Towards an RRG treebank of Biblical Hebrew

Christian Canu Højgaard
Nicolai Winther-Nielsen
Fjellhaug International University College

This paper will introduce and discuss the ongoing construction of a treebank of Biblical Hebrew based on Role and Reference Grammar. The work is anchored in the TreeGraSP research project which aims to combine linguistic descriptions and data-driven approaches (Bladier et al. 2018; 2019; 2020). Biblical Hebrew is an ancient Semitic language preserved over millennia in the Hebrew Bible. As such, Biblical Hebrew offers stable and ideal data to test how well an RRG description applies to an ancient language. The corpus is stored in the so-called BHSA database (*Biblia Hebraica Stuttgartensia Amstelodamensis*), created and maintained by the Eep Talstra Centre of Bible and Computer at Vrije Universiteit, Amsterdam (https://etcbc.github.io/bhsa/). The database is richly annotated with morphological, lexical, and syntactic features, but no semantic description. The purpose of the present project is to enrich the database with an RRG-description of Biblical Hebrew sentences that allows Old Testament researchers to apply the insights of RRG in linguistic and exegetical interpretation. The project is a continuation of previous work on building a computational parser for Biblical Hebrew (Winther-Nielsen 2008; 2009). An RRG description of Hebrew sentences and verbal aspect is also the topic of recent research (Højgaard 2021). For the present paper, the agenda is twofold: Firstly, we will present the automatic parsing of Hebrew sentences into an RRG-like format (cf. Figure 1 for an example).

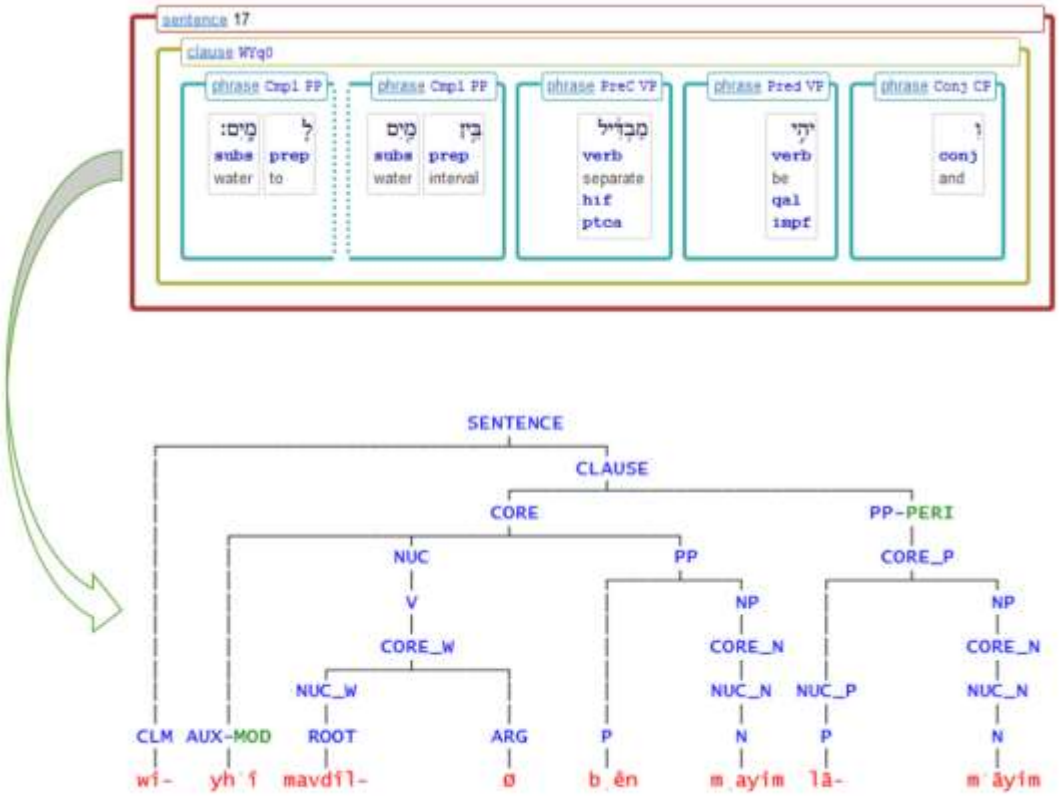

Figure 1: Transformation of BHSA syntax to RRG-trees (Genesis 1:6)

Secondly, we will discuss the concrete challenges for creating trees of a heavily inflected language such as Biblical Hebrew.

References:

Bladier, Tatiana, Andreas van Cranenburgh, Kilian Evang, Laura Kallmeyer, Robin Möllemann, and Rainer Osswald. 2018. "RRGbank: A Role and Reference Grammar Corpus of Syntactic Structures Extracted from the Penn Treebank." In *Proceedings of the 17th International Workshop on Treebanks and Linguistic Theories*. Oslo University, Norway.

Bladier, Tatiana, Kilian Evang, Laura Kallmeyer, Robin Möllemann, and Rainer Osswald. 2019. "Creating RRG Treebanks through Semi-Automatic Conversion of Annotated Corpora." Presented at the International Conference on Role and Reference Grammar 2019, University of Buffalo, NY, August 19.

Bladier, Tatiana, Laura Kallmeyer, Rainer Osswald, and Jakup Waszchuk. 2020. "Automatic Extraction of Tree-Wrapping Grammars for Multiple Languages." In *Proceedings of the 19th International Workshop on Treebanks and Linguistic Theories*, 55–61. Düsseldorf: Association for Computational Linguistics.

Højgaard, Christian C. 2021. "Roles and Relations in Biblical Law: A Study of Participant Tracking, Semantic Roles, and Social Networks in Leviticus 17–26." Ph.D. diss., Amsterdam: Vrije University.

Winther-Nielsen, Nicolai. 2008. "A Role-Lexical Module (RLM) for Biblical Hebrew: A Mapping Tool for RRG and WordNet." In *Investigations of the Syntax-Semantics-Pragmatics Interface*, edited by Robert D. Van Valin Jr., 455–78. Studies in Language Companion Series 105. Amsterdam ; Philadelphia, PA: John Benjamins.

———. 2009. "Biblical Hebrew Parsing on Display: The Role-Lexical Module (RLM) as a Tool for Role and Reference Grammar." *Hiphil* 6: 1–51.