HEBREW MULTIWORD EXPRESSIONS Lexical Representation and Morphological Processing

Alon Itai^T Shuly Wintner^H

Department of Computer Science, Technion, Haifa, Israel

Department of Computer Science, University of Haifa, Israel

COST Action IC1207 PARSEME Meeting 17 September 2013 Warsaw, Poland

IN MEMORIAM: IVAN SAG November 9, 1949 – September 9, 2013



SCIENTIFIC OBJECTIVES (1ST YEAR) FROM AGATA'S INTRODUCTORY PRESENTATION YESTERDAY

- Better understanding of
 - Linguistic properties of MWEs
 - The potential of frameworks and methodologies wrt. parsing MWEs
 - Challenges behind the MWE annotation in treebanks
- Studies towards
 - Enhanced usability of MWE lexicons and valence dictionaries in parsing
 - A better coverage of MWEs in LRT
 - Evaluation capacity for MWE parsing

3 / 47

SCIENTIFIC OBJECTIVES (1ST YEAR) FROM AGATA'S INTRODUCTORY PRESENTATION YESTERDAY

- Better understanding of
 - Linguistic properties of MWEs
 - The potential of frameworks and methodologies wrt. parsing MWEs
 - Challenges behind the MWE annotation in treebanks
- Studies towards
 - Enhanced usability of MWE lexicons and valence dictionaries in parsing
 - A better coverage of MWEs in LRT
 - Evaluation capacity for MWE parsing

- An architecture for lexical specification of MWEs in morphologically-complex languages
- An implementation for (Modern) Hebrew
 - A solution for storing MWEs in an existing large-scale lexicon (Itai et al., 2006)
 - A protocol for integrating MWEs in an existing morphological processing system (Itai and Wintner, 2008)
- A survey of the variety and diversity of Hebrew MWEs within a computational setup

OUTLINE

1 INTRODUCTION

- **2** LINGUISTIC BACKGROUND
- 3 THE HEBREW MORPHOLOGICAL PROCESSOR
- LEXICAL REPRESENTATION OF HEBREW MWES
- **5** MORPHOLOGICAL PROCESSING OF MWES

6 CONCLUSION

HEBREW ORTHOGRAPHY

- Most vowels are not explicit
- Many particles (the definite article h "the", the frequent prepositions b "in", k "as", l "to" and m "from", the coordinating conjunction w "and" and the subordinating conjunctions š "that" and kš "when") attach to the word which immediately follows them

EXAMPLE (THE PREPOSITION **m** "FROM")

- Combines with nouns but not with adverbs
- The same rules govern the combination of Hebrew prefix particles with MWEs, but these combinations are constrained by the syntactic category of the whole expression, rather than its first word
- *m+ph* "from mouth"
- $m+ph \ l+awzn$ "from mouth to ear" \Rightarrow "through the grapevice"
- **m* "from" + *ph axd* "mouth one" \Rightarrow "unanimously"

HEBREW MORPHOLOGY

- Nominals (nouns, adjectives and numerals) inflect for number and gender
- Nominals have three distinct states: the absolute (citation) state; the definite state, which is indicated by the prefix h "the"; and the construct state, which is typically used in genitive (possessive) constructions
- Nominals (in the construct state) take pronominal suffixes, sometimes referred to as clitics, which are interpreted as possessives

EXAMPLE

sirh "boat"; h+sirh "the+boat"; sirt+i "my boat"; sirt mnwy "boat-of engine" ⇒ "speedboat"

HEBREW MORPHOLOGY

- Verbs inflect for number, gender and person, and also for a combination of tense/aspect and mood
- A single verb lemma can yield dozens of inflected forms
- Prepositions can combine with pronominal affixes that are interpreted as the objects of the preposition, and inflect for number, gender, and person

EXAMPLE

lid "near"; *lid+i* "near me"

HEBREW SYNTAX

- The standard constituent order of Hebrew is Subject-Verb-Object, although many other orders are possible, and some are highly frequent
- Within the noun phrase, constituents tend to occur in a fixed order
- Various elements of a noun phrase may be marked as definite; all elements of the noun phrase must agree with respect to definiteness

EXAMPLE

h+sirh h+iph h+zaw "the+boat the+nice the+this" \Rightarrow "this nice boat"

• Hebrew has three different possessive constructions

OUTLINE

- **1** INTRODUCTION
- 2 LINGUISTIC BACKGROUND
- **3** The Hebrew morphological processor
- 4 LEXICAL REPRESENTATION OF HEBREW MWES
- **(5)** MORPHOLOGICAL PROCESSING OF MWES

6 CONCLUSION

The Hebrew morphological processor

THE HEBREW MORPHOLOGICAL PROCESSOR (WINTNER AND YONA, 2003; ITAI AND WINTNER, 2008)



MORPHOLOGICALLY ANALYZED TEXT

EXAMPLE (šbi)

```
<sentence id="1">
 <token id="1" surface="šbi">
   <analysis id="1">
      <base lexiconPointer="1541" transliterated="šbh">
        <verb gender="feminine" number="singular" person="2" tense="imperative"/>
     </hase>
   </analysis>
   <analysis id="2">
      <base lexiconPointer="1636" transliterated="išb">
       <verb gender="feminine" number="singular" person="2" tense="imperative"/>
      </base>
   </analysis>
   <analysis id="5">
      <base lexiconPointer="7863" transliterated="šbi">
       <noun definiteness="false" gender="masculine" number="singular" state="absolute"/>
      </base>
   </analysis>
   <analysis id="6">
      <prefix function="relativizer/subordinatingConjunction" id="1" surface="š"/>
      <base lexiconPointer="26553" transliterated="b">
       <preposition />
      </hase>
      <suffix function="pronominal" gender="masculine and feminine" number="singular" person="1"/>
   </analysis>
 </token>
</sentence>
```

THE HEBREW MORPHOLOGICAL PROCESSOR

- The morphological processor operates on a token-by-token basis
- Tokens are acquired from the tokenizer which uses only blanks and punctuation for segmentation
- The tokenizer is completely independent of the lexicon
- The lexicon includes single-word tokens only, and the morphological analyzer is completely unaware of MWEs

14 / 47

OUTLINE

- **1** INTRODUCTION
- 2 LINGUISTIC BACKGROUND
- **3** The Hebrew morphological processor
- 4 LEXICAL REPRESENTATION OF HEBREW MWES
- **5** MORPHOLOGICAL PROCESSING OF MWES

6 CONCLUSION

BASICS

- Each MWE is represented as an item in the lexicon, which encodes its morphological and syntactic properties
- A MWE lexical entry includes an element that specifies that the item is a MWE, followed by its POS

EXAMPLE (*niw iwrq* "NEW YORK")

WORD-LEVEL PROPERTIES

FROZEN FORM Constituents can appear in one fixed form, disallowing all inflections. This form can be the citation form:

EXAMPLE

ain lw <u>id</u> bdbr "does not have a <u>hand</u> in the thing" \Rightarrow "is uninvolved" kptwr wprx "button and flower" \Rightarrow "fantastic"

It can also be some inflected form:

EXAMPLE

<u>hxlwnwt</u> hgbwhim "the+windows the+high" ⇒ "upper echelon"

WORD-LEVEL PROPERTIES

PARTIAL INFLECTION In some cases, constituents undergo a (strict) subset of the full set of inflections that they would undergo in isolation

EXAMPLE

hlk axri lbw "walk after his+heart" ⇒ "follow one's heart" hlkw axri lbm "they followed their heart" *hlkw axri lbbwtihm "they followed their hearts"

EXAMPLE

bit xwlim "house-of sick-people" ⇒ "hospital" **bit xwlik* "house-of sick-people+your"

REPRESENTING MWE CONSTITUENTS

- Each MWE constituent is realized as an atom
- Atoms represent morphemes, rather than words
- To support partial inflections (including frozen forms): ATOM Defines a constituent along with all its possible inflected forms. Atoms have the following optional sub-elements:
 - **PREFIX** Specifies that the constituent is a prefix that is an inherent part of the MWE
 - **INFLECT** Restricts the possible inflections of the constituent to those specified
 - **SUFFIX** Specifies that the constituent is a pronominal suffix that attaches to the previous atom

Representing MWE constituents

EXAMPLE (mcd šni "FROM SIDE SECOND" \Rightarrow "ON THE OTHER HAND") <item id="29000 transliterated="mcd šni"> <MWE pos="adverb"/> <atom id="1" lexiconPointer="10418"> <!-- m --><prefix/> </atom> <atom id="2" lexiconPointer="20473"> <!--><inflect state="absolute" definiteness="false" number="singular"/> </atom> <atom id="3" lexiconPointer="3561"> <!-- šni --> <inflect state= "absolute" definiteness="false" number="singular"</pre> gender="masculine"/> </atom> </item>

PARTIAL MORPHOLOGICAL INFLECTIONS

EXAMPLE (THE LEXICAL ENTRY OF *ywrk din* "LAWYER")

```
<item id="28579" transliterated="ywrk din" hprefix="true">
  <MWE pos="noun"/>
  <atom id="1" lexiconPointer="8174">
                                                  <!-- ywrk -->
   <inflect state="construct"/>
  </atom>
                                                   <!-- din -->
  <atom id="2" lexiconPointer="5208">
    <inflect number="singular"/>
  </atom>
  <atom id="3" lexiconPointer="0"> <!-- pronominal suffix -->
    <suffix/>
  </atom>
</item>
```

22 / 47

WORD-LEVEL PROPERTIES

FOSSIL WORDS Constituents that only occur in MWEs

EXAMPLE

<u>kmTxwwi</u> qšt "a stone's throw" abd yliw <u>hklx</u> "outdated" lit man dplig "without dispute"

FOSSIL WORDS

```
EXAMPLE (kmTxwwi qšt "STONE'S THROW")
<item id="27000" transliterated="kmTxwwi">
      <fossil/>
</item>
<item id="23999 transliterated="kmTxwwi qšt">
  <MWE pos="adverb"/>
  <atom id="1" lexiconPointer="27000">
                                               <!-- kmTxwwi -->
  </atom>
  <atom id="2" lexiconPointer="3507">
                                                   <!-- qšt -->
    <inflect definiteness="false" state="absolute"</pre>
             number="singular"/>
  </atom>
</item>
```

MORPHO-SYNTACTIC PROPERTIES

RETRIEVING MORPHOLOGICAL FEATURES Often, MWEs inherit some of their morphological features from those of their constituents

EXAMPLE

```
ywrk din "editor-of law" \Rightarrow "a lawyer"
ywrkt din "editor-of law" \Rightarrow "a female lawyer"
ywrki din "editors-of law" \Rightarrow "lawyers"
ywrk hdin "editor-of the+law" \Rightarrow "the lawyer"
hywrk din "the+editor-of law" \Rightarrow "the lawyer"
```

RETRIEVING MORPHOLOGICAL FEATURES

EXAMPLE (*ywrk din* "LAWYER")

```
<item id="28579" transliterated="ywrk din" hprefix="optional">
  <MWE pos="noun"
       definiteness="2" state="2" number="1" gender="1"/>
  <atom id="1" lexiconPointer="8174">
                                                  <!-- ywrk -->
   <inflect state="construct"/>
  </atom>
  <atom id="2" lexiconPointer="5208">
                                                   <!-- din -->
    <inflect number="singular"/>
  </atom>
  <atom id="3" lexiconPointer="0"> <!-- pronominal suffix -->
   <suffix/>
  </atom>
</item>
```

MORPHO-SYNTACTIC PROPERTIES

AGREEMENT AMONG CONSTITUENTS Some MWEs require agreement between the morphological features of some of their constituents

EXAMPLE

milh nrdpt "word chased" \Rightarrow "synonym"

VIOLATED AGREEMENT In some MWEs, constituents that generally agree in some morphological features violate the agreement constraints

EXAMPLE

yin hry "eye the+evil" \Rightarrow "evil eye"

AGREEMENT AMONG CONSTITUENTS

EXAMPLE (*milh nrdpt* "WORD CHASED" \Rightarrow "SYNONYM") <item id="39991" transliterated="milh nrdpt" <MWE pos="noun" state="absolute" gender="feminine" definiteness="1" number="1"/> <atom id="1" lexiconPointer="3265"> <!-- milh --> <inflect state="absolute"/> </atom> <atom id="2" lexiconPointer="10097"> <!-- nrdpt --> <inflect tense="participle" type="adjective" state="absolute" gender="feminine" definiteness="1" number="1"/> </atom> </item>

ACCOUNTING FOR SYNTACTIC FLEXIBILITY

COMPOSITIONALITY Some MWEs contain **open slots**, which can be filled by a variety of complements

EXAMPLE

išb yl X šbyh "sit on X seven (days)" \Rightarrow "mourn" išb yliw šbyh "sat on-him seven" \Rightarrow "mourn him"

CONSTITUENT ORDER The order of the constituents in most MWEs tends to be fixed, but some, especially verb phrases, still exhibit some flexibility

EXAMPLE

išb šbyh yl abiw "sat seven on his-father"

ACCOUNTING FOR SYNTACTIC FLEXIBILITY

- We add a set of attributes and elements in order to account for syntactic variability
- By default, all the constituents must appear consecutively in the order determined by the *atoms*
- If other orders are possible, all the allowed permutations are prescribed within *perm* items

29 / 47

ACCOUNTING FOR SYNTACTIC FLEXIBILITY

```
EXAMPLE (yšh imim klilwt "MADE DAYS LIKE-NIGHTS" \Rightarrow "WORK INTENSIVELY")
```

```
<item id="39991" transliterated="yšh imim klilwt">
  <MWE pos="VP" tense="1" person="1" number="1" gender="1"/>
  <atom id="1" lexiconPointer="376"><inflect/></atom> <!-- všh -->
  <atom id="2" lexiconPointer="9475">
                                                         <!-- iwm -->
    <inflect state="absolute" definiteness="false" number="plural"/>
  </atom>
  <atom id="3" lexiconPointer="20001"><prefix/></atom> <!-- k -->
                                                     <!-- lilh -->
  <atom id="4" lexiconPointer="8024">
   <inflect state="absolute" definiteness="false" number="plural"/>
  </atom>
  <perms>
   <perm id="1" order="1 2 3 4"/>
   <perm id="2" order="1 4 3 2"/>
  </perms>
</item>
```

OPEN SLOTS

EXAMPLE (*akl at X bli mlx* "EAT X WITHOUT SALT" \Rightarrow "DEFEAT")

```
<item id="23986" transliterated="akl at + bli mlx">
   <MWE pos="VP" person="1" number="1" gender="1" tense="1"/>
   <atom id="1" lexiconPointer="8442">
        <inflect/>
        </atom>
        <atom id="2" lexiconPointer="3382"/>
        <itom id="3" lexiconPointer="0">< <!-- at -->
   <atom id="3" lexiconPointer="0">< <!-- at -->
   <atom id="3" lexiconPointer="0">< <!-- at -->
   <atom id="4" lexiconPointer="0">< <!-- bli -->
   <atom id="4" lexiconPointer="21542"/>
   <atom id="5" lexiconPointer="608"/>
   </atom</pre>
```

OPEN SLOTS

EXAMPLE (*akl at X bli mlx* "EAT X WITHOUT SALT" \Rightarrow "DEFEAT")

<perms>

<perm< th=""><th>id="1"</th><th>order="1</th><th>2 3</th><th>4</th><th>5"/></th><th><!--</th--><th>akl awtw</th><th>bli</th><th>mlx</th><th>></th></th></perm<>	id="1"	order="1	2 3	4	5"/>	</th <th>akl awtw</th> <th>bli</th> <th>mlx</th> <th>></th>	akl awtw	bli	mlx	>
<perm< td=""><td>id="2"</td><td>order="2</td><td>3 1</td><td>4</td><td>5"/></td><td><!--</td--><td>awtw akl</td><td>bli</td><td>mlx</td><td>></td></td></perm<>	id="2"	order="2	3 1	4	5"/>	</td <td>awtw akl</td> <td>bli</td> <td>mlx</td> <td>></td>	awtw akl	bli	mlx	>
<perm< td=""><td>id="3"</td><td>order="2</td><td>34</td><td>5</td><td>1"/></td><td><!--</td--><td>awtw bli</td><td>mlx</td><td>akl</td><td>></td></td></perm<>	id="3"	order="2	34	5	1"/>	</td <td>awtw bli</td> <td>mlx</td> <td>akl</td> <td>></td>	awtw bli	mlx	akl	>
<perm< td=""><td>id="4"</td><td>order="4</td><td>5 1</td><td>2</td><td>3"/></td><td><!--</td--><td>bli mlx a</td><td>kl a</td><td>wtw</td><td>></td></td></perm<>	id="4"	order="4	5 1	2	3"/>	</td <td>bli mlx a</td> <td>kl a</td> <td>wtw</td> <td>></td>	bli mlx a	kl a	wtw	>
<perm< td=""><td>id="5"</td><td>order="4</td><td>52</td><td>3</td><td>1"/></td><td><!--</td--><td>bli mlx a</td><td>wtw</td><td>akl</td><td>></td></td></perm<>	id="5"	order="4	52	3	1"/>	</td <td>bli mlx a</td> <td>wtw</td> <td>akl</td> <td>></td>	bli mlx a	wtw	akl	>
<perm< td=""><td>id="6"</td><td>order="1</td><td>2 +</td><td>4</td><td>5"/></td><td><!--</td--><td>akl at +</td><td>bli</td><td>mlx</td><td>></td></td></perm<>	id="6"	order="1	2 +	4	5"/>	</td <td>akl at +</td> <td>bli</td> <td>mlx</td> <td>></td>	akl at +	bli	mlx	>
<perm< td=""><td>id="7"</td><td>order="2</td><td>+ 1</td><td>4</td><td>5"/></td><td><!--</td--><td>at + akl</td><td>bli</td><td>mlx</td><td>></td></td></perm<>	id="7"	order="2	+ 1	4	5"/>	</td <td>at + akl</td> <td>bli</td> <td>mlx</td> <td>></td>	at + akl	bli	mlx	>
<perm< td=""><td>id="8"</td><td>order="2</td><td>+ 4</td><td>5</td><td>1"/></td><td><!--</td--><td>at + bli</td><td>mlx</td><td>akl</td><td>></td></td></perm<>	id="8"	order="2	+ 4	5	1"/>	</td <td>at + bli</td> <td>mlx</td> <td>akl</td> <td>></td>	at + bli	mlx	akl	>
<perm< td=""><td>id="9"</td><td>order="4</td><td>5 1</td><td>2</td><td>+"/></td><td><!--</td--><td>bli mlx a</td><td>kl a</td><td>it +</td><td>></td></td></perm<>	id="9"	order="4	5 1	2	+"/>	</td <td>bli mlx a</td> <td>kl a</td> <td>it +</td> <td>></td>	bli mlx a	kl a	it +	>
<perm< td=""><td>id="10'</td><td>order="4</td><td>15</td><td>2 +</td><td>+ 1"/></td><td><!--</td--><td>bli mlx a</td><td>it +</td><td>akl</td><td>></td></td></perm<>	id="10'	order="4	15	2 +	+ 1"/>	</td <td>bli mlx a</td> <td>it +</td> <td>akl</td> <td>></td>	bli mlx a	it +	akl	>
	>									

PROPER NAMES

```
EXAMPLE (hnri wiliam pwrd "HENRY WILLIAM FORD")
<item id="28605" transliterated="hnri wiliam pwrd">
 <MWE pos="pName" type="person"
      number="singular" gender="masculine"/>
 <atom id="1" lexiconPointer="7356"/>
                                                <!-- Henry -->
                                              <!-- William -->
 <atom id="2" lexiconPointer="2266"/>
 <atom id="3" lexiconPointer="222"/>
                                                   <!-- W. -->
                                                <!-- Ford -->
 <atom id="3" lexiconPointer="8544"/>
 <perms>
   <perm id="1" order="1 2 4"/>
                                   <!-- Henry William Ford -->
   <perm id="2" order="1 3 4"/>
                                        <!-- Henry W. Ford -->
   <perm id="3" order="1 4"/>
                                           <!-- Henry Ford -->
                                                 <!-- Ford -->
   <perm id="4" order="3"/>
 </perms>
</item>
```

OUTLINE

- **1** INTRODUCTION
- 2 LINGUISTIC BACKGROUND
- **3** The Hebrew morphological processor
- LEXICAL REPRESENTATION OF HEBREW MWES
- **5** MORPHOLOGICAL PROCESSING OF MWES

6 CONCLUSION

34 / 47

MORPHOLOGICAL PROCESSING OF MWES

- The morphological generator embodies vast linguistic knowledge which is applicable to MWEs and to single words alike
- But the analyzer operates on a token-by-token basis
- We therefore decided not to interfere with the generator and analyzer, and instead to add a **post-processing** layer
- First, the existing morphological analyzer is applied to all the tokens of a sentence
- Then, the post-processor identifies MWEs in the analyzed output using information derived from the MWE lexicon

Morphological processing of MWEs

THE EXTENDED MORPHOLOGICAL SYSTEM



MORPHOLOGICAL PROCESSING OF MWES

- The **MWE lexicon** reflects all the information associated with MWEs
- For each MWE we choose an **anchor** word which helps identify it in the text
- When applied to the anchor, the generator produces not only all the inflected forms of that word, but also an additional analysis, as a component of the MWE that this word anchors
- This additional analysis is associated with the ID of the MWE

```
EXAMPLE (A PARTIAL ANALYSIS OF ywrkwt hdin "THE (FEMALE) LAWYERS" BEFORE POST-PROCESSING)
```

```
<token id="1" surface="ywrkwt">
 <analysis id="1">
    <base lexiconPointer="8174" transliterated="ywrk">
      <noun state="absolute" definiteness="false"
            gender="feminine" number="plural"/>
   </base>
 </analysis>
  <analysis id="2">
   <base lexiconPointer="8174" transliterated="ywrk">
      <noun state="construct" definiteness="false"
            gender="feminine" number="plural"/>
   </base>
 </analysis>
```

```
</token>
```

```
EXAMPLE (A PARTIAL ANALYSIS OF ywrkwt hdin "THE (FEMALE) LAWYERS" BEFORE POST-PROCESSING)
```

```
<token id="2" surface="hdin">
 <analysis id="1">
   <base lexiconPointer="5208" transliterated="din">
      <noun state="absolute" definiteness="true"
            gender="masculine" number="singular"/>
   </base>
 </analysis>
  <analysis id="2"/>
   <base lexiconPointer="28579" transliterated="din">
      <MWE lexiconPointer="28579" atom="2" definiteness="true"/>
   </base>
 </analysis>
</token>
```

- The post processor works on a sentence-by-sentence basis
- It checks the analyses of the tokens in the sentence to find analyses as anchors of MWEs
- For each such anchor the post-processor retrieves the entry of the corresponding MWE from the MWE lexicon
- This record contains the IDs of the remaining constituents, thereby enabling the post-processor to search for them in the sentence and verify that they satisfy the agreement and order requirements of the MWE
- Thus only one database search is needed for each anchor

EXAMPLE (A PARTIAL ANALYSIS OF *ywrkwt hdin* "THE (FEMALE) LAWYERS" **AFTER** POST-PROCESSING)

```
<token id="2" surface="hdin">
...
<analysis id="2" />
<base lexiconPointer="28579" transliterated="ywrk din">
<MWE pos="noun" definiteness="true" number="plural"
gender="feminine"/>
</base>
</analysis>
</token>
```

OUTLINE

- **1** INTRODUCTION
- 2 LINGUISTIC BACKGROUND
- **3** The Hebrew morphological processor
- LEXICAL REPRESENTATION OF HEBREW MWES
- **(5)** MORPHOLOGICAL PROCESSING OF MWES

6 CONCLUSION

- These modifications were implemented as part of the MILA tools (Itai and Wintner, 2008), and are currently part of the lexicon and the morphological processor
- The current MWE lexicon includes a total of 3718 MWEs:

POS	Noun	Adj	Prep	Adv	Intrjct	PropN	Other	Total
Count	1950	105	23	248	38	1215	139	3718

More entries are constantly added

CONCLUSION

- We focus on the special needs of Hebrew, but this architecture is in principle appropriate for several morphologically interesting languages
- The architecture satisfies many of the properties listed by Savary (2008)
- But not all MWEs can be represented
 - Constraints on the syntactic structure of potential fillers of the open slot
 - More intricate interactions of MWEs with productive syntactic structure

FUTURE DIRECTIONS

- Extension to other languages
- Automatic acquisition of Hebrew MWEs from corpora (Al-Haj and Wintner, 2010; Tsvetkov and Wintner, 2010, 2011, 2012, Forthcoming)

ACKNOWLEDGEMENTS

- Hassan Al-Haj
- Israel Science Foundation



Hassan Al-Haj, Alon Itai, Shuly Wintner. Lexical Representation of Multiword Expressions in Morphologically-complex Languages. *International Journal of Lexicography*, Under Review.

© Shuly Wintner (University of Haifa)

Conclusion

BIBLIOGRAPHY I

- Hassan Al-Haj and Shuly Wintner. Identifying multi-word expressions by leveraging morphological and syntactic idiosyncrasy. In **Proceedings of the 23rd International Conference on Computational Linguistics (COLING 2010)**, pages 10–18, Beijing, China, August 2010. Coling 2010 Organizing Committee. URL http://www.aclweb.org/anthology/C10–1002.
- Alon Itai and Shuly Wintner. Language resources for Hebrew. Language Resources and Evaluation, 42(1):75–98, March 2008.
- Alon Itai, Shuly Wintner, and Shlomo Yona. A computational lexicon of contemporary Hebrew. In Proceedings of The fifth international conference on Language Resources and Evaluation (LREC-2006), Genoa, Italy, May 2006.
- Agata Savary. Computational inflection of multi-word units, a contrastive study of lexical approaches. Linguistic Issues in Language Technology, 1:1–53, 2008.
- 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. Identification of multi-word expressions by combining multiple linguistic information sources. In Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, pages 836–845, Edinburgh, Scotland, UK, July 2011. Association for Computational Linguistics. URL http://www.aclweb.org/anthology/D11-1077.
- 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.
- Yulia Tsvetkov and Shuly Wintner. Identification of multi-word expressions by combining multiple linguistic information sources. **Computational Linguistics**, Forthcoming.
- Shuly Wintner and Shlomo Yona. Resources for processing Hebrew. In **Proceedings of the MT-Summit IX workshop on** Machine Translation for Semitic Languages, pages 53–60, New Orleans, September 2003.