025)*, pages 2875–2879. IFAAMAS, 2025.
- [17] François Olivier and Zied Bouraoui. Towards a neurosymbolic reasoning system grounded in schematic representations. In *Proceedings of the 19th International Conference on Neurosymbolic Learning and Reasoning (NeSy 2025)*, 2025.
- [18] François Olivier and Carl Schultz. Here-and-there with constraints for spatial reasoning. In *Proceedings of the Sixth Workshop on Trends and Applications of Answer Set Programming (TAASP’23)*, 2023.
- [19] OpenAI. Gpt-4 technical report. *arXiv preprint arXiv :2303.08774*, 2023.
- [20] Liangming Pan, Alon Albalak, Xinyi Wang, and William Wang. Logic-lm : Empowering large language models with symbolic solvers for faithful logical reasoning. In *Findings of the Association for Computational Linguistics : EMNLP 2023*, pages 3806–3824, 2023.
- [21] Atsushi Shimojima. *Semantic Properties of Diagrams and Their Cognitive Potentials*. CSLI Publications, Stanford, California, 2015.
- [22] A Srivastava, A Rastogi, A Rao, AAM Shoeb, A Abid, A Fisch, AR Brown, A Santoro, A Gupta, A Garriga-Alonso, et al. Beyond the imitation game : Quantifying and extrapolating the capabilities of language models (arxiv : 2206.04615). arxiv, 2023.
- [23] Ronen Tamari, Chen Shani, Tom Hope, Miriam R L Petruck, Omri Abend, and Dafna Shahaf. Language (re)modelling : Towards embodied language understanding. In Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault, editors, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6268–6281, Online, July 2020. Association for Computational Linguistics.
- [24] Przemysław Andrzej Wałęga, Carl Schultz, and Mehul Bhatt. Non-monotonic spatial reasoning with answer set programming modulo theories. *Theory and Practice of Logic Programming*, 17(2) :205–225, 2017.