Can we preserve the polynomial parsability of restricted graph grammars when adding contextuality?

Henrik Björklund, Johanna Björklund, Frank Drewes, Anna Jonsson

Department of Computing Science, Umeå University, Sweden

Abstract (describing work-in-progress)

The management of semantic representations lies at the heart of many high-level natural language tasks, e.g., question answering and knowledge mining. Motivated by the interest in graph-based semantic representations, we explore two current graph-generating formalisms with the goal of combining them to obtain a formalism that can express well-structured semantic graphs while at the same time being efficiently parsable. Specifically, the semantic representation that we aim to express is the *abstract meaning representation* (AMR) by Banarescu et al. [1].

The first formalism that we consider is the *order-preserving dag grammars* (OPDGs) presented by Björklund et al. [2], a type of graph grammar based on *hyperedge replacement grammars* (HRGs [3], see, e.g., [4] for a survey) that is efficiently parsable by imposing a number of structural restrictions. Intuitively, these restrictions combine to ensure that each generated graph can be uniquely represented by a term in a particular graph algebra. Under this lens, the graph language essentially becomes a tree language of terms. The second formalism is the *contextual HRG*, developed by Drewes and Hoffmann [5]. This formalism extends the ordinary HRGs with so-called *contextual* rules, which allow for isolated nodes in their left-hand-sides. Contextual rules can access previously generated nodes, that are not attached to the replaced nonterminal hyperedge, and add structure to them.

Despite their restrictions, OPDGs seem to be able to describe central structural properties of AMR, but their limitation lies in the modelling of reantrancies. We propose to distinguish between two types of reentrancies: we speak of (i) *structural* reentrancies when they are syntactically governed, e.g. by control or coordination, and of (ii) *non-structural* reentrancies when they represent coreferences, which can in principle refer to any antecedent, essentially disregarding the structure of the graph. An example of (i) could be subject control as in "They persuaded him to talk to her", where the person who does the talking must be the same person who was persuaded, whereas "[...], but she liked them" is an example of (ii) since antecedent of "them" may be picked from anywhere in "[...]" for the semantic representation to be a valid one. Type (i) can to a large extent be modelled using OPDGs, but modelling type (ii) cannot be done (except for in very limited cases) since it requires adding edges in an unstructured way, which cannot be achieved using only context-freeness.

We have previously shown that contextual HRGs can handle both structural and non-structural reentrancies [6]. Moreover, a parser generator has been provided by Drewes, Hoffmann, and Minas for contextual HRGs [7] that, when it succeeds in producing a parser, guarantees that the parser will run in quadratic (and in the common case, linear) time on its input. However, for some input HRGs the generator discovers a parsing conflict and fails to output a parser. It is unknown whether the class of contextual HRGs on which the parser generator succeeds is sufficient to cover all AMR languages, but it is known that these are incomparable to OPDGs. For these reasons, we are now looking to combine OPDGs with the contextual mechanism used in [7], so as to be able to efficiently parse graph languages that contain reentrancies of both type (i) and (ii).

References

[1] Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. Abstract meaning representation for sembanking. In *Proc. of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, 2013.

[2] Henrik Björklund, Frank Drewes, and Petter Ericson. Parsing weighted order-preserving hyperedge replacement grammars. In *Proceedings of the 16th Meeting on the Mathematics of Language*, pages 1–11, 2019.

[3] Annegret Habel. *Hyperedge replacement: grammars and languages*, volume 643. Springer Science & Business Media, 1992.

[4] Frank Drewes, Hans-Jörg Kreowski, and Annegret Habel. Hyperedge replacement graph grammars. In *Handbook of Graph Grammars and Computing by Graph Transformation*. 1997.

[5] Frank Drewes and Berthold Hoffmann. Contextual hyperedge replacement. *Acta Informatica*, 52(6):497–524, 2015.

[6] Frank Drewes and Anna Jonsson. Contextual hyperedge replacement grammars for abstract meaning representations. In *Proceedings of the 13th International Workshop on Tree Adjoining Grammars and Related Formalisms*, pages 102–111, Umeå, 2017. Association for Computational Linguistics.

[7] Frank Drewes, Berthold Hoffmann, and Mark Minas. Extending predictive shift-reduce parsing to contextual hyperedge replacement grammars. In Esther Guerra and Fernando Orejas, editors, *Graph Transformation*, pages 55–72, Cham, 2019. Springer International Publishing.