Abstract

Nonverbal predicates in Modern Hebrew (MH) have been the subject of investigation in a number of studies. However, to our knowledge, none of them was corpus-based. Corpus searches reveal that the nonverbal constructions which are most commonly addressed in the literature are not the most commonly used ones. Once a broader range of data is considered additional issues are raised. Our analysis addresses these issues, 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.

1 Introduction

This paper contributes to the longstanding discussion of nonverbal predicates and the copula. Our starting point is a corpus investigation of this phenomenon in Modern Hebrew (MH).[1] Although nonverbal predicates in MH have been the subject of investigation in a number of studies (e.g., Doron 1983; Falk 2004; Greenberg 2008), to our knowledge, none of the existing studies have conducted a corpus investigation. We will show that an empirical corpus-based examination of this phenomenon reveals patterns which have not been previously considered yet which should be taken into account when proposing a comprehensive analysis.

The standard data items that illustrate these constructions in the literature are:

(1) a. dani (hu) more/nexmad
dani (he) teacher.SM/nice.SM
'Dani is a teacher/nice.'

b. ha-yeladim (hem) al ha-gag
the-kids (they.M) on the-roof
'The kids are on the roof.'

Here, the predicates consist of NPs (1a), AdjPs (1a), and PPs (1b). The copula linking the subject and the predicate is homonymous with 3rd person pronouns (hence the gloss) and agrees with the subject. The pronominal 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.

(2) a. dina hayta mora/nexmada
dina was.3SF teacher.SF/nice.SF
'Dina was a teacher/nice.'

b. dani ve-dina yihiyu morim/nexmadim
danny and-dina will.be.3P teachers.PM/nice.PM
'Danny and Dina will be teachers/nice.'

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

[1]For this study we use a 60-million token WaCky corpus of Hebrew (Baroni et al., 2009).

In what follows we will refer to the present tense copula as 'Pron' to distinguish it from the past and future tense forms of *haya*.

The sentences above constitute the prototypical examples of the copular construction, as is reflected by the majority of the papers that address this construction (in MH as well as in other languages). The subject in most 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 searches[2] 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 is the use of two additional types of predicates: infinitival VPs and finite clauses. These are illustrated in (3)–(4).

(3) ha-matara hi lehenot.
the-goal.3SF she to.enjoy
'The goal is to have fun.'

(4) ha-matara hi she-dani yehene.
the-goal.3SF she that-dani will.enjoy
'The goal is 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 commonly discussed constructions above. In the following section we will investigate the syntactic properties of all types of nonverbal constructions in MH.

2 Syntactic Properties of Nonverbal Constructions in MH

In this section we will focus on a number of syntactic properties of the constructions, which will ultimately be accounted for in the proposed analysis. More specifically, we will address the issues of subject-predicate agreement, word order alternations, unbounded dependency constructions, and the categorial identity of Pron.

2.1 Agreement Patterns

Among the different nonverbal predicates identified in MH, the only relevant ones to consider in terms of subject-predicate agreement are NPs and AdjPs, since both nouns and adjectives in MH are specified for number and gender. Animate nouns are inflected for natural gender and number. Inanimate and abstract nouns have grammatical gender. Adjectives obligatorily agree with whatever they modify; attributive adjectives agree with the nominal head in an NP and predicative adjectives with their subjects (cf. (1a) & (2)).

[2]For the English data, The BYU British National Corpus was used (Davies, 2004-).

The agreement patterns between NP subjects and NP predicates are not as straightforward as those with AdjP predicates. When the NP predicate is an animate noun that has a full number-gender inflectional paradigm the two NPs exhibit full agreement.

(5) a. ha-more hu sporta'i
the-teacher.SM he athlete.SM

b. ha-mora hi sporta'it
the-teacher.SF she athlete.SF
'The teacher is an athlete.'

(6) a. ha-morim hem sporta'im
the-teachers.PM they.M athletes.PM

b. ha-morot hen sporta'iyot
the-teacher.PF they.F athletes.PF
'The teachers are athletes.'

With inanimate nouns, which are marked with grammatical gender, gender agreement is irrelevant. Thus in (7) below the masculine subject appears with a feminine predicate. Number, however, does play a role with singular concrete nouns, as singular NP subjects are incompatible with plural NP predicates.

(7) ha-sefer hu matana/*matanot
the-book.SM he gift.SF/gifts.PF
'The book is a gift.'

There are, however, examples of number mismatches when abstract nouns are involved. In (8), a singular NP subject appears with a plural NP predicate.

(8) ha-hesber ha-yexidi le-nicxono hu ha-havtaxot
the-explanation.SM the-single.SM to-his.victory he the-promises.PF
ha-mafligot she-hu natan le-boxarav
the-overarching.PF that-he gave to-his.voters
'The only explanation for his victory is the over-arching promises that he gave his voters.'

Furthermore, with plural subjects the cardinality of the NP predicate encodes a collective vs. distributive distinction, where singular predicates produce a collective reading (9a), and plural predicates a distributive one (9b).

(9) a. ha-sfarim ha-'ele hem matana mi-axi
the-books.PM the-these.PM they.M present.SF from-my.brother
'These books are a present from my brother.'

b. ha-sfarim ha-'ele hem matanot mi-xaverai
the-books.PM the-these.PM they.M presents.PF from-my.friends
'These books are presents from my friends.'

As the English translations of these example sentences indicate, these number agreement patterns are similar in the two languages.

2.2 Word Order Alternations

The unmarked word order of clauses in MH is SVO, and this applies to clauses with nonverbal predicates as well. Nevertheless, a number of word order alternations were attested in the corpus. In one such alternation, exemplified by (10), Pron follows the predicate.

(10) shtei ha-yecirot makbilot hen mi-bxinat ha-mivne
two the-pieces.PF parallel.PF they.F from-aspect the-structure
'The two pieces are parallel in terms of their structure.'

Moreover, there are attested examples of predicate-initial constructions (aka inverse copular constructions), where Pron optionally appears between the predicate and the subject (11a).[3] However, no occurrences of predicate-initial and Pron-final clauses (Pred-S-Pron) were attested (cf. (11b)).

(11) a. me'atim (hem) ha-nos'im ha-ma'asikim et ha-siyax
few.PM (they.M) the-topics.PM the-occupying.PM ACC the-discourse
ha-ciburi
the-public
'Few are the topics which occupy the public discourse.'

b. * me'atim ha-nos'im hem
few.PM the-topics.PM they.M
Intended meaning: 'The topics are few.'

Finally, one position where Pron does not occur is clause-initially, regardless of the relative ordering of the subject and predicate.

2.3 Unbounded Dependency Constructions

"Extraction" from clauses with nonverbal predicates is possible with subjects and predicates. When predicates are extracted Pron can optionally appear (12). When subjects are extracted Pron is absent (13).

(12) a. eifo (hu) ha-oto?
where the-car
'Where is the car?'

b. adayin lo barur ma (hi) ha-siba
still NEG clear what (she) the-reason.PF
'It is still not clear what is the reason.'

[3] It appears that the "weight" of the subject NP plays a role in the licensing of this construction.

(13) mi (*hu) ayef?
who he tired.SM
'Who is tired?'

When subjects of nonverbal predicates are relativized the relativizer *she-* 'that' is prefixed to the predicate, an AdjP in the following example. Similarly to subject wh-questions, Pron cannot occur in this construction.

(14) zehu sug ha-mahalaxim she-ofyaniyim lo
this-he type the-moves.PM that-typical.PM to-him
'This is the type of moves which are typical of him.'

2.4 What is Pron?

Pron is identical in form to 3rd person personal pronouns and it obligatorily agrees in number and gender with the subject. Nevertheless, we assume that it is not a pronoun. First, Pron is not assigned a semantic role. Second, it can be used as a copula in a construction where the subject is a personal pronoun (15) and the person features of the two elements are mismatched.

(15) ani hu ha-manhig
I he the-leader
'I am the leader.'

Third, the wh-words *ma* 'what' and *mi* 'who' have variants that are inflected for number and gender. Inflected wh-forms are used only in the present tense copular construction (cf. (16b) & (17b)), while the pronoun in the periphrastic form can be either Pron or a pronoun (cf. (16a) & (17a)).

(16) a. ma hi ha-be'aya?
what she the-problem.SF
b. mahi ha-be'aya?
what-she the-problem.SF
'What is the problem?'

(17) a. ma hi amra?
what she said.3SF
b. * mahi amra?
what-she said.3SF
'What did she say?'

The cliticized form of Pron can also appear with the demonstrative *ze*, while personal pronouns cannot.

(18) zehu axi
this-he axi
'This is my brother.'

(19) * zehu
this-he
'It's him.'

Nevertheless, Pron cannot be classified as a clitic since there are no adjacency requirements between it and the subject or predicate. Adverbs can be placed in the two positions.

(20) a. ha-seret hu be'ecem ma'agali lexalutin
the-movie.SM he actually circular.SM completely
'The movie is actually completely circular.'

b. kol ha-mishkal be'ecem hu ba-beten
all the-weight.SM actually he in.the-stomach
'All the weight is actually in the stomach.'

The similar function of present tense Pron and past/future tense *haya* 'be' suggests that Pron may be a type of a verb. However, as Doron (1983) notes, the distribution of Pron is distinct from verbs in general, and *haya* in particular. For example, V2-like constructions occur in MH with 'real' verbs (21), but not with Pron (22).

(21) a. hayom Dani roce banana
today Dani wants.SM banana

b. hayom roce Dani banana
today wants.SM Dani banana
'Today, Danny wants a banana.'

(22) a. hayom Dani hu more
today Dani he teacher

b. * hayom hu Dani more
today he Dani more
'Today, Danny is a teacher.'

An additional distinction involves the placement of the negative *lo*:

(23) a. Dani lo roce banana
dani NEG wants.SM banana
'Danny doesn't want a banana.'

b. Dani lo haya more
dani NEG was.3SM teacher
'Danny wasn't a teacher.'

(24) Dani hu lo more
dani he NEG teacher
'Danny isn't a teacher.'

Moreover, in subject extraction constructions such as (13) and (14) above, where Pron is not licensed, the inflected forms of *haya* appear in past and future tense.

To summarize, the types of predicates involved, the agreement patterns between subjects and predicates, the word order alternations, and the categorial identity of Pron and its apparent optionality, are all issues which need to be addressed when proposing an account of the data. In what follows we first review previous HPSG-based analyses of copular constructions and then present our proposal.

3 Nonverbal Constructions in HPSG

3.1 The Role of the Copula

The 'canonical' HPSG analysis of nonverbal predicate constructions views the copula as a type of a subject raising verb that structure-shares the subject requirement of its predicative complement, and combines with the subject to realize this requirement and form a clause. Importantly, in this analysis the copula does not contribute to the semantics of the clause. These properties are captured in the abbreviated description of the copula *be* (Pollard & Sag, 1994, 147).

(25) $$\left[\begin{array}{l}\text{CAT} \mid \text{SUBCAT}\left\langle \boxed{1}\text{NP}, \text{XP}\left[\text{+PRD}, \text{SUBCAT}\left\langle \boxed{1} \right\rangle\right]:\boxed{2} \right\rangle \\ \text{CONTENT}\ \boxed{2}\end{array}\right]$$

This approach to the construction requires that nonverbal predicates select for NP subjects and be marked as +PRD. To this end, a lexical rule takes as input a 'regular' noun lexeme and outputs a predicative noun with a non-empty SUBCAT list (Pollard & Sag 1994; Ginzburg & Sag 2000). 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, page 66).

This analysis adequately accounts for the constructions commonly considered in the literature. Nevertheless, when more data and other languages are investigated 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. Moreover, he provides evidence against a raising analysis which identifies the subject of the copula with the subject of the predicate. One such case is the infinitival VP predicate, such as illustrated in (3) above. The unexpressed subject of the VP 'to enjoy' has arbitrary reference which cannot be equated with the subject of the clause 'the goal'. This observation holds for the MH data above, its English translation, and the Dutch examples given by Van Eynde (2008). Moreover, when clausal

predicates are involved (e.g., (4)) no unexpressed subject exists, since the subject of the clausal predicate is realized within the predicate itself.

Van Eynde (2008; 2009; 2012) proposes a Montagovian treatment of predicative complements, which shifts the burden from the predicate to the copula (or the selecting verb). 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.

(26)
$$\begin{bmatrix} \text{PHON} \left\langle \textit{be} \right\rangle \\ \text{ARG-ST} \left\langle \text{NP}_{\boxed{1}}, \text{XP}_{\boxed{2}} \right\rangle \\ \text{SS} \mid \text{LOC} \mid \text{CONTENT} \mid \text{NUCLEUS} \begin{bmatrix} \textit{be-rel} & \\ \text{THEME} & \boxed{1}\ \textit{index} \\ \text{ATTRIBUTE} & \boxed{2}\ \textit{index} \end{bmatrix} \end{bmatrix}$$

One argument for the necessity of ascribing semantic content to the copula comes from the assignment of the EXPERIENCER role in sentences such as *This book is too expensive for me*. If the copula is semantically vacuous, Van Eynde asks, what assigns the EXPERIENCER role to the PP *for me*?

3.2 Copula Omission

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 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.

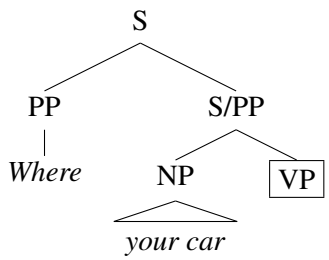

Figure 1: Bender's analysis of copula absence in AAVE

Müller (2002; to appear) 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.

A construction-based analysis of zero copula is proposed by Henri & Abeillé (2007) for the copular construction in Mauritian. They show that the Mauritian copula does not behave like the AAVE copula. Its realization or omission are construction-dependent, as the copula appears only in extraction contexts. This particular behavior, they argue, lends itself well to a construction-based analysis. Moreover, they conclude that a null form analysis is warranted only in cases where the distribution of the copula and zero copula are not complementary.

4 An Analysis of MH Nonverbal Predicates

The analysis we propose provides a unified account of the three types of constructions identified for MH: the present tense construction, with and without Pron, and the past/future construction with verb-like inflected *haya* 'be'. It accounts for the entire range of nonverbal predicates of MH: NP, AdjP, PP, AdvP, VPinf, and S.

Similarly to Van Eynde (2008; 2009; 2012), 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., (3), (4)). In addition, a raising construction is not compatible with cases of agreement mismatches, such as in (7) above.

Moreover, contrary to 'standard' HPSG analyses, we concur with Van Eynde in assuming that the semantic content of the copula is not vacuous. Rather, its function is to link the subject and the predicate. Nevertheless, unlike Van Eynde's analysis, ours does not require all predicates to be of type *scope-object*. Predicates are linked to the subject by the *copula-rel* depending on their type.

Finally, we assume the existence of a phonologically empty Pron element which shares the syntactic and semantic properties of the overt Pron and the verbal *haya* 'be'. In doing so we can account for cases of complement extraction, which were shown by Bender (2001) (for AAVE) to require the stipulation of an empty element. Moreover, this approach enables us to provide a unified account of the three constructions and to "capture the facts in an insightful way", in the words of Müller (to appear, page 103).

4.1 Grammar Design

The grammar we design makes a distinction between the parse tree and the constituent structure (Haugereid & Morey, 2012). The parse tree is left-branching, and is built incrementally (Figure (2)). 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.

The grammar that generates the parse tree implements a stack that stores the necessary features of some ('matrix') constituent(s) while a dependent structure is generated. The grammar has tree types of rules; (i) *emb-struct* rules, (ii) *in-const-struct* rules, and (iii) *pop-struct* rules.

The *emb-struct* rules push constituent structures onto the stack, attaching the first word of the dependent. This is illustrated in ((27)), where the feature 'F $\boxed{1}$' represents the features of the matrix constituent, the feature 'F $\boxed{2}$' represents the features of the dependent, and the feature S represents the stack. The *in-const-struct* rules add words to constituents that have already been initiated. This type of rules is illustrated in ((28)), where the feature S, representing the stack, is transferred from initial daughter to mother. The rules used in our account of copulas in Hebrew are of this type, and as will be shown, this kind of rules can be unary-branching. Finally, *pop-struct* rules explicate dependent constituent structures, popping their matrix structure off the stack. This is illustrated in ((29)).

(27)
$$\begin{bmatrix} \textit{emb-struct} \\ \text{F} \quad \boxed{2} \\ \text{S} \quad \left\langle \begin{bmatrix} \text{F} & \boxed{1} \end{bmatrix} \right\rangle \end{bmatrix}$$
$$\begin{bmatrix} \textit{struct} \\ \text{F} \quad \boxed{1} \\ \text{S} \quad \langle\rangle \end{bmatrix} \quad \begin{bmatrix} \textit{word} \\ \text{F} \quad \boxed{2} \end{bmatrix}$$

(28)
$$\begin{bmatrix} \textit{in-const-struc} \\ \text{S} \quad \boxed{1} \end{bmatrix}$$
$$\begin{bmatrix} \textit{struct} \\ \text{S} \quad \boxed{1} \end{bmatrix} \quad \left(\textit{word} \right)$$

(29)
$$\begin{bmatrix} \textit{pop-struct} \\ \text{F} \quad \boxed{1} \\ \text{S} \quad \langle\rangle \end{bmatrix}$$
$$\begin{bmatrix} \textit{struct} \\ \text{S} \quad \left\langle \begin{bmatrix} \text{F} \, \boxed{1} \end{bmatrix} \right\rangle \end{bmatrix}$$

An analysis of a simple transitive sentence (*The boy is eating a fish*) is shown in Figure 2. A rule of type *vbl-struct*, which is a subtype of *in-const-struct*, adds the verb which is selected via the VBL feature (tagged $\boxed{2}$). The VBL value of the verb is transferred to the mother *vbl-struct*, thus constraining whether or not an additional verb is expected. In Hebrew, there will only be one verb, hence the VBL value transferred to the mother will always be *anti-synsem*.

Although the parse tree is strictly left-branching, a corresponding constituent structure which encodes the appropriate semantic structure can be extracted declaratively from the AVM resulting from the parse. For each node in the parse tree, the path to the root node of the corresponding constituent tree is reflected by the stack. When a dependent structure is introduced in the parse tree, with an *emb-struct* rule, a bracket is opened in the constituent tree, and when a dependent is com-

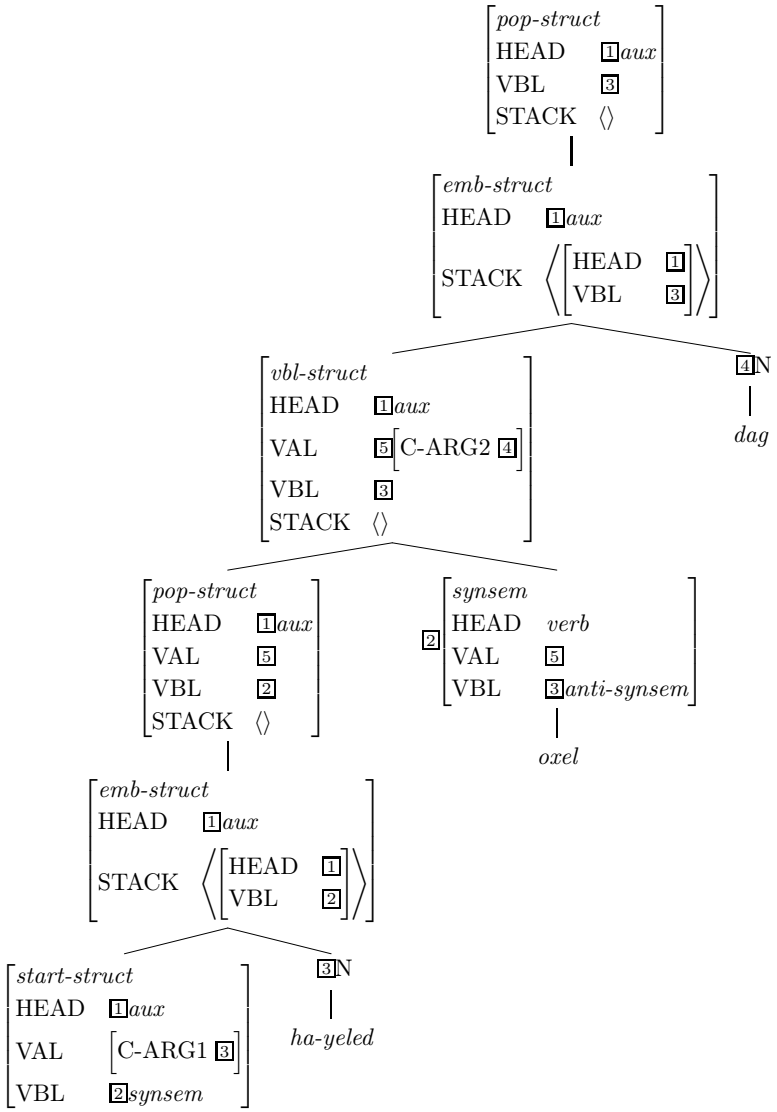

Figure 2: Parse tree for transitive sentence in Hebrew

pletely identified, with a *pop-struct* rule, a bracket is closed in the constituent tree. The constituent structure we assume is relatively flat; see Figure (3).

4.2 The Type Hierarchy

In accounting for nonverbal constructions in MH we distinguish between two cases: the present tense construction, with and without Pron, and the past/future construction with inflected *haya* 'be'. Nevertheless, we recognize a set of properties that the two types have in common. This is reflected in the type hierarchy in Figure 4.

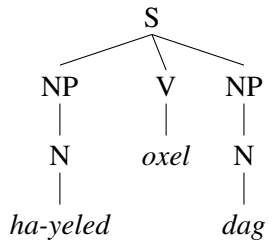

Figure 3: Constituent tree for transitive sentence in Hebrew

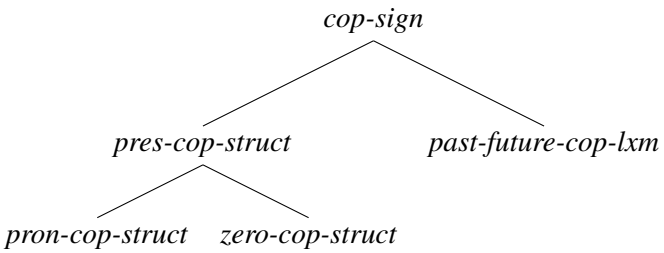

Figure 4: Type hierarchy of copula signs

The type *cop-sign* subsumes all copula signs and specifies their valence requirements and semantic content.

(30)
$$\left[\begin{array}{ll} \textit{cop-sign} & \\ \text{VAL} & \left[\begin{array}{ll} \text{C-FRAME} & \textit{arg12-14} \\ \text{C-ARG1} & \left[\begin{array}{ll} \text{CAT} \mid \text{HEAD} & \textit{noun} \\ \text{CONT} \mid \text{INDEX} & \boxed{3} \end{array}\right] \\ \text{C-ARG2} & \left[\text{CAT} \mid \text{HEAD} \quad \textit{c-i-n}\right] \\ \text{C-ARG4} & \left[\begin{array}{ll} \text{CAT} \mid \text{HEAD} & \textit{adj-adv-p} \\ \text{LKEYS} \mid \text{KEYREL} \mid \text{ARG1} & \boxed{3} \end{array}\right] \end{array}\right] \\ \text{VBL} & \textit{anti-synsem} \\ \text{KEYREL} & \textit{cop-rel} \end{array}\right]$$

The syntactic relationship between the subject and the predicate is defined in the VAL feature. The value *arg12-14* of C-FRAME indicates a transitive frame. C-ARG1 is associated with the subject, and the different kinds of predicates are distributed over two alternating features: C-ARG2 is associated with direct object-like arguments (of type *c-i-n*: S, VPinf, and NP), and C-ARG4 is associated with delimiters (AdjP, AdvP and PP). With C-ARG4 predicates an additional constraint is stated: the INDEX value of the subject (tagged $\boxed{3}$) is structure-shared with the ARG1 of the key relation denoted by the predicate. This constraint introduces a semantic predication relation between the two elements. Moreover, it ensures the full agreement between subjects and AdjP predicates, since adjectives are required

to agree with the noun they modify (i.e., the ARG1 of their key relation), whether predicationally or attributively.

Similarly to Van Eynde, and contrary to previously mentioned HPSG analyses, the semantic content of the copula is not vacuous. Rather, the semantic function of the copula in all three constructions is to link the INDEX of the subject with that of the predicate in *copula-rel*.[4] Unlike standard accounts, KEYREL is specified in the type hierarchy, rather than the lexicon, so that all the three types of copulas inherit this constraint.

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, while the present tense copula, Pron, is the realization of agreement features.

The lexeme type of *haya* inherits the syntactic and semantic characterizations of *cop-sign*. The lexeme-specific information defined for it are its category, tense specification, and the structure-sharing of its RELS feature with KEYREL.[5]

(31)
$$\begin{bmatrix} \textit{past-future-cop-lxm} & \\ \text{HEAD} & \textit{verb} \\ \text{CONT} & \begin{bmatrix} \text{INDEX} & \begin{bmatrix} \text{TENSE} & \textit{past-fut} \end{bmatrix} \\ \text{RELS} & \left\langle ! \boxed{1} ! \right\rangle \end{bmatrix} \\ \text{KEYREL} & \boxed{1} \end{bmatrix}$$

The type *pres-cop-struct* is of subtype of *in-const-struct*, and, as such, licenses the combination of a parsed structure with the next word. The constraints on this type require that the HEAD and VAL features of the structure parsed so far be 'passed up' to the newly parsed structure. Moreover, it constrains tense to be 'present'. Naturally, this information is inherited by the two subtypes, which account for the alternation between copular and copula-less constructions.

(32)
$$\begin{bmatrix} \textit{pres-cop-struct} & \\ \text{HEAD} & \boxed{1}\textit{aux-subcompl} \\ \text{VAL} & \boxed{2} \\ \text{INDEX} & \begin{bmatrix} \text{TENSE} & \textit{present} \end{bmatrix} \\ \text{KEYREL} & \boxed{3} \\ \text{ARGS} & \left\langle \begin{bmatrix} \text{HEAD} & \boxed{1} \\ \text{VAL} & \boxed{2} \\ \text{VBL} & \textit{synsem} \end{bmatrix}, \ldots \right\rangle \\ \text{C-CONT} \mid \text{RELS} & \left\langle ! \boxed{3} ! \right\rangle \end{bmatrix}$$

[4] This is illustrated in the MRS structures discussed in section 4.3.

[5] Angle brackets with exclamation marks ($\langle ! \cdots ! \rangle$) are used for representing difference lists in HPSG grammar implementations.

One subtype, *pron-cop-struct*, licenses the combination of a parsed structure (the first element in ARGS) with Pron (*pron-cop*, the second element in ARGS). In addition to all the constraints inherited from its supertypes, this particular type ensures number-gender agreement between the subject (C-ARG1 in the parsed structure) and Pron.

(33)
$$\begin{bmatrix} \textit{pron-cop-struct} \\ \text{ARGS} \left\langle \begin{bmatrix} \text{VAL} \mid \text{C-ARG1} \begin{bmatrix} \text{INDEX} \begin{bmatrix} \text{NUM} & \boxed{1} \\ \text{GEN} & \boxed{2} \end{bmatrix} \end{bmatrix} \end{bmatrix}, \begin{bmatrix} \textit{pron-cop} \\ \text{INDEX} \begin{bmatrix} \text{NUM} & \boxed{1} \\ \text{GEN} & \boxed{2} \end{bmatrix} \end{bmatrix} \right\rangle \end{bmatrix}$$

The Pron element that appears in the ARGS list was shown here to be distinct from verbs, pronouns, and clitics. For this reason we define for it a separate category with a specific head feature: *pron-cop*. The definition of the singular-masculine Pron *hu* 'he' is given in (34).

(34)
$$\begin{bmatrix} \textit{sgm-pron-cop} \\ \text{STEM} \quad \langle \textit{hu} \rangle \\ \text{CAT} \quad \left[\text{HEAD } \textit{pron-cop} \right] \\ \text{CONT} \begin{bmatrix} \text{INDEX} & \begin{bmatrix} \text{NUM } \textit{sg} \\ \text{GEN } \textit{masc} \end{bmatrix} \\ \text{RELS} & \langle \text{!!} \rangle \end{bmatrix} \end{bmatrix}$$

The licensing of zero copula is achieved by the second subtype, *zero-cop-struct*, 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*) on the parsed structure without attaching a phonologically realized Pron. It should be noted that although we do not define a phonologically empty Pron in the lexicon, in employing such a unary rule we are in fact proposing the existence of an underlyingly present empty Pron.

(35)
$$\begin{bmatrix} \textit{zero-cop-struct} \\ \text{ARGS} \quad \langle \textit{struct} \rangle \end{bmatrix}$$

The type hierarchy proposed here captures the similarities and differences between the three different nonverbal predicate constructions found in MH: the present tense construction, with and without Pron, and the past/future construction with inflected *haya* 'be'. The non-standard grammar design adopted here introduces into the type inventory of the grammar a new type *structure*, which licenses the incremental construction of a representation. The proposed grammar makes

use of binary parsing rules which incorporate 'the next word' into the structure, and unary parsing rules, which in effect incorporate empty elements and which impose constraints on the structure by way of type inheritance.[6] As will be shown in the next section, in adopting an incremental parsing design we do not lose information regarding the constituent structure and semantic relations between the different components of the sentence. This information is reflected in the constituent tree and MRS representation that are produced by the grammar.

4.3 Example Analyses

In order to illustrate the proposed analysis and its implementation we will first consider the following simple zero-copula construction:

(36) ha-oto (hu) po
the-car (he) here
'The car is here.'

The parse tree produced for this sentence is shown in Figure 5. The parse begins similarly to the standard transitive structure illustrated in Figure 2 above. However, once the parser consumes the first NP constituent the analysis diverges. In the simple transitive clause case, the *vbl-struct* type licenses the introduction of the verb. In the zero-copula construction there is no phonologically realized verb or copula. Instead the parser assumes a null copula, a step which is licensed by the *zero-cop-struct*, a unary rule, and then proceeds to consume the adverbial predicate *po* 'here'.

In the constituent structure produced for this sentence (Figure 6) a phonologically null Pron is represented. As was discussed earlier, the application of the unary *zero-cop-struct* rule on the parsed structure imposes the syntactic and semantic properties associated with the copular construction. Consequently, the feature structure which is associated with the fully parsed sentence captures the semantic relations between the subject and predicate, regardless of the occurrence or absence of a copular element. The MRS of the sentence in (36) given in (37).

(37)
$$\begin{bmatrix} \textit{mrs} & \\ \text{LTOP} & \boxed{h1}\,h \\ \text{INDEX} & \boxed{e2}\,e \\ \text{RELS} & \left\langle \begin{bmatrix} \textit{\_car\_n\_1\_rel} & \\ \text{LBL} & \boxed{h3}\,h \\ \text{ARG0} & \boxed{x4}\,x \end{bmatrix}, \begin{bmatrix} \textit{def\_q\_rel} & \\ \text{LBL} & \boxed{h5}\,h \\ \text{ARG0} & \boxed{x4} \\ \text{RSTR} & \boxed{h6}\,h \\ \text{BODY} & \boxed{h7}\,h \end{bmatrix}, \begin{bmatrix} \textit{copula\_v\_rel} & \\ \text{LBL} & \boxed{h8}\,h \\ \text{ARG0} & \boxed{e2} \\ \text{ARG1} & \boxed{x4} \\ \text{ARG4} & \boxed{h9}\,h \end{bmatrix}, \begin{bmatrix} \textit{\_here\_a\_rel} & \\ \text{LBL} & \boxed{h9} \\ \text{ARG0} & \boxed{e10}\,e \\ \text{ARG1} & \boxed{x4} \end{bmatrix} \right\rangle \\ \text{HCONS} & \left\langle \begin{bmatrix} \textit{qeq} & \\ \text{HARG} & \boxed{h6} \\ \text{LARG} & \boxed{h3} \end{bmatrix} \right\rangle \end{bmatrix}$$

[6] It should be noted that the use of this empty element is constrained to a very specific context.

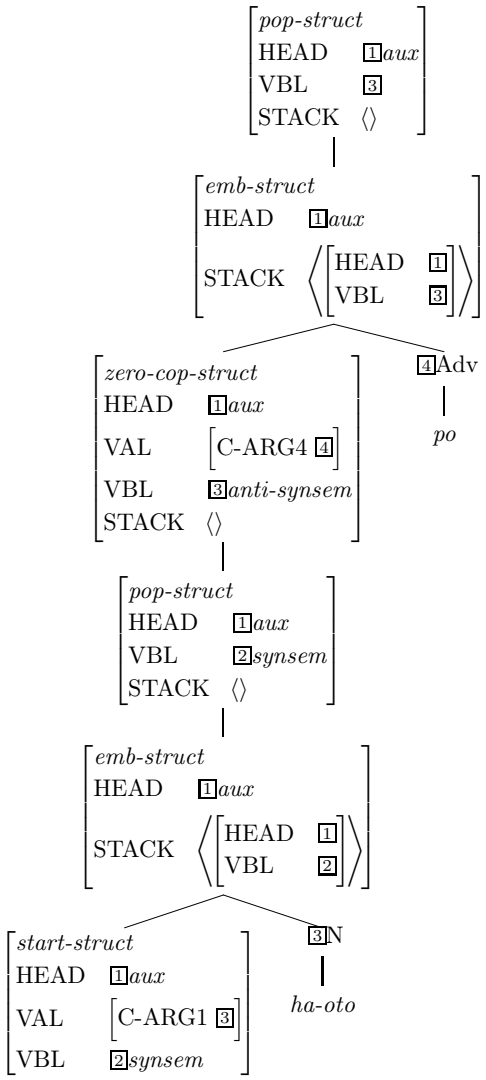

Figure 5: Parse tree for copula sentence with zero copula in Hebrew

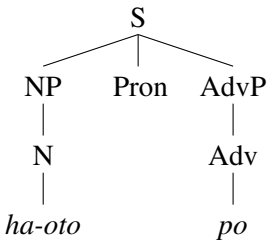

Figure 6: Constituent tree for copula sentence with zero copula in Hebrew

The *copula_v_rel* relation links ARG1 with ARG4, where the value of the former is the INDEX of the subject (tagged x4), and the latter is the label of the relation denoted by the predicate (tagged h9). Moreover, the value of ARG1 in the Adverbial *here_a_rel* relation is structure-shared with the INDEX of the subject. These links are defined in the supertype *cop-sign*.

A different case is VPinf predicates, which are realized as ARG2 complements. The following MRS is a representation of the semantics of the sentence given in (3) ('The goal is to enjoy.').

(38)
$$\begin{bmatrix} \textit{mrs} & \\ \text{LTOP} & \boxed{h1}\, h \\ \text{INDEX} & \boxed{e2}\, e \\ \text{RELS} & \left\langle \begin{bmatrix} \textit{\_goal\_n\_1\_rel} & \\ \text{LBL} & \boxed{h3}\, h \\ \text{ARG0} & \boxed{x4}\, x \end{bmatrix}, \begin{bmatrix} \textit{def\_q\_rel} & \\ \text{LBL} & \boxed{h5}\, h \\ \text{ARG0} & \boxed{x4} \\ \text{RSTR} & \boxed{h6}\, h \\ \text{BODY} & \boxed{h7}\, h \end{bmatrix}, \begin{bmatrix} \textit{copula\_v\_rel} & \\ \text{LBL} & \boxed{h8}\, h \\ \text{ARG0} & \boxed{e2} \\ \text{ARG1} & \boxed{x4} \\ \text{ARG2} & \boxed{e9}\, e \end{bmatrix}, \begin{bmatrix} \textit{\_enjoy\_v\_1\_rel} & \\ \text{LBL} & \boxed{h10}\, h \\ \text{ARG0} & \boxed{e9} \\ \text{ARG1} & \boxed{u11}\, u \end{bmatrix} \right\rangle \\ \text{HCONS} & \left\langle \begin{bmatrix} \textit{qeq} & \\ \text{HARG} & \boxed{h6} \\ \text{LARG} & \boxed{h3} \end{bmatrix} \right\rangle \end{bmatrix}$$

Here, in *copula_v_rel* ARG1 is the INDEX of the subject (tagged x4) and ARG2 is the event INDEX of the *_enjoy_v_1_rel* relation, denoted by the VP (tagged e9).

A more complex case is the construction which led Bender (2001) to conclude that empty elements are necessary in order to account for copula absence in AAVE: complement extraction. The key example which Bender used to illustrate this challenge is *Where your car?*, where a predicative phrase is extracted, yet there is no phonologically realized element where this extraction can be registered. An additional complication to this construction is the case exemplified in (39), where a predicate is extracted from a zero-copula construction, leaving behind an adverb. The zero copula in this case not only records the extraction but is also accessible for modification.

(39) eifo ata xoshev she-ha-oto axshav?
where you.2SM think.SF that-the-car now
'Where do you think the car is now?'

Our grammar handles such cases and produces the correct analysis. The constituent tree constructed for this sentence is given in Figure 7. Note the empty Pron and the AdvP_i, indicating the extraction site.

4.4 Implementation and Evaluation

The analysis proposed here is implemented with the LKB system (Copestake, 2002) as part of HeGram, a computational grammar of Modern Hebrew. HeGram is

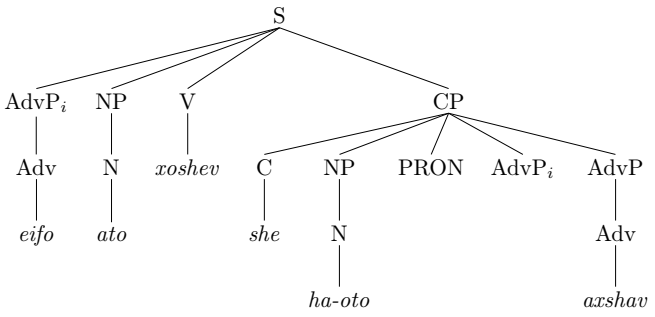

Figure 7: Complement extraction

based on the Norsyg grammar (http://moin.delph-in.net/NorsygTop), and is a part of the DELPH-IN effort (http://www.delph-in.net/). We integrated into the grammar a wide-coverage morphological processor of Hebrew (Itai & Wintner, 2008), thereby obtaining broad coverage and robustness. Consequently our lexicon now includes over 30 thousand lemmas, or some 150,000 inflected forms. The grammar covers basic clause structures such as main clauses, subordinate clauses, relative clauses and infinitival VPs. It handles long distance dependencies, modification, word order, agreement and object marking.

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 AdjP predicate and between the subject and Pron, 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.

Not covered by the grammar are: the interface between syntax, semantics, and pragmatics[7]; constraints on the choice between copula and zero copula; the interaction between copular constructions and the existentials; the complex agreement between NP subjects and NP predicates; and copular constructions with *ze* as copula[8]. These issues are left for future research. In addition, we plan to investigate similar constructions in Standard Arabic and explore the possibility of adapting the MH grammar to account for them.

[7]Müller (2009) distinguishes between three types of copular constructions (equational, predicational, and specificational) and argues that a uniform analysis of all three is inappropriate.

[8]The differences and similarities between the two types of pronominal copulas are discussed by Greenberg (2008).

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(3). 209–226.

Bender, Emily M. 2001. *Syntactic variation and linguistic competence: The case of AAVE copula absence*: Stanford University dissertation. http://faculty.washington.edu/ebender/dissertation/.

Copestake, Ann. 2002. *Implementing typed feature structure grammars*. Stanford: CSLI publications.

Davies, Mark. 2004-. BYU-BNC. (Based on the British National Corpus from Oxford University Press) Available online at http://corpus.byu.edu/bnc/.

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.

Greenberg, Yael. 2008. Predication and equation in Hebrew (nonpseudocleft) copular sentences. In Sharon Armon-Lotem, Gabi Danon & Susan Rothstein (eds.), *Current issues in generative Hebrew linguistics*, 161–196. Amsterdam: John Benjamins Publishing.

Haugereid, Petter & Mathieu Morey. 2012. A left-branching grammar design for incremental parsing. In Stefan Müller (ed.), *Proceedings of the 19th international conference on HPSG*, 181–194. http://cslipublications.stanford.edu/HPSG/2012/haugereid-morey.pdf.

Henri, Fabiola & Anne Abeillé. 2007. The syntax of copular construction in Mauritian. In Stefan Müller (ed.), *The proceedings of the 14th international conference on HPSG*, 130–149. Stanford: CSLI Publications. http://cslipublications.stanford.edu/HPSG/8/.

Itai, Alon & Shuly Wintner. 2008. Language resources for Hebrew. *Language Resources and Evaluation* 42(1). 75–98.

Müller, Stefan. 2002. Multiple frontings in German. In Gerhard Jäger, Paola Monachesi, Gerald Penn & Shuly Winter (eds.), *Proceedings of Formal Grammar 2002*, Trento. http://hpsg.fu-berlin.de/~stefan/Pub/mehr-vf.html.

Müller, Stefan. 2009. On predication. In *Proceedings of the 16th international conference on HPSG*, 213–233. Stanford: CSLI Publications. http://hpsg.fu-berlin.de/~stefan/Pub/predication-hpsg.html.

Müller, Stefan. to appear. Elliptical constructions, multiple frontings, and surface-based syntax. In G. Penn G. Jäger, P. Monachesi & S. Wintner (eds.), *Proceedings of Formal Grammar 2004*, Nancy.

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* (CSLI Lecture Notes 13). Stanford University: CSLI Publications. Distributed by University of Chicago Press.

Pollard, Carl J. & Ivan A. Sag. 1994. *Head-driven phrase structure grammar*. Chicago: University of Chicago Press.

Van Eynde, Frank. 2008. Predicate complements. In Stefan Müller (ed.), *The proceedings of the 15th international conference on HPSG*, 253–273. Stanford: CSLI Publications. http://cslipublications.stanford.edu/HPSG/9/.

Van Eynde, Frank. 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. http://cslipublications.stanford.edu/HPSG/2009/.

Van Eynde, Frank. 2012. On the agreement between predicative complements and their target. In Stefan Müller (ed.), *Proceedings of the 19th international conference on HPSG*, 348–366. http://cslipublications.stanford.edu/HPSG/2012/vaneynde.pdf.