Hebrew Verbal Multi-Word Expressions

Livnat Herzig Sheinfux
University of Haifa

Tali Arad Greshler
University of Haifa

Nurit Melnik
The Open University of Israel

Shuly Wintner
University of Haifa

Proceedings of the 22nd International Conference on
Head-Driven Phrase Structure Grammar

Nanyang Technological University (NTU), Singapore

Stefan Müller (Editor)

2015

CSLI Publications

pages 122–135

http://csli-publications.stanford.edu/HPSG/2015

Herzig Sheinfux, Livnat, Arad Greshler, Tali, Melnik, Nurit, & Wintner, Shuly. 2015. Hebrew Verbal Multi-Word Expressions. In Müller, Stefan (Ed.), *Proceedings of the 22nd International Conference on Head-Driven Phrase Structure Grammar, Nanyang Technological University (NTU), Singapore*, 122–135. Stanford, CA: CSLI Publications.

Abstract

Multi-word expressions (MWEs) are challenging for grammatical theories and grammar development since they blur the traditional distinction between the lexicon and the grammar, and vary in the degree of idiosyncrasy with respect to their semantic, syntactic, and morphological behavior. Nevertheless, the need to incorporate MWEs into grammars is unquestionable, especially in light of estimates claiming that MWEs account for approximately half of the entries in the lexicon. In this study we focus on verbal MWEs in Modern Hebrew: we consider different types of this class of MWEs, and propose an analysis in the framework of HPSG. Moreover, we incorporate this analysis into HeGram, a deep linguistic processing grammar of Modern Hebrew.

1 Introduction

Multi-word expressions (MWEs) in Modern Hebrew (MH), as in other languages, are not simple to characterize, since they vary in the degree of idiosyncrasy with respect to their semantic, syntactic, and morphological behavior. In this study we focus on verbal MWEs: we consider different types of this class of MWEs, and propose an analysis in the framework of HPSG (Pollard & Sag, 1994). Moreover, we incorporate this analysis in HeGram (Herzig Sheinfux et al., 2015), a deep linguistic processing grammar of Modern Hebrew.

Our motivation is twofold. First, the need to incorporate MWEs into the grammar is unquestionable, especially in light of estimates claiming that MWEs account for approximately half of the entries in the lexicon (Sag et al., 2002). Second, we view MWEs as a challenging test case for the innovative architecture implemented in HeGram.

2 Multi-word expressions

MWEs are lexical units that consist of more than one word. They tend to be semantically idiosyncratic. Consider, for example, (1) and (2), in which the idiomatic reading cannot be derived from the idioms' literal parts. One would only understand the meaning if the MWE was already known to him.

(1) *dan yaca me-ha-kelim*
Dan came.out from-the-tools
Literal: 'Dan came out of the tools.'
Idiomatic: 'Dan lost his temper.'

(2) *dan higdil rof*
Dan made.grow head
'Dan took.on responsibility.'

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

In addition, MWEs are characterized by having constrained syntactic behavior. Namely, MWEs can't necessarily be passivized, or undergo wh-questions about the idiomatic arguments ((3) and (4), respectively). However, wh-questions about the literal arguments can occur (5).

(3) *dan huca me-ha-kelim*
**Dan was.taken.out from-the-tools*
'Dan was taken out of the tools' (only odd literal)

(4) *mi-ma dan yaca*
**from-what Dan came.out*
'What did Dan come out of?' (only literal)

(5) *mi yaca me-ha-kelim*
who came.out from-the-tools
'Who lost his temper?'

MWEs are challenging for grammatical theories and grammar development, but as they account for approximately half of the entries in the lexicon (Sag et al., 2002), incorporating them into grammars is important. Moreover, identifying MWEs is important for natural language processing applications – if MWEs are not identified as such, that will probably cause problems further down the processing pipeline.

3 Verbal MWEs in Hebrew

3.1 The Patterns

Hebrew verbal MWEs vary with respect to the specificity of the arguments they take and the relations that hold among them. We identify the following patterns:

Idiomatic NP & PP complements

MWEs can be headed by verbs which lexically select for a particular NP complement (2) or for a PP headed by a particular preposition and complemented by a particular NP (6).

(6) *dan yarad me-haʕec*
Dan went.down from-the.tree
'Dan conceded.'

Possessive idioms

Some MWEs are headed by verbs which select for possessive NPs, either as complements of the verb (7) or as complements in the PP complement of the verb (8), and impose agreement between the possessor and one of the verb's dependents:

(7) *dan*$_i$ *ṭaman* *yad-o*$_i$ *ba-calaħat*
*Dan buried.*3SM *hand-his in.the-plate*
'Dan refrained from acting.'

(8) *dan*$_i$ *yaca* *mi-kelav*$_i$
Dan came.out from-tools.his
'Dan lost his temper.'

Idioms with "empty slots"

MWEs can include "empty slots", filled by non-idiomatic and unrestricted complements (e.g., *Dana* in (9)).

(9) *dan heʕemid et dana*$_i$ *ʕal ṭaʕut-a*$_i$
Dan made.stand ACC Dana on mistake-her
'Dan proved Dana wrong.'

3.2 The Challenges

The occurrence of verbal MWEs poses a number of challenges to any linguistic theory. Following are a number of challenges which we observe in our data and which we account for in our grammar.

The verbs which head most verbal MWEs play a dual function in language as both literal and idiomatic expressions. One challenge is to capture the commonalities of the different instantiations, while accounting for their differences. As an example, consider the following sentences illustrating a literal and an idiomatic *hoci* ('*take.out*').

(10) a. *dan hoci et ha-sefer (me-ha-argaz)*
Dan took.out ACC the-book (from-the-box)
'Dan took the book out (of the box).'

b. *dan hoci et dana*$_i$ *me-ha-kelim / mi-keleiha*$_i$
Dan took.out ACC Dana from-the-tools / from-tools.her
'Dan made Dana lose her temper.'

Most of the characteristics of the literal and idiomatic instantiations of the verb *hoci* ('*take.out*') are shared. The verb semantically selects two complements, *Theme* and *Source*, which are realized as NP and PP, respectively, with the PP headed by the preposition *me-* ('*from*'). Moreover, the syntactic structure of the two instantiations is identical.

The two senses diverge in a number of ways. As expected, the idiomatic sense is more restrictive in terms of its selectional restrictions. The *Source* argument can only be realized by an NP headed by the idiomatic plural definite noun *ha-kelim* ('*the tools*'). Moreover, the *Source* NP can optionally appear with a possessor

suffix, provided that it is co-indexed with the *Theme* argument of the verb. Another difference is that the *Source* argument is obligatory in the idiomatic sense, and optional in the literal one. Any divergence from these restrictions eliminates the idiomatic reading.

While MWEs are quite specific with respect to their lexical selection, in some cases, they do allow for some flexibility. Consider, for example, the plural subject counterpart of (7):

(11) a. *ha-anaʃim$_i$ ṭamnu yad-am$_i$ ba-calaħat*
*the-people buried.*3P *hand**.S**-their in.the-plate**.S***

b. *ha-anaʃim$_i$ ṭamnu yadei-hem$_i$ ba-calaħat*
*the-people buried.*3P *hand**.P**-their in.the-plate**.S***
'The people refrained from acting.'

(12) *ha-anaʃim$_i$ ṭamnu yadei-hem$_i$ ba-calaħot*
*the-people buried.*3P *hand**.P**-their in.the-plate**.P***
'The people buried their hands in the plates.' (only odd literal)

With the MWE *ṭaman yad-o ba-calaħat* ('*buried his hand in the plate*'), plural subjects can either bury their singular *hand* (11a) or plural *hands* (11b) in the (singular) *plate*. Neverthless, once *plate* becomes plural (12), the idiomatic reading is lost. These constraints, of course, are expression-specific and need to be specified in the lexicon.

A different case of constrained flexibility involves internal modification. Internal modifiers can be adverbs, which, in MH, can intervene between the verb and its complement (e.g., (13)).

(13) *dan yarad* ***ba-sof*** *me-haʕec*
Dan went.down at.the-end from-the.tree
'Finally Dan conceded.'

Alternatively, internal modifiers can be adjectives which syntactically modify one of the complements, as in (14) and (15).

(14) *ha-cibur nafal ba-paħ* ***ha-pirsumi***
the-public fell in.the-bin the-advertising
'The public was tricked by advertisement.'

(15) *ha-irgunim ha-lahaṭabim mehadqim et ha-ħagora*
the-organizations the-LGBT are tightening ACC
ha-vruda *ʃelahem*
the-belt the-pink their
'The LGBT organizations are tightening their pink belt.'[1]

[1]This is an attested MH counterpart to Manfred Sailer's (p.c.) example: *They had to tighten their Gucci belts*.

Note that in all three cases the modifier is optional. Nevertheless, its occurrence inside an idiomatic verb phrase rules out the possibility of analyzing idioms under a 'word with spaces' account.

A final challenge is posed by non-local selection phenomana, of which there are two types: In the case of PP complements, such as that in (10b), there is a chain of lexical selection, where a verb selects for a PP with a particular prepositional head, which in turn selects for an NP with a particular nominal head; Additional non-local constraints are imposed in the case of possessive idioms, which require the obligatory co-indexation between possessors and arguments. For example, in (10b) the possessor of the NP complement in the *Source* PP *mi-keleiha* ('*from-her.tools*') must be co-indexed with the *Theme* NP *Dana*. Consequently, in order for this relation to hold, the index of a possessor within an NP must be "visible" at the level of the PP of which the NP is a complement.

4 The incorporation of MWEs into the grammar

4.1 HeGram

Our proposed analysis is cast in the context of HeGram (Herzig Sheinfux et al., 2015), a deep linguistic processing grammar of Modern Hebrew, which is based on a starter grammar created with the Lingo Grammar Matrix customization system (Bender et al., 2002) and implemented in the LKB (Copestake, 2002) and ACE systems. Morphology is handled outside the grammar, as the lexicon is comprised of automatically analyzed forms.

HeGram currently covers a variety of phenomena, including case marking, subject-verb and noun-adjective agreement, SVO and V2 word order, relatively free complement order, multiple subcategorization frames, selectional restrictions of verbs on their PP complements, topicalization, wh-questions, passive and unaccusative verbs, control verbs, raising verbs, and the copular construction (including zero copula). HeGram is developed in parallel with AraGram (see Arad Greshler et al., 2015), a grammar of Modern Standard Arabic.

The architecture of HeGram embodies significant changes to the way argument structure is standardly viewed in HPSG. The main one is that it distinguishes between semantic selection and syntactic selection, and provides a way of stating constraints regarding each level separately. Moreover, one lexical entry can account for multiple subcategorization frames, including argument optionality and the realization of arguments with different syntactic phrase types (e.g., *want food* vs. *want to eat*). This involves the distribution of valence features across ten categories.[2] Each valence category is characterized in terms of its semantic role, as well as the types of syntactic phrases which can realize it (referred to as *syntactic realization classes*). Consequently, the semantic relations denoted by predicates

[2]Our restructuring of the VALENCE complex is inspired by Haugereid's packed argument frames (Haugereid, 2012).

consist of coherent argument roles, which are consistent across all predicates in the language.

Table 1 presents the ten valence categories used in HeGram, along with the corresponding semantic roles and syntactic realization phrases.[3] For example, Arg2 corresponds to the *Theme* semantic role, and can be realized in MH as an NP, an infinitive VP, a CP or a PP.

Label	**Semantic Selection**	**Syntactic Realization**
Arg1	Actor, Perceiver, Causer	NP, PP
Arg2	Theme	NP, VP_{inf}, CP, PP
Arg3	Affectee, Benefactive, Malfactive , Recipient	NP, PP
Arg4	Attribute	AdjP, AdvP, PP, NP, $VP_{beinoni}$
Arg5	Source	PP
Arg6	Goal	PP
Arg7	Location	PP, AdvP
Arg8	Topic of Communication	PP
Arg9	Instrument	PP
Arg10	Comitative	PP

Table 1: Semantic roles and realization classes in HeGram

Each predicative lexical type in our grammars inherits from types which specify the possible semantic roles of its dependents and their possible syntactic realizations. As an example, consider the lexical type which licenses the (literal) MH verb *hoci* (‘*took out*’).

(16) MH *hoci* (‘*took out*’):

```
arg12-125_n_p := arg1_n & arg2_n & arg5_p &
      [ SYNSEM.LOCAL.CAT.VAL.R-FRAME arg12-125 ].
```

The verb semantically selects three arguments: an *Actor* (arg1), a *Theme* (arg2), and a *Source* (arg5). Moreover, it requires that its *Actor* and *Theme* roles be syntactically realized, yet allows for the omission of the *Source*. This is captured by the value of its lexical type’s R(EALIZATION)-FRAME feature, *arg12-125*, which lists the different realization frames in which the verb can appear, separated by dashes; *arg12* is a transitive syntactic frame and *arg125* represents the realization of all three semantic arguments.

The syntactic realization of the semantic arguments is defined via inheritance. The lexical type in (16) inherits from three subtypes, each pertaining to one of its semantic arguments, and each determining the syntactic category of the phrases which realize that semantic role (noun, noun, and preposition, respectively). The

[3]This architecture is similar in spirit to work done on Polish by Przepiórkowski et al. (2014).

name of this type (i.e., *arg12-125_n_p*) reflects the different realization frames, as well as the syntactic category of its dependents (since Arg1 is always realized as an NP, its syntactic realization is omitted from the name of the type).

The association between semantic roles and syntactic phrases is based on corpus investigation of MH, which included at least 100 randomly selected examples of sentences containing each of the 50 most frequent verb lemmas in the 60-million token WaCky corpus of Modern Hebrew (Baroni et al., 2009). Whereas the semantic classes are expected to be more or less universal, some language-specific differences are expected in the syntactic realizations. Corpus investigations on Modern Standard Arabic in the context of the development of AraGram confirmed these expectations (for more elaboration, see Arad Greshler et al., 2015).

4.2 Verbal MWEs in HeGram

The example sentence in (10b) repeated here as (17), poses most of the challenges described above.

(17) *dan hoci et dana$_i$ mi-keleiha$_i$*
Dan took.out ACC Dana from-tools.her
'Dan made Dana lose her temper.'

It is an "empty slot" MWE, with an idiomatic PP complement with a possessed NP whose possessor is obligatorily co-indexed with the literal NP complement filling the "slot". In what follows we use this example to illustrate our approach to the analysis of verbal MWEs.

4.2.1 Verbs with dual instantiations and their selectional restrictions

Verbs which can head VP MWEs can also occur in "standard" VP constructions. The degree of overlap between the behavior of the verb in its standard guise and in its idiomatic role is mostly verb-specific. Nevertheless, regardless of the degree, our lexical inheritance hierarchy enables us to distinguish between shared properties and those which differ in the two instantiations.

The subcategorization properties of the literal instantiation of *hoci* ('*take.out*') are expressed in its VALENCE (see Figure 1), which includes the three relevant arguments: DEP1 (*Actor*), DEP2 (*Theme*) and DEP5 (*Source*) (the rest are suppressed for space reasons). Moreover, the value of its R(EALIZATION)-FRAME is *arg12-125*, indicating that while the *Actor* and *Theme* arguments are obligatory, the *Source* argument is optional. These characteristics are all a result of the fact that the literal instantiation is an instance of the type *arg12-125_n_p_past_le* (for further elaboration, see (16) in section 4.1).

The idiomatic instantiation of *hoci* ('*take.out*') is an instance of a distinct, yet very similar type, *arg125_n_pi_xarg25_past_le*. Its syntactic selection properties are identical to its literal counterpart. However, in contrast to the literal *hoci* ('*take.out*'), the idiomatic one has a different R(EALIZATION)-FRAME value,

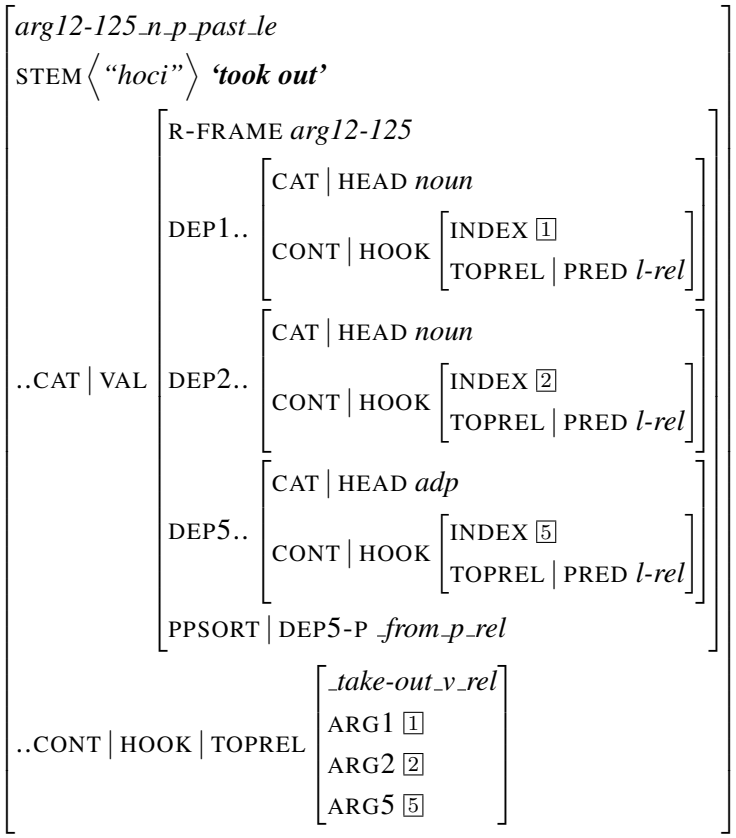

Figure 1: The literal *hoci* ('*take.out*')

arg125, indicating that all arguments are obligatory (an abbreviated description is shown in Figure 5).

The main distinction between the two variants is in their semantic content, and semantic selection. In order to distinguish between literal and idiomatic words, and to control their distribution, semantic relations are divided into *l(iteral)-rel*s and *i(diomatic)-rel*s (Copestake, 1994; Sag et al., 2002; Kay & Sag, 2012, among others). Consequently, the semantic relation denoted by the literal verb, *_take-out_v_rel*, is a subtype of *l-rel*, and the one denoted by the idiomatic verb, *_i-take_out-cause_lose_v_rel*, is a subtype of *i-rel*.[4] The TOPREL feature is a pointer to the main semantic relation (in RELS) denoted by a lexeme (for more about TOPREL, see the following section).

Selectional restrictions of verbs in HeGram are specified in the respective DEP feature. The literal verb requires the *Source* (Arg5) PP to be headed by the specific preposition *me* (or *mi*). This requirement is defined in the PPSORT feature

[4]Please note that our analysis does not distinguish decomposable from non-decomposable idioms, as we only have a relatively superficial semantic representation of MWEs. All the idiomatic components of an MWE have separate idiomatic entries in the lexicon, which include an approximated paraphrase of their idiomatic meaning.

complex, under DEP5-P, whose value is set to *_from_p_rel*. Naturally, its idiomatic counterpart is more selective. It requires that its *Source* PP be headed by a specific idiomatic *_from_tools_ip_rel* relation. This selective preposition *me* (or *mi*), in turn, selects for an NP with an idiomatic *_i-tools-temper_n_rel* relation. This notwithstanding, the *Theme* argument of the idiomatic *hoci* ('*take.out*') is an "open slot" and can be filled by any NP complement, provided that it is not idiomatic (i.e., has an *l-rel*), and the same applies to the NP *Actor* in subject position.

4.2.2 A chain of lexical selection

Indirect non-local lexical selection such as the one described above, where a verb selects for a preposition which selects for a noun, forms a type of a chain, where heads of phrases select heads of other phrases. This mechanism is supported by the TOPREL feature, an independently motivated feature in HeGram, which identifies the main semantic relation denoted by a lexeme. Idiomatic selectors target this feature, which percolates from head daughter to the "mother" phrase.[5]

The AVM in Figure 2 illustrates the selection chain which characterizes the idiomatic form of the preposition *mi*, which is selected by the idiomatic *hoci* ('*take.out*'). The co-indexation of XARG will become relevant in the next section.

$$\left[\begin{array}{l} \textit{poss-raise-adposition-lex-np-i} \\ \text{STEM}\left\langle \textit{"mi"} \right\rangle \\ \ldots \left[\begin{array}{l} \text{CONT}\left[\begin{array}{l} \text{TOPREL } \textit{\_from\_tools\_ip\_rel} \\ \text{HOOK} \mid \text{XARG } \boxed{1} \end{array}\right] \\ \text{CAT} \mid \text{VAL} \mid \text{DEP2} \mid \ldots \left[\begin{array}{l} \text{XARG } \boxed{1} \\ \text{TOPREL} \mid \text{PRED } \textit{\_i-tools-temper} \end{array}\right] \end{array}\right] \end{array}\right]$$

Figure 2: The idiomatic form of the preposition *mi*

Admittedly, using a selection chain to ensure that idiomatic verbs that select specific PPs only combine with the correct complements introduces some redundancy to the lexicon. However, this solution does solve the non-local selection problem.[6]

Semantic selection via the TOPREL of dependents is instrumental in accounting for cases of internal modification (e.g., (14) and (15)). The TOPREL of a phrase is identical to the main relation of the head, regardless of whether it is modified or not. This is illustrated in Figure 3.

[5] Kay & Sag (2012) suggest a similar feature, LEXICAL-ID (LID).

[6] Although there is no independent evidence for the existence of an idiomatic form of prepositions, usage patterns diverge: the one used in an MWE selects for a specific complement, whereas the standard preposition does not.

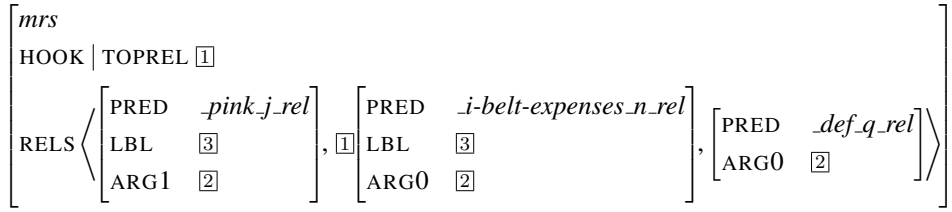

Figure 3: The MRS of the idiomatic *the pink belt*

4.2.3 Possessive idioms

Possessive idioms present a second type of non-local selection. In such idioms the possessor of NP dependents, or NP complements of PP dependents, must be co-indexed with the verb's subject or complement (depending on the MWE). This requires that the index of the possessor be "visible" at the NP, and even PP level. The feature which projects the lower possessor to this higher level is the XARG feature (Kay & Sag, 2012; Bond et al., 2015).

The account of possessive idioms builds on our analysis of possessive nouns. Consider as an example the (literal) noun *keleiha* ('*her.tools*'), shown in Figure 4. The agreement property of this particular noun is 3rd-person-plural-masculine, and this is defined in its PNG feature, which is structure-shared with CNCRD (tagged 2). Its possessor is realized by the 3rd-person-single-feminine pronominal clitic *ha*. This information is represented in the semantic XARG feature. Finally, the semantic relations denoted by the NP include *tool-rel*, which is the main relation (structure-shared with TOPREL), and *poss-rel*, which identifies the possessor (1) and possessed (3).

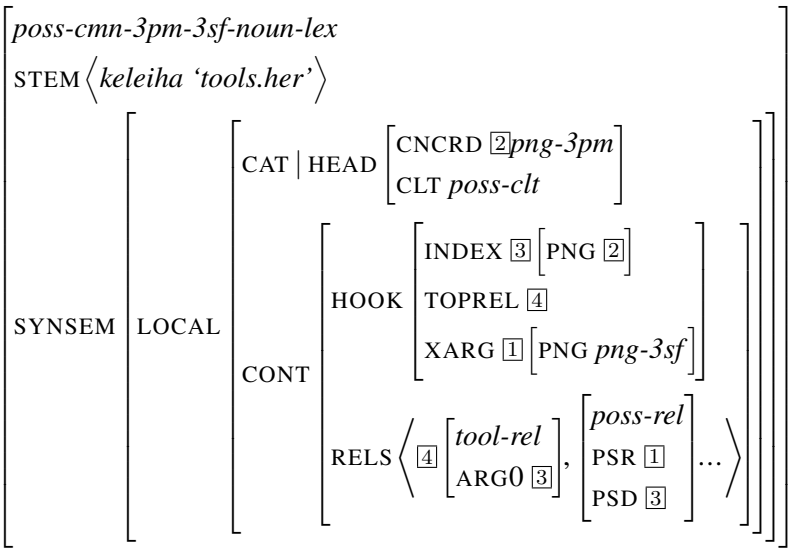

Figure 4: A possessive noun

The XARG feature exposes the INDEX features of the "inner" possessor at the NP level, and thus makes it visible to an idiomatic selector. When a possessed NP is a complement of a preposition, the XARG features of its possessor percolate to the PP level. This is illustrated in the AVM describing the idiomatic form of the preposition *mi* in Figure 2.

Different idiomatic MWEs have different patterns of co-indexed possession, so the exact structure-sharing pattern is lexically specified per verb type.[7] In (7) the subject must be co-indexed with the possessor of the NP *Theme* complement (Arg2), while in (8) it must be co-indexed with the possessor of the NP complement inside the PP. In (10b) it is the NP complement which is co-indexed with the possessor of the NP complement inside the PP. Each one of these co-indexation relations between arguments is represented in the grammar by a lexical type, from which the relevant lexemes inherit. For example, the idiomatic *hoci* ('*take.out*') is an instance of a general lexical type *arg125_n_pi_xarg25_past_le*, which requires the co-indexation between the Arg2 complement and the possessor within the Arg5 argument.

The different components of the analysis of the MWE in the example sentence in (17) are shown together in Figure 5.

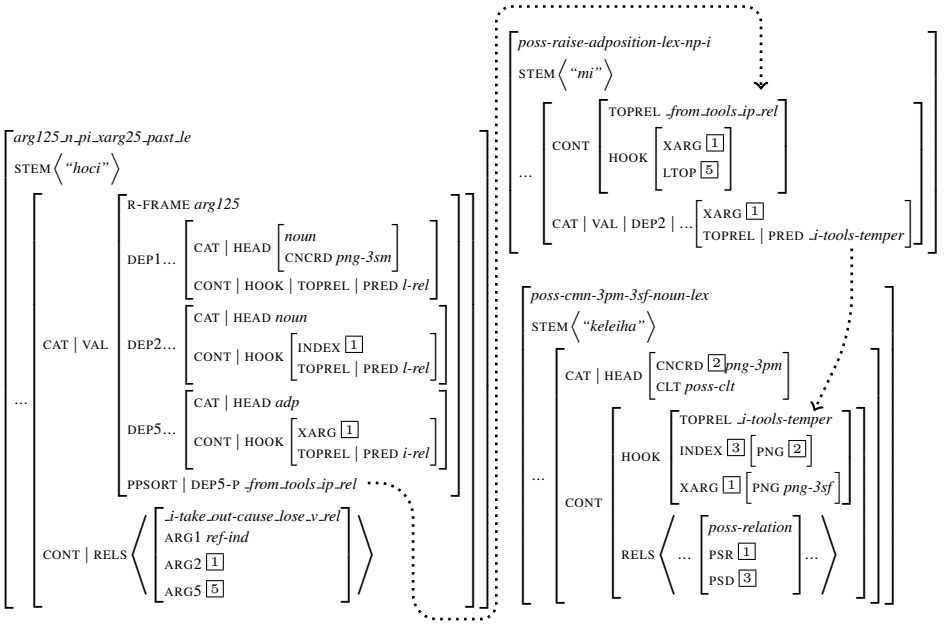

Figure 5: The selection chain in possessive idioms

[7] Bond et al. (2015) introduce an extra *identity* relation to the semantics of idiomatic verbs, which identifies the possessor and the index of the appropriate argument. This solution requires post-processing with MRS rewriting rules, which are not needed in our analysis.

5 Conclusion

We presented an account of Hebrew verbal MWEs in an existing HPSG grammar. The analysis covers a multitude of MWE types, including challenging phenomena such as (possessive) co-indexation and internal modification. Moreover, the grammar now produces two analyses for most MWEs, corresponding to their idiomatic and literal readings.

MWEs are challenging because they blur the traditional distinction between the lexicon and the grammar. In our analysis, support of MWEs required minimal changes to the grammar: most crucially, the division of *rels* to either *i-rels* or *l-rels*. All other changes involve the lexicon: we make extensive use of HPSG's type hierarchies in order to state generalizations over lexical types.

The main contribution of this work is of course the extension of the coverage of HeGram to verbal MWEs. To the best of our knowledge, this is the first account of Hebrew MWEs in a linguistically-motivated grammar. Moreover, the mechanisms that we advocate are fully applicable to other languages, and can be incorporated into existing HPSG grammars with minimal effort.

In the future we intend to explore syntactic constraints on MWEs and account for their full behavior. This includes phenomena such as topicalization, wh-questions, coordination, etc.

References

Arad Greshler, Tali, Livnat Herzig Sheinfux, Nurit Melnik & Shuly Wintner. 2015. Development of maximally reusable grammars: Parallel development of Hebrew and Arabic grammars. In Stefan Müller (ed.), *Proceedings of the 22nd international conference on Head-Driven Phrase Structure Grammar, Singapore*, 27–40. Stanford, CA: CSLI Publications.

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., Dan Flickinger & Stephan Oepen. 2002. The grammar matrix: an open-source starter-kit for the rapid development of cross-linguistically consistent broad-coverage precision grammars. In *Coling-02 workshop on grammar engineering and evaluation*, 1–7. Morristown, NJ, USA: Association for Computational Linguistics. doi:http://dx.doi.org/10.3115/1118783.1118785.

Bond, Francis, Jia Qian Ho & Daniel Flickinger. 2015. Feeling our way to an analysis of English possessed idioms. In Stefan Müller (ed.), *Proceedings of the 22nd international conference on Head-Driven Phrase Structure Grammar, Singapore*, 61–75. Stanford, CA: CSLI Publications.

Copestake, Ann. 1994. Representing idioms. Paper presented at The 20th International Conference on HPSG, Copenhagen.

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

Haugereid, Petter. 2012. A grammar design accommodating packed argument frame information on verbs. *International Journal of Asian Language Processing* 22(3). 87–106.

Herzig Sheinfux, Livnat, Nurit Melnik & Shuly Wintner. 2015. Representing argument structure in computational grammars. Submitted.

Kay, Paul & Ivan A. Sag. 2012. A lexical theory of phrasal idioms. Unpublished manuscript, Stanford University.

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

Przepiórkowski, Adam, Elżbieta Hajnicz, Agnieszka Patejuk, Marcin Woliński, Filip Skwarski & Marek Świdziński. 2014. Walenty: Towards a comprehensive valence dictionary of Polish. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk & Stelios Piperidis (eds.), *Proceedings of the ninth international Conference on Language Resources and Evaluation, LREC 2014*, 2785–2792. Reykjavik, Iceland: ELRA. http://www.lrec-conf.org/proceedings/lrec2014/index.html.

Sag, Ivan A., Timothy Baldwin, Francis Bond, Ann Copestake & Dan Flickinger. 2002. Multiword expressions: A pain in the neck for NLP. In *Proceedings of the third international conference on intelligent text processing and computational linguistics (cicling 2002)*, 1–15. Mexico City, Mexico.