

Nonverbal Predicates in Modern Hebrew

Petter Haugereid^{1,2,3}, Nurit Melnik² and Shuly Wintner³

¹Department of Linguistic, Literary and Aesthetic Studies, University of Bergen, Norway

²Department of Literature, Language and the Arts, The Open University of Israel, Israel

³Department of Computer Science, University of Haifa, Israel

1 Introduction

Nonverbal predicates in Modern Hebrew (MH) have been the subject of investigation in a number of studies (e.g., Doron 1983; Falk 2004). However, to our knowledge, none of the existing studies was corpus-based. We show that an empirical examination of this phenomenon reveals patterns which have not been previously considered yet which should be taken into account when proposing a comprehensive analysis. Our analysis indeed accounts for all these patterns, unifying the treatment of three types of copular constructions that we identify in MH. The analysis is implemented as part of a larger-scale grammar, and is extensively tested.

The standard data items which appear in the literature regarding the constructions are the following:

- | | | |
|---|---|---|
| (1) <i>dani (hu) more</i>
dani (he) teacher.SM
'Dani is a teacher.' | (2) <i>dani (hu) nexmad</i>
dani (he) nice.SM
'Dani is nice.' | (3) <i>ha-yeladim (hem) al ha-gag</i>
the-kids (they) on the-roof
'The kids are on the roof.' |
|---|---|---|

Here, the predicates consist of NPs (1), AdjPs (2), and PPs (3). The copula linking the subject and the predicate is homophonous with the 3rd person pronoun (hence the gloss) and agrees with the subject. These forms of the copula are only used in present tense, and they are sometimes optional. In past and future tense an inflected form of the verb *haya* 'be' is obligatorily used. The present tense form of *haya* is missing from the MH inflectional paradigm.

- | | |
|---|---------------------|
| (4) <i>dina hayta/tihiye</i>
dina was.3SF/will.be.3SF teacher.SF/nice.SF
'Dina was/will be a teacher/nice.' | <i>mora/nexmada</i> |
|---|---------------------|

The examples above constitute the prototypical examples of the copular construction. In the majority of the papers which address this construction (in MH as well as in other languages), the subject in the data items is an animate NP. Less frequent subjects are concrete nouns (e.g., *The books are on the table.*). Yet it is particularly rare to find abstract nouns as subjects. Nevertheless, corpus searches¹ reveal that abstract nouns are in fact more frequent, and, perhaps more importantly, exhibit additional properties which are often overlooked when only animate nouns are examined.

One such property involves agreement. While AdjP predicates obligatorily agree with the subject ((2) and (4)), the agreement relationship between subjects and NP predicates is more complex. Animate nouns are inflected for natural gender and number. Inanimate and abstract nouns have grammatical gender. When subjects and nominal predicates in the copular construction are animate they agree in

¹In this study we use a 60-million token WaCky corpus of Hebrew (Baroni *et al.* 2009).

number and gender ((1) and (4)). Although it seems like a subject-predicate agreement relationship mandated by the grammar, this constraint only holds when the nominal predicate has natural gender. Nominals with grammatical gender do not inflect and consequently agreement mismatches such as those in (5) occur.

- (5) *ha-sefer hu matana*
the-book.SM he gift.SF
'The book is a gift.'

In addition to the often discussed NP, AdjP, and PP predicates, two additional types of predicates are revealed, when one considers subjects that are realized as *abstract* nouns: infinitival VPs and finite clauses. These are illustrated in (6).

- (6) *ha-matara hi lehenot / she-dani yehene.*
the-goal.3SF she to.enjoy / that-dani will.enjoy
'The goal is to have fun / that Dani will have fun.'

These types of predicates are absent from the literature on copular constructions in MH, yet seem to belong to the same category as the more prototypical constructions above.

To summarize, the apparent optionality of the copula and its categorial identity, the types of predicates involved, and the different agreement patterns are all issues which need to be addressed when proposing an account of the data.

2 Nonverbal Constructions in HPSG

Nonverbal predicates are not prototypical selecting predicates. Phrases that can function as nonverbal predicates can also appear as subjects or complements of finite verbs. This raises the question of whether there are any differences between these phrases in their predicational vs. non-predicational function.

The 'canonical' HPSG analysis of nonverbal predicates involves a lexical rule which takes as input a 'regular' noun lexeme and outputs a predicative noun (Pollard & Sag 1994; Ginzburg & Sag 2000). The predicative noun is marked as such by the binary feature PRD, its semantic content shifts from *scope-obj* to *psoa*, and it selects for an NP subject, thus creating a link between it and the clausal subject. The requirement in English for a finite verb to form a clause is fulfilled in this case by a finite form of the copula *be*, which combines with its predicative complement to form a VP. As a subject raising predicate, the copula structure-shares the subject requirement of its predicative complement, and combines with the subject to form a clause. Importantly, in this analysis the copula is semantically vacuous.

The motivation for identifying predicational phrases as such extends beyond the copular construction, since the same class of predicates have a similar distribution (Pollard & Sag 1987:66). Nevertheless, when more data and other languages are considered the analysis faces some problems. Van Eynde (2008) raises a number of arguments against the lexical rule analysis. He objects to the systematic ambiguity which the noun rule introduces and provides evidence against the identification of the subject of the copula with the subject of the predicate.

Van Eynde (2008; 2009; 2012) shifts the burden from the predicate to the copula. According to his analysis, the copula is not devoid of semantic content. Rather, the semantic link between the subject and the predicate is captured in the lexical entry of the copula.

While Van Eynde places the burden of the licensing of the construction on the copula, MH as well as other languages allow its omission in certain contexts. Although Van Eynde (2009:368) argues that this "is not by itself an argument for semantic vacuity" the (sometimes optional) omission of a copula is a challenge to an analysis in any framework. In the HPSG framework phonologically empty elements

are generally avoided. Nevertheless, several accounts of copular constructions in a variety of languages do assume empty elements.

Bender (2001) proposes a “silent verb analysis” to account for copula absence in African American Vernacular English (AAVE). Her crucial argument is based on complement extraction. When the predicative phrase is extracted in copula-less clauses (e.g., *Where your car?*) there needs to be a place to register the extraction site. This, she argues, can be done either by reintroducing traces or by the use of a silent copula. Both solutions require the stipulation of phonologically empty elements.

Müller (2002; 2004) argues that in certain constructions the use of empty elements is advantageous. One such case is the omission of the copula in declarative sentences in German, where assuming a phonologically empty copula preserves the topological fields. Moreover, he demonstrates that the avoidance of empty elements may lead to the stipulation of additional rules and schemata, as well as a linguistically less insightful grammar.

Finally, the agreement relationship between subjects and predicates is shown by Van Eynde (2012) to be neither CONCORD agreement nor INDEX agreement, in the sense of Kathol (1999) and Wechsler & Zlatić (2003). Instead of having predicates structure-share their CONCORD properties with the INDEX features of their subject, he proposes that their INDEX features be matched. When predicates have identical CONCORD and INDEX features, INDEX agreement between them and their subjects is reflected morphologically (e.g., adjective agreement in (1) and (4)). When predicates are underspecified for INDEX agreement features they structure-share those features with their subject, while their morphological agreement properties can either match or mismatch those of the subject.

3 Our Analysis

Three types of constructions with nonverbal predicates are identified for MH: the present tense construction, with and without a copula, and the past/future construction with inflected *haya* ‘be’. Our analysis accounts for the commonalities as well as the distinct properties of each construction by proposing the type hierarchy in Figure 1.

The type *cop-sign* subsumes all copula signs and specifies their valence requirements and semantic content. Similarly to Van Eynde, we propose that nonverbal constructions do not involve raising. While raising requires identity between the subject of the copula and the subject of the predicate, this is not the case with infinitival VPs or clauses (e.g., (6)). Moreover, a raising construction is not compatible with cases of agreement mismatches, such as in (5) above. The syntactic relationship between the subject and the predicate, then, is defined in the VAL feature, which consists of two arguments. The subject (C-ARG1) is an NP, while the predicate (C-ARG2) can be either AdjP, AdvP, S, VPinf, NP or PP.

The treatment of agreement is adopted from Van Eynde (2012) and involves the structure-sharing of just the INDEX features of the two arguments. Morphological agreement occurs only with predicates whose INDEX and CONCORD features are structure-shared (e.g., animate nouns and AdjPs in (1), (2) and (4)). Details are suppressed for brevity.

The semantic function of the copula in all three constructions is to link the index of the subject with that of the predicate. Unlike existing proposals, we account for this function in the type hierarchy rather than in the lexicon. Consider *cop-sign* (Figure 1); KEYREL specifies the key relation of signs, and its value, *cop-rel*, specifies a binary relation with two arguments.² The value *arg12* of C-FRAME indicates a transitive frame, guaranteeing linking between C-ARG1 and C-ARG2 and the arguments of *cop-rel*.

The similarity between the two types of constructions ends when the categorial status of the copular element is considered, and thus two immediate subtypes are defined: *pres-cop-struct* and *past-future-cop-lxm*. Following Doron (1983) we posit that the past and future tense forms of *haya* are truly verbal,

²Note that unlike standard accounts, here KEYREL is specified in the type hierarchy, rather than the lexicon, so that all the three types of copulas inherit this constraint.

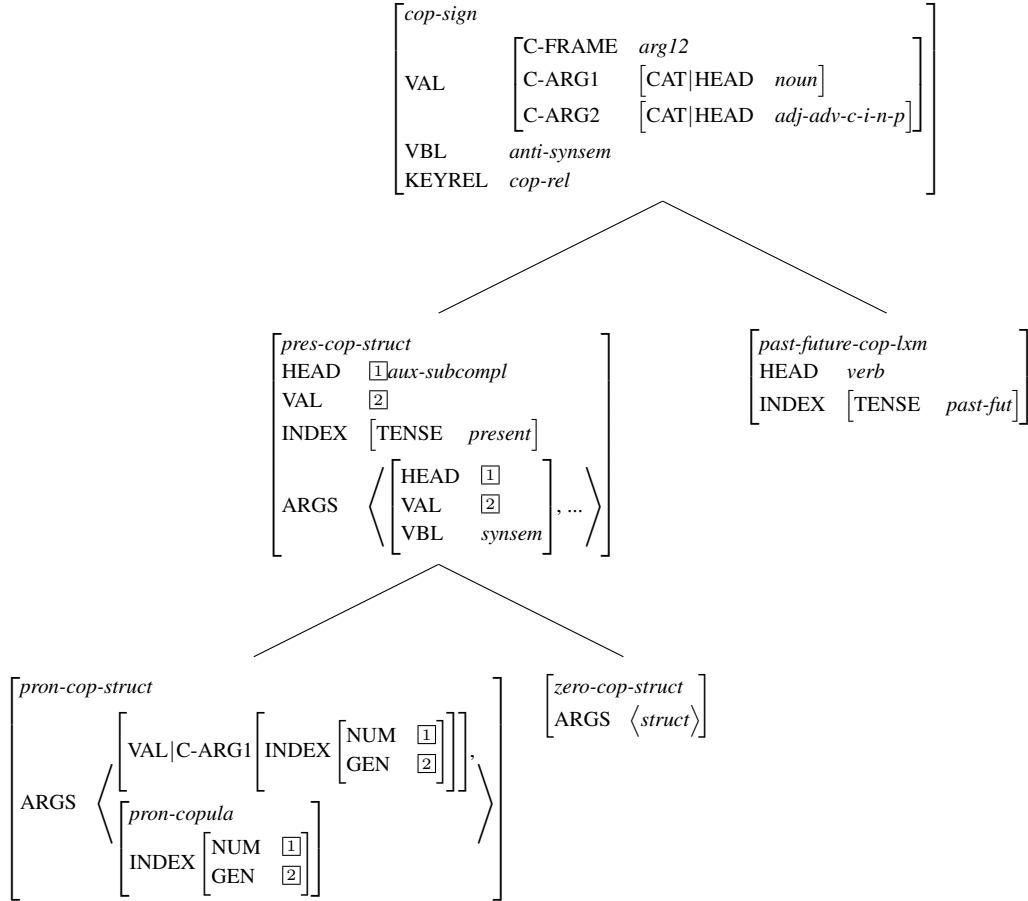


Figure 1: Type hierarchy of copula signs

while the present tense copula, *Pron*, is the realization of agreement features. Inflected forms of *haya* behave like ordinary transitive verbs: they take a subject and a complement, exhibit full agreement with the subject, and are licensed in the same word order configurations in which finite verbs are licensed. The syntactic behavior of *Pron* is different in all the aforementioned contexts. Thus we conclude that *Pron* is not a verb. Moreover, while all *Prons* are homophonous with 3rd person pronouns, they have several distinctive semantic and syntactic properties, which lead us to propose a separate category. Consequently we view *Pron* as an agreement marker which is associated with a specific head feature: *pron-cop*. The definition of the *Pron hu* ‘he’ is given in (7).

$$(7) \left[\begin{array}{l} \text{sgm-pron-cop} \\ \text{STEM} \langle hu \rangle \\ \text{CAT} \left[\text{HEAD } \textit{pron-cop} \right] \\ \text{CONT} \left[\begin{array}{l} \text{INDEX} \left[\begin{array}{l} \text{NUM } \textit{sg} \\ \text{GEN } \textit{masc} \end{array} \right] \\ \text{RELS} \langle !! \rangle \end{array} \right] \end{array} \right]$$

The alternation between copular and copula-less constructions is accounted for by the two subtypes of *pres-cop-struct*. *Pron-cop-struct* ensures the agreement between the subject (C-ARG1) and *Pron* (the second element in ARGS). *Zero-cop-struct* is a unary rule which introduces the empty *Pron*. In practice

this means that the rule imposes the constraints of *cop-sign* (and *pres-cop-struct*) without attaching *Pron*.

This analysis of zero copulas enables us to account for cases of complement extraction such as those discussed by Bender (2001), as well as for sentences such as (8) below where a predicate is extracted from a zero-copula construction, leaving behind an adverb.

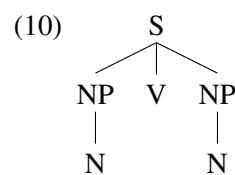
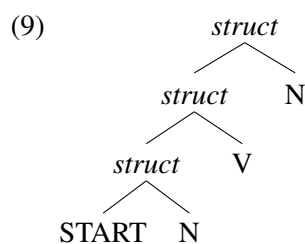
- (8) *eifo ata xoshev she-dani axshav?*
 where you.2SM think.SF that-Dani now

‘Where do you think Dani is now?’

We thus provide a unified analysis for the three types of copulas in Modern Hebrew (*Pron*, the verb-like *haya*, and the empty copula) in terms of agreement, constituent structure, and semantics.

The analysis is implemented with the LKB system (Copestake 2002). We integrated into the grammar a wide-coverage morphological processor of Hebrew (Itai & Wintner 2008), thereby obtaining broad coverage and robustness. The grammar now includes over 30 thousand lemmas, or some 150,000 inflected forms.

The grammar design employed in our implementation makes a distinction between the parse tree and the constituent structure (Haugereid & Morey 2012). The parse tree is left-branching, and is built incrementally (9). Each step of the parse is licensed by a *structure*, a subtype of *sign*. *Structures* have an ARGS list consisting of either one or two members. The first member describes the string parsed so far, and various features of this string are used to constrain the properties of the following word. These constraints are encoded in the second member of the ARGS list. An additional feature, *VBL*, is used to record whether a verb is still required, and if so, to pose constraints on that verb. Although the parse tree is strictly left-branching, the grammar builds a (standard) constituent structure that encodes the appropriate semantic structure (10). The constituent structure we assume is relatively flat.



To test the grammar, we created a test suite of positive and negative items in the format of [incr tsdb()] (Oepen & Flickinger 1998). The suite tests agreement between the subject and the predicate, agreement between the subject and the copula, empty copula constructions, word order alternations, and subject and predicate extraction (including extraction from subordinate clauses). Our grammar fully covers the positive items, assigning the correct expected syntactic and semantic structures to all of them. In terms of negative examples, the grammar slightly overgenerates. This is due to the fact that the binary copula construction is allowed to insert a *Pron* after the predicate and the subject. While this is grammatical for regular verbs, including *haya* and its inflected forms, it is strongly questionable for *Pron*. We have not yet decided whether or not this should be ruled out by the grammar.

Acknowledgments

This research was supported by THE ISRAEL SCIENCE FOUNDATION (grant No. 505/11).

References

- BARONI, MARCO, SILVIA BERNARDINI, ADRIANO FERRARESI, & EROS ZANCHETTA. 2009. The WaCky wide web: a collection of very large linguistically processed web-crawled corpora. *Language Resources And Evaluation* 43.209–226.
- BENDER, EMILY M., 2001. *Syntactic Variation and Linguistic Competence: The Case of AAVE Copula Absence*. Stanford University dissertation.
- COPESTAKE, ANN. 2002. *Implementing Typed Feature Structure Grammars*. CSLI publications.
- DORON, EDIT, 1983. *Verbless predicates in Hebrew*. University of Texas, Austin dissertation.
- FALK, YEHUDA N. 2004. The hebrew present-tense copula as a mixed category. In *Proceedings of the Lexical Functional Grammar 04 Conference (LFG-04)*.
- GINZBURG, JONATHAN, & IVAN A. SAG. 2000. *Interrogative Investigations: the form, meaning, and use of English interrogatives*. Stanford, California: CSLI Publications.
- HAUGEREID, PETTER, & MATHIEU MOREY. 2012. A left-branching grammar design for incremental parsing. In *Proceedings of the 19th International Conference on HPSG*, ed. by Stefan Müller, 181–194.
- ITAI, ALON, & SHULY WINTNER. 2008. Language resources for Hebrew. *Language Resources and Evaluation* 42.75–98.
- KATHOL, ANDREAS. 1999. Agreement and the syntax-morphology interface in HPSG. In *Studies in Contemporary Phrase Structure Grammar*, ed. by Robert Levine & Georgia Green, 223–274. Cambridge University Press.
- MÜLLER, STEFAN. 2002. Multiple frontings in german. In *Proceedings of Formal Grammar 2002*, ed. by Gerhard Jäger, Paola Monachesi, Gerald Penn, & Shuly Winter, Trento.
- . 2004. Elliptical constructions, multiple frontings, and surface-based syntax. In *Proceedings of Formal Grammar*, ed. by G. Penn G. Jäger, P. Monachesi & S. Wintner.
- OEPEN, STEPHAN, & DANIEL P. FLICKINGER. 1998. Towards systematic grammar profiling. Test suite technology ten years after. *Journal of Computer Speech and Language* 12 # 4 Special Issue on Evaluation.411–436.
- POLLARD, CARL J., & IVAN A. SAG. 1987. *Information-based Syntax and Semantics, Vol. 1*. Number 13 in CSLI Lecture Notes. Stanford University: CSLI Publications. Distributed by University of Chicago Press.
- , & ———. 1994. *HPSG*. Chicago: University of Chicago Press.
- VAN EYNDE, FRANK. 2008. Predicate complements. In *The Proceedings of the 15th International Conference on HPSG*, ed. by Stefan Müller, 253–273, Stanford. CSLI Publications.
- . 2009. On the copula: from a fregean to a montagovian treatment. In *Proceedings of the 16th International Conference on HPSG*, 359–375, Stanford. CSLI Publications.
- VAN EYNDE, FRANK. 2012. On the agreement between predicative complements and their target. In *Proceedings of the 19th International Conference on HPSG*, ed. by Stefan Müller, 348–366.
- WECHSLER, STEPHEN, & LARISA ZLATIC. 2003. *The Many Faces of Agreement*. Stanford Monographs in Linguistics. Stanford, CA: CSLI Publications.