Ismaïl Baaj, Jean-Philippe Poli and Wassila Ouerdane

Some Insights Towards a Unified Semantic Representation of Explanation for eXplainable Artificial Intelligence (XAI)

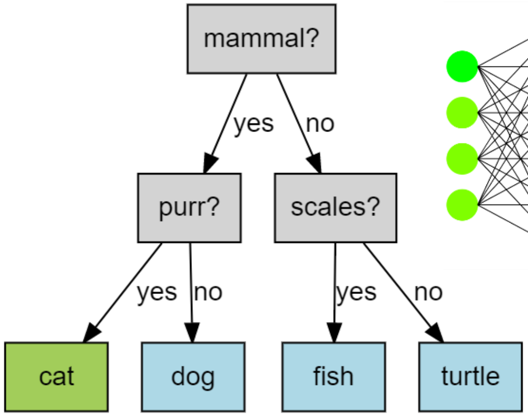
NL4XAI workshop | 12th International Conference on NLG | 29th October 2019

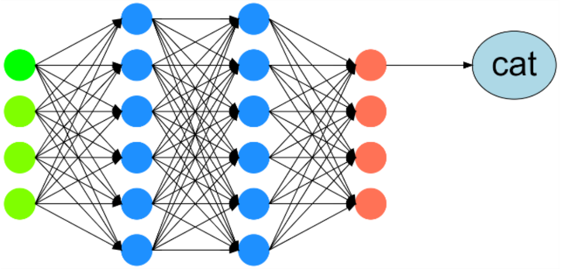
OUTLINE

- Need for a semantic representation of explanation
- Conceptual graph structures
- Example of representation : automatic image annotation explanation
- Conclusion

NEED FOR A SEMANTIC REPRESENTATION OF EXPLANATION

CHALLENGES OF XAI

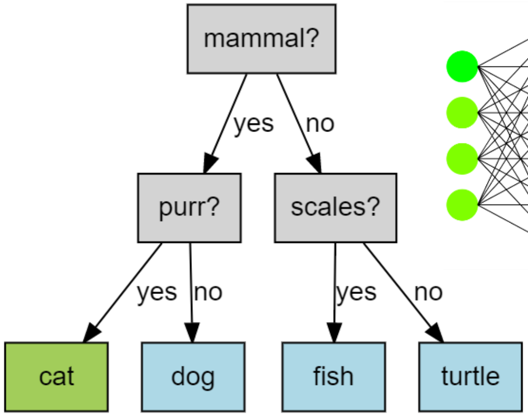


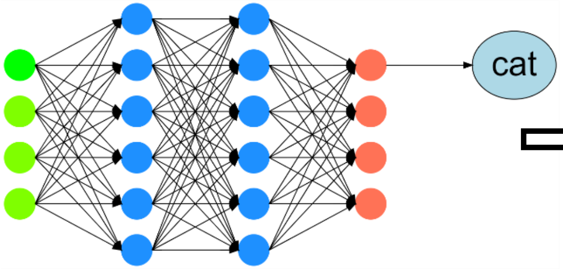


Step 1: XAI systems justify their decisions by selecting clues of their reasoning

NEED FOR A SEMANTIC REPRESENTATION OF EXPLANATION

CHALLENGES OF XAI





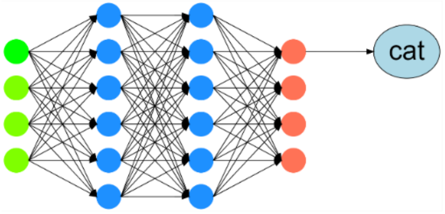
This animal is a cat.
It is a mammal and it purrs.

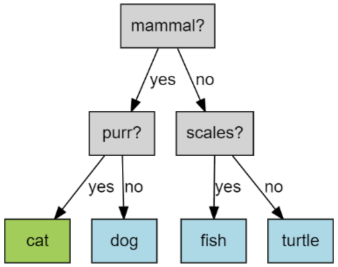
Step 1: XAI systems justify their decisions by selecting clues of their reasoning

Step 2: XAI systems use surface realizer to produce textual explanations

NEED FOR A SEMANTIC REPRESENTATION OF EXPLANATION

XAI ARCHITECTURE





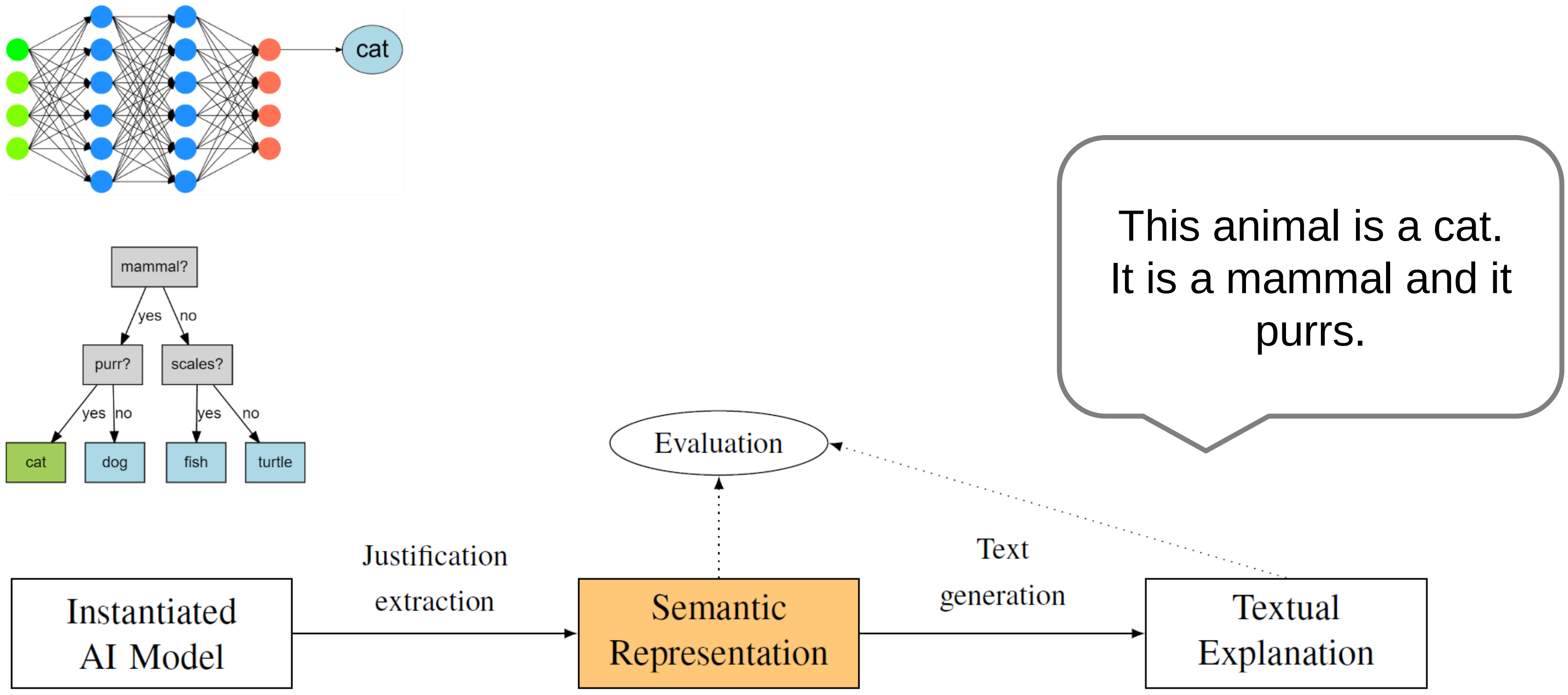
This animal is a cat. It is a mammal and it purrs.

Instantiated AI Model

Textual Explanation

NEED FOR A SEMANTIC REPRESENTATION OF EXPLANATION

XAI ARCHITECTURE



This animal is a cat. It is a mammal and it purrs.

NEED FOR A SEMANTIC REPRESENTATION OF EXPLANATION

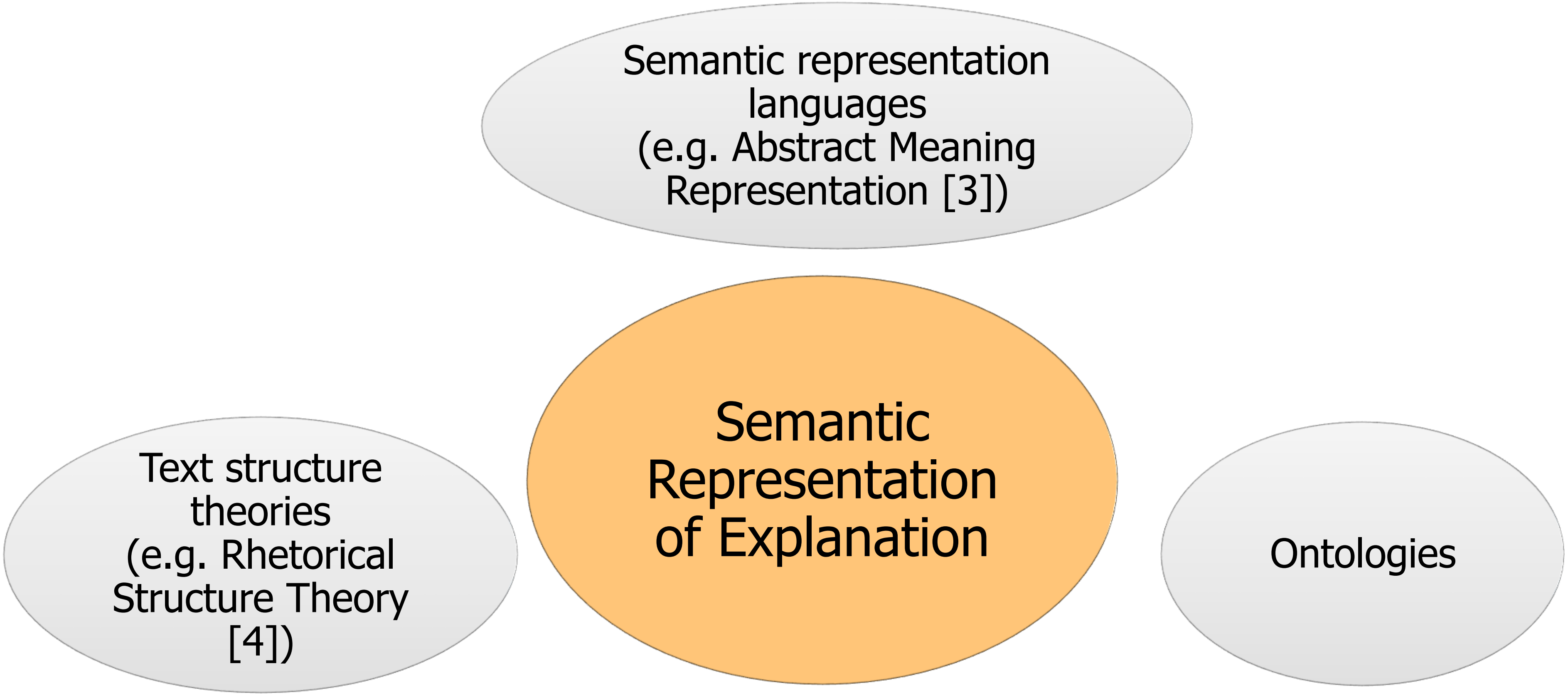
REQUIREMENTS

- Expressing **temporal, spatial, causal, agents** and **their intentions** knowledge [1]
- Representing basic **logical inference**
- Guaranteeing a sufficient level of **granularity**
- Highlighting important **aspects of explanations** (e.g. contrast) [2]
- …

[1] Zwaan, R.A. and Radvansky, G.A., 1998. Situation models in language comprehension and memory. *Psychological bulletin*, *123*(2), p.162.

[2] Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*. vol. 267,pp.1–38, 2019.

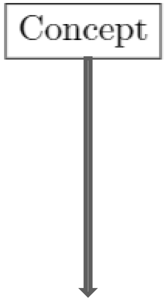
NEED FOR A SEMANTIC REPRESENTATION OF EXPLANATION


Semantic representation languages (e.g. Abstract Meaning Representation [3])
Text structure theories (e.g. Rhetorical Structure Theory [4])
Semantic Representation of Explanation
Ontologies

[3] Banarescu, L., Bonial, C., Cai, S., Georgescu, M., Griffitt, K., Hermjakob, U., Knight, K., Koehn, P., Palmer, M. and Schneider, N., 2013, August. Abstract meaning representation for sembanking. In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse* (pp. 178-186).
[4] Mann, W.C. and Thompson, S.A., 1988. Rhetorical structure theory: Toward a functional theory of text organization. *Text-interdisciplinary Journal for the Study of Discourse*, *8*(3), pp.243-281.

CONCEPTUAL GRAPH STRUCTURES

5 TYPES OF NODES



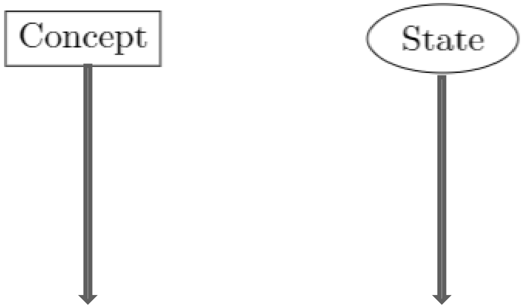
Concept

“driver”
“car”

[5] Arthur C Graesser, Peter Wiemer-Hastings, and Katja Wiemer-Hastings. 2001. Constructing inferences and relations during text comprehension. *Text representation: Linguistic and psycholinguistic aspects*, 8:249–271

CONCEPTUAL GRAPH STRUCTURES

5 TYPES OF NODES



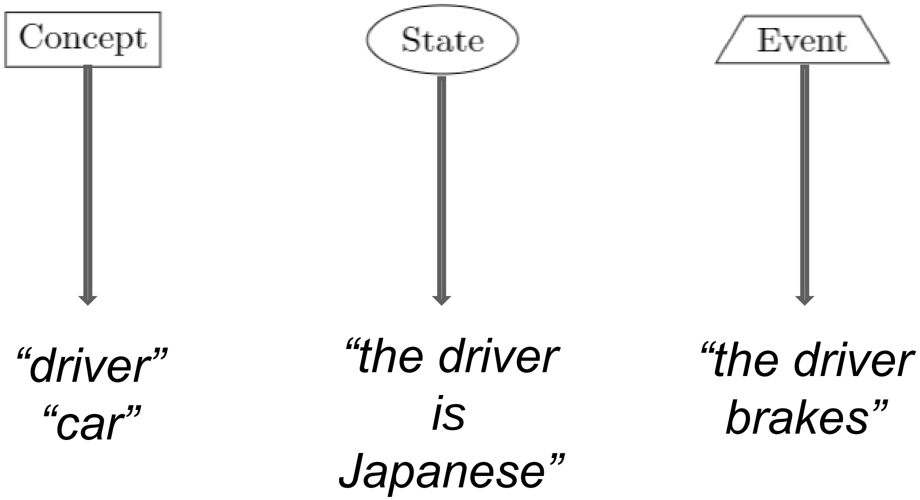
“driver”
“car”

“the driver
is
Japanese”

[5] Arthur C Graesser, Peter Wiemer-Hastings, and Katja Wiemer-Hastings. 2001. Constructing inferences and relations during text comprehension. *Text representation: Linguistic and psycholinguistic aspects*, 8:249–271

CONCEPTUAL GRAPH STRUCTURES

5 TYPES OF NODES



“driver”
“car”

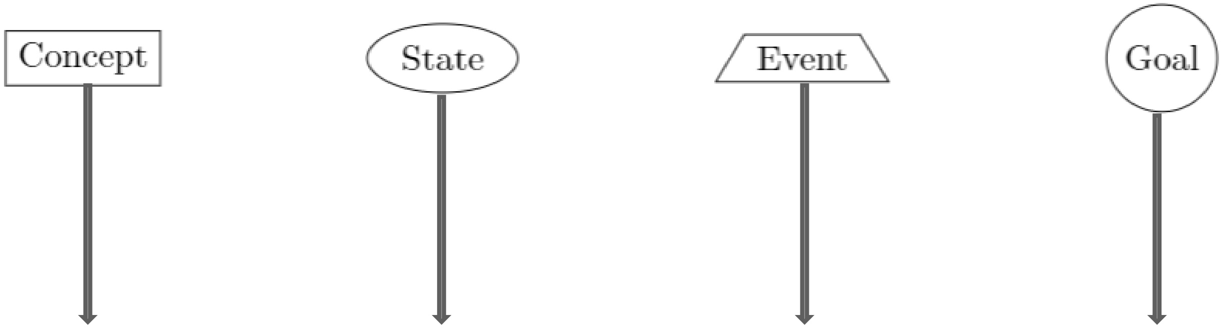
“the driver is Japanese”

“the driver brakes”

[5] Arthur C Graesser, Peter Wiemer-Hastings, and Katja Wiemer-Hastings. 2001. Constructing inferences and relations during text comprehension. *Text representation: Linguistic and psycholinguistic aspects*, 8:249–271

CONCEPTUAL GRAPH STRUCTURES

5 TYPES OF NODES



“driver”
“car”

“the driver
is
Japanese”

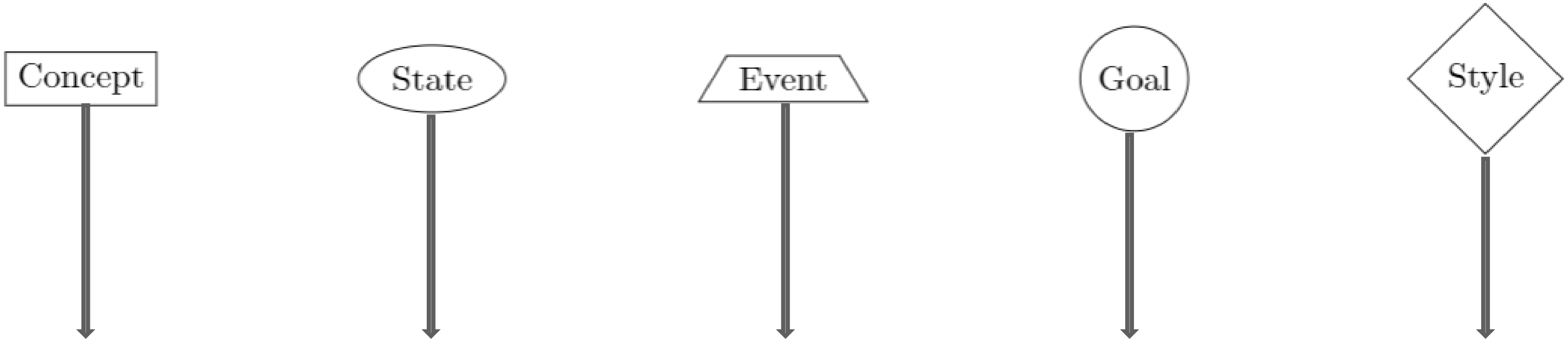
“the driver
brakes”

“the driver
wants to
arrive at
9:00pm”

[5] Arthur C Graesser, Peter Wiemer-Hastings, and Katja Wiemer-Hastings. 2001.
Constructing inferences and relations during text comprehension.
Text representation: Linguistic and psycholinguistic aspects, 8:249–271

CONCEPTUAL GRAPH STRUCTURES

5 TYPES OF NODES



"driver"
"car"

"the driver
is
Japanese"

"the driver
brakes"

"the driver
wants to
arrive at
9:00pm"

"slowly"

[5] Arthur C Graesser, Peter Wiemer-Hastings, and Katja Wiemer-Hastings. 2001. Constructing inferences and relations during text comprehension. *Text representation: Linguistic and psycholinguistic aspects*, 8:249–271

CONCEPTUAL GRAPH STRUCTURES

RELATIONS

- **22 relations with composition rules and definitions**

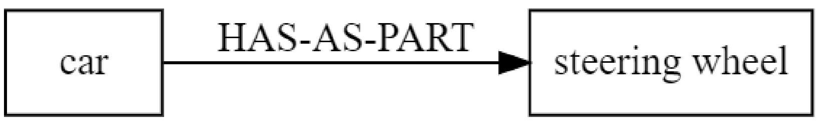
Relation: HAS-AS-PART	Relation: INITIATES
• Synonym: HAS-COMPONENT • Inverse: IS-A-PART-OF • Definition: *A has a part or component of B* • Composition rule: • (concept) – HAS-AS-PART → (concept)	• Synonym: ELICITS • Inverse: CONDITION, CIRCUMSTANCE, SITUATION • Negation: DISABLES • Definition: *A initiates or elicits a goal B* • Composition rule: • (event|state|style) – INITIATES → (goal)

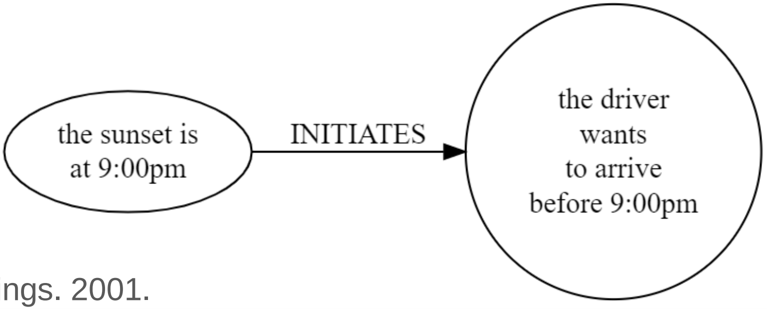
[5] Arthur C Graesser, Peter Wiemer-Hastings, and Katja Wiemer-Hastings. 2001. Constructing inferences and relations during text comprehension. *Text representation: Linguistic and psycholinguistic aspects*, 8:249–271

CONCEPTUAL GRAPH STRUCTURES
RELATIONS

- **22 relations with composition rules and definitions**

Relation: HAS-AS-PART	Relation: INITIATES
• Synonym: HAS-COMPONENT • Inverse: IS-A-PART-OF • Definition: *A has a part or component of B* • Composition rule: • (concept) – HAS-AS-PART → (concept)	• Synonym: ELICITS • Inverse: CONDITION, CIRCUMSTANCE, SITUATION • Negation: DISABLES • Definition: *A initiates or elicits a goal B* • Composition rule: • (event|state|style) – INITIATES → (goal)





[5] Arthur C Graesser, Peter Wiemer-Hastings, and Katja Wiemer-Hastings. 2001. Constructing inferences and relations during text comprehension. *Text representation: Linguistic and psycholinguistic aspects*, 8:249–271

EXAMPLE OF REPRESENTATION

AUTOMATIC IMAGE ANNOTATION EXPLANATION

- **Explanations produced from a fuzzy constraint satisfaction problem (FSCP) [6] :**
  - A set of **variables** X
  - A set of **domains** D
  - A set of **fuzzy constraints** C

[6] Régis Pierrard, Jean-Philippe Poli, and Céline Hudelot. 2019. A new approach for explainable multiple organ annotation with few data. In *Proceedings of the Workshop on Explainable Artificial Intelligence (XAI) 2019 co-located with the 28th International Joint Conference on Artificial Intelligence*, XAI@IJCAI 2019, pages 107–113. IJCAI.

EXAMPLE OF REPRESENTATION

AUTOMATIC IMAGE ANNOTATION EXPLANATION

- **Explanations produced from a fuzzy constraint satisfaction problem (FSCP) [6] :**
  - A set of **variables** X : organs to label
  - A set of **domains** D : region of the image
  - A set of **fuzzy constraints** C : e.g. right lung above liver

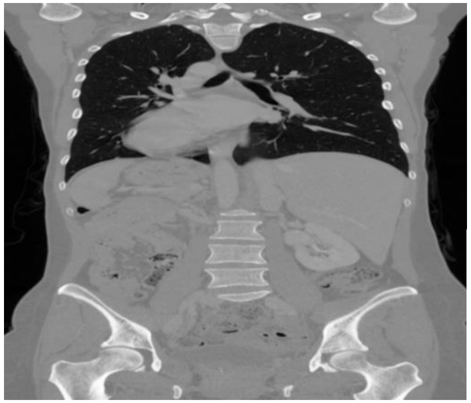
Instantiated AI Model

[6] Régis Pierrard, Jean-Philippe Poli, and Céline Hudelot. 2019. A new approach for explainable multiple organ annotation with few data. In *Proceedings of the Workshop on Explainable Artificial Intelligence (XAI) 2019 co-located with the 28th International Joint Conference on Artificial Intelligence*, XAI@IJCAI 2019, pages 107–113. IJCAI.

EXAMPLE OF REPRESENTATION

AUTOMATIC IMAGE ANNOTATION EXPLANATION

- **Explanations produced from a fuzzy constraint satisfaction problem (FSCP) [6] :**
  - A set of **variables** X : organs to label
  - A set of **domains** D : region of the image
  - A set of **fuzzy constraints** C : e.g. right lung above liver



“Organ 1 is very likely to be annotated as the left lung because it is to the left of the right lung (organ 2), it is symmetrical to the right lung and it is above the spleen (3).”

Instantiated AI Model

Textual Explanation

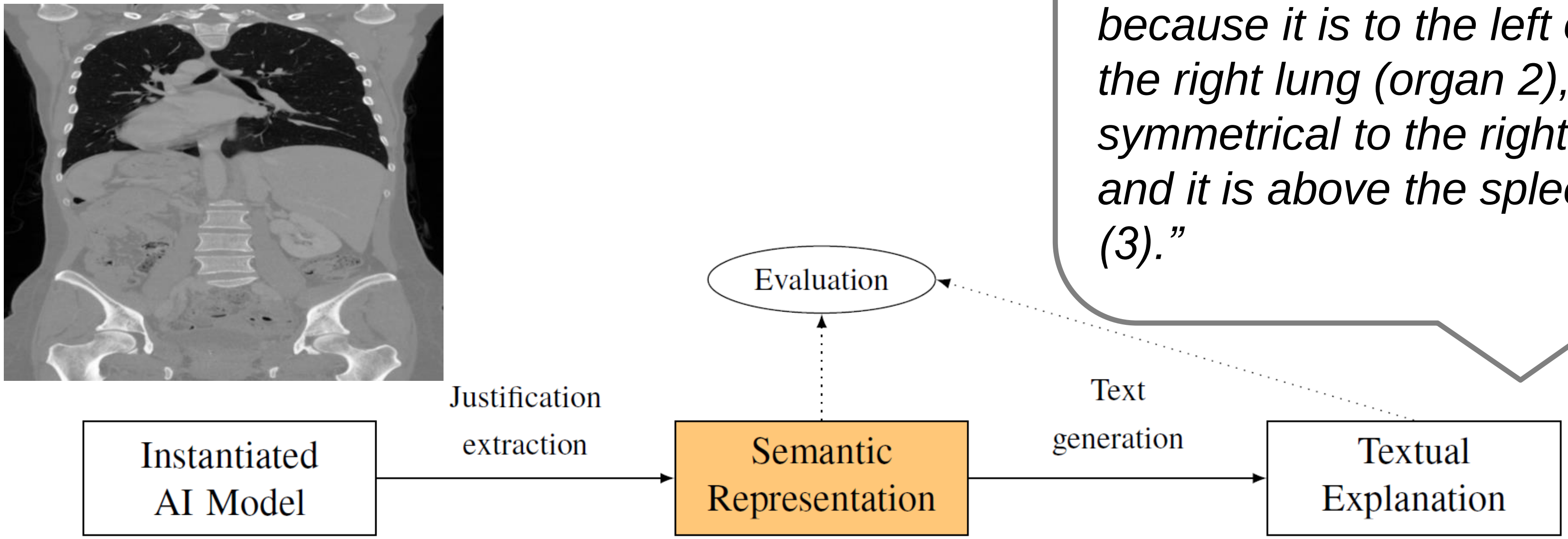
[6] Régis Pierrard, Jean-Philippe Poli, and Céline Hudelot. 2019. A new approach for explainable multiple organ annotation with few data. In *Proceedings of the Workshop on Explainable Artificial Intelligence (XAI) 2019 co-located with the 28th International Joint Conference on Artificial Intelligence*, XAI@IJCAI 2019, pages 107–113. IJCAI.

EXAMPLE OF REPRESENTATION

AUTOMATIC IMAGE ANNOTATION EXPLANATION

- **Explanations produced from a fuzzy constraint satisfaction problem (FSCP) [6] :**
  - A set of **variables** X : organs to label
  - A set of **domains** D : region of the image
  - A set of **fuzzy constraints** C : e.g. right lung above liver

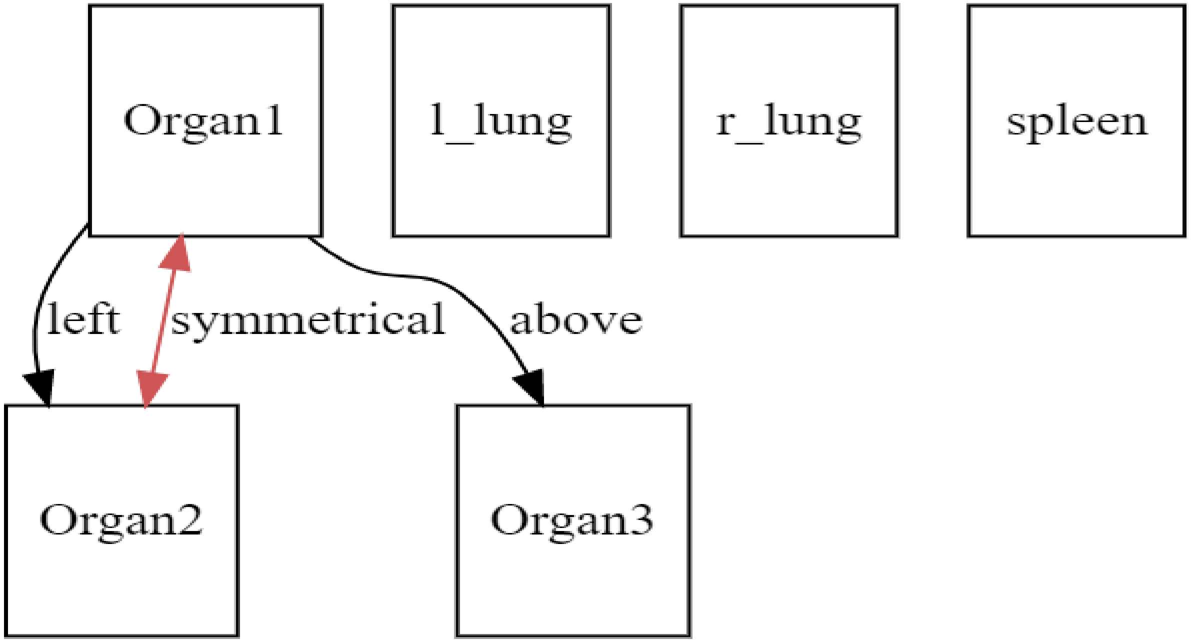
“Organ 1 is very likely to be annotated as the left lung because it is to the left of the right lung (organ 2), it is symmetrical to the right lung and it is above the spleen (3).”



[6] Régis Pierrard, Jean-Philippe Poli, and Céline Hudelot. 2019. A new approach for explainable multiple organ annotation with few data. In *Proceedings of the Workshop on Explainable Artificial Intelligence (XAI) 2019 co-located with the 28th International Joint Conference on Artificial Intelligence*, XAI@IJCAI 2019, pages 107–113. IJCAI.

EXAMPLE OF REPRESENTATION

AUTOMATIC IMAGE ANNOTATION EXPLANATION

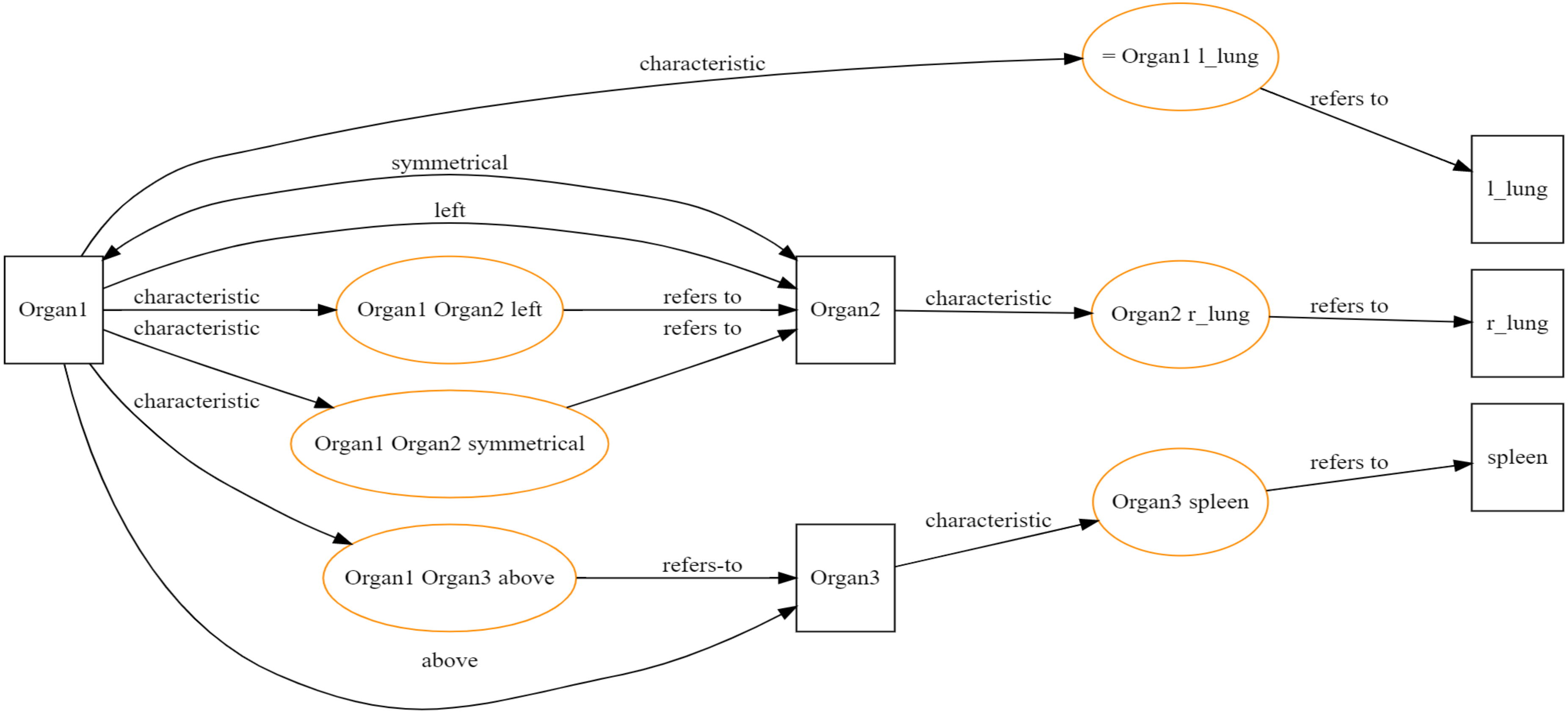


- **Explanation :**

“Organ 1 is very likely to be annotated as the left lung because it is to the left of the right lung (organ 2), it is symmetrical to the right lung and it is above the spleen (3).”

EXAMPLE OF REPRESENTATION

AUTOMATIC IMAGE ANNOTATION EXPLANATION

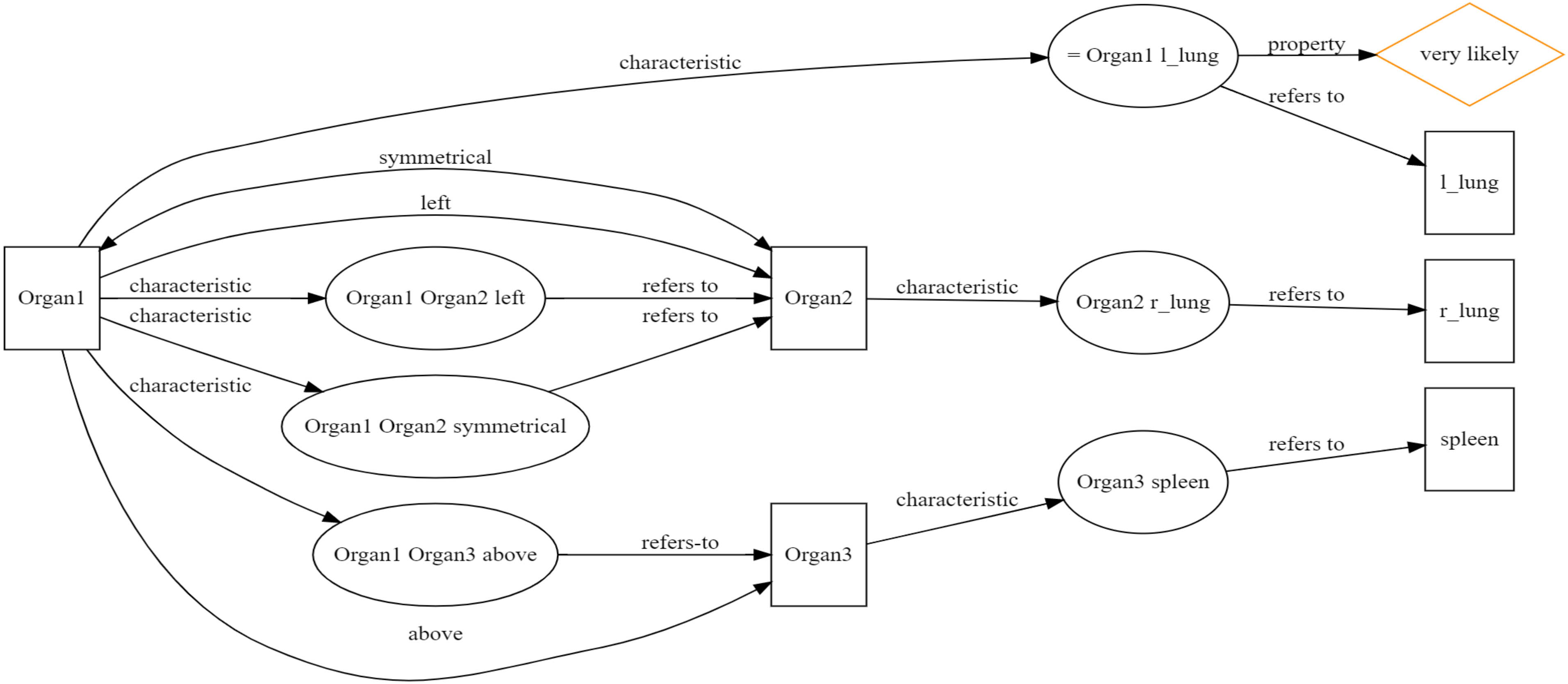


- **Explanation :**

“Organ 1 is very likely to be annotated as the left lung because it is to the left of the right lung (organ 2), it is symmetrical to the right lung and it is above the spleen (3).”

EXAMPLE OF REPRESENTATION

AUTOMATIC IMAGE ANNOTATION EXPLANATION

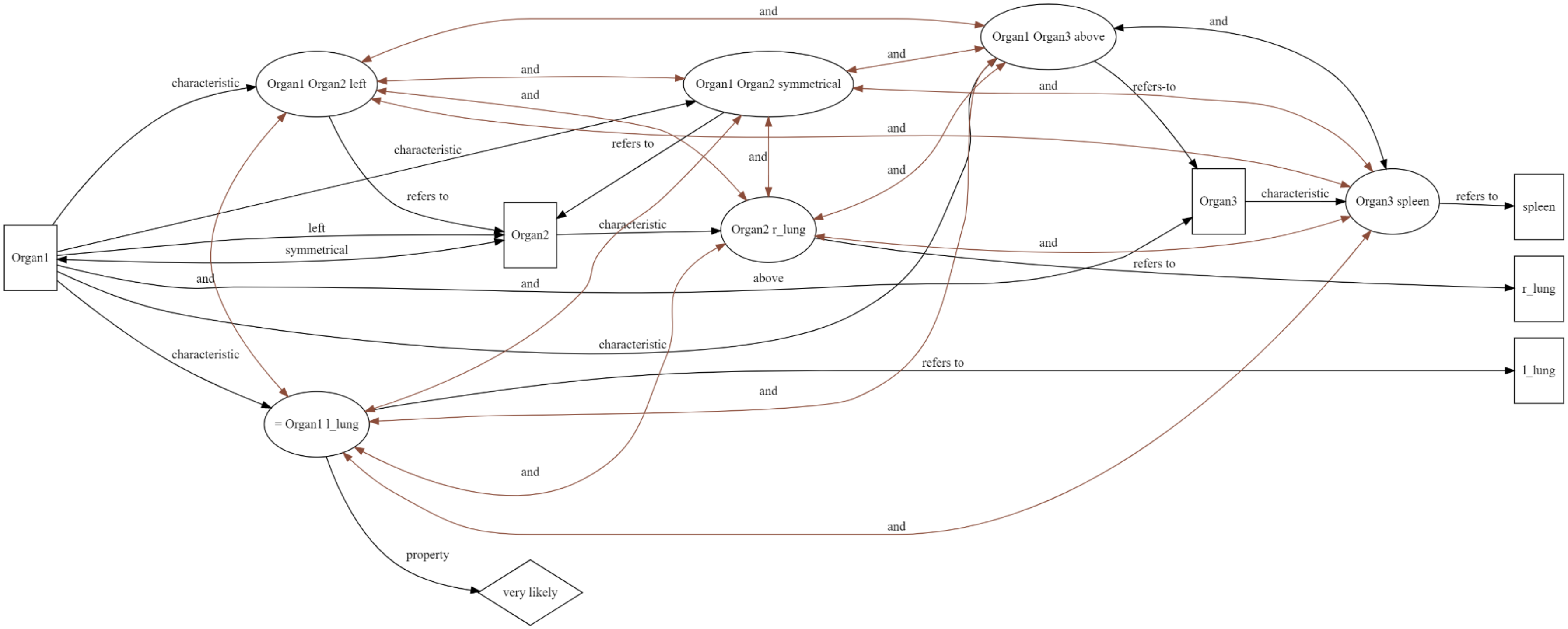


- **Explanation :**

“Organ 1 is very likely to be annotated as the left lung because it is to the left of the right lung (organ 2), it is symmetrical to the right lung and it is above the spleen (3).”

EXAMPLE OF REPRESENTATION

AUTOMATIC IMAGE ANNOTATION EXPLANATION

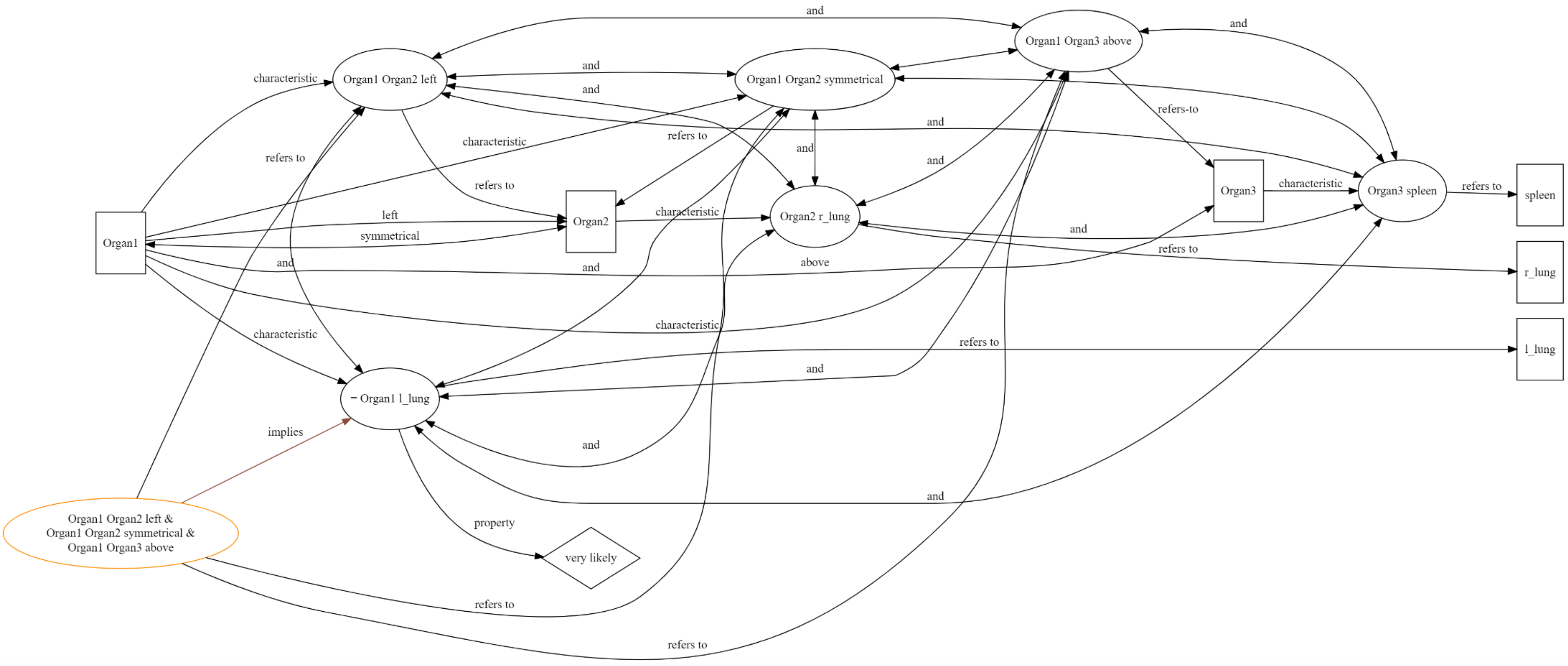


- **Explanation :**

“Organ 1 is very likely to be annotated as the left lung because it is to the left of the right lung (organ 2), it is symmetrical to the right lung and it is above the spleen (3).”

EXAMPLE OF REPRESENTATION

AUTOMATIC IMAGE ANNOTATION EXPLANATION

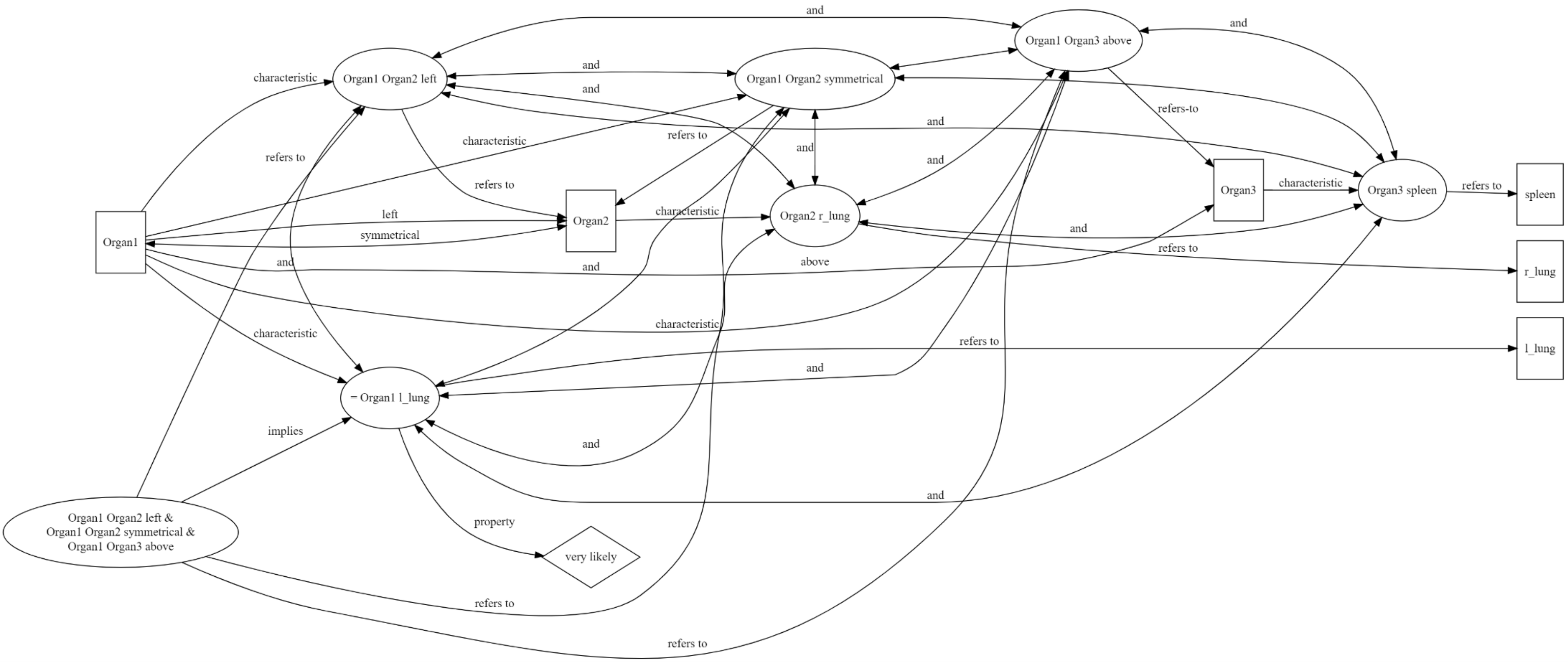


- **Explanation :**

“Organ 1 is very likely to be annotated as the left lung because it is to the left of the right lung (organ 2), it is symmetrical to the right lung and it is above the spleen (3).”

EXAMPLE OF REPRESENTATION

AUTOMATIC IMAGE ANNOTATION EXPLANATION



- **Other textual explanation :**

“Organ 1 is to the left of the right lung (organ 2), symmetrical to the right lung and it is above the spleen (3) so it must be very likely the left lung.”

CONCLUSION AND PERSPECTIVES

- **A semantic representation of explanation could unify XAI research works**

- **Conceptual graph structures are expressive and will be a source of our further developments**

REFERENCES

[1] Zwaan, R.A. and Radvansky, G.A., 1998. Situation models in language comprehension and memory. *Psychological bulletin*, *123*(2), p.162.

[2] Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*. vol. 267,pp.1–38, 2019.

[3] Banarescu, L., Bonial, C., Cai, S., Georgescu, M., Griffitt, K., Hermjakob, U., Knight, K., Koehn, P., Palmer, M. and Schneider, N., 2013, August. Abstract meaning representation for sembanking. In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse* (pp. 178-186).

[4] Mann, W.C. and Thompson, S.A., 1988. Rhetorical structure theory: Toward a functional theory of text organization. *Text-interdisciplinary Journal for the Study of Discourse*, *8*(3), pp.243-281.

[5] Arthur C Graesser, Peter Wiemer-Hastings, and Katja Wiemer-Hastings. 2001. Constructing inferences and relations during text comprehension. *Text representation: Linguistic and psycholinguistic aspects*, 8:249–271

[6] Régis Pierrard, Jean-Philippe Poli, and Céline Hudelot. 2019. A new approach for explainable multiple organ annotation with few data. In *Proceedings of the Workshop on Explainable Artificial Intelligence (XAI) 2019 co-located with the 28th International Joint Conference on Artificial Intelligence*, XAI@IJCAI 2019, pages 107–113. IJCAI.

Thank you !

ismail.baaj@cea.fr

jean-philippe.poli@cea.fr

wassila.ouerdane@centralesupelec.fr

Commissariat à l’énergie atomique et aux énergies alternatives
Institut List | CEA Saclay | Digiteo – Bât. 565 – PC192
91191 Gif-sur-Yvette Cedex – FRANCE
www-list.cea.fr

Établissement public à caractère industriel et commercial | RCS Paris B 775 685 019

Département Métrologie, Instrumentation
et Information