Frame Dynamics in Knowledge Graphs

Aldo GANGEMI [a]
[a] *University of Bologna and ISTC-CNR, Rome, Italy*

Keywords. frame semantics, compositionality, natural language, knowledge graphs

1. Semantic Compositionality

The compositionality principle is a classical problem in semantics: "is the meaning of a structure *entirely* determined by the meaning of its constituents?". However, the very *object* of semantics is elusive: is it about symbols, things, world states, cognitive or neural states within an individual, or emerging structures in culture or communities or agents? Or all of that, e.g. as a result of interaction between agents and their environments? (cf. the heterogeneous contributions on the compositionality principle, ranging from embodied to symbolic views, collected in [6]).

Reasonably, compositionality could only be studied after fixing what are the constituents of a structure. Let's concentrate on natural language semantics: if we fix constituents to syntactic phrase-based grammars, and we assume a semantic theory for phrases, compositionality effects on phrase semantics can be evaluated on the basis of grammatical rules, as with the sentence *The cigar-shaped asteroid Oumuamua and the hypothesis of Harvard: "It can be an alien spacecraft"*. Multiple compositional phenomena appear: *cigar-shaped*, *cigar-shaped asteroid*, *hypothesis of Harvard*, *alien spacecraft*, *cigar-shaped asteroid Oumuamua and the hypothesis of Harvard*, *the hypothesis of Harvard: "It can be an alien spacecraft'*, etc.

Unfortunately, current tools for natural language processing do not give us complete accounts of compositionality even at a phrase level: noun-noun compounds have opaque semantic relations, adjectival modifiers follow unpredictable patterns, parataxis (as provided by conjunctions and punctuation) is semantically underspecified, metonymy (as in *Harvard*) or metaphor (as in *cigar-shaped*) require knowledge beyond the typical one associated with lexical constituents, etc.

2. Robust vs. Analytic approaches

Indeed, these problems have emerged very early in artificial intelligence, e.g. robust parsing [7], extensively used in speech recognition, tries to use linguistic constituents as hints to approximate a pragmatic task, e.g. the *intent* of questions, the *category* of a text, a possible *causal relation* expressed in a claim, etc. The robust approach is typical of computational approaches that can be described as "directed at optimising a cost function", in this case out of linguistic processing time and resources.

However, optimising a cost function is not necessarily, or not only, what scientific research is about. In the case of semantic compositionality, science may want to address what are the possible constituents of semantics proper, and if compositionality relations correspond to what we make sense of when interpreting e.g. a linguistic construction.

3. Knowledge Graphs

Given such premises, the contribution of knowledge graphs, and especially linked data and ontologies, is to make it explicit, and eventually test, hypotheses about what constitutes semantics in a certain context, and link it to public identities of individuals, concepts, relations, models, and things in general. Once represented into knowledge graphs, candidate semantic constructions become the object of investigation, and patterns may emerge, and can be associated with other constructions represented under different constraints. For example, multiple news describing a same event, or a series of related events, may be reconciled, and the dynamics of events, their storytelling, judgments, participants, can be investigated on a rigorous basis.

FRED [5][1] is a state-of-the-art knowledge graph extraction tool that transforms natural language texts into knowledge graphs, and links them to existing linked data, so playing the role of a *semiotic hub* that partly simulates human interpretation of a text during reading (Fig. 1).

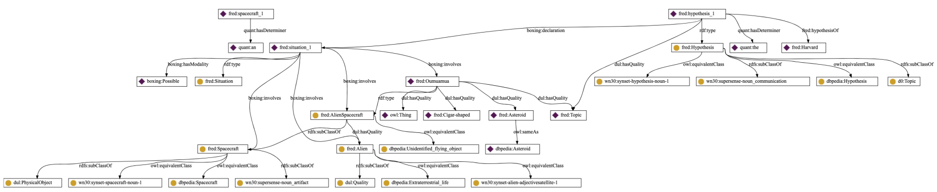

Figure 1. A diagrammatic representation of the knowledge graph extracted automatically by FRED for the example sentence.

In the knowledge graph perspective, compositionality becomes a matter of model reconciliation, evolution, and its related dynamics. However, current ontologies and data hardly address higher levels of meaning, e.g. multi-modality, grounding, attitudes, metaphor, sentiment, emotion, normative principles, storytelling, interpersonal issues, etc.

4. Frame Dynamics and Compositionality

A semantic theory that provides a clean, simple solution in order to talk about different meaning levels is *frame semantics* [2], which basically holds that situation patterns (frames) have multiple aspects (roles) filled by certain types of entities. In [4], a semantic-web-ready formalisation of frame semantics is proposed, which is able to unify semantics

[1] `http://wit.istc.cnr.it/stlab-tools/fred/demo`

across factual knowledge graphs such as DBpedia or YAGO, and linguistic resources, such as WordNet, FrameNet, or BabelNet, after having them represented as knowledge graphs. The theory treats roles as binary, and types as unary projections, of frames. This seems enough to obtain the factual-linguistic interoperability. Framester [3] is a large knowledge graph that contains a frame-based unification of linguistic and factual resources.

The compositionality arising from the frames evoked in a *a-priori* way in linguistic resources, or detected in large corpora, has been started to be investigated by using hybrid symbolic and statistical tools [1], so counting on both knowledge extraction and frame semantics. Despite all that, the gap between automating frame detection and frame dynamics in natural language, and achieving a human-like interpretation of compositionality at the right level of engagement and value for humans is still huge.

References

[1] Mehwish Alam, Diego Reforgiato Recupero, Misael Mongiovi, Aldo Gangemi, and Petar Ristoski. Event-based knowledge reconciliation using frame embeddings and frame similarity. *Knowledge-Based Systems*, 135:192–203, 2017.

[2] C. F. Baker, C. J. Fillmore, and J. B. Lowe. The Berkeley Framenet project. In *Proceedings of the 1998 International Conference on Computational Linguistics (COLING-ACL)*, 1998.

[3] Aldo Gangemi, Mehwish Alam, Luigi Asprino, Valentina Presutti, and Diego Reforgiato Recupero. Framester: a wide coverage linguistic linked data hub. In *European Knowledge Acquisition Workshop*, pages 239–254. Springer, 2016.

[4] Aldo Gangemi and Carola Catenacci. What's in a Schema? In Chu-Ren Huang, Nicoletta Calzolari, Aldo Gangemi, Alessandro Lenci, Alessandro Oltramari, and Laurent Prevot, editors, *Ontologies and the Lexicon*. Cambridge University Press, Cambridge, UK, 2007.

[5] Aldo Gangemi, Valentina Presutti, Diego Reforgiato Recupero, Andrea Giovanni Nuzzolese, Francesco Draicchio, and Misael Mongiovì. Semantic web machine reading with fred. *Semantic Web*, 8(6):873–893, 2017.

[6] James A Hampton and Yoad Winter. *Compositionality and Concepts in Linguistics and Psychology*, volume 3. Springer, 2017.

[7] Stephanie Seneff. Robust parsing for spoken language systems. In *[Proceedings] ICASSP-92: 1992 IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 1, pages 189–192. IEEE, 1992.