Noname manuscript No.
(will be inserted by the editor)

A Hebrew Verb–Complement Dictionary

Hanna Fadida · Alon Itai · Shuly Wintner

Received: November 20, 2013/ Accepted: November 20, 2013

Abstract We present a verb-complement dictionary of Modern Hebrew, automatically extracted from text corpora. Carefully examining a large set of examples, we defined ten types of verb complements that cover the vast majority of the occurrences of verb complements in the corpora. We explored several collocation measures as indicators of the strength of the association between the verb and its complement. We then used these measures to automatically extract verb complements from corpora. The result is a wide-coverage, accurate dictionary that lists not only the likely complements for each verb, but also the likelihood of each complement. We evaluated the quality of the extracted dictionary both intrinsically and extrinsically. Intrinsically, we showed high precision and recall on randomly (but systematically) selected verbs. Extrinsically, we showed that using the extracted information is beneficial for two applications, PP attachment disambiguation and Arabic-to-Hebrew machine translation.

Keywords Verb subcategorization · Hebrew · Lexicography

1 Introduction

The core of syntactic structure, according to most contemporary syntactic theories and for most languages, revolves around verbs and their complements. The relations between verbs and their complements are syntactic in nature, but they reflect semantic relations that hold between the action or state denoted

H. Fadida
Department of Computer Science, Technion. E-mail: fadida@cs.technion.ac.il

A. Itai
Department of Computer Science, Technion. E-mail: itai@cs.technion.ac.il

S. Wintner
Department of Computer Science, University of Haifa. E-mail: shuly@cs.haifa.ac.il

by the verb, and the entities that participate in this action. Several linguistic theories map verb-complement constructions to such semantic relations: understanding the syntactic relations between verbs and their complements is thus instrumental for understanding the meaning of natural language sentences. Correctly identifying verb complements in naturally-occurring texts is therefore important both theoretically, for linguistic investigations, and practically, for natural language processing (NLP) applications. A crucial resource needed for this task is a dictionary listing the number and types of complements that are most likely to occur with each verb, ideally with some statistical measure of the strength of the relation between the verb and each of its complements.

The importance of such a resource must not be underestimated. Briscoe and Carroll (1993) observed that half of the errors made by a parser tested on unseen data were due to inaccurate subcategorization information in a manually-compiled dictionary. Briscoe and Carroll (1997) then described a novel way for extracting such a dictionary from corpora, and showed a small improvement in the accuracy of parsing. Carroll et al. (1998) repeated the same experiment, on a much larger scale, and demonstrated that the use of subcategorization frames can improve both precision and recall in the task of inducing bracketing on sentences, whereas for the task of assigning grammatical relations, precision improved by 9 percentage points (from 79% to 88%), at the cost of only half a point drop in recall. A similar improvement is also observed when *dependency* parsers are concerned (Zeman, 2002).

This problem holds for modern parsers, too. Kummerfeld et al. (2012), experimenting with a large number of English parsers, demonstrated that prepositional phrase (PP) attachment "is the largest contributor to errors, across all parsers". Kummerfeld et al. (2013) conducted a similar investigation of Chinese parsers, and revealed that the error types were quite different from the errors made by the English parsers. We experimented with the state-of-the-art Hebrew parser of Goldberg (2011); this is a dependency parser based on the EasyFirst parsing algorithm (Goldberg and Elhadad, 2010), enhanced with morphology-based features which greatly improve its accuracy. It was trained on the Hebrew dependency treebank, which was automatically converted from a manually constructed constituent-structure treebank (Sima'an et al., 2001; Guthmann et al., 2009). Testing on 25 newspaper sentences containing 51 verb occurrences, the F-score of the parser on identifying the correct verb complements was below 0.85. Indeed, Goldberg (2011) specifically mentions (page 93) that "nodes... which are assigned an incorrect head are predominantly PPs" and that "All parsers have a hard time dealing with PP attachment."

Other applications that have been shown to benefit from information on subcategorization include automatic classification of verbs to *semantic* classes (Schulte im Walde and Brew, 2002), information extraction (Surdeanu et al., 2003), machine translation (Hajič et al., 2004) and more. Knowledge of verb subcategorization frames also affects human sentence processing (Garnsey et al., 1997), and a subcategorization dictionary is therefore highly useful for psycholinguistic experimentation (Lapata et al., 2001; Baldewein, 2004).

We introduce the first automatically-created verb-complement dictionary of Modern Hebrew, extracted from large text corpora. Using available resources, we morphologically analyzed and syntactically parsed the corpora. We employed standard collocation measures to assess the degree to which potential complements tend to combine with each verb, focusing on a small set of potential complement types, which covers the vast majority of complement instances in the corpora. More importantly, we did not attempt to construct full subcategorization frames; rather, we viewed each complement type in isolation, and determined its likelihood to combine with the verb. We also favored high-frequency complements that can be extracted with high precision, at the expense of potentially lower recall.

The result is a wide-coverage dictionary of almost 3,000 verb lemmas, listing more than 6,500 verb-complement pairs, each with a statistically-derived score. We evaluated the quality of the dictionary both intrinsically and extrinsically. Intrinsically, we manually constructed a set of representative verbs and their canonical complements; our automatically-extracted dictionary achieves high precision and recall on this test set. Extrinsically, we incorporated linguistic knowledge derived from our verb-complement dictionary in two computational tasks: reducing the ambiguity of PP-attachment and translating from Arabic to Hebrew. We demonstrated that knowledge derived from our dictionary is instrumental in significantly improving the accuracy of these two tasks. The contribution of this work is thus a digital, freely-available, wide-coverage and accurate verb-complement dictionary of Hebrew.

After reviewing related work in the next section, we outline the structure of verb–complement constructions in Hebrew in Section 3. We describe the research methodology, as well as the required language resources, in Section 4. The results are discussed in Section 5, followed by an evaluation of their quality in Section 6. We conclude with a discussion and suggestions for future research.

2 Related Work

The *verb subcategorization frame* (Chomsky, 1965) determines the number and type of verb complements. Clearly, subcategorization frames are part of the syntax-semantic interface (Levin, 1993): they determine the syntactic structure of verb phrases, but they are typically shaped by the verb's argument structure, reflecting the verb's meaning. Indeed, subcategorization frames can be used to categorize verbs semantically (Sun et al., 2008a,b; Sun and Korhonen, 2009), and semantic knowledge can improve the extraction of subcategorization frames (Korhonen, 2000, 2002b; Korhonen and Preiss, 2003).

Syntactic theories tend to distinguish between *complements* and *adjuncts*. The former are phrases that are syntactically required by the verb; without them, verb phrases are incomplete. The latter are more general verb modifiers, that are typically optional, and may occur more than once with a given verb instance. Generally, there is a mapping of complements to verb arguments: the meanings of verb complements are considered crucial to understanding the meaning of a predicate, whereas other verb modifiers are less central semantically.

Clearly, the distinction between complements and adjuncts is vague (Huddleston and Pullum, 2002, Chapter 4, Section 1.2). There is overlap in the syntactic realization of both types of verb modifiers, and both are frequently realized as prepositional phrases. For example, the prepositional phrase in *I slept on the floor* is considered an adjunct, whereas in *I rely on you* it is considered a complement. While in some cases there is a clear distinction between complements and adjuncts, the degree of relationship between a verb and its modifiers is often gradual, and no clear-cut line distinguishes complements from adjuncts in many cases. Having said that, distnguishing between complements and adjuncts has been found useful for applications such as prepositional phrase attachment (Merlo and Ferrer, 2006), see Section 6.3.1.

One of the first works to address automatic acquisition of subcategorization frames introduces the *Lerner* system (Brent, 1991, 1993). Lerner uses large unannotated corpora and no dictionary, identifying English verbs through simple morphological cues. It focuses on six subcategorization frames using a simple, finite-state grammar of English. Complements include direct objects, clauses and infinitival verb phrases, but no prepositional phrases. Hypothesis testing, with *binomial frequency* data as the collocation measure, is used to determine whether a candidate is indeed a member of the subcategorization frame of a verb. Different settings of the parameters can yield perfect recall at 50% precision, and vice versa, on a set of 193 manually selected verbs; for clauses and infinitival verb phrases, accuracy is much higher.

The Lerner system was a pioneering work, and we follow its spirit by using hypothesis testing as the main tool for determining the statistical validity of a decision on a verb–complement pair. We extended the scope of our study to more (ten) complement types, particularly prepositional phrases, and we experimented with several collocation measures. We also took advantage of the availability of morphologically analyzed and syntactically parsed corpora of Hebrew.

With the proliferation of language resources, contemporary approaches to subcategorization frame induction tend to use as many data sources as are available. For example, Sarkar and Zeman (2000), working on Czech, used the manually-constructed Prague Dependency Treebank for this task. Since word order in Czech is much freer than in English, complements are sought around the verb, and not just immediately following it. Induction was done iteratively: first, a wide subcategorization frame was assumed for each verb, with all elements that are dependent on the verb in the training material as candidate complements. Then, iteratively, the wide frame was replaced by proper subsets thereof, based on statistical considerations. Sarkar and Zeman (2000) employed three hypothesis testing measures, likelihood ratio, t-test and binomial frequency, yielding three different subcategorization dictionaries. For each measure, candidate complements that did not pass the test were removed from the frame, until convergence.

Again, our method is very similar, but unlike Czech, Hebrew only has a very small treebank (see Section 4.2; the Hebrew treebank includes some 6,000 sentences, compared with over 115,000 for Czech). We also experimented with more than one collocation measure. Unlike Sarkar and Zeman (2000), we did not begin with the widest frame observed in a treebank (since our treebank is so small). Rather, we treated each potential complement independently of other complements, and only attempted to determine the strength of its own attachment to the verb. This approach is similar to the one used by Dębowski (2009), who also collected single possible complements from a (Polish) corpus; however, Dębowski (2009) also defined a second *filtering* step in which full frames are constructed.

When a treebank is unavailable, automatically parsed data can be used instead. This is the main resource used by Briscoe and Carroll (1997), who inspired many subsequent works for Chinese (Han et al., 2004), English (Li and Brew, 2005) French (Chesley and Salmon-alt, 2006), Italian (Ienco et al., 2008), Portuguese (Zanette et al., 2012), and other languages. Again, the approach is to first assume a wide frame for each verb, sometimes based on existing, manually-created lexicons, and sometimes on assumed frames observed in the corpus. Then, a refinement step constrains the set of possible complements for each verb, based on the strength of the association between the verb and the candidate complement. Briscoe and Carroll (1997) used binomial frequency; Korhonen (2002a) compared three different collocation measures (binomial frequency, log-likelihood ratio and raw frequency), and showed that raw frequency produces the best results.

Korhonen et al. (2006) extended the work of Korhonen (2002a) to a full subcategorization dictionary of English verbs, called *Valex*. The dictionary includes not just the complements of each verb, but also the likelihood of their realization, as obtained from a large corpus. Each verb was associated with a subset of the full set of 163 subcategorization frame types, with an average of 33 frames per verb. A similar dictionary for French, *LexSchem*, was compiled by Messiant et al. (2008).

More recently, approaches based on topic modeling have been used to automatically induce lexical semantic information, in particular selectional preferences (Ritter et al., 2010; Ó Séaghdha, 2010). These are completely unsupervised methods, which may be useful for the task at hand, but are beyond the scope of this paper.

The only available information source on Hebrew subcategorization so far is the verb dictionary of Stern (1994), which includes 833 verbs and 1,430 subcategorization frames. While the information was manually collected from corpora of written and spoken language, news articles, literary texts, etc., being a manually compiled dictionary its coverage is obviously limited. For comparison, the MILA Hebrew computational lexicon (Itai and Wintner, 2008) contains almost 5,000 verbs. The advantages of automatic extraction of verb subcategorization frames are obvious: not only does the method provide better coverage (at the expense of reduced precision, of course), but it also facilitates adaptation of the extracted dictionary to a specific genre, domain, register, etc.

Furthermore, our approach provides probabilistic estimates of the likelihood of various complements, which may be more useful than the deterministic information listed in Stern (1994).

3 Hebrew Verb Subcategorization

In light of the discussion in Section 2, we refrained from making a clear distinction between complements and adjuncts in this work. Rather, we used a working definition whereby a particular type of modifier was considered part of the subcategorization frame of a verb if it frequently occured with the verb in text corpora. According to this definition, both argument-denoting complements and frequent adjuncts may be listed along with each verb. We did not include subjects in this investigation, as they are trivially required by all verbs.

Verbs can be ambiguous, with more than one subcategorization frame, often reflecting semantic variability. For example, when the verb[1] *nicx* takes a direct object, its meaning is "win"; when its complement is a prepositional phrase headed by *yl* "on", its meaning is "conduct, direct". In this work we made no attempt to distinguish between the various meanings of ambiguous verbs. If our system inferred that the verb *nicx* is strongly associated both with a prepositional phrase headed by *yl* "on" and with a direct object, then both were included in the complement dictionary for the verb.

Syntactically, we addressed four types of complements in this work: noun phrases, prepositional phrases, clauses and infinitival verb phrases. We now exemplify these complements and the way they are realized in text corpora. We discuss the various types of complements (Section 3.1), the number of complements (Sections 3.2, 3.3), and their order (Section 3.4). Finally, Section 3.5 provides some numerical data on the actual realization of verb complements in our corpora.

3.1 Types of complements

In what follows, we use *slanted font* for natural language examples, **bold face** for the target verb and <u>underline</u> for the target complement.

3.1.1 Noun phrases

Direct objects, clearly verb complements, are realized as noun phrases in Hebrew:

[1] For readability, we use a straight-forward, one-to-one transliteration of Hebrew in which the letters, in traditional Hebrew alphabetic order, are represented by *abgdhwzxTiklmnsypcqršt*.

(1) *h-xil* ***qibl*** *Tipwl rpwai* *bmqwm*
the+soldier received treatment medical in+the+place
"The soldier was treated on location"

When the direct object is a *definite* noun phrase (and only when it is definite), it must be introduced by an accusative marker, *at*, which we gloss as "ACC". The accusative marker behaves like a typical preposition; in particular, it can combine with a pronominal enclitic:

(2) *hšwpTt* ***zimnh*** *at iw"r hhstdrwt*
the+judge summoned ACC chairperson the+union
lhmšk *hdiwn*
to+continuation-of the+proceedings
"The judge summoned the Union chairperson for the pursuant proceedings"

(3) *hšwpTt* ***zimnh*** *awtw* *lhmšk*
the+judge summoned ACC+3pSgMasc to+continuation-of
hdiwn
the+proceedings
"The judge summoned him for the pursuant proceedings"

3.1.2 Prepositional phrases

Indirect objects, as well as several types of adverbials, are realized as prepositional phrases.

(4) *ynn* *hapr* *hwwlqni* ***mny*** *m-440 Tiswt*
cloud-of the+ash the+volcanic prevented from+440 flights
lhmria *mairlnd*
to-take-off from+Ireland
"The volcanic ash cloud prevented 440 Ireland flights from taking off"

(5) *ywbdi* *mšrd* *hpnim* ***ptxw*** *byicwmim*
employees-of ministry-of the+interior opened in+strike
"Ministry of the Interior employees started a strike"

(6) ***grti*** *bdirh škwrh*
I-lived in+apartment rented
"I lived in a rented apartment"

(7) *ild* *qTn* ***htprc*** *lxdr*
child small burst to+the+room
"A small child burst into the room"

3.1.3 Clauses

Several verbs take clausal complements. These are often introduced by the subordinating conjunctions *ki* "that" and *š* "that":

(8) *Thrn* ***hwdiyh*** *ki tpsiq lmkwr npT*
Teheran announced that will-stop to-sell oil
"Teheran announced it will stop selling oil"

Clausal complements include also quoted speech and clauses introduced by relative/interrogative pronouns and question words.

3.1.4 Infinitival verb phrases

Sometimes, verbs take complements that are realized as verb phrases in the infinitive:

(9) *alpi anšim* ***nalcw*** *lyzwb at hmtxm*
thousands-of people were-forced to-leave the+site
"Thousands of people were forced to leave the site"

3.2 Number of complements

The number of elements on a verb's subcategorization list varies. Some verbs, traditionally known as *intransitive*, take no complements:

(10) ***mt*** *hmlxin whmbqr bnimin brym*
died the+composer and+the+critic Benjamin Baram
"The composer and music critic Benjamin Baram died"

We view such verbs as having *empty* subcategorization lists.

Other verbs take one, two or even three complements, as in the case of *htyrb* "bet":

(11) *bxwdš šybr* ***htyrb*** *masq yl miliwn dwlr*
in+the+month that+passed bet Mask on million dollar
ym ktb hrkb šl wwl strit g'wrnl
with reporter-of vehicle of Wall Street Journal
ki hrkb ica bmwydw hmtwknn
that the+car will-come-out in+its-time the+planned
"Last month Mask bet the Wall Street Journal transportation reporter one million dollars that the car will come out at the expected time"

Sometimes, complements that have the same function can be realized in more than one way. This is common with cognitive verbs, which can have noun phrase, clausal, or verb phrase complements:

(12) *hwa* ***rch*** *lcat aith*
he wanted to-go-out with+her
"He wanted to date her"

(13) *ališy* ***rch*** *mzkrt mqprisin*
Elisha wanted souvenir from+Cyprus
"Elisha wanted a souvenir from Cyprus"

(14) *la* ***rcinw*** *šanšim idagw*
not we-wanted that+people will-worry
"We didn't want people to worry"

Other verbs can only take one or two variants of these complements. For example, *msr* "report" can take a clause or a noun phrase, but not a verb phrase; *hxliT* "decide" can take a clause or a verb phrase, but not a noun phrase. Consequently, we view the three possible complements of cognitive verbs as different, and treat them independently of one another.

3.3 Internal complementation

Certain verbs that are typically intransitive, with empty subcategorization frames, can sometimes be complemented by an object whose meaning is very close to the meaning of the verb (often, a nominalization of the verb itself). Such complements, known as *internal objects*, are not considered part of the verb's subcategorization frame:

(15) *ym išral* ***išn*** *at šntw blilwt*
people Israel sleep ACC its-sleep in+the+nights
"The Israelis are fast asleep"

Similar examples include *rqd Tngw* "dance the tango", *xik xiwk rxb* "smile a wide smile", *mt mwwt šqT* "die a peaceful death", etc.

3.4 Linear precedence inside the Hebrew verb phrase

Hebrew constituent order is relatively free; in particular, verb complements and modifiers, including the subject, can both precede and succeed the verb (Belletti and Shlonsky, 1995):

(16) *hwa* ***hymid*** *at hild yl kisa*
he stood ACC the+child on chair
"He stood the child on a chair"

(17) *hwa* ***hymid*** *yl kisa at hild*
he stood on chair ACC the+child
"He stood the child on a chair"

Having said that, sometimes the order is fixed, or there is a strong tendency towards a particular order. This is the case with idiomatic expressions, which tend to occur in a fixed order:

(18) *hwa* ***hymid*** *at hild* *yl Tywtw*
he stood ACC the+child on his+error
"He corrected the child"

Here, the alternative order would be awkward.

As in other languages (Ross, 1967), a strong tendency towards a particular order can result from heaviness considerations; when one of the complements is a pronoun, for example, it strongly tends to precede a full noun phrase:

(19) *hwa* ***hymid*** *awtw* *yl kisa*
he stood him on chair
"He stood him on a chair"

(20) *hwa* ***hymid*** *yliw* *at hild*
he stood on+it ACC the+child
"He stood the child on it"

Again, the two alternative orders are awkward.

3.5 Realization of verb complements in text corpora

We listed above general, textbook properties of Hebrew verb subcategorization. Often, language use differs significantly from such generalizations. We therefore used text corpora to quantify some of these properties. Specifically, we used a treebank of 5,281 sentences (Sima'an et al., 2001), which include 1,423 verb types and 7,561 verb tokens, to which 9,486 complement tokens are attached (in this section, when the number of complements is mentioned, it refers to tokens, not types). Section 4.2 provides more detail about the treebank. The observations that we drew from the treebank drove the design of our subcategorization frame extraction algorithm (Section 4.1).

We first addressed the distance, in words, between the verb and (the first word of) its complements. Table 1 lists the data; as can be seen, the vast majority (87.46%) of the complements *follow* the verb, and as many as 52.42% of the complements *immediately* follow the verb. Consequently, we focused on complements that occur immediately after the verb in our extraction algorithm.

Next, we addressed the types of complements. The treebank includes 5,983 prepositional phrases that modify verbs, headed by as many as 129 preposition types. Table 2 lists the number of prepositional phrase complements corresponding to the various prepositions in the treebank. The two most frequent prepositions, *b* "in" and *l* "to", account for more than half of the instances, and the top six prepositions account for almost 80% of them. All other prepositions

Distance	Number	%	Distance	Number	%
1	4,973	52.42	-1	11	0.12
2	685	7.22	-2	228	2.40
3	796	8.39	-3	231	2.44
4	580	6.11	-4	148	1.56
5	300	3.16	-5	126	1.33
more than 5	962	10.14	less than -5	446	4.70

Table 1 The distance, in words, between the verb and its complement. **Number** is the number of complement tokens that occur at distance **Distance** from the verb.

account for no more than one percent of the complements each. Consequently, our algorithm focuses only on prepositional phrases headed by the six most frequent prepositions.

Preposition	Number	%
b "in"	2,378	39.75
l "to"	1,161	19.40
m "from"	452	7.55
yl "on"	451	7.54
ym "with"	136	2.27
al "to"	69	1.15

Table 2 The number of prepositional phrase complements headed by various prepositions.

Noun phrases occur as verb complements in the treebank 2,207 times. Of these, 1,141 instances (51.7%) are indefinite, and occur without the accusative marker. Consequently, we considered both bare noun phrases and noun phrases that were introduced by the accusative marker as potential complements.

Since in Hebrew the verb may precede the subject, some noun phrases that immediately follow verbs are in fact subjects. Of the 3,692 verb tokens in the treebank that have a subject, the subject immediately follows the verb in 476 instances (12.89%). In order to filter out some of the noise introduced by subjects that immediately follow their verbs, we considered as complements only noun phrases that do *not* agree with the target verb in number, gender or person. Since in Hebrew the subject must agree with the verb, such candidates are obviously not subjects.

Turning now to clausal complements, we observed that of the 554 instances of such complements in the treebank, 281 (50.72%) were introduced by *š* "that", whereas 273 (49.28%) were introduced by *ki* "that". We therefore addressed both of these subordinating conjunctions in our algorithm.

Table 3 summarizes the distribution of complement types in the treebank, listing only the types of complements we addressed in this work. Of the 9,486 complements in the corpus, the ones we addressed constitute 84.67%.

Finally, we addressed the number of verb complements in the treebank. As is evident from Table 4, the vast majority (93.61%) of the verb occurrences in the treebank have at most one complement. Consequently, we focused on

Type	Number	%
Preposition (top 6)	4,647	48.99
Preposition (other)	1,337	14.09
Noun phrase	2,207	23.27
Infinitival verb phrase	624	6.58
Clause	554	5.84
Other	117	1.23

Table 3 The distribution of complement types in the treebank.

extracting each complement independently of other potential complements of the same verb.

Number of Complements	Number	%
0	2,862	37.85
1	4,216	55.76
2	479	6.34
3	4	0.05

Table 4 The distribution of the number of complements in the treebank.

4 Methodological issues

4.1 Task definition

In light of the observations of Section 3.5, we defined our task as follows. First, we focused on ten types of complements only: (1-6) prepositional phrases with the six most frequent prepositions; (7) noun phrases (with or without the accusative marker); (8) infinitival verb phrases; (9) clauses headed by *š* "that" and *ki* "that"; and (10) the empty subcategorization frame, indicating that the verb phrase does not require a complement. Then, our goal was to associate each verb in a given Hebrew corpus with a measure of the strength of the association between the verb and each of the ten complements. We also defined a threshold such that only complements whose association strength exceeds the threshold were included in the dictionary.

Notice that this definition is somewhat different from the traditional definition of subcategorization frames; in particular, we have no way to distinguish between a subcategorization frame of two complements (e.g., *hsbir* "explain", which subcategorizes for a noun phrase and a prepositional phrase headed by *l* "to") and two single-complement frames for a single verb. In light of Table 4, however, this does not seem to be a major drawback, as very few verbs have more than one complement.

4.2 Resources

We used the MILA corpus of written Modern Hebrew (Itai and Wintner, 2008), consisting of newspaper articles, newswire items and parliament proceedings. The total number of tokens in the corpus is over 40 million, with 1.8 million verb tokens reflecting 4,358 verb lemmas.

The corpus is morphologically analyzed and disambiguated using the MILA tools (Itai and Wintner, 2008). The current disambiguation module does not always fully resolve the ambiguity of some forms. For example, when two analyses differ only in the lemma, they remain ambiguous. This happens quite frequently with verbs, where two different analyses differ only in the pattern (*binyan*), as in *xšbh*, which can mean either "she thought" in one pattern or "she calculated" in another (this latter form is spelled irregularly, but this spelling is frequent). We use a simple heuristic that prefers certain patterns over others in order to fully resolve the ambiguity in such cases. We refer to this corpus as the *morphologically analyzed corpus*. Recently, two syntactic parsers have been made available for Hebrew, facilitating automatic computation of constituent- and dependency-structures (Goldberg, 2011). We applied the dependency parser to the same corpus; we refer to the result as the *syntactically parsed corpus*.

We also used the much smaller Hebrew Treebank (Sima'an et al., 2001; Guthmann et al., 2009). This is a set of 6,219 sentences from the HaAretz newspaper, which were manually parsed and then semi-automatically converted to a dependency representation (Goldberg and Elhadad, 2009). The treebank lists three types of syntactic relations between the verb and its modifiers: OBJect, used for direct objects; COMplement, indicating other subcategorized complement; and DEPendency, used for adjuncts. We considered only the first two as complements. Of the 6,219 sentences we used only 5,281; we filtered out sentences with quoted speech, since verb dependents were not accurately indicated in such sentences. We divided this set into *development* and *test* subsets, as indicated in Table 5.

	Sentences	**Verb lemmas**	**Verb tokens**
Development	1,057	676	1,536
Test	4,224	1,301	6,025
Total	5,281	1,423	7,561

Table 5 The Hebrew treebank, development and test sets.

4.3 Hypothesis testing

We employed four different statistical measures to assess the strength of the association between a verb and its complement: raw frequency (RF), log likelihood ratio (LLR), *t*-test and pointwise mutual information (PMI).

Let v be a verb lemma[2] and c, c' be specific complements (of the ten complement types listed above). Given a corpus with N tokens, of which V are (possibly inflected forms of) verbs, we define the following counts:

$n_{v,c}$ The number of occurrences of any inflected form of v in the corpus, with c as a complement. When the corpus is morphologically analyzed, we count c only if it immediately follows v. When the corpus is parsed, c must be marked as a subcategorized dependent of v. When c is the empty frame, $n_{v,c}$ is the number of occurrences of v that are not complemented by any of the other nine complement types.

n_v The number of occurrences of any inflected form of v in the corpus; i.e., $n_v = \sum_c n_{v,c}$.

n_c The number of occurrences of verb complements in the corpus; i.e., $n_c = \sum_v n_{v,c}$.

$n_{\neg v,c}$ The number of occurrences of any inflected form of any verb other than v in the corpus, with c as a complement.

$n_{\neg v}$ The number of occurrences of any inflected form of verbs other than v in the corpus, $V - n_v$.

Using maximum likelihood, we estimate:

$$\begin{aligned} p_{c|v} &= \frac{n_{v,c}}{n_v} \\ p_{c|\neg v} &= \frac{n_{\neg v,c}}{n_{\neg v}} \\ p_c &= \frac{n_c}{\sum_{c'} n_{c'}} \\ p_v &= \frac{n_v}{N} \\ p_{v,c} &= \frac{n_{v,c}}{N} \end{aligned}$$

With these, we define:

Raw frequency The likelihood of a complement c to occur with a verb v. This is exactly $p_{c|v}$.

Log likelihood ratio Following Dunning (1993), we define log-likelihood ratio as

$$\begin{aligned} LLR(v,c) = 2 \times (\ & L(p_{c|v}, n_{v,c}, n_v) + L(p_{c|\neg v}, n_{\neg v,c}, n_{\neg v}) \\ & - L(p_c, n_{v,c}, n_v) - L(p_c, n_{\neg v,c}, n_{\neg v})), \end{aligned}$$

where for all p, k, and n,

$$L(p,k,n) = (k \times \log p) + ((n-k) \times \log(1-p)).$$

***T*-score** We use the adapted definition of Sarkar and Zeman (2000):

$$T(v,c) = \frac{p_{c|v} - p_{c|\neg v}}{\sqrt{\sigma^2(n_v, p_{c|v}) + \sigma^2(n_{\neg v}, p_{c|\neg v})}},$$

[2] Different *binyanim* (verb patterns) constitute different lemmas.

where for all n, p,

$$\sigma^2(n,p) = n \times p \times (1-p).$$

Pointwise mutual information Following Church and Hanks (1990), we define PMI as

$$I(v,c) = \log \frac{p_{v,c}}{p_v \times p_c}.$$

4.4 Thresholds

Each of the four association measures defined above provides a way to estimate the strength of the association between a verb and its complement. To determine whether this association is strong enough for the complement to be included in the dictionary, we needed to set thresholds for each measure; only complements whose association is higher than the threshold were considered part of the subcategorization frame of the verb.

To set the threshold, we used the development part of the treebank described in Section 4.2. For each collocation measure, independently, we searched for a threshold that maximized the F-score of the task of identifying the correct complements of verbs in the development set. We used an exhaustive search with finely separated thresholds (obtained by dividing the full range of values the test can yield into 100 evenly-spaced intervals), and obtained the following threshold values: for RF, 0.11; LLR, 544.17; PMI, 0.12; and for t-score, 0.12 (t-score values are multiplied by 1,000 for readability).

In spite of the relatively small size of the development set, the accuracy is not highly sensitive to the precise value of the threshold. Figure 1 shows the F-score on the development set with respect to varying thresholds, for PMI and t-score. Clearly, significant changes in the value of the threshold (for PMI, from 0 to 0.75; for RF, from 0.05 to 0.2) result in minor changes to the F-score. The other measures are similarly robust.

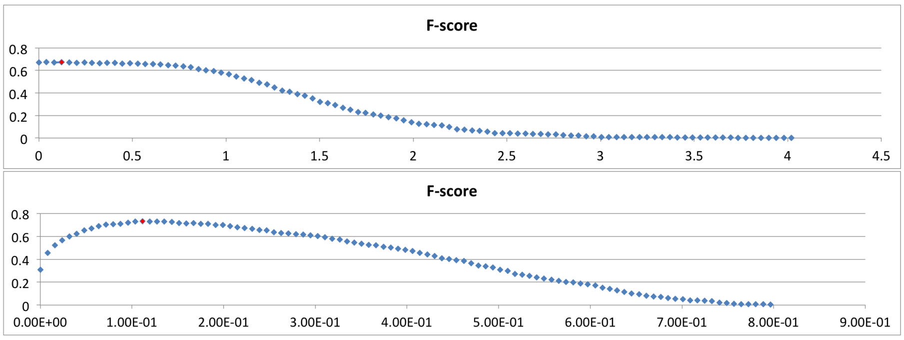

Fig. 1 Sensitivity of the accuracy to the threshold, PMI (above) and RF (below).

We used these threshold values in the remainder of this work: whenever the association by a verb and a potential complement was higher than the threshold, the complement was considered a member of the subcategorization frame of the verb according to the specific measure.

5 Results

5.1 Extracting a verb dictionary from a morphologically analyzed corpus

We applied the association measures defined in Section 4.3, with the thresholds defined in Section 4.4, to the entire morphologically analyzed corpus (Section 4.2). For each occurrence of a verb in the corpus, we considered the token that immediately follows the verb. As shown in Section 3.5, most complements occur immediately after their head verb, so we restricted the search to this position. Furthermore, we only considered as candidates instances in which the token immediately following the verb is either (1) one of the six prepositions and two conjunctions we focus on, or (2) an infinitival verb, or (3) a noun phrase that either is preceded by the accusative marker, or does *not* agree with the verb in either number, gender, or person (to rule out potential subjects). In all other cases, we considered the empty subcategorization frame as a candidate.

We then computed the association between the verb and its candidate complement, according to each of the association measures defined in Section 4.3. If it was above the threshold, the candidate was considered a complement of the verb.

The result of this process is a large set of verb–complement (V–C) pairs for each association measure; we refer to these sets as the *verb dictionaries*. Data on the four verb dictionaries extracted from the morphologically analyzed corpus are listed in Table 6.

Measure	**Verbs**	**V–C pairs**	**Pairs per verb**
RF	3,393	7,392	2.18
LLR	413	706	1.71
T-score	3,390	5,633	1.66
PMI	3,393	7,392	2.18

Table 6 Verb–complement pairs extracted from the morphologically analyzed corpus.

To exemplify the results, we now list some of the entries in the verb dictionary induced by PMI; as we will show below, PMI turns out to be a good association measure, providing not only wide coverage but also relatively high accuracy.

– The verb *tm* “expire, be finished” was correctly associated only with the empty frame. All other complements were much lower than the threshold.

- The verb *amr* "say" was correctly associated with *l* "to" and with a clause. The correct noun phrase complement was not included in the dictionary (false negative); additionally, the empty frame *was* included (false positive). This is most likely due to the fact that in journalistic writing, the subject often *follows* the verb *amr* "say" (even in English).
- The verb *hxliT* "decide" was a complete success: all and only the correct complements (clause, infinitival verb phrase and *yl* "on") were included in the dictionary.
- The same applies to the verb *pwnh* "be evacuated", for which prepositional phrases with *al* "to", *l* "to", and *m* "from" were listed.
- The verb *nhr* "stream, flow" was correctly associated with *al* "to", but also incorrectly with the empty subcategorization frame. We attribute this mismatch to errors in the morphological tagging, since the verb is homographic with the noun *nhr* "river".
- The verb *ndbq* is ambiguous; it can mean "stick" (typically with *l* "to" or *al* "to"), or "be infected", in which case both *m* "from" and *b* "with" are included in the subcategorization frame. As noted above, our methods do not distinguish between the different meanings, nor to distinguish between two single complements and a frame that includes both. Indeed, our PMI verb dictionary included all the four possible complement types.

The complete verb dictionaries induced by each of the association measures, including numeric values reflecting the strength of the association, are available for download from the MILA website.[3]

5.2 Extracting a verb dictionary from a syntactically parsed corpus

We used the syntactically parsed corpus (Section 4.2) for the task of extracting verb complements. Recall that this is the corpus used above, but each sentence was parsed with the dependency parser of Goldberg (2011). The reported accuracy of the parser is approximately 80%, so many errors can be expected.

For each verb occurrence in the corpus, we considered as potential complements all the phrases that depend on the verb with any of the three dependency labels: OBJ, COM and DEP. We considered the empty subcategorization frame as a candidate if none of the other complement types are dependent on the verb. We then used the same association measures as in Section 4.3, and the same thresholds as in Section 4.4, to determine whether the candidate should be included in the dictionary entry of the verb.

Again, the result is a set of four verb dictionaries, one for each association measure. Data on the dictionaries are listed in Table 7.

The main difference between the parsed corpus and the morphologically analyzed one is that the former facilitates a wider scope in which to search for complements, as it is not limited to complements occurring immediately after the verb. This improved recall but potentially harmed precision. This also

[3] Under 'Lexicons', at `http://www.mila.cs.technion.ac.il`

Measure	Verbs	Verb-complement pairs	Average # of pairs per verb
RF	2,955	8,180	2.77
LLR	2,955	20,516	6.94
T-score	2,955	5,453	1.85
PMI	2,955	6,545	2.21

Table 7 Verb–complement pairs extracted from the syntactically parsed corpus.

prevented verbs in which the subject tends to immediately follow the verb (e.g., *amr* "say") from being wrongly associated with the empty frame. Indeed, using the parsed corpus, the algorithm decided to include a potential complement in the dictionary much more frequently than with the morphologically analyzed corpus, except for the empty frame, as can be seen in Table 8.

Complement type	Parsed corpus	Morph. analyzed corpus
Noun phrase	367,391	131,055
al "to"	9,156	3,646
b "in"	333,745	106,957
l "to"	190,598	78,418
m "from"	66,946	21,912
yl "on"	76,869	26,140
ym "with"	21,771	1,424
Infinitival verb phrase	92,483	58,371
Clause	108,871	50,696
Empty	326,120	443,349

Table 8 The number of complements determined using both corpora, by complement type.

6 Evaluation

To assess the quality of the verb dictionaries discussed in Section 5, we conducted both intrinsic (Sections 6.1, 6.2) and extrinsic (Section 6.3) evaluations. In all experiments we used the verb–complement dictionaries but ignored the strength of the association between the verb and its complements (as long as it was above threshold, of course.)

6.1 Intrinsic evaluation

Our evaluation measure is F-score. Define:

TP The number of extracted verb–complement pairs that are indeed correct.
TN The number of verb–complement candidates that are not extracted, and are indeed not correct.
FP The number of extracted pairs that are not correct.
FN The number of correct pairs that are not extracted.

Then the *precision* is $P = TP/(TP + FP)$, the *recall* is $R = TP/(TP + FN)$, and the F-score is their harmonic mean, $F = 2 \times P \times R/(P + R)$.

In order to apply these measures, one needs to determine what is indeed "correct". For this, we manually constructed the verb dictionaries of 58 verbs, with various subcategorization frames and frequencies. The verbs in the evaluation set were chosen randomly (and uniformly) according to their distribution in the entire corpus. Many of them are consequently highly frequent (e.g., *amr* "say", 39,356 occurrences, *hgiy* "arrive", 16,818), while others are rarer (e.g., *htrgz* "get angry", 31 occurrences only, or *crx* "scream", 41). We asked two lexicographers to specify the complements of each verb in this list. The instructions given to the annotators were:

> "Determine whether the verb has a complement of the specific construction (e.g., a prepositional phrase with the specific preposition). A subcategorized complement denotes an argument of the verb; its meaning is necessary for complete understanding of the meaning of the verb. Syntactically, a subcategorized complement can only occur once, and if it is omitted the verb phrase is conceived as incomplete."

Several examples, both positive and negative, were also specified. Since we were concerned with 10 complement types and 58 verbs, there were 580 decisions to be made; the two annotators disagreed on 93 (16%) of them. The annotators then discussed each of the disagreements and consolidated the differences. The result is a gold set of manually annotated complements for 58 verbs, a subset of which is listed in Table 9.

Verb	**"Correct" complements**
aišr "confirm"	noun phrase, *l* "to", verb phrase, clause
asr "forbid"	noun phrase, *l* "to", clause
biqr "visit"	noun phrase, *b* "in"
dwbr "be spoken"	*b* "in", *yl* "on", clause
drš "request"	noun phrase, *m* "from", verb phrase, clause
hamin "believe"	*b* "in", *l* "to", clause
hgiy "arrive"	*l* "to", *m* "from", empty
hwbil "lead"	noun phrase, *b* "in", *l* "to", *yl* "on"
hwgš "be served"	*l* "to"
hxliT "decide"	*yl* "on", verb phrase, clause

Table 9 A sample of the manually-annotated verb dictionary for evaluation.

Table 10 shows the number of verb lemmas in the gold set that are associated with each of the ten complement types. Evidently, with the exception of *ym* "with", all complement types are well represented in the gold set.

As a baseline, we performed the following experiment: we considered the syntactically parsed corpus as a gold standard, and viewed every verb–complement pair that occured in this corpus as correct. In other words, if the parsed corpus included a sentence in which a complement c was annotated as a dependent of the verb v, and the dependency was labeled either OBJ or

Complement type	Verbs	%
Noun phrase	24	41
al "to"	7	12
b "in"	23	40
l "to"	23	40
m "from"	12	21
yl "on"	18	31
ym "with"	1	2
Infinitival verb phrase	17	29
Clause	15	26
Empty	19	33

Table 10 The number of verb lemmas in the gold set that are associated with each complement type.

COM, we considered the pair v–c as correct. For each verb in the test set we thus obtained a list of complements, and compared it to the gold annotations exemplified in Table 9. This yielded high recall at the expense of precision, of course. The baseline results on our test set are listed as the first row of Tables 11 and 12.

We then used the morphologically analyzed corpus, and extracted, for each verb in the test set, the above-threshold candidates. Next we compared them to the gold annotations of Table 9. The results are listed in Table 11. We found that PMI is the best association measure, yielding both the highest recall and the second highest precision on this set. While recall is lower than the baseline, precision is much higher, resulting in a much higher F-score. T-score and RF both do well, while LLR is below the baseline.

Measure	TP	TN	FP	FN	R (%)	P (%)	F
Baseline	158	36	385	1	99.37	29.10	45.01
RF	89	369	52	70	55.97	63.12	59.33
LLR	54	371	50	105	33.96	51.92	41.06
T-score	79	398	23	80	49.69	77.45	60.54
PMI	100	384	37	59	62.89	72.99	**67.57**

Table 11 Intrinsic evaluation results on the test set, morphologically analyzed corpus.

We repeated the evaluation with the verb dictionaries that were extracted from the syntactically parsed corpus; the results are listed in Table 12. Surprisingly, the F-score of the PMI dictionary does not improve. By contrast, all other measures actually improve somewhat, though PMI remains the best measure. Our conclusion is that the use of the syntactically parsed corpus does not contribute significantly to this task.

We also performed a weighted evaluation with the manual annotation, where the contribution of each verb in the gold set is proportional to the verb's frequency in the corpus. The results are very similar, typically within one percentage point difference from the balanced evaluation results.

Measure	TP	TN	FP	FN	R (%)	P (%)	F
Baseline	158	36	385	1	99.37	29.10	45.01
RF	109	348	73	50	68.55	59.89	63.93
LLR	56	381	40	103	35.22	58.33	43.92
T-score	87	397	24	72	54.72	78.38	64.44
PMI	99	383	38	60	62.26	72.26	**66.89**

Table 12 Intrinsic evaluation results on the test set, syntactically parsed corpus.

Our results clearly indicate that PMI is the best collocation measure for the task at hand. Other works that dealt with induction of verb valence and subcategorization frames promoted different collocation measures. For example, Korhonen et al. (2000) compared two methods for hypothesis testing, *binomial hypothesis test* and *log-likelihood ratio*, with a threshold on the relative frequencies of frames, using maximum likelihood to estimate probabilities. The latter performed best, and Korhonen et al. (2000) suggested as an explanation the Zipfian distribution of subcategorization frames. However, for many other tasks that involve the induction of collocations, PMI seems to be a preferred test (Chang et al., 2002; Villavicencio et al., 2007; Tsvetkov and Wintner, 2010, 2012), and Pecina (2005) actually advocated the combination of several collocation measures. We conclude that it is always important to experiment with more than one measure, as we do here.

6.2 Error analysis

The association measures we used reflect the frequency of the various complements as they immediately follow the verb. Sometimes this frequency is clear-cut. For example, 98% of the occurrences of the verb *tm* "expire, be finished" in the corpus are with the empty subcategorization frame; all the measures thus correctly associated this verb with the empty frame. Other verbs, however, do not behave as nicely. Consider *amr* "say"; especially in the journalistic genre that constitutes a significant portion of our corpus, this verb tends to occur with a post-verbal subject (55% of the instances). Our simplistic method considered such occurrences as instances of the empty subcategorization frame, and neglected to find the object, which is often pre-verbal.

The cause of many other errors is the lower frequency of the expected complement immediately after the verb. Consider the verb *nhr* "stream, flow", which should be complemented by *l* "to" or *al* "to". As mentioned above, many "instances" of this verb are in fact mistagged instances of the noun *nhr* "river". In addition, the subject of the verb frequently occurs post-verbally. Consequently, only 9% of the instances of this verb in the corpus are immediately followed by the expected preposition.

Similarly, the verb *aixl* "wish, congratulate" takes two complements, a noun phrase and a *l* "to" prepositional phrase, but in the typical order the prepositional phrase precedes the noun phrase (possibly because the *l* "to" preposition typically combines with a pronoun, and thus is lighter). Consequently, our method found the preposition complement but not the noun phrase.

Another major source of errors is data sparsity. The verb *htiiyc* "consult" occurs 219 times in the corpus. This frequency is sufficient for all the collocation measures to extract the complement *ym* "with", but not to identify the less frequent, and typically second, complement *yl* "about". The ambiguous verb *mxh* can either mean "protest", in which case it takes *yl* "on"; or "erase, wipe", in which case it takes a noun phrase. The former meaning is far more frequent, and hence all association measures yielded the *yl* "on" complement, but none yielded the noun phrase.

For a more quantitative error analysis, refer to Table 13, which depicts the accuracy of identifying the complements of verbs in the gold set, using the PMI dictionary extracted from the morphologically analyzed corpus, broken down by complement type.

Complement type	**Verbs**	**TP**	**TN**	**FP**	**FN**	**R (%)**	**P (%)**	**F**
Noun phrase	24	7	48	3	0	100.00	70.00	82.35
al "to"	7	17	32	2	7	70.83	89.47	79.07
b "in"	23	12	30	5	11	52.17	70.59	60.00
l "to"	23	18	32	3	5	78.26	85.71	81.82
m "from"	12	9	43	3	3	75.00	75.00	75.00
yl "on"	18	12	35	5	6	66.67	70.59	68.57
ym "with"	2	1	53	4	0	100.00	20.00	33.33
Infinitival verb phrase	17	7	40	1	10	41.18	87.50	56.00
Clause	15	8	42	1	7	53.33	88.89	66.67
Empty	19	9	29	10	10	47.37	47.37	47.37

Table 13 Accuracy by complement type, PMI dictionary, morphologically parsed corpus.

The highest accuracy was obtained, not surprisingly, for noun phrase complements, the most common complement type in our gold set. It is surprising, however, that the recall of identifying noun phrase complements is perfect; since we quite aggressively filtered out NP complements that agree with the verb, one would expect many such complements to be mistakenly filtered out, resulting in a lower recall. Evidently, this was not the case. Table 13 also shows that the empty frame is not identified accurately; this is consistent with the discussion above: often, unrelated prepositional phrases are mistaken to be complements, and data sparseness can cause actual complements not to be identified.

6.3 Extrinsic evaluation

The intrinsic evaluation of Section 6.1 is necessarily limited to a small set of verbs. As a more robust evaluation, we used the automatically extracted dictionaries in two natural language processing tasks, and showed a significant improvement in the performance of those tasks.

It is worth noting that our verb-complement dictionaries are extracted in a way that favors high-frequency data, whereby frequent complements are more likely to be recorded (at the expense of lower recall). For the two applications we discuss below, higher-precision frequent data most probably suffice for improving the performance, hence the gains we witness.

6.3.1 PP attachment

First, we address the problem of *prepositional phrase attachment*: in Hebrew, as in many other languages, prepositional phrases can be attached both to verbs and to nouns. Determining the correct attachment of prepositional phrases is challenging, and can significantly affect the accuracy of parsing (Lin, 1998; Goldberg, 2011). The task has attracted much interest, and several works attempt to address it, using pure statistical methods (Resnik and Hearst, 1993; Hindle and Rooth, 1993; Ratnaparkhi et al., 1994), or through approaches that incorporate additional linguistic knowledge (Wilks et al., 1985; Dahlgren and McDowell, 1986; Jensen and Binot, 1987; Hirst, 1988). In particular, several works showed that information on verb subcategorization is beneficial for this task (Stetina and Nagao, 1997; Yeh and Vilain, 1998; Pantel and Lin, 2000; Volk, 2002). More specifically, Merlo and Ferrer (2006) argued that a distinction between complements and adjuncts is needed in order to properly attach prepositional phrases, and suggested a (supervised) machine-learning-based classification method for the task. They found that "both linguistic diagnostics of argumenthood and lexical semantic classes are useful."

Subcategorization information can indeed help determine the correct attachment: when a prepositional phrase is a subcategorized complement of a verb, its occurrence is likely to be attached to the verb, rather than to some intervening noun. Consider the following examples (again, subcategorized complements are underlined):

(21) *la* ***nmcaw*** *ršiwnwt lnšiat nšq*
not were-found licenses to+carrying-of weapons
"Weapon carrying licenses were not found"

(22) *hnhg* ***hrah*** *ršiwnwt lšwTr*
the+driver showed licenses to+the-policeman
"The driver showed the police officer his license"

Both sentences involve a verb, followed by the noun *ršiwnwt* "licenses", followed by a prepositional phrase with *l* "to". Since such a prepositional phrase is subcategorized by the verb *hrah* "show", but not by *nmca* "be found", the prepositional phrase is more likely to attach to the verb in the second example, but to the noun in the first example. This is indeed the correct attachment.

For evaluation, we used the test subset of the treebank (Section 4.2). We focused on constructions of the form verb–noun–preposition, allowing any number of words from other POS classes to intervene between the verb and the noun and between the noun and the preposition. The test set included 323 such constructions, with 204 different verbs. The task was to determine whether the preposition attaches to the verb or to the noun.[4]

The baseline was obtained by always attaching the preposition to the noun. A better-informed baseline uses the syntactically parsed corpus as a gold standard, and considers each verb–complement pair as correct (as in Section 6.1). This turns out to be worse than the naïve baseline, probably because the parser seems to have a clear preference toward attaching PPs to the verb (a tendency which we do not fully understand).

To improve upon the baselines we used information from the verb dictionaries, albeit in a very simplistic way: if the preposition was strongly associated with the verb (above the threshold), we attached it to the verb. We compared this decision with the correct attachment, as reflected by the treebank's annotation.

The results, with each of the verb dictionaries corresponding to the four association measures, are listed in Table 14. We report *accuracy (Acc)*, defined as $(TP+TN)/(TP+FP+TN+FN)$, as well as *error rate reduction (ERR)*. Evidently, all the verb dictionaries are instrumental in this task; the best accuracy, obtained by t-score, is over 65% (reflecting a reduction of almost 30% of the errors compared with the baseline), but the PMI and RF dictionaries also reduce the error rate by more than 20%. Implementing a voting mechanism among the four dictionaries did not improve the results. We also repeated this evaluation with the dictionaries that were extracted from the syntactically parsed corpus; accuracy improved for PMI (from 61.61% to 65.02%), but deteriorated slightly for the three other measures. Again, as in the case of the intrinsic evaluation, the morphologically analyzed corpus yielded higher accuracy than the syntactically parsed corpus.

Measure	**TP**	**TN**	**FP**	**FN**	**Acc (%)**	**ERR (%)**
Baseline	0	167	0	156	51.70	
Baseline (parser)	145	15	152	11	49.54	
RF	70	132	35	86	62.54	22.44
LLR	48	136	31	108	56.97	10.90
T-score	72	140	27	84	**65.63**	**28.85**
PMI	84	115	52	72	61.61	20.51

Table 14 Extrinsic evaluation results: PP attachment.

6.3.2 Machine translation

As another method of extrinsic evaluation, we incorporated knowledge extracted from the verb dictionaries into a transfer-based machine translation

[4] Since we did not use a parser to determine the prepositional phrases to be attached, we are immune in this experiment to the criticism of Atterer and Schütze (2007), whereby using an "oracle" distorts the actual performance of the attachment module.

(MT) system, and showed improved results.[5] Specifically, we used translation from Arabic to Hebrew (Shilon et al., 2010, 2012b); the system was developed in the framework of Stat-XFER (Lavie, 2008), which facilitates the explicit expression of synchronous (extended) context-free transfer rules.

Prepositions are hard to translate, especially when they are required by their governing verb, since in such cases the choice of preposition tends to be arbitrary. In fact, the choice of preposition can vary among synonymous verbs even in the same language. Thus, Hebrew *hkh* "hit" takes the accusative preposition *at*, whereas the synonymous *hrbic* "hit" takes *l* "to". While Hebrew and Arabic are both Semitic languages, and several verbs and prepositions in the two languages are cognate, there is no clear mapping of subcategorization frames from one language to another. Clearly, then, prepositions cannot be translated literally, and the head that they modify, as well as the object of the preposition, have to be taken into account when a preposition is chosen to be generated.

We used the (PMI-induced) verb dictionary in a transfer-based MT system as follows. The system uses a morphological generator to generate inflected forms of lemmas obtained from a bilingual dictionary. Each such form is associated with a feature structure that describes some properties of the form (e.g., its gender, number and person). To the feature structures of verbs we added an additional feature, ALLOWED_PREPS, whose value is the list of prepositions licensed by the verb, as determined by the verb dictionary. In this way, verbs were specified for the prepositions with which they are most likely to occur.

As the MT system is transfer-based, it allows the specification of synchronous rules that map local syntactic structures between the source and the target languages. We thus implemented several transfer rules that map verb–complement constructions between Arabic and Hebrew. When these rules are applied, they have access to (the surface form of) the actual preposition in the source and target phrases. To these rules we added constraints that only allow them to fire when the actual preposition is indeed licensed by the verb to which it is attached. For example, the rule that combines a verb with a prepositional phrase in Arabic, to yield a verb phrase, is synchronized with a similar rule that combines a verb with a PP in Hebrew. We added a requirement that the actual preposition that heads the Hebrew PP be licensed by the Hebrew verb (as determined by the verb dictionary). See Shilon et al. (2012a) for the details.

To evaluate the contribution of the verb dictionary, we created a test set of 300 sentences from newspaper texts, which were manually translated by three human translators. Of those, we selected short sentences (up to 10 words), for which the bilingual dictionary used by the system had full lexical coverage. This resulted in a set of 28 sentences (still with three reference translations each), which allowed us to focus on the actual contribution of the preposition-mapping solution rather than on other limitations of the MT system. (Unfortunately, evaluation on the entire test set of 300 sentences without accounting

[5] This experiment was reported in Shilon et al. (2012a).

for full lexical coverage yields such poor translations that the comparison between different configurations of the system is meaningless.) As a baseline system, we used exactly the same setup, but withheld all the verb–preposition association knowledge. Table 15 lists the BLEU (Papineni et al., 2002) and METEOR (Denkowski and Lavie, 2011) scores of both systems.

	BLEU	METEOR
Baseline	0.325	0.526
With prepositions	0.370	0.560

Table 15 Extrinsic evaluation results: machine translation

The system that incorporates linguistic knowledge on prepositions significantly ($p < 0.05$) outperformed the baseline system, as Table 15 shows. A detailed analysis of the obtained translations revealed that the baseline system generated prepositions that were not licensed by their head verb, and the language model failed to choose the hypothesis with the correct preposition, if such a hypothesis was generated at all.

7 Conclusions

We presented an automatically-created verb dictionary of Hebrew, specifying the most likely complements to occur with each verb, along with a quantitative degree of the strength of the association between the complement and the verb. As it is extracted from a large corpus, the dictionary has wide coverage, and its accuracy is satisfying. It was proven beneficial for two natural language processing applications, and we trust that is will be useful for various other purposes in the future.

This is a preliminary work. Specifically, it views each complement of a verb in isolation, and does not attempt to construct full subcategorization frames. While the current dictionary is still useful, in the future we would like to refine it by extending the verb–complement relations to full, multi-complement subcategorization frames. We are also interested in developing methods for disambiguation: when a verb has more than one meaning, with different subcategorization frames, we would like to be able to obtain multiple frames from the the extraction procedure.

As more and more corpora become available, we plan to generate domain- and corpus-specific dictionaries, for more focused applications. We are particularly keen on developing such a dictionary for a corpus of *spoken* Hebrew that is currently being compiled (Nir et al., 2010; Albert et al., 2012). We would also like to extend the extracted relations to triplets, including also the noun that heads the object of the preposition. Such triplets can often indicate multi-word expressions, such as *hbia b+xšbwn* “brought in+calculation ⇒ consider”, or *ymd yl dytw* “stood on his-mind ⇒ insist”; as such, they can be instrumental for the construction of a multi-word dictionary of Hebrew. We leave these directions for future research.

Acknowledgments

This research was supported by THE ISRAEL SCIENCE FOUNDATION (grants No. 1269/07, 505/11). We are grateful to Reshef Shilon for his help with the machine translation experiments, and to Yoav Goldberg for his help with the Hebrew parser. Thanks are also due to Kayla Jacobs for several useful comments. We benefitted greatly from the constructive comments of three anonymous reviewers.

References

Aviad Albert, Brian MacWhinney, Bracha Nir, and Shuly Wintner. A morphologically annotated Hebrew CHILDES corpus. In *Proceedings of the Workshop on Computational Models of Language Acquisition and Loss*, pages 20–22, Avignon, France, April 2012. Association for Computational Linguistics. URL `http://www.aclweb.org/anthology/W/W12/W12-0904`.

Michaela Atterer and Hinrich Schütze. Prepositional phrase attachment without oracles. *Computational Linguistics*, 33(4):469–476, December 2007. ISSN 0891-2017. doi: 10.1162/coli.2007.33.4.469. URL `http://dx.doi.org/10.1162/coli.2007.33.4.469`.

Ulrike Baldewein. Modeling attachment decisions with a probabilistic parser: The case of head final structures. In *Proceedings of the 26th Annual Conference of the Cognitive Science Society*, pages 73–78. Erlbaum, 2004.

Adriana Belletti and Ur Shlonsky. The order of verbal complements: A comparative study. *Natural Language and Linguistic Theory*, 13(3):489–526, 1995.

Michael R. Brent. Automatic acquisition of subcategorization frames from untagged text. In *Proceedings of the 29th annual meeting on Association for Computational Linguistics*, pages 209–214, Stroudsburg, PA, USA, 1991. Association for Computational Linguistics. doi: http://dx.doi.org/10.3115/981344.981371. URL `http://dx.doi.org/10.3115/981344.981371`.

Michael R. Brent. From grammar to lexicon: Unsupervised learning of lexical syntax. *Computational Linguistics*, 19(2):243–262, 1993.

Ted Briscoe and John Carroll. Generalised probabilistic LR parsing of natural language (corpora) with unification-based grammars. *Computational Linguistics*, 19(1):25–59, 1993.

Ted Briscoe and John Carroll. Automatic extraction of subcategorization from corpora. In *Proceedings of the 5th ACL Conference on Applied Natural Language Processing*, pages 356–363, 1997.

John Carroll, Guido Minnen, and Ted Briscoe. Can subcategorisation probabilities help a statistical parser? In *Proceedings of the 6th ACL/SIGDAT Workshop on Very Large Corpora*, pages 118–126, 1998.

Baobao Chang, Pernilla Danielsson, and Wolfgang Teubert. Extraction of translation unit from Chinese-English parallel corpora. In *Proceedings of the first SIGHAN workshop on Chinese language processing*, pages 1–5, Morristown, NJ, USA, 2002. Association for Computational Linguistics. doi: http://dx.doi.org/10.3115/1118824.1118825.

Paula Chesley and Susanne Salmon-alt. Automatic extraction of subcategorization frames for French. In *Proceedings of the Language Resources and Evaluation Conference, LREC 2006*, pages 253–258. European Language Resources Association (ELRA), 2006.

Noam Chomsky. *Aspects of the theory of syntax.* MIT Press, 1965.

Kenneth Ward Church and Patrick Hanks. Word association norms, mutual information, and lexicography. *Computational Linguistics*, 16(1):22–29, 1990. ISSN 0891-2017.

Kathleen Dahlgren and Joyce P. McDowell. Using commonsense knowledge to disambiguate prepositional phrase modifiers. In Tom Kehler, editor, *Proceedings of the 5th National Conference on Artificial Intelligence*, pages 589–593. Morgan Kaufmann, 1986.

Łukasz Dębowski. Valence extraction using EM selection and co-occurrence matrices. *Language Resources and Evaluation*, 43(4):301–327, 2009.

Michael Denkowski and Alon Lavie. Meteor 1.3: Automatic metric for reliable optimization and evaluation of machine translation systems. In *Proceedings of the Sixth Workshop on Statistical Machine Translation*, pages 85–91. Association for Computational Linguistics, July 2011. URL `http://www.aclweb.org/anthology/W11-2107`.

Ted Dunning. Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*, 19:61–74, 1993.

Susan M. Garnsey, Neal J. Pearlmutter, Elizabeth Myers, and Melanie A. Lotocky. The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language*, 37 (1):58–93, 7 1997.

Yoav Goldberg. *Automatic Syntactic Processing of Modern Hebrew.* PhD thesis, Ben Gurion University of the Negev, Israel, 2011.

Yoav Goldberg and Michael Elhadad. Hebrew dependency parsing: Initial results. In *Proceedings of the 11th International Workshop on Parsing Technologies (IWPT-2009), 7-9 October 2009, Paris, France*, pages 129–133. The Association for Computational Linguistics, 2009.

Yoav Goldberg and Michael Elhadad. An efficient algorithm for easy-first non-directional dependency parsing. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 742–750, Stroudsburg, PA, USA, 2010. Association for Computational Linguistics. ISBN 1-932432-65-5. URL `http://dl.acm.org/citation.cfm?id=1857999.1858114`.

Noemie Guthmann, Yuval Krymolowski, Adi Milea, and Yoad Winter. Automatic annotation of morpho-syntactic dependencies in a Modern Hebrew treebank. In *Proceedings of Trees in Linguistic Theory (TLT-2009)*, January 2009.

Jan Hajič, Martin Čmejrek, Bonnie Dorr, Yuan Ding, Jason Eisner, Daniel Gildea, Terry Koo, Kristen Parton, Gerald Penn, Dragomir Radev, and Owen Rambow. Natural language generation in the context of machine translation. Technical report, Center for Language and Speech Processing, Johns Hopkins University, March 2004. URL http://cs.jhu.edu/~jason/papers/ws02. Final report from 2002 CLSP summer workshop (87 pages).

Xiwu Han, Tiejun Zhao, Haoliang Qi, and Hao Yu. Subcategorization acquisition and evaluation for Chinese verbs. In *Proceedings of the 20th international conference on Computational Linguistics (COLING '04)*, Stroudsburg, PA, USA, 2004. Association for Computational Linguistics. doi: 10.3115/1220355.1220459. URL http://dx.doi.org/10.3115/1220355.1220459.

Donald Hindle and Mats Rooth. Structural ambiguity and lexical relations. *Computationa Linguistics*, 19(1):103–120, March 1993. ISSN 0891-2017.

Graeme Hirst. Semantic interpretation and ambiguity. *Artificial Intelligence*, 34(2):131–177, 1988.

Rodney Huddleston and Geoffrey K. Pullum. *The Cambridge Grammar of the English Language*. Cambridge University Press, 2002.

Dino Ienco, Serena Villata, and Cristina Bosco. Automatic extraction of subcategorization frames for Italian. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*. European Language Resources Association (ELRA), may 2008. ISBN 2-9517408-4-0. URL http://www.lrec-conf.org/proceedings/lrec2008/.

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

Karen Jensen and Jean-Louis Binot. Disambiguating prepositional phrase attachments by using on-line dictionary definitions. *Computational Linguistics*, 13(3-4):251–260, 1987. ISSN 0891-2017.

Anna Korhonen. Using semantically motivated estimates to help subcategorization acquisition. In *Proceedings of the 2000 Joint SIGDAT conference on Empirical methods in natural language processing and very large corpora*, pages 216–223, Stroudsburg, PA, USA, 2000. Association for Computational Linguistics. doi: 10.3115/1117794.1117821. URL http://dx.doi.org/10.3115/1117794.1117821.

Anna Korhonen. *Subcategorisation acquisition*. PhD thesis, Computer Laboratory, University of Cambridge, 2002a. Techical Report UCAM-CL-TR-530.

Anna Korhonen. Semantically motivated subcategorization acquisition. In *Proceedings of the ACL-02 workshop on Unsupervised lexical acquisition*, pages 51–58, Stroudsburg, PA, USA, 2002b. Association for Computational Linguistics. doi: 10.3115/1118627.1118634. URL http://dx.doi.org/10.3115/1118627.1118634.

Anna Korhonen and Judita Preiss. Improving subcategorization acquisition using word sense disambiguation. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics*, pages 48–55, Stroudsburg, PA, USA, 2003. Association for Computational Linguistics. doi: 10.3115/1075096.1075103. URL http://dx.doi.org/10.3115/1075096.1075103.

Anna Korhonen, Genevieve Gorrell, and Diana McCarthy. Statistical filtering and subcategorization frame acquisition. In *Proceedings of the 2000 Joint SIGDAT conference on Empirical methods in natural language processing and very large corpora*, pages 199–206, Stroudsburg, PA, USA, 2000. Association for Computational Linguistics. doi: 10.3115/1117794.1117819. URL `http://dx.doi.org/10.3115/1117794.1117819`.

Anna Korhonen, Yuval Krymolowski, and Ted Briscoe. A large subcategorization lexicon for natural language processing applications. In *Proceedings of the Language Resources and Evaluation Conference, LREC 2006*, pages 1015–1020. European Language Resources Association (ELRA), 2006.

Jonathan K. Kummerfeld, David Hall, James R. Curran, and Dan Klein. Parser showdown at the Wall Street Corral: An empirical investigation of error types in parser output. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 1048–1059, Jeju Island, South Korea, July 2012. Association for Computational Linguistics. URL `http://www.aclweb.org/anthology/D12-1096`.

Jonathan K. Kummerfeld, Daniel Tse, James R. Curran, and Dan Klein. An empirical examination of challenges in Chinese parsing. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 98–103, Sofia, Bulgaria, August 2013. Association for Computational Linguistics. URL `http://www.aclweb.org/anthology/P13-2018`.

Mirella Lapata, Frank Keller, and Sabine Schulte im Walde. Verb frame frequency as a predictor of verb bias. *Journal of Psycholinguistic Reseach*, 30 (4):419–435, 2001.

Alon Lavie. Stat-XFER: A general search-based syntax-driven framework for machine translation. In Alexander F. Gelbukh, editor, *CICLing*, volume 4919 of *Lecture Notes in Computer Science*, pages 362–375. Springer, 2008. ISBN 978-3-540-78134-9.

Beth Levin. *English Verb Classes and Alternations: A Preliminary Investigation*. University of Chicago Press, Chicago, 1993. ISBN 9780226475332.

Jianguo Li and Chris Brew. Automatic extraction of subcategorization frames from spoken corpora. In *Proceedings of the Interdisciplinary Workshop on the Identification and Representation of Verb Features and Verb Classes*, pages 74–79, 2005.

Dekang Lin. Dependency-based evaluation of MINIPAR. In *Proceedings of the Workshop on the Evaluation of Parsing Systems*, pages 317–330. Springer, 1998.

Paola Merlo and Eva Esteve Ferrer. The notion of argument in prepositional phrase attachment. *Computational Linguistics*, 32(3):341–378, September 2006. ISSN 0891-2017. doi: 10.1162/coli.2006.32.3.341.

Cédric Messiant, Thierry Poibeau, and Anna Korhonen. LexSchem: a large subcategorization lexicon for French verbs. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*. European Language Resources Association (ELRA), May 2008. ISBN 2-

9517408-4-0. URL http://www.lrec-conf.org/proceedings/lrec2008/.

Bracha Nir, Brian MacWhinney, and Shuly Wintner. A morphologically-analyzed CHILDES corpus of Hebrew. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, pages 1487–1490. European Language Resources Association (ELRA), May 2010. ISBN 2-9517408-6-7.

Diarmuid Ó Séaghdha. Latent variable models of selectional preference. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 435–444, Stroudsburg, PA, USA, 2010. Association for Computational Linguistics. URL http://dl.acm.org/citation.cfm?id=1858681.1858726.

Patrick Pantel and Dekang Lin. An unsupervised approach to prepositional phrase attachment using contextually similar words. In *Proceedings of the 38th Annual Meeting on Association for Computational Linguistics*, pages 101–108, Stroudsburg, PA, USA, 2000. Association for Computational Linguistics. doi: 10.3115/1075218.1075232.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. BLEU: a method for automatic evaluation of machine translation. In *ACL '02: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, pages 311–318, Morristown, NJ, USA, 2002. Association for Computational Linguistics. doi: http://dx.doi.org/10.3115/1073083.1073135.

Pavel Pecina. An extensive empirical study of collocation extraction methods. In *Proceedings of the ACL Student Research Workshop*, pages 13–18, Ann Arbor, Michigan, June 2005. Association for Computational Linguistics. URL http://www.aclweb.org/anthology/P/P05/P05-2003.

Adwait Ratnaparkhi, Jeff Reynar, and Salim Roukos. A maximum entropy model for prepositional phrase attachment. In *Proceedings of the workshop on Human Language Technology*, pages 250–255, Stroudsburg, PA, USA, 1994. Association for Computational Linguistics. ISBN 1-55860-357-3. doi: 10.3115/1075812.1075868.

Philip Resnik and Marti A. Hearst. Structural ambiguity and conceptual relations. In *Proceedings of the Workshop on Very Large Corpora: Academic and Industrial Perspectives*, pages 58–64, June 1993.

Alan Ritter, Mausam, and Oren Etzioni. A latent dirichlet allocation method for selectional preferences. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 424–434, Stroudsburg, PA, USA, 2010. Association for Computational Linguistics. URL http://dl.acm.org/citation.cfm?id=1858681.1858725.

John Robert Ross. *Constraints on variables in syntax*. PhD thesis, Massachusetts Institute of Technology, Department of Modern Languages and Linguistics, 1967.

Anoop Sarkar and Daniel Zeman. Automatic extraction of subcategorization frames for Czech. In *Proceedings of the 18th conference on Computational linguistics*, pages 691–697, Stroudsburg, PA, USA, 2000. Association for Computational Linguistics. doi: http://dx.doi.org/10.3115/992730.992746. URL http://dx.doi.org/10.3115/992730.992746.

Sabine Schulte im Walde and Chris Brew. Inducing German semantic verb classes from purely syntactic subcategorisation information. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 223–230, Philadelphia, PA, 2002.

Reshef Shilon, Nizar Habash, Alon Lavie, and Shuly Wintner. Machine translation between Hebrew and Arabic: Needs, challenges and preliminary solutions. In *Proceedings of AMTA 2010: The Ninth Conference of the Association for Machine Translation in the Americas*, November 2010.

Reshef Shilon, Hanna Fadida, and Shuly Wintner. Incorporating linguistic knowledge in statistical machine translation: Translating prepositions. In *Proceedings of the Workshop on Innovative Hybrid Approaches to the Processing of Textual Data*, pages 106–114, Avignon, France, April 2012a. Association for Computational Linguistics. URL http://www.aclweb.org/anthology/W/W12/W12-0514.

Reshef Shilon, Nizar Habash, Alon Lavie, and Shuly Wintner. Machine translation between Hebrew and Arabic. *Machine Translation*, 26:177–195, 2012b. ISSN 0922-6567. URL http://dx.doi.org/10.1007/s10590-011-9103-z.

Khalil Sima'an, Alon Itai, Yoad Winter, Alon Altman, and Noa Nativ. Building a tree-bank of Modern Hebrew text. *Traitement Automatique des Langues*, 42(2):247–380, 2001.

Neftali Stern. *Milon ha-Poal*. Bar Ilan University, 1994. ISBN 965-226-164-5. In Hebrew.

Jiri Stetina and Makoto Nagao. Corpus based PP attachment ambiguity resolution with a semantic dictionary. In Joe Zhou and Kenneth W. Church, editors, *Proceedings of the Fifth Workshop on Very Large Corpora*, pages 66–80, 1997.

Lin Sun and Anna Korhonen. Improving verb clustering with automatically acquired selectional preferences. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 638–647, Stroudsburg, PA, USA, 2009. Association for Computational Linguistics. ISBN 978-1-932432-62-6.

Lin Sun, Anna Korhonen, and Yuval Krymolowski. Verb class discovery from rich syntactic data. In *Proceedings of the 9th international conference on Computational linguistics and intelligent text processing*, pages 16–27, Berlin, Heidelberg, 2008a. Springer-Verlag. ISBN 3-540-78134-X, 978-3-540-78134-9.

Lin Sun, Anna Korhonen, and Yuval Krymolowski. Automatic classification of English verbs using rich syntactic features. In *Proceedings of the Third International Joint Conference on Natural Language Processing*, pages 769–774, 2008b. URL http://aclweb.org/anthology-new/I/I08/I08-2107.pdf.

Mihai Surdeanu, Sanda Harabagiu, John Williams, and Paul Aarseth. Using predicate-argument structures for information extraction. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics*, pages 8–15, Stroudsburg, PA, USA, 2003. Association for Computational Linguistics. doi: 10.3115/1075096.1075098. URL http://dx.doi.org/10.

3115/1075096.1075098.

Yulia Tsvetkov and Shuly Wintner. Extraction of multi-word expressions from small parallel corpora. In *Proceedings of the 23rd International Conference on Computational Linguistics (COLING 2010)*, pages 1256–1264, August 2010.

Yulia Tsvetkov and Shuly Wintner. Extraction of multi-word expressions from small parallel corpora. *Natural Language Engineering*, 18(4):549–573, October 2012. doi: 10.1017/S1351324912000101. URL http://dx.doi.org/10.1017/S1351324912000101.

Aline Villavicencio, Valia Kordoni, Yi Zhang, Marco Idiart, and Carlos Ramisch. Validation and evaluation of automatically acquired multiword expressions for grammar engineering. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 1034–1043, 2007. URL http://www.aclweb.org/anthology/D/D07/D07-1110.

Martin Volk. Combining unsupervised and supervised methods for PP attachment disambiguation. In *Proceedings of the 19th international conference on Computational linguistics*, volume 1, pages 1–7, Stroudsburg, PA, USA, 2002. Association for Computational Linguistics. doi: 10.3115/1072228.1072232.

Yorick Wilks, Xiuming Huang, and Dan Fass. Syntax, preference, and right attachment. In *Proceedings of the 9th international joint conference on Artificial intelligence*, volume 2 of *IJCAI'85*, pages 779–784, San Francisco, CA, USA, 1985. Morgan Kaufmann Publishers Inc. ISBN 0-934613-02-8, 978-0-934-61302-6.

Alexander S. Yeh and Marc B. Vilain. Some properties of preposition and subordinate conjunction attachments. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics*, volume 2, pages 1436–1442, Stroudsburg, PA, USA, 1998. Association for Computational Linguistics. doi: 10.3115/980691.980803.

Adriano Zanette, Carolina Scarton, and Leonardo Zilio. Automatic extraction of subcategorization frames from corpora: an approach to portuguese. In *Proceedings of PROPOR 2012: International Conference on Computational Processing of the Portuguese Language*, May 2012. URL http://www.propor2012.org/demos/DemoSubcategorization.pdf.

Daniel Zeman. Can subcategorization help a statistical dependency parser? In *Proceedings of the 19th international conference on Computational linguistics (COLING-02)*, pages 1156–1162, Stroudsburg, PA, USA, 2002. Association for Computational Linguistics. doi: 10.3115/1072228.1072346. URL http://dx.doi.org/10.3115/1072228.1072346.