

# Named Graphs for Semantic Representations

**Richard Crouch**

A9.com

dick.crouch@gmail.com

**Aikaterini-Lida Kalouli**

University of Konstanz

aikaterini-lida.kalouli@uni-konstanz.de

## Abstract

A position paper arguing that purely graphical representations for natural language semantics lack a fundamental degree of expressiveness, and cannot deal with even basic Boolean operations like negation or disjunction, let alone intensional phenomena. Moving from graphs to named graphs leads to representations that stand some chance of having sufficient expressive power. Named  $\mathcal{FL}_0$  graphs are of particular interest.

## 1 Introduction

Graphs are popular for both (semantic web) knowledge representation (Screiber and Raimond, 2014; Dong et al., 2014; Rospocher et al., 2016) and natural language semantics (Banarescu et al., 2013; Perera et al., 2018; Wities et al., 2017). The casual observer might assume there is substantial overlap between the two activities, but there is less than meets the eye. This paper attempts to make three points:

1. Knowledge graphs, which are designed to represent facts about the world rather than human knowledge, are not well set up to represent negation, disjunction, and conditional or hypothetical contexts. Arguably, the world contains no negative, disjunctive, or hypothetical facts; just positive facts that make them true. Natural language semantics has to deal with more partial assertions, where all that is known are the negations, disjunctions, or hypotheticals, and not the underlying facts that make them true.

2. Named graphs (Carroll et al., 2005) are an extension of RDF graphs (Screiber and Raimond, 2014), primarily introduced to record provenance information. They are worthy of further study, since they promise a way of bridging between the relentlessly positive world of knowledge representation and the more partial, hypothetical world of natural language. In RDF-OWL graphs (Hitzler et al., 2012), the subject-predicate-object triples

forming the nodes and arcs of the graph correspond to atomic propositions. Beyond conjunction, no direct relations between these propositions can be expressed. Named graphs allow sub-graphs (i.e. collections of atomic propositions) to be placed in relationships with other sub-graphs, and thus allow for negative, disjunctive and hypothetical relations between complex propositions.

3. Named graphs illustrate a certain way of factoring out complexity, in this case between predicate-argument structure and Boolean / modal structure. As a semantic representation, the predicate-argument structure is correct, but not complete. Adding a named, Boolean layer requires no adjustment to the syntax or semantics of the predicate-argument structure; it just embeds it in a broader environment. This often is not the case; e.g. in moving from unquantified predicate logic to first-order quantified logic, to first-order modal logic, to higher-order intensional logic.

After reviewing RDF graphs and named graphs, we discuss how they could be applied to a (somewhat incestuous) family of layered, graphical, semantic representations (Boston et al., forthcoming; Shen et al., 2018; Bobrow et al., 2007) (see our companion paper (Kalouli and Crouch, 2018) for an introduction to these representations). This offers the prospect of a formal semantics that takes graphs to be first class semantic objects, which differs from approaches like AMR (Banarescu et al., 2013), where the graphs are descriptions of underlying semantic objects.

## 2 Graphs and Named Graphs

A graph is a collection of binary relationships between entities. Since any  $n$ -ary relationship can be decomposed into  $n + 1$  binary relationships through the introduction of an extra entity that serves as a “pivot” (this is the basis of neo-Davidsonian event semantics (Parsons, 1990)), all  $n$ -ary relationships can be represented in graphical

form as a collection of entity-relation triples.

## 2.1 RDF

This graphical approach to  $n$ -ary relationships has seen perhaps its fullest use in the Resource Description Framework (RDF) (Screiber and Raimond, 2014), where subject-relation-object triples can be stored to treat complex ontologies as graphs. But since the triples form a conjunctive set, RDF has to go through some contortions to emulate negation and disjunction.

Unadorned, RDF is lax about what kinds of entity can occur in triples, and individuals, relations, and classes can intermingle freely. One can state facts about how classes relate to other classes (e.g. one is a subclass of the other), how relations relate to other relations, and how individual relate to relations and classes. Successive restrictions, such as RDFS (Brickley and Guha, 2014) and OWL (Hitzler et al., 2012) tighten up on this freedom of expression, for the resulting gain in inferential tractability.

OWL provides a number of class construction operations that mimic Booleans at a class level: complement (negation), intersection (conjunction) and union (disjunction). One could therefore assert that Rosie is not a cat by saying that she is an instance of the cat-complement class, and one could assert that Rosie is a cat or a dog by asserting that she is an instance of the class formed by taking the union of cats and dogs. Additionally, OWL and RDFS allow negative properties as a way of stating that a particular relation does not hold between two entities (i.e. a form of atomic negation).

The semantic web is geared toward capturing positive facts about what is known. Two positive facts can establish a negative, e.g. that cats and dogs are disjoint classes and that Rosie is a dog establishes that Rosie is not a cat. But the need to *assert* a negative rarely arises: better to wait until the corresponding incompatible positive is known, or as a last resort make up a positive fact that is incompatible with negative (e.g. that Rosie is a non-cat). Natural language, by contrast, is full of negative, disjunctive, and hypothetical assertions for which the justifying positive facts are not known. And these Boolean and modal assertions express relationships between propositions (i.e. collections of triples), and not between classes.

Moreover, Gardenförs (2014) makes the case for restricting semantics to natural concepts within

a conceptual space. A conceptual space consists of a set of quality dimensions (c.f. dimensions in word vectors). A point in the space is a particular vector along these dimensions. A natural concept is a region (collection of points in the space) that is connected and convex. This essentially means that the shortest path from one sub-region of a natural concept to another does not pass outside of the region defined by the concept: natural concepts are regions that are not gerrymandered. OWL unions of classes can arbitrarily combine disconnected regions, whereas complements can tear holes in the middle of regions: they can produce gerrymandering that would make the most partisan blush.

## 2.2 Named Graphs

Named graphs were introduced by Carroll et al. (2005) as a small extension on top of RDF, primarily with the goal of recording provenance metadata for different parts of a complex graph, such as source, access restrictions, or ontology versions. However, applications to stating propositional attitudes and capturing logical relationships between graphs were also mentioned in passing. A named graph simply associates an extra identifier with a set of triples. For example, a propositional attitude like *Fred believes John does not like Mary* could be represented as follows<sup>1</sup>:

```
:g1 { :john :like :mary }
:g2 :not :g1
:fred :believe :g2
```

where `:g1` is the name given to the graph expressing the proposition that *John likes Mary*, and `:g2` to the graph expressing its negation. Disjunction likewise can be expressed as a relationship between named graphs:

```
:g1 { :john :like :mary }
:g2 { :fred :like :mary }
:g0 :or :g1
:g0 :or :g2
```

where the graph `:g0` expresses the disjunction of `:g1` and `:g2`.

The graph semantics for named graphs is a simple extension of the basic semantics (Carroll et al., 2005). The meaning of a named graph is the meaning of the graph, and sub-graph relations between named graphs must reflect the underlying relations between the graphs that are named. But significantly, named graphs are not automatically asserted — there is no presumption that the triples

<sup>1</sup>Using the TriG format (Bizer and Cyganiak, 2014), with the prefix definition for `:` omitted for brevity.

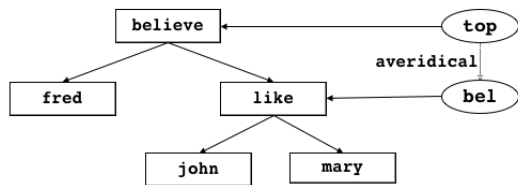


Figure 1: GKR for *Fred believes John likes Mary*

occurring in a named graph are true. This is somewhat inconvenient if your main goal is to assert positive, true facts. But this looks ideal for dealing with negation, disjunction and hypotheticals in natural language. In particular, the named graph above asserts neither that John likes Mary nor that John doesn't like Mary.

Reification in RDF was an earlier approach to dealing with provenance meta-data (Screiber and Raimond, 2014). This turns every triple into four triples that describe it, so that `:john :like :mary` becomes `:t :type :statement; :t :subj :john; :t :pred :like; :t :obj :mary`. The reified graph is graphical description of the original graph. Naming preserves the underlying graph in a way that reification does not.

### 3 Layered Graphs and the Graphical Knowledge Representation

A recent proposal for semantic representation has made use of so-called “layered-graphs” (GKR, (Kalouli and Crouch, 2018), see also (Boston et al., forthcoming; Shen et al., 2018)), with the claim that this gives a good way of handling Boolean, hypothetical, and modal relations<sup>2</sup>. The proposal is based on earlier work on an Abstract Knowledge Representtions (AKR, (Bobrow et al., 2007)), which imposes a separation between conceptual / predicate-argument structure and contextual structure. The GKR representation (simplified) for *Fred believes John likes Mary* is shown in Figure 1. This comprises two sub-graphs: a concept/predicate-argument graph on the left, and a context graph on the right. The concept graph can be read conjunctively as stating the following, but where variables range over (sub)concepts and not over individuals:

$$\exists b, l, f, j, m.$$

<sup>2</sup>This section does not attempt to motivate GKR in terms of coverage of phenomena or support of natural language inference: see the cited papers for this.

```
believe(b) & like(l) & fred(f)
& sub(b,f) & comp(b,l)
& mary(m) & john(j)
& subj(l,j) & obj(l,m)
```

Thus `b` denotes a sub-concept of `believe`, that is further constrained to have as a subject role some sub-concept of `fred`, and as a complement some sub-concept of `like`. The concept graph makes no assertions about whether any of these concepts have individuals instantiating them: it asserts neither that Fred has a belief, nor that John likes Mary. It is a level at which semantic similarity can be assessed, but not one at which — on its own — logical entailments can be judged. The concept graph is a correct characterization of the sentence, but an incomplete one.

Entailment requires existential commitments that are introduced by the context graph shown on the right of Figure 1. There are two contexts. The top level, “true”, context `top` states the commitments of the sentence’s speaker. The arc connecting it to the `believe` node means that the speaker is asserting that there is an instance of the `believe` concept. The second context, `bel` is lexically induced by the word “believes”. The arc from `bel` to the `like` node means that in this context there is asserted to be an instance of the `like` context. However, `bel` is marked as being *averidical* with respect to `top`. This means that we cannot lift the existential commitments of `bel` up into `top`. Hence Figure 1 does not entail that John likes Mary (nor that he doesn’t).

Other words introduce different context relations. For example *know* creates a *veridical* lower context, which means that the lower existential commitments can be lifted up. Whereas *negation* creates an *anti-veridical* lower context, which specifically says that the concept that is instantiated in the lower context is uninstantiated in the upper one. Following the work of (Nairn et al., 2006), these instantiation raising rules allow complex intensional inference to be drawn (see (Boston et al., forthcoming) for a fuller description).

### 4 Named $\mathcal{FL}_0$ Graphs

The semantics for GKR has yet to be clearly laid out. Our claim is that the layered graphs are better seen in terms of named graphs. First, that the context graph simply expresses relationships between named concept graphs, so that contexts

$$C, D \Rightarrow A \mid C \sqcap D \mid \forall R.C$$

where  $A \in \text{Atomic concept}$   
 $R \in \text{Atomic role}$   
 $C, D \in \text{concept}$

Figure 2: Concept Construction in  $\mathcal{FL}_0$

are nothing more than concept (sub-)graphs. Second, that the concept graph corresponds to the  $\mathcal{FL}_0$  description logic (Baader and Nutt, 2003), for which subsumption is decidable in polynomial time.  $\mathcal{FL}_0$  is a simple logic that is generally regarded as too inexpressive to deal with interesting language-related phenomena. But in combination with graph naming it becomes much more expressive.

The  $\mathcal{FL}_0$  description logic allows concepts to be constructed as shown in Figure 2. Given a stock of atomic concepts, complex concepts can be formed by (i) intersection, e.g.  $(Adult \sqcap Male \sqcap Person) \equiv Man$ ; and ii) slot/role restriction, e.g.  $Bite \sqcap \forall subj.Dog \sqcap \forall obj.Man$  (the class of bitings by dogs of men). The concept graphs of GKR correspond to the application of  $\mathcal{FL}_0$  operations to atomic lexical concepts. The concept of the `like` node in Figure 1 is thus  $like \sqcap \forall subj.j \sqcap \forall obj.m^3$ .

In order to keep the concept graphs of GKR within  $\mathcal{FL}_0$ , it is important that context nodes are not allowed to participate in role restrictions. This rules out the kind of free intermingling of graph nodes and other nodes that was presented in Section 2.2. The GKR treatment of *Fred believes John likes Mary* is shown in Figure 1. Expressed as a named graph, this corresponds to:

```
top: {b ⊑ Believe
      f ⊑ Fred
      b ⊑ ∀subj.f
      b ⊑ ∀comp.l }
bel: {l ⊑ Like
      j ⊑ John
      m ⊑ Mary
      l ⊑ ∀subj.j
      l ⊑ ∀obj.m }
:top averidical :bel
```

This named-graph formulation of GKR inherits a standard graph semantics, as described by Carroll et al. (2005). The graph semantics is complementary to the kind of truth-conditional semantics set out for AKR (and by analogy, GKR) by Bobrow

<sup>3</sup>GKR extends its concept graph with a property graph that captures the effect of morpho-syntactic features like cardinality. This corresponds to extending the logic to  $\mathcal{FLN}_0$  by introducing cardinality restrictions.

et al. (2005). More work, however, is needed to explore the connections between the graph and truth-conditional semantics.

## 5 Abstract Meaning Representation (AMR)

AMR is best seen as a graphical notation for describing logical forms, which is the view taken by Bos (2016) and Stabler (2017) in their augmentations of AMR to increase its expressive power. This can be seen by considering the AMR for *Fred believes John likes Mary*:

```
(b / believe
  :arg0 (f / Fred)
  :arg1 (l / like
        :arg0 (j / John)
        :arg1 (m / Mary)))
```

Expressed as a set of triples, this becomes

```
b instance believe;  l instance like;
f instance Fred;    j instance John;
b :arg0 f;          m instance Mary;
b :arg1 l;          l :arg0 j;
                   l :arg1 m;
```

Since a graph is a conjunction of triples, and because  $A \wedge B \models A$ , all the triples on the left can be validly eliminated to leave those on the right, which correspond to the graph for *John likes Mary*. The inference from *Fred believes John likes Mary* to *John likes Mary* is clearly not semantically valid. Consequently the AMR triples cannot be interpreted as stating semantic-web style facts; rather they state sub-formulas of a logical form.

There is nothing wrong in having a more habitable, graphical notation for logical formulas, especially if large amounts of annotation are to be done. But this is different from a goal of having graphs as first class semantic objects.

## 6 Concluding Observations

This paper attempts to make the case for named graphs as an interesting tool for natural language semantics. The first task in exploring this further would be to provide a truth-conditional, graph-based semantics for GKR. A positive outcome would enable closer links between semantic and knowledge graphs.

By naming graphs, it appears that an inexpressive, conjunctive concept logic,  $\mathcal{FL}_0$ , can be employed to handle a wide variety of more complex phenomena including Booleans and hypotheticals. However, one should not assume that the inferential tractability of  $\mathcal{FL}_0$  carries across to a system that combines it with named graphs.

We conjecture that the restriction of concept formation to  $\mathcal{FL}_0$  will satisfy (Gardenförs, 2014) requirements on the connectedness and convexity of concepts. Additionally, the restricted operations may be better for the operations inherent in dealing with vector spaces used in distributed semantic representation; it is currently unclear what corresponds to negation in vector spaces, though see (Bowman et al., 2015). The strategy of having a correct but incomplete conceptual structure may make it easier to reconcile logical and distributional accounts of semantics if distributional semantics is relieved of the burden of having to account for Boolean structure.

Naming a graph essentially boxes it off, to be evaluated or asserted within a different context. GKR focuses on the analogue between these contexts and switching assignments to possible worlds in standard Kripke semantics for modal logics. With regard to distributional quantification it observes that assignments to variables in standard first-order logic plays a similar role, and suggests using this to account for quantifier scope via contexts. This does not exhaust the space of evaluative contexts. Named graphs were primarily motivated by the desire to record (provenance) meta-data about triples. They provide an ideal means of associating meta-data with semantic relationships, such as the confidence that a particular role restriction is correct. This can be extended to record inter-dependencies between collections of ambiguous relationships, using the packing mechanism of (Maxwell and Kaplan, 1993): choices between alternate interpretations also set up different evaluation contexts.

The embedding of boxes in Discourse Representation Theory (Kamp and Reyle, 1993) is strongly reminiscent of embedding sub-graphs. We speculate that DRT could be given a graph-based semantics, in which discourse representation structures (DRSs) are seen as first class graphical and semantic objects. However, one difference between DRT and GKR is that GKR imposes a strict separation between concepts and contexts. This essentially means that contexts cannot be referred to in conceptual predicate-argument structures. In DRT, this would correspond to not permitting DRSs to serve as arguments of predicates.

With regard to AMR, naming some of the graphs and expressing context relations between them seems a relatively conservative extension in

terms of notation. But doing so offers the prospect of lifting AMRs out of being graphical descriptions of some other semantic object (like a logical form), and becoming much closer to RDF graphs as first-class semantic objects.

## Acknowledgements

This paper does not describe any work done by the first author in his role at A9. We are grateful to Tagyoung Chung and Valeria de Paiva for helpful comments on presentation, though they should not be held responsible for the views expressed.

## References

- Franz Baader and Werner Nutt. 2003. The description logic handbook. chapter Basic Description Logics, pages 43–95. Cambridge University Press, New York, NY, USA.
- Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. 2013. Abstract Meaning Representation for Sembanking. In *Proceedings of the Linguistic Annotation Workshop*.
- Chris Bizer and Richard Cyganiak. 2014. RDF 1.1 TriG, <https://www.w3.org/tr/trig/>.
- Danny G. Bobrow, Cleo Condoravdi, Richard Crouch, Valeria de Paiva, Lauri Karttunen, Tracy Holloway-King, Rowan Nairn, Charlotte Price, and Annie Zaenen. 2007. Precision-focused Textual Inference. In *Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing*, RTE '07, pages 16–21, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Danny G. Bobrow, Richard Crouch, Valeria de Paiva, Ron Kaplan, Lauri Karttunen, Tracy King-Holloway, and Annie Zaenen. 2005. A basic logic for textual inference. In *Proceedings of the AAAI Workshop on Inference for Textual Question Answering*.
- Johan Bos. 2016. Expressive Power of Abstract Meaning Representations. *Computational Linguistics*, 42(3):527–535.
- Marisa Boston, Richard Crouch, Erdem Özcan, and Peter Stubbley. forthcoming. Natural language inference using an ontology. In Cleo Condoravdi, editor, *Lauri Karttunen Festschrift*.
- Samuel R. Bowman, Christopher Potts, and Christopher D. Manning. 2015. Recursive neural networks can learn logical semantics. In *Proceedings of the 3rd Workshop on Continuous Vector Space Models and their Compositionality*.

- Dan Brickley and R. V. Guha. 2014. RDF Schema 1.1, <https://www.w3.org/tr/rdf-schema/>.
- Jeremy Carroll, Christian Bizer, Pat Hayes, and Patrick Stickler. 2005. Named Graphs. *Web Semantics: Science, Services and Agents on the World Wide Web*, 3(4).
- Xin Luna Dong, Evgeniy Gabrilovich, Geremy Heitz, Wilko Horn, Ni Lao, Kevin Murphy, Thomas Strohmann, Shaohua Sun, and Wei Zhang. 2014. Knowledge Vault: A Web-Scale Approach to Probabilistic Knowledge Fusion. In *The 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '14, New York, NY, USA - August 24 - 27, 2014*, pages 601–610.
- Peter Gardenförs. 2014. *The Geometry of Meaning: Semantics Based on Conceptual Spaces*. MIT Press.
- Pascal Hitzler, Markus Krötzsch, Bijan Parsia, Peter F. Patel-Schneider, and Sebastian Rudolph. 2012. OWL 2 Web Ontology Language Primer, <https://www.w3.org/tr/owl2-primer>.
- Aikaterini-Lida Kalouli and Richard Crouch. 2018. GKR: the graphical knowledge representation. In *SemBEaR-18*.
- Hans Kamp and Uwe Reyle. 1993. *From Discourse to Logic. Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*. Kluwer, Dordrecht.
- John T. Maxwell and Ronald M. Kaplan. 1993. The interface between phrasal and functional constraints. *Computational Linguistics*, 19(4).
- Rowan Nairn, Cleo Condoravdi, and Lauri Karttunen. 2006. Computing relative polarity for textual inference. *Inference in Computational Semantics (ICoS-5)*, pages 20–41.
- Terence Parsons. 1990. *Events in the semantics of English*. MIT Press.
- Vittorio Perera, Tagyoung Chung, Thomas Kollar, and Emma Strubell. 2018. Multi-Task Learning for parsing the Alexa Meaning Representation Language. In *Proc AAAI*.
- Marco Rospocher, Marieke van Erp, Piek Vossen, Antske Fokkens, Itziar Aldabe, German Rigau, Aitor Soroa, Thomas Ploeger, and Tessel Bogaard. 2016. Building event-centric knowledge graphs from news. 37.
- Guus Sreiber and Yves Raimond. 2014. RDF 1.1 primer, <https://www.w3.org/tr/rdf-primer/>.
- Jiaying Shen, Henk Harkema, Richard Crouch, Ciaran O'Reilly, and Peng Yu. 2018. Layered semantic graphs for dialogue management. In *SigDial*.
- Ed Stabler. 2017. Reforming AMR. In *Formal Grammar 2017. Lecture Notes in Computer Science*, volume 10686. Springer.
- Rachel Wities, Vered Shwartz, Gabriel Stanovsky, Meni Adler, Ori Shapira, Shyam Upadhyay, Dan Roth, Eugenio Martínez-Cámara, Iryna Gurevych, and Ido Dagan. 2017. A consolidated open knowledge representation for multiple texts. In *Proceedings of the 2nd Workshop on Linking Models of Lexical, Sentential and Discourse-level Semantics*, pages 12–24.