A Layered Approach to Encoding Semantic Representation

Siyana Pavlova Maxime Amblard Bruno Guillaume
Université de Lorraine, CNRS, Inria, LORIA, F-54000 Nancy, France
{firstname.lastname}@loria.fr

RÉSUMÉ

Une approche stratifiée pour l'encodage de la représentation sémantique

Nous présentons une première version de notre approche de la représentation sémantique. Bien qu'il existe de nombreux formalismes sémantique, parvenir à rassembler plusieurs informations sémantiques tout en restant lisible reste une tâche difficile. Nous proposons de nous appuyer sur la structure prédicat-argument (PA) utilisée dons les AMR et d'étendre la représentation avec de nouveux type de sommets et d'arrêtes. Nous introduisons la notion de "caractéristiques" qui représente des phénomènes sémantiques nouvellement encodés sous forme de couches. Des arrêtes font le lien avec les caractéristiques sémantiques et la structure PA, ou dans le cas de phénomènes sémantiques en interaction avec d'autres caractéristiques. Notre approche permet de conserver la traditionnelle structure PA et de se focaliser sur des phénomènes particuliers et leurs interactions. Un avantage explicite est aussi de permettre de rendre compte de phénomènes complexes comme la portée (négation, quantification, etc.).

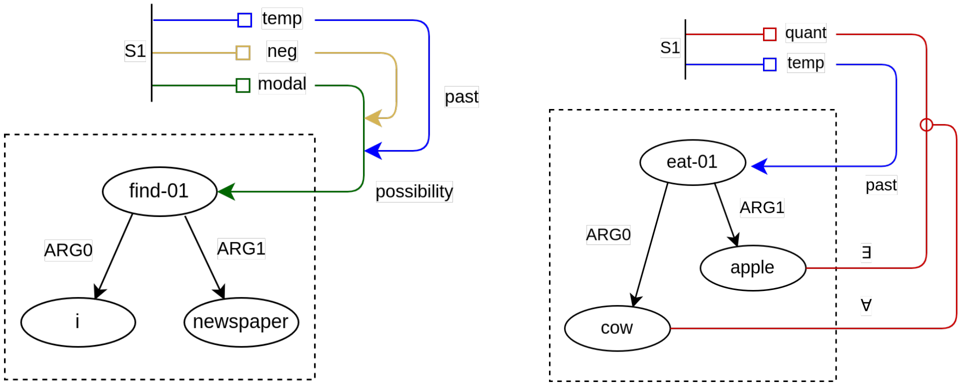
ABSTRACT

In this article, we present a first version of a layered approach we take to semantic representation. Our representation derives its core predicate-argument (PA) structure from Abstract Meaning Representation (AMR), but is then extended with a number of types of nodes and edges. We introduce "features" - nodes for each semantic phenomenon we wish to encode. Features represent various layers of our representation. Edges are then attached to each feature node, that can link to nodes from the PA-structure of the representation or, in the case of interacting semantic phenomena, to edges from other layers. An advantage of our approach is the possibility to exclude layers from the representation easily, while still being able to represent phenomena such as scope that are difficult for most graph-oriented formalisms.

MOTS-CLÉS : formalismes de représentation sémantique.

KEYWORDS: semantic representation formalisms, layered semantic representation.

Introduction Current semantic representation formalisms can be split into two broad categories - those stemming from a logic-based perspective (Kamp & Reyle, 1993; Montague, 1970), and those that are graph-based (Banarescu *et al.*, 2013; Abend & Rappoport, 2013;



(a) “I could not find the newspaper.”

(b) “Every cow ate an apple.”

Figure 1: Representation for two sentences

White *et al.*, 2016; Van Gysel *et al.*, 2021). While powerful in terms of encoding, logic-based representations are not easy to read without prior training in Logic. Graph-based ones, on the other hand, are easier to read and annotate, but often lack when it comes to expressing scope or being compositional.

We propose a way to extend the simple graph representation by adding explicit layers to the structure. Our goal is to provide users (and annotators) with a formalism that can encode various phenomena (Pavlova *et al.*, 2023b), but at the same time keep the basic representation simple. Each layer is designed to encode a different semantic phenomenon, allowing a straightforward way to “switch off” the layer. This avoids cluttering of the representation when certain phenomena are not the focus of a given task.

1 Proposal

We present here our proposal to extend graph structures in order to encode various semantic phenomena in specifically dedicated layers. For this purpose, we add a number of node types and relation types. We will explore these with the help of the two examples shown in Figure 1: for the sentence “I could not find the newspaper” (Figure 1a) with negation and modality, and “Every cow ate an apple” (Figure 1b), with the usual scope ambiguity.

The predicate-argument structure of each sentence is represented as a standard graph, with nodes (from a set of nodes **V**) and edges (from a set of edges **E**). The nodes are predicates and concepts. The edges are argument roles. The predicates and the argument roles have been taken from PropBank (Palmer *et al.*, 2005). This is the standard structure we see in Abstract Meaning Representation (AMR) (Banarescu *et al.*, 2013) and AMR-derived formalisms.

However, we restrict the usual AMR notation to the PA structure of the main verb of the sentence. By doing this we assume to follow a neo-davidsonian representation of semantics. In this line, we extend this by first adding a node **S** to represent the sentence itself (*S1* in the two figures).

Then, a set of features **F** is added, each representing a semantic phenomenon that we wish to encode in the structure. Each feature is linked to the sentence node. In Figure 1a, we represent temporality, modality and negation. In Figure 1b, in addition to temporality, we also encode quantification.

Next, we introduce a set of edges, **E'**, which link a feature to a node from **V**, as we see with `modal` and `find-01` in Figure 1a, or with `temp` and `eat-01` in Figure 1b.

Figure 1a demonstrates that there are cases where it is not enough to link features only to nodes from **V**. In that sentence, it is not the *finding* event that is expressed in the past, but the speaker's *ability* to do so. Thus, we introduce another set of edges, which link a feature to an edge from **E'**. This is also useful for negation as we can see in the same sentence.

Finally, as demonstrated in Figure 1b, when more than one quantifier is present in a sentence, we often need to specify which one takes scope over which other ones. To be able to do this, in the quantification layer we introduce another type of edges, which can link an edge from **E'** to a node from **V**.

One of the advantages of the thus described structure is that while various phenomena can be encoded, it is easy to choose to not represent a layer if the phenomenon it represents is not relevant for a given task. In Figure 1a, even though it is possible, we have chosen to not represent quantification.

2 Discussion and Future Work

In this work, we do not provide all the definitions of the framework, but we have explored its ability to encode a lot of different phenomena and how they could interact in the representation, namely: temporality, aspect, modality, negation, quantifier scope, definiteness, plurality and generics. Thus, we believe our representation has the potential to encode the phenomena encoded by logic-based formalisms. However, with our proposal on how to visualise it, it is also more easy to read and annotate.

The next steps for this project are to give the formal definition of our proposal, expand the list of phenomena we encode, annotate a first exploration corpus, and provide annotation guidelines. This work will be followed by transformation experiments between the thus proposed formalism and other semantic representation formalisms in the line of our previous work (Pavlova *et al.*, 2023a). Furthermore, to allow for this representation to be used for semantic parsing and tasks further up the NLP pipeline, we want to develop an equivalent textual representation in the spirit of PENMAN notation (Matthiessen & Bateman, 1991).

References

ABEND O. & RAPPOPORT A. (2013). Universal Conceptual Cognitive Annotation (UCCA). In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, p. 228–238, Sofia, Bulgaria: Association for Computational Linguistics.

BANARESCU L., BONIAL C., CAI S., GEORGESCU M., GRIFFITT K., HERMJAKOB U., KNIGHT K., KOEHN P., PALMER M. & SCHNEIDER N. (2013). Abstract Meaning Representation for sembanking. In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, p. 178–186, Sofia, Bulgaria: Association for Computational Linguistics.

KAMP H. & REYLE U. (1993). *From discourse to logic: Introduction to modeltheoretic semantics of natural language, formal logic and discourse representation theory*. Dordrecht. Kluwer.

MATTHIESSEN C. M. & BATEMAN J. A. (1991). Text generation and systemic-functional linguistics: experiences from english and japanese.

MONTAGUE R. (1970). English as a formal language.

PALMER M., GILDEA D. & KINGSBURY P. (2005). The Proposition Bank: An annotated corpus of semantic roles. *Computational Linguistics*, **31**(1), 71–106. DOI : 10.1162/0891201053630264.

PAVLOVA S., AMBLARD M. & GUILLAUME B. (2023a). Bridging Semantic Frameworks: mapping DRS onto AMR. In *The 15th International Conference on Computational Semantics (IWCS 2023)*, Nancy, France. HAL : hal-04129563.

PAVLOVA S., AMBLARD M. & GUILLAUME B. (2023b). Structural and Global Features for Comparing Semantic Representation Formalisms. In *The 4th International Workshop on Designing Meaning Representation*, Nancy, France. HAL : hal-04129557.

VAN GYSEL J. E., VIGUS M., CHUN J., LAI K., MOELLER S., YAO J., O'GORMAN T., COWELL A., CROFT W., HUANG C.-R. *et al.* (2021). Designing a uniform meaning representation for natural language processing. *KI-Künstliche Intelligenz*, **35**(3-4), 343–360.

WHITE A. S., REISINGER D., SAKAGUCHI K., VIEIRA T., ZHANG S., RUDINGER R., RAWLINS K. & VAN DURME B. (2016). Universal decompositional semantics on universal dependencies. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, p. 1713–1723.