

Semantic mapping of participants in legal discourse

Christian Canu Højgaard
Vrije Universiteit, Amsterdam
cch@dbi.edu

Abstract: In this paper, I will present a new method of applying open source digital resources for investigating patterns of participants in the legal material of the so-called Holiness Code (Lev 17-26). Although several scholars have tried to explain the internal relationships between the participants in the Holiness Code, the large corpus complicates a systematic and consistent analysis. Developments within computational corpus-linguistics, however, provide new ways of systematic investigations into the discourse of Biblical legal texts. Accordingly, this paper will demonstrate how open source technologies and resources, such as the ETCBC database, text-fabric, R, and Jupyter Notebook, can benefit discourse analysis of the Holiness Code. Using Leviticus 25:23-28 as a case-study, the paper presents a three-step method of creating a semantic map of the participants and their internal relationships. First, semantic role labels are distributed to all participants according to the Role and Reference Grammar theory of the thematic relationship between verb and arguments. Second, all linguistic participant references are tracked and linked to their respective referents. Finally, the formal, linguistic participant tracking and the semantic role labelling are combined to create a semantic map of participants.

Keywords: computational methods, corpus-linguistics, semantic roles, Role and Reference Grammar, Leviticus 17–26.

Leviticus 17-26, often called the Holiness Code (H), has a recurring set of participants, namely the Lord, Moses, the addressees, the foreigners, the brother, the neighbor etc. Traditionally, the unexpected shifts of participant references, e.g. between singular and plural references, were seen as signs of “seems” in the text, indicating a presumed historical growth of the law code. More recently, scholars have increasingly read the law code synchronically and treated the participant references as rhetorical devices constituting the discourse of the law code (e.g. Joosten 1996; Meyer 2004). Regardless of how the participant references are appreciated, it is complicated to map the participants in a text spanning ten chapters. Therefore, there is a need to develop more systematic approaches to tracking and mapping participants and their semantic roles.

In this paper, I will demonstrate how this task can be performed with open-source, digital resources. The corpus-linguistic analysis presented in this paper uses the ETCBC database of the Hebrew Bible which can be accessed via the Python3 package text-fabric (Roorda et al. 2018). The extraction and modelling of data is performed with open source Jupyter Notebook¹ which supports the programming languages Python and R used in this project.

The method presented in this paper has three steps. First, the semantic roles of the participants are interpreted, assisted by a computational algorithm. Second, all participant references are clustered into groups of participants and added to a participant concordance. Finally, the semantic roles and the participant tracking are combined in order to create a visual map of the participants, allowing for further interpretation. For the present paper, Leviticus 25:23-28 has been chosen as a case study to demonstrate the method outlined. The programming codes used in the project are available at GitHub.²

¹ <http://jupyter.org/>

² <https://github.com/ch-jensen/Semantic-mapping>

1. Semantic classification of verbs and arguments

The semantic role labelling (SRL) of participants in this case study is based on the Role and Reference Grammar theory (RRG) of the semantic relationship between a verb and its arguments (Van Valin 2005; Van Valin and LaPolla 1997). Van Valin (2005) suggests a three-step method for SRL:

1. Determine the predicate class (*Aktionsart*) of the verb
2. Determine the logical structure of the verb
3. Determine the thematic relation between the verb and its argument(s)

The strength of this procedure is its consistency. Rather than arbitrarily choosing semantic roles, the procedure moves from a close inspection of the *Aktionsart* of the verb (i.e., the inherent temporal aspects of the verb) towards an informed interpretation of the thematic relationship between the verb and its arguments. Moreover, the procedure can be computer-assisted in order to further enhance consistency. In what follows, I will present a computer-assisted annotation of *Aktionsart* and semantic roles according to the procedure just outlined.

1.1 Determining the predicate class (*Aktionsart*)

Van Valin (2005) proposes 12 predicate classes to account for all verbal temporal aspects. These are as follows (all six classes below have causative counterparts, e.g. Causative state, Causative activity, etc.):

- | | |
|---------------------------|---|
| 1. State: | [+static] [–dynamic] [–telic] [–punctual] |
| 2. Activity: | [–static] [+dynamic] [–telic] [–punctual] |
| 3. Achievement: | [–static] [–dynamic] [+telic] [+punctual] |
| 4. Semelfactive: | [–static] [±dynamic] [–telic] [+punctual] |
| 5. Accomplishment: | [–static] [–dynamic] [+telic] [–punctual] |
| 6. Active accomplishment: | [–static] [+dynamic] [+telic] [–punctual] |

Given the fact that *Aktionsart* properties are mutually exclusive, e.g. a verb cannot be both static and active, Winther-Nielsen (2016; 2008) has suggested an algorithm for decomposing Hebrew verbs using test questions to specify the *Aktionsart*. In what follows, I will present an adaptation of Winther-Nielsen's algorithm in four steps, cf. Table 1.

The algorithm is constructed to divide after test 2. If the verb is identified as progressive, tests 3a and possibly 3b are carried out to further classify the verb. On the contrary, tests 4a and 4b are used to classify non-progressive verbs. Classes in bold signal the resulting *Aktionsart* and the end of the algorithm.

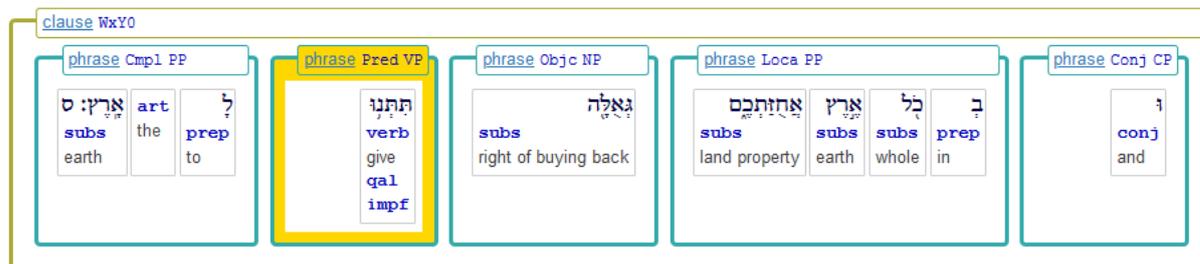
The algorithm, being coded in Jupyter Notebook, has several benefits. The algorithm forces the researcher to answer one question at a time by posing the relevant question constrained by answers previously submitted. Moreover, the program is designed to work with texts in their context, thus starting from the first clause of a predefined text and working through the text clause by clause. At the same time, the entire clause is printed in the output of the program in order for the researcher to observe the verb in its context (cf. Figure 1). Finally, the verb classifications are saved in a global file which is always consulted when analyzing a verb, so that the researcher can select between previously classifications of the verb – if present – or choose to make a new classification. To sum up, the classification program assists the researcher in answering the right test questions at the right time and

making the right conclusions, and by storing the results ready for further investigation as well as in a global file containing Hebrew verbs and *Aktionsart*.

Table 1. Algorithm for determining *Aktionsart*.

# Test	Test question	Yes	No
1. ± Causal	Is this verb inherently causative, that is, can it be paraphrased with ‘caused’?	+ CAU	
2. ± Progression ³	Is this situation compatible with a progressive condition, i.e. ‘is ...ing’?	ACC, ACT, AA	STA, ACH, SEM
<i>If progression = Yes:</i>			
3a. ± Dynamic	Can the verb occur with adverbs like ‘vigorously’ or ‘actively’?	ACT, AA	ACC
3b. ± Telic ⁴	Does the event reach a result state after some time, e.g. ‘in an hour’?	AA	ACT
<i>If progression = No:</i>			
4a. ± Punctual	Can the verb occur with adverbs like ‘instantly’?	ACH, SEM	STA
4b. ± Telic	Does the event endure for some time, e.g. ‘for an hour’?	SEM	ACH

Leviticus 25:24



Is this verb inherently causative, that is, can it be paraphrased with 'caused'?

Figure 1. Computer-assisted annotation of *Aktionsart* with text-fabric.

1.2 Determining logical structure

The next step of the SRL procedure is to determine the logical structure of the verb. The *Aktionsart* is regarded as a cross-linguistic phenomenon instantiated by a language-specific predicate, i.e. the event of, e.g., saying can be expressed in any language with different lexicalization. This combination of universal semantics and language-specific lexicalization is expressed in the logical structure of the verb. In the following table of logical structures, **predicate** represents the language-specific verbal lexeme while **do** represents the semantics of an activity verb. x, y, and z represent the semantic arguments of the verb:⁵

³ Van Valin (2005) uses Progression to distinguish between verbs in various combinations of dynamism, telicity, and punctuality. Progression can be formalized as [±dynamic][±telic][-punctual].

⁴ Tests 3b and 4b examine two different types of telicity, namely, temporal endpoint (3b) and duration (4b).

⁵ INGR means ingressive and conceptualizes a punctual change from state to another. SEML means punctual and continuous, and BECOME represents Accomplishment.

Table 2. Logical representations for *Aktionsart* classes (Van Valin 2005, 45).

Aktionsart class	Logical structure
State	predicate' (x) or (x, y)
Activity	do' (x, [predicate' (x) or (x, y)])
Achievement	INGR predicate' (x) or (x, y), or INGR do' (x, [predicate' (x) or (x, y)])
Semelfactive	SEML predicate' (x) or (x, y), or SEML do' (x, [predicate' (x) or (x, y)])
Accomplishment	BECOME predicate' (x) or (x, y), or BECOME do' (x, [predicate' (x) or (x, y)])
Active accomplishment	do' (x, [predicate' ₁ (x, (y))]) & INGR predicate' ₂ (z, x) or (y)
Causative	α CAUSE β , where α , β are logical structures of any type

The logical structures will prove efficient in the final step towards classification of participant roles.

1.3 Classification of participant of roles

The logical structure of the verb and its semantic arguments provides the necessary foundation for classifying participant roles, which is the final step of the SRL procedure. Van Valin (2005, 55) has suggested a list of 39 semantic roles that a semantic argument can take, depending on the its position in the logical structure (x, y, or z) and the lexical meaning of the verb. Unlike the classification of *Aktionsart*, there is no strict procedure, such as test questions, to determine the participant role. Therefore, the semantic role must be identified on the basis of the lexical meaning of the verb itself. This vagueness is also evident in the computer-assisted annotation.

The computational algorithm works by looping sequentially over all phrases in the text, except for certain phrase types which are not likely to be participant references, such as adverbial phrases. For each relevant phrase, the researcher is required to select the most reasonable semantic role. The inherent vagueness of this procedure is countered by providing information regarding verbal *Aktionsart* and the logical structure of the clause. This information enhances the quality of the annotation.

2. Participant tracking

For a comprehensive semantic mapping of discourse participants in a text, it goes without saying that it is imperative to be able to keep track of the participants, even in sentences where the participants are only referred to with anaphors or are implied, such as an implicit subject. The basic principle is co-referentiality, that is, linking of linguistic references referring to the same extra-textual referent. For several years, Eep Talstra has developed advanced algorithms for tracking participants in the Hebrew Bible. In a study of Ex 19, Talstra (2016; cf. 2018) describes his eight-step procedure as a bottom-up methodology starting by linking co-referring entities within the same clause, e.g. subject and verb, followed by linking references across textual domains, e.g. narrative and discourse.

For the purposes of this case-study, I have tracked the participants manually and collected the data in a small data set comprising the phrase nodes according to the ETCBC database, the transliterated text and the presumed referent.

3. Semantic mapping

The overall objective of the preceding data collection is to generate a map of all participants and their semantic roles. A semantic mapping will in turn allow for an extensive statistical analysis of the participants and their roles. In contrast to more traditional studies of the participants in the Hebrew Bible, which have primarily focused on individual participants and limited textual examples, a semantic mapping includes all references and all semantic roles and thereby enhances the quality and consistency of the interpretation. Apart from qualitative interpretation, quantitative measures can be taken into consideration.

The purpose of this paper is not to explore the quantitative and statistical implications of a semantic mapping, but merely to illustrate how textual participants can be abstracted into a visual mapping. Further studies will need to explore the interpretational potential of a semantic mapping.

A visual mapping of the participants is given in Figure 4. The sizes of the bubbles refer to the frequency of any particular combination of participant and role, while the colors illustrate the group of *Aktionsart* to which each predicate class belongs.⁶ The colors correspond to the position of the bubbles in the diagram. Bubbles at the top of the diagram refer to the most agentive roles in contrast to the bottom of the diagram containing the least agentive roles. In the present case, for instance, “man” and “purchaser” are both more agentive than “brother”. This makes sense, because the brother is the participant being sold as a debt slave and we would expect him to be less agentive than, e.g., his purchaser.

The semantic map, therefore, has a great potential for discourse analysis. The mapping shows the roles attributed to each participant in the text, and it allows for distinguishing agentive or dominant participants from less agentive or dominant participants. In the end, a semantic mapping of the participants allows for a more informed analysis of the discourse-centrality and role of the participants.

⁶ For *Aktionsart*-“groups”, see Van Valin (2005, 57–58).

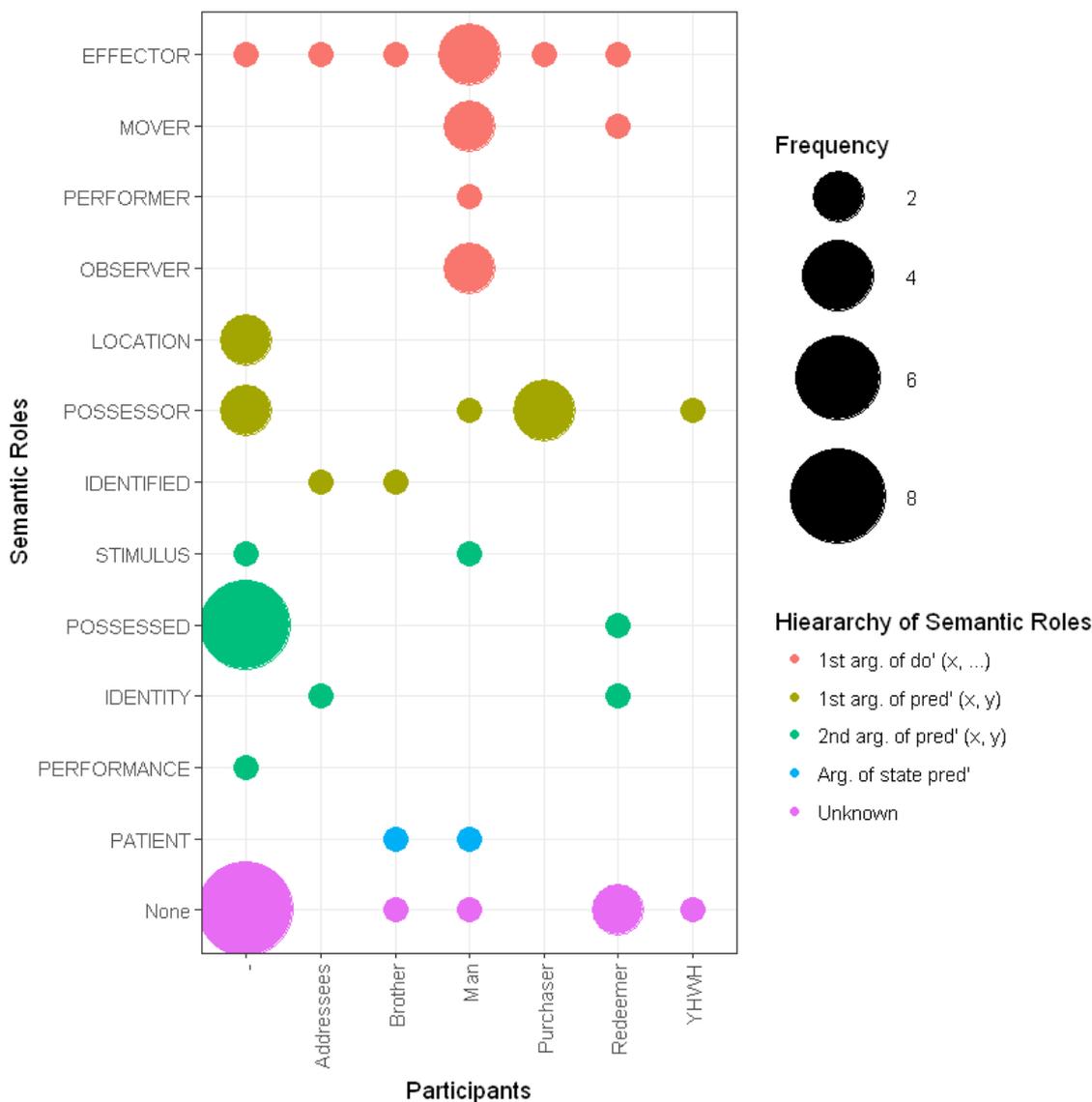


Figure 2. Semantic map of participant roles and references in Lev 25:23-28.

4. Conclusion

The method described here is a three-step method of 1) classifying the semantic roles of participants, 2) participant tracking, and 3) combining step 1 and 2 into a joint semantic map. This approach enhances the consistency of the participant analysis and, thereby, the understanding of the discourse of the text. Throughout this project, open-source digital resources have been used in order to provide freely accessible data and programming codes for transparency and reproduction of the data.

It is important to keep in mind that the approach described here is statistical. This means that even though a participant may frequently be ascribed a certain semantic role, this role may not be as important to the discourse of the text as another, less frequent role. Some propositions carry more weight than others, often signaled by occupying a prominent place in the text or by being restated in close repetition. Therefore, a fuller investigation of the relationship between semantic roles and the discourse of the text will also need to include an analysis of the structure of the text. The benefit of this statistical approach, however, is that it may reveal otherwise unknown patterns to be investigated. In

this respect, a statistical analysis can be a fruitful first step for analyzing participants in (legal) discourse.

5. Bibliography

- Joosten, Jan. 1996. *People and Land in the Holiness Code: An Exegetical Study of the Ideational Framework of the Law in Leviticus 17-26*. Supplements to Vetus Testamentum 67. Leiden ; New York: E.J. Brill.
- Meyer, Esias Engelbertus. 2004. "The Jubilee in Leviticus 25: A Theological Ethical Interpretation from a South African Perspective." University of Stellenbosch.
- Roorda, Dirk, Cody Kingham, Christiaan Erwich, and SeHoon Park. 2018. "Etcbc/Bhsa: Default Text Format For Lexemes." Zenodo. <https://doi.org/10.5281/zenodo.1446393>.
- Talstra, Eep. 2016. "Approaching the Mountain of Exodus 19: Thou Shalt Explore Syntax First." *HIPHIL Novum* 3 (1): 2–24.
- . 2018. "Who is who in Zechariah 1:1-6? Text Linguistics, Participant Tracking and the Reading of Biblical Texts." In *Reading and listening: meeting one God in many texts: Festschrift for Eric Peels on the occasion of his 25th jubilee as professor of Old Testament Studies*, edited by Jaap Dekker and Geert Kwakkel, 16:151–58. Amsterdamse Cahiers voor Exegese van de Bijbel en zijn Tradities, Supplement Series. Bergambacht: 2VM.
- Van Valin, Robert D., Jr. 2005. *Exploring the Syntax-Semantics Interface*. Cambridge: Cambridge University Press.
- Van Valin, Robert D., Jr., and Randy J. LaPolla. 1997. *Syntax: Structure, Meaning, and Function*. Cambridge Textbooks in Linguistics. Cambridge, U.K. ; New York, NY: Cambridge University Press.
- 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: John Benjamins Pub.
- . 2016. "How to Classify Hebrew Verbs: Plotting Verb-Specific Roles." In *Contemporary Examinations of Classical Languages (Hebrew, Aramaic, Syriac, and Greek): Valency, Lexicography, Grammar, and Manuscripts*, edited by Timothy Martin Lewis, Alison Salvesen, and Beryl Turner, 67–94. Perspectives on Linguistics and Ancient Languages 8. Piscataway, NJ: Gorgias Press.