

PDRT-SANDBOX: An implementation of Projective Discourse Representation Theory

Noortje J. Venhuizen

University of Groningen

n.j.venhuizen@rug.nl

Harm Brouwer

Saarland University

brouwer@coli.uni-saarland.de

Abstract

We introduce PDRT-SANDBOX, a Haskell library that implements Projective Discourse Representation Theory (PDRT) (Venhuizen et al., 2013), an extension of Discourse Representation Theory (DRT) (Kamp, 1981; Kamp and Reyle, 1993). The implementation includes a translation from PDRT to DRT and first-order logic, composition via different types of merge, and unresolved structures based on Montague Semantics (Muskens, 1996), defined as Haskell functions.

1 Introduction

The semantic property of projection, traditionally associated with presuppositions, has challenged many structure-driven formal semantic analyses. Linguistic content is said to project if it is interpreted outside the scope of an operator that syntactically subordinates it. In semantic formalisms, this behaviour has often been treated as a *deviation* from standard meaning construction, despite the prevalence of expressions exhibiting it (van der Sandt, 1992; Geurts, 1999; Beaver, 2001). By contrast, we have proposed a formalism that *centralizes* the property of projection as a strategy for integrating material into the foregoing context. This formalism is called Projective Discourse Representation Theory (PDRT) (Venhuizen et al., 2013), and is an extension of the widely used framework Discourse Representation Theory (DRT) (Kamp, 1981; Kamp and Reyle, 1993). In PDRT, all linguistic material is associated with a *pointer* to indicate its interpretation site. In this way, an explicit distinction is made between the surface form of an utterance, and its logical interpretation. The formalism can account for various projection phenomena, including presuppositions (Venhuizen et al., 2013) and Potts' (2005) conventional implicatures (Venhuizen et al., 2014),

and has already been integrated into the Groningen Meaning Bank (Basile et al., 2012).

Critically, adding projection pointers to all linguistic material affects the formal properties of DRT non-trivially; the occurrence of projected material at the interpretation site results in non-hierarchical variable binding, and violates the traditional DRT notion of context accessibility, thereby compromising the basic construction mechanism. Here, we present an updated construction mechanism as part of a Haskell library called PDRT-SANDBOX that implements PDRT, as well as standard DRT. The implementation incorporates definitions for building and combining structures, translating Projective Discourse Representation Structures (PDRSs) to Discourse Representation Structures (DRSs) and first-order logic (FOL) formulas, and dealing with unresolved structures via lambda abstractions (Muskens, 1996). Moreover, it allows for various input and output representations, and is highly modular, thereby providing a full-fledged toolkit for use in other NLP applications.

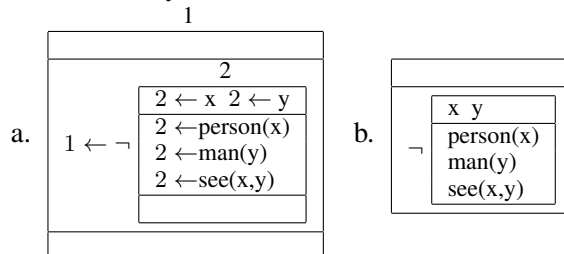
2 Projective Discourse Representation Theory

PDRSs carry more information than DRSs; in addition to the structural and referential content of a DRS, a PDRS also makes the information structure of a discourse explicit by keeping linguistic content at its introduction site, and indicating the interpretation site via a projection variable. That is, each PDRS introduces a *label* that can be used as an identifier, and all of its referents and conditions are associated with a *pointer*, which is used to indicate in which context the material is *interpreted* by means of binding it to a context label.

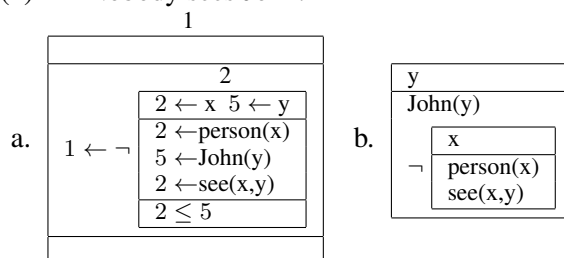
Examples (1) and (2) show two PDRSs and their corresponding DRSs. An important addition to the PDRS definitions described in Venhuizen et al. (2013), is the introduction of *Min-*

imally Accessible Projection contexts (MAPs) in the footer of each PDRS. These MAPs pose minimal constraints on the accessibility of projection contexts, creating a partial order over PDRS contexts (Reyle, 1993; Reyle, 1995).

(1) Nobody sees a man.



(2) Nobody sees John.



In the PDRS in (1a), all pointers are bound by the label of the PDRS in which the content is introduced, indicating *asserted material*. As shown in (1b), this representation is identical to the standard DRT representation of this sentence, except for the addition of labels to PDRSs and pointers to all referents and conditions. In (2), on the other hand, the proper name “John” triggers a presupposition about the existence of someone called ‘John’. The pointer associated with the referent and condition describing this presupposition indicates *projected material*; it occurs free, as it is not bound by the label of any accessible PDRS. This means that no antecedent has been found yet. In the corresponding DRS in (2b) the presupposition is accommodated at the most global accommodation site. Note that in contrast to the DRT representation, the accommodation site of the presupposition is not determined in the PDRS; (2a) only stipulates that the accommodation site should be accessible from the introduction site of the presupposition. This flexibility of interpretation increases the compositionality of PDRT, since more context may become available later on in which the presupposition becomes bound. In combination with MAPs, this property can also be exploited to account for the projection behaviour of conventional implicatures (Venhuizen et al., 2014).

3 Playing in the PDRT-SANDBOX

We implemented the formal definitions for the construction and manipulation of the structures of PDRT and standard DRT in a Haskell library called PDRT-SANDBOX. For a full description of all definitions, see Venhuizen et al. (in prep). The library provides the following core features:

- **Definitions for building and combining (P)DRSs.** The binding and accessibility definitions in DRT and PDRT are fully worked out, and applied as conditions on combining (*merging*) structures and resolving them. Two different types of merge are defined for PDRT: *projective merge* and *assertive merge* (Venhuizen et al., 2013).
- **Translations.** PDRSs can be translated to DRSs, FOL-formulas, and flat (non-recursive) representations called P-Tables.
- **Lambda abstractions.** Unresolved structures obtain Montague-style representations, following Muskens (1996). The implementation exploits Haskell’s lambda-theoretic foundations by formalising unresolved structures as Haskell functions, thereby profiting from all existing associated functionality.
- **Various input and output formats.** As (P)DRS output format, the standard “boxes” representation is available, as well as a linear representation of the boxes, a set-theoretic representation, and the internal syntax for (P)DRSs. The latter two are also recognised as input formats, along with the Prolog syntax from Boxer (Bos, 2003).

4 Conclusion

PDRT-SANDBOX is a full-fledged NLP library for constructing and manipulating the discourse structures from DRT and PDRT, which can be used as part of a larger NLP architecture. One direction would be combining the implementation with a syntactic parser, resulting in a tool-chain similar to the one created by the C&C tools and Boxer (Curran et al., 2007). Furthermore, the representations produced by PDRT-SANDBOX may be applied in a separate model checker, QA system, or any other NLP tool that uses deep semantic representations. PDRT-SANDBOX is freely available (under the Apache License, Version 2.0) at: <http://hbrouwer.github.io/pdrt-sandbox/>

References

- Valerio Basile, Johan Bos, Kilian Evang, and Noortje Joost Venhuizen. 2012. Developing a large semantically annotated corpus. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 3196–3200, Istanbul, Turkey. European Language Resources Association (ELRA).
- David I. Beaver. 2001. *Presupposition and Assertion in Dynamic Semantics*. CSLI Publications, Stanford, CA.
- Johan Bos. 2003. Implementing the binding and accommodation theory for anaphora resolution and presupposition projection. *Computational Linguistics*, 29(2):179–210.
- James R. Curran, Stephen Clark, and Johan Bos. 2007. Linguistically motivated large-scale NLP with C&C and Boxer. In *Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions*, pages 33–36. Association for Computational Linguistics.
- Bart Geurts. 1999. *Presuppositions and pronouns*. Elsevier.
- 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.
- Hans Kamp. 1981. A theory of truth and semantic representation. In J.A.G. Groenendijk, T.M.V. Janssen, and M.B.J. Stokhof, editors, *Formal Methods in the Study of Language 135*, pages 277–322. Mathematisch Centrum.
- Reinhard Muskens. 1996. Combining Montague Semantics and Discourse Representation. *Linguistics and Philosophy*, 19(2):143–186.
- Christopher Potts. 2005. *The Logic of Conventional Implicatures*. Oxford University Press, USA.
- Uwe Reyle. 1993. Dealing with ambiguities by underspecification. *Journal of Semantics*, 10(2):123–179.
- Uwe Reyle. 1995. On reasoning with ambiguities. In *Proceedings of the seventh European chapter of the Association for Computational Linguistics*, pages 1–8. Morgan Kaufmann Publishers Inc.
- Rob van der Sandt. 1992. Presupposition projection as anaphora resolution. *Journal of Semantics*, 9:333–377.
- Noortje J. Venhuizen, Johan Bos, and Harm Brouwer. 2013. Parsimonious semantic representations with projection pointers. In Katrin Erk and Alexandre Koller, editors, *Proceedings of the 10th International Conference on Computational Semantics (IWCS 2013) – Long Papers*, pages 252–263, Potsdam, Germany, March. Association for Computational Linguistics.
- Noortje J. Venhuizen, Johan Bos, Petra Hendriks, and Harm Brouwer. 2014. How and why conventional implicatures project. In *Proceedings of the 24th Semantics and Linguistic Theory Conference (SALT 24)*, New York, May 30 – June 1. New York University.
- Noortje J. Venhuizen, Johan Bos, Petra Hendriks, and Harm Brouwer. in prep. Harnessing projection: implementing Projective Discourse Representation Theory. Manuscript in preparation.