Automatic Thesaurus Construction for Modern Hebrew

°Chaya Liebeskind, *Ido Dagan, *Jonathan Schler

*Computer Science Department, Jerusalem College of Technology, Lev Academic Center, Israel *Computer Science Department, Bar-Ilan University, Israel

lieb chaya @gmail.com, dagan @cs.biu.ac.il, schler @gmail.com

Abstract

Automatic thesaurus construction for Modern Hebrew is a complicated task, due to its high degree of inflectional ambiguity. Linguistics tools, including morphological analyzers, part-of-speech taggers and parsers often have limited in performance on *Morphologically Rich Languages* (MRLs) such as Hebrew. In this paper, we adopted a schematic methodology for generating a co-occurrence based thesaurus in a MRL and extended the methodology to create distributional similarity thesaurus. We explored three alternative levels of morphological term representations, surface form, lemma, and multiple lemmas, all complemented by the clustering of morphological variants. First, we evaluated both the co-occurrence based method and the distributional similarity method using Hebrew WordNet as our gold standard. However, due to Hebrew WordNet's low coverage, we completed our analysis with a manual evaluation. The results showed that for Modern Hebrew corpus-based thesaurus construction, the most directly applied statistical collection, using linguistics tools at the lemma level, is not optimal.

Keywords: Hebrew, Morphologically Rich Language, thesaurus, co-occurrence, distributional similarity

1. Introduction and Background

A thesaurus is a lexical resource which groups words together by *semantic similarity*. The notion of semantic similarity includes semantic relations, such as synonyms (*car-automobile*), hyperonyms (*vehicle-car*), hyponyms (*car-vehicle*), meronyms (*wheel-car*), and antonyms (*acceleration-deceleration*).

Generally, two statistical approaches for corpus-based thesaurus construction were explored: a first-order, cooccurrence-based approach which assumes that words that frequently occur together are topically related (Schütze and Pedersen 1997) and a second-order, distributional similarity approach (Hindle 1990; Lin 1998; Gasperin et al. 2001; Weeds and Weir 2003; Kotlerman et al. 2010) which suggests that words occurring within similar contexts are semantically similar (Harris 1968).

The aim of this research was to construct a thesaurus for Modern Hebrew. Modern Hebrew lacks repositories of machine-readable knowledge (i.e, lexical resources) fundamental to many Natural Language Processing (NLP) tasks, such as machine translation (Irvine and Callison-Burch 2013), question answering (Yang et al. 2015), and word sense disambiguation (Chen et al. 2014). Therefore, in the last decade, a few semantic resources for modern Hebrew have been constructed. Examples for such resources are Hebrew WordNet¹ (Ordan et al. 2007), which groups words into sets of synonyms, and Hebrew FrameNet (Hayoun and Elhadad 2016), which defines formal structures for semantic frames, and various relationships between and within them.

Furthermore, Hebrew is a *Morphologically Rich Language* (MRL), a language in which significant information about syntactic units and relations is expressed at word-level. Due to the rich and challenging morphology of MRLs, tools, such as part-of-speech taggers, and parsers, often perform poorly. In MRLs, commonly used statistical extraction at the lemma level, using a morphological analyzer and tagger (Lindén and Piitulainen 2004; Peirsman et al. 2008; Rapp 2008), might not be the optimal choice.

Therefore, Liebeskind et al. (2012) suggested a methodology for generating a co-occurrence based thesaurus in MRL. They explored three options for term representation: surface form, lemma, and multiple lemmas, all supplemented by term variant clustering. While the default lemma representation is dependent on tagger performance, the two other representations avoid choosing the right lemma for each word occurrence. Instead, the multiple-lemma representation assumes that the correct analysis will accumulate enough statistical prominence throughout the corpus; while by clustering term variants at the end of the extraction process, the surface representation solves morphological disambiguation "in retrospect". The input is a thesaurus entry (target term) in one of the possible term representations (surface, best, or all). Their algorithm is as follow: for each target term, retrieve all the corpus documents where the target term appears. Then, Liebeskind et al. (2012) define a set of candidate terms that consists of all the terms that appear in all these documents (this again for each of the three possible term representations). Next, a co-occurrence score (e.g., Dice coefficient (Smadja et al. 1996), Pointwise Mutual Information (PMI) (Church and Hanks 1990) and log-likelihood test (Dunning 1993)) between the target term and each of the candidates is calculated. Then, candidates are sorted, and the highest rated candidate terms are clustered into lemma-oriented clusters. Finally, the clusters are ranked by their members' co-occurrence scores and the highest rated clusters become related terms in the thesaurus. The two choices for term representation are independent, resulting in nine possible configurations of the algorithm for representing both the target term and the candidate terms. The methodology provides a generic scheme for exploring the alternative representation levels, each corpus and language-specific tool set might yield a different optimal configuration.

In this paper, we adopt and extend Liebeskind et al. (2012) algorithmic scheme to deal with second-order distributional similarity methods. In the second-order statistical approach, there is an additional dimension, feature representation. For each term in the corpus, a

¹ http://cl.haifa.ac.il/projects/mwn/index.shtml

feature vector is constructed by collecting terms that appear in its context. Each feature term is assigned a weight indicating its association to the given term. Then, secondorder similarity is calculated between the target term and all the other terms in the corpus. (e.g., Jaccard's coefficient (Gasperin et al. 2001), Cosine-similarity (Salton and McGill 1983; Ruge 1992; Caraballo 1999; Gauch et al. 1999; Pantel and Ravichandran 2004) and Lin's mutual information metric (Lin 1998)). Therefore, each term in the corpus is a potential candidate term. Yet, the ranked list of terms is considered as the candidate terms list. The three choices for term representation are independent, resulting in 27 configurations which cover the range of possibilities for term representation in second-order thesaurus. Exploring all of them in a systematic manner should reveal the best configuration for a particular setting.

In Section 2, we aim to describe our corpora, target term collection, and the experimental setting. Section 3 elaborates on the results of the algorithmic scheme for both first-order and second-order statistical extraction. We complete this section with a post-hoc manual evaluation and an error analysis. In Section 4, we suggest directions for future research.

2. Modern Hebrew Thesaurus Construction

2.1 Corpora

The MILA Knowledge Center² has acquired a number of Hebrew corpora from various domains. We used four corpora for the thesaurus construction experiment: 2 news corpora: HaAretz (11,097,790 word tokens) and Arutz 7 (15,107,618 word tokens), a financial newspaper: TheMarker (692,919 word tokens), and the sessions protocols of the Israeli parliament, HaKnesset (15,066,731 word tokens).

2.2 Target Terms Collection

We used Hebrew WordNet (Ordan et al. 2007) as our goldstandard. English WordNet (Fellbaum 1998) is a large lexical database of English. Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept. Synsets are interlinked by means of conceptual semantic and lexical relations. WordNet superficially resembles a thesaurus; words are grouped based on their meanings.

Following the success of the English WordNet, parallel networks have been developed for a variety of languages. In particular, researchers from Italy have developed a methodology for parallel development of multilingual WordNets (Bentivogli et al. 2000). The system, called MultiWordNet³, contains information on several aspects of multilingual dictionaries, including lexical relationships between words, semantic relations between lexical concepts, and several mappings of lexical concepts in different languages, etc.

Hebrew WordNet uses the MultiWordnet methodology (Bentivogli et al. 2000) and is thus aligned with English, Italian, Spanish, Portuguese, Romanian and Latin. The Hebrew WordNet currently contains 5261 synsets, with an average of 1.47 synonyms per synset, where nouns are much more frequent than other parts of speech (almost 78 percent).

Our target terms list consisted of all 3362 Hebrew WordNet terms with at least 5 appearances in our Modern Hebrew corpora. The target terms were of different parts of speech, including nouns, verbs, adjectives and adverbs.

2.3 Experimental Setting

2.3.1 Morphological Tools

We used the MILA's automatically morphologically analyzed texts. The texts are tokenized and tagged with all possible morphological analyses. Therefore, we directly extracted all the three possible term representations from the analyzed texts. The morphological analysis of the texts was performed by the MILA Hebrew Morphological Analyzer (Itai and Wintner 2008; Yona and Wintner 2008) and POS tagging was performed by the Bar Haim et al. (2008) tagger. The reported accuracy of the tagger for Modern Hebrew is 90%.

2.3.2 Evaluation measures

Following previous works (Lin 1998; Riedl and Biemann 2013; Melamud et al. 2014), we used Hebrew Wordnet to construct a large scale gold standard for semantic similarity evaluations. In our experiments, we compared the performance of our algorithms by Relative Recall (RR) and Mean Average Precision (MAP). Since each related terms list is of different length, these measures ensure that any recall increase is reflected in both scores. The recall scores are *micro-averaged*; we sum the number of extracted WordNet related terms for all the target terms for each configuration and then divide the sum by the total number of related terms for all target terms.

2.3.3 Methods

We applied the Liebeskind et al. (2012) methodology for generating a first-order co-occurrence based thesaurus and extended the methodology for generating a second-order distributional similarity thesaurus as described in Section 1. The Pointwise mutual information (Church and Hanks 1990) was used as our co-occurrence measure for both firstorder similarity and feature weighting in second-order similarity. In this paper, we focused on the word window context representation, which is the most common one for lexical similarity extraction. This representation maintains comparable performance for alternative context representations (Bullinaria and Levy 2012; Kiela and Clark 2014; Lapesa and Evert 2014)). We used a sliding window of 3 words on each side of the represented word, not crossing sentence boundaries. At the end of the extraction process, we grouped the top 50 related term variants and ranked them based on the summation approach. The summation approach adds up the group members' scores as the group score. The approach accounts for the cumulative impact of all group members, which corresponds to the morphological variants of each candidate term.

3. Results

In general, due to the low coverage of Hebrew WordNet, we expected the measured precision of our methods to be low. For example, the statistical methods extracted related

² http://www.mila.cs.technion.ac.il/resources corpora.html

³ http://multiwordnet.fbk.eu/english/home.php

terms such as m`lh⁴ (degree) and clsiws (Celsius) for the target term TmprTwrh ((physics) temperature). However, these related terms do not appear in Hebrew WordNet and thus would be judged irrelevant. In addition, WordNet makes abstract generalizations that we did not expect to be extract by statistical measures. For example, the related terms: mišhw (someone), xi (living thing) and adm (person) for the target term ab (father), or the related terms: p`wlh (action) and m`šh (deed) for the target term ktivh (writing). Therefore, our measured recall was expected to be low. Still, even though the absolute values of our evaluation measures was expected to be low.

Candidate►	_	Surface	Best	All
Target ▼				
Surface	RR	0.0172	0.0122	0.0133
	MAP	0.0057	0.0022	0.0026
	RR	0.0208	0.0111	0.0124
Best	MAP	0.0064	0.0024	0.003
	RR	0.0169	0.0092	0.0098
All	MAP	0.0051	0.0019	0.0019

Table 1: Results for first-order method

Hebrew WordNet is still a valuable resource for comparing performance of different statistical methods. In addition, we performed a post-hoc manual evaluation, which better indicated our performance level in absolute terms. The manual evaluation also shows the contribution of statistical methods to the construction and enrichment of lexical resources for resource-poor languages, such as Hebrew. In the next subsections, we report the results of the firstorder and second-order thesaurus construction methods. We examine the thee levels of term representations: surface form (Surface), best lemma, as identified by a POS tagger (Best), and all possible lemmas produced by a morphological analyzer (All).

3.1 First-order Results

Table 1 compares the performance of all nine term representation configurations for the first-order method. The lemma-based representations of the target term outperformed its surface representation. The second-best results were obtained from the candidate representation at the surface level, which was complemented by grouping term variants to lemmas in the grouping phase. The improvement over the common default best lemma representation, for both target and candidate, is statistically significant at the 0.01 level for both MAP and RR.

Target►		Surface			Best			All		
Candidate►		Surface	Best	All	Surface	Best	All	Surface	Best	All
Feature ▼	RR	0.0413	0.0137	0.0125	0.0452	0.0203	0.0216	0.0434	0.0216	0.0245
Surface	MAP	0.0174	0.0103	0.0107	0.0204	0.0181	0.0143	0.0271	0.0181	0.0209
	RR	0.0452	0.0161	0.017	0.048	0.0208	0.0269	0.0454	0.0244	0.0259
Best	MAP	0.0164	0.0127	0.0119	0.0197	0.0064	0.0147	0.0261	0.0183	0.0197
	RR	0.0414	0.0155	0.0151	0.0488	0.0209	0.0254	0.0441	0.0244	0.0245
All	MAP	0.0164	0.0127	0.01	0.0189	0.0124	0.0147	0.0266	0.018	0.0197

Table 2: Results for second-order met	hod
---------------------------------------	-----

Target term	WordNet related terms	DistSim related terms
akiph (enforcement)	bcw` (execution)	piqwx (supervision), bqrh (inspection)
bkwx (aggressively)	xzq (strong)	`wcmh (strength)
zkrwn (memory)	qwgnicih (psychology) cognition),	andrTh (monument), azkrh (memorial)
	zkr memory (of someone or something)	
TmprTwrh	qr (cold), xm (hot)	m_lh (degree), clsiws (Celsius)
((physics) temperature)		
nwrmTibi (normative,	pwsqni ((linguistics) normative)	nwrmli (normal), tqin (in order), nwrmh (norm)
regular)		
crxh (scream)	tq_srt (communication), cwxh	c`q (to shout), dibwr (speaking)
	(shout), amirh (saying)	
šitwq (paralysis)	štwq (paralysis)	qipawn (standstill), hšbth (stoppage)
ti`wb (disgust)	mšTmh ((flowery) hatred),	bwz (contempt), slidh (disgust), ginwi (denunciation)
	aibh (loathing), šnah (hatred)	

Table 3: WordNet vs. distributional similarity (DistSim) related terms

⁴ To facilitate readability, we use a transliteration of Hebrew using Roman characters; the letters used, in Hebrew lexicographic order, are abgdhwzxTiklmns`pcqršt.

3.2 Second-order Results

The performance of all the 27 representation configurations is presented in Table 2. The best MAP score of the secondorder method was obtained from target term representation at the all-lemma level, feature representation at the surface level and candidate representation at the surface level. The improvement over the common default best lemma representation, for target, feature and candidate, is statistically significant at the 0.01 level for both MAP and RR. The common default best lemma representation has relatively low performance. Its recall is ranked 20 out of the 27 configurations and its MAP is the lowest, while, even the default surface representation for both target, feature and candidate has higher rank (9 and 13 respectively).

Rapp (2002) observed that there is a relationship between the type of computation performed (first-order versus second-order) and the type of the extracted association (syntagmatic versus paradigmatic). Whereas the results of the second order computations are exclusively paradigmatic, the first order computations are a combination of both syntagmatic and paradigmatic associations. Since we limited preferences to lexical similarity, a thesaurus based on second order associations is better than one based on first-order associations.

We analyzed the results of the best configuration (*all-surface-surface*) and found that due to the limitations of the Hebrew WordNet some truly related terms found by our system were judged irrelevant. Since these terms do not appear in Hebrew WordNet, they were counted as false positives and decreased the overall MAP. Table 3 shows examples of target terms with their WordNet related terms and additional related terms that were extracted by the second-order algorithm.

3.3 Manual evaluation

Due to the low coverage of Hebrew WordNet and the poor results in the previous analysis, we performed a post-hoc manual evaluation to assess our preliminary findings. We randomly selected 30 target terms and annotated their results.

First, we compared the related term extracted by WordNet to the related terms extracted by the best distributional similarity configuration (*all-surface-surface*). While WordNet relative recall was only 0.23, the MAP of the best configuration was 0.38 and its relative recall was 0.78. The intersection between the two sources of related terms was low.

Next, we chose 4 configurations; 2 first-order configurations and 2 second-order configurations. In each configuration pair, the first was the best performing configuration and the second was the default best lemma representation for all the term levels. The comparison results are presented in Table 4.

The ranking of the four configurations is consistent with the full ranking of all the possible configurations of our previous evaluation (Tables 1 and 2). Since we did not manually compare all the configurations, we cannot guarantee that the full rankings are consistent. However, we can still assess the benefits of our methodology by assessing its statistical significance. We used the one-sided Wilcoxon signed-rank test (Wilcoxon 1945) and observed that the advantage of the best first-order configuration over the default best lemma representation is statistically significant at the 0.05 level for MAP. While the advantage of the best second-order configuration over the default best lemma representation is statistically significant at the 0.01 level for both RR and MAP. The relatively high MAP score of the best second-order configuration is consistent with the results in prior research for statistical thesaurus construction methods.

The best results for both the first and second order methods were obtained from candidate representation at the surface level, which was complemented by grouping term variants to lemmas in the grouping phase. The effectiveness of solving morphological disambiguation "in retrospect" was significant. Considering the relatively high accuracy of the tagger on Modern Hebrew corpora (90%), this finding is somewhat surprising.

Since available morphological tools were designed for processing modern Hebrew, we hardly recognized lemmatization errors in the grouping phase. The lemmatizer's lexicon covers modern Hebrew well and no invalid derivations were found in the modern corpora.

We also analyzed the results of the best configuration, the *all-surface-surface* configuration of the second-order method. We observed that about 15% of the related terms groups had a different lemma than the target term, but shared a similar root with the target term, often corresponding to a morphological derivation rather than an inflection. For example: kwšl (failing) - kišlwn (failure), akilh (eating) - makl (food), hmcah (invention) - mmcia (inventor), and dliqh (fire) - dliq (flammable).

In addition, we performed an error analysis and found that 62% of the related terms groups were indeed irrelevant. However, 38% of the groups shared a broader context with the target term and were judged irrelevant in the current use case. This broader contextual group may include verbs associated with the target term or its related terms, or different nouns associated with the related terms. For example: akilh (eating) - dg (fish), swkr (sugar); bkwx (aggressively) – xiil (soldier); bki (crying) - xrdh (fear); and lqixh (taking) - hlwwah (loan), pir`wn (redemption).

Due to the problematic gold-standard (Hebrew Wordnet), we do not have a decisive conclusion on the best configuration for term representation. However, we did demonstrate the importance of the methodological scheme by showing that the default configuration is definitely not the optimal one.

Configuration	RR	MAP
best-surface	0.18	0.05
best-best	0.15	0.03
all-surface-	0.55	0.28
surface		
best-best-best	0.19	0.1

Table 4: Manual comparison of 4 configurations

4. Conclusions and Future work

We constructed a Modern Hebrew thesaurus by adopting and extending a first-order co-occurrence based methodological method to second-order distributional similarity method. We showed that our methodology is effective for constructing a more comprehensive thesaurus for MRLs. Our automatic thesaurus construction tool outperforms the current term representation and yields an optimal configuration. We plan to extend our methodology to deal with Multi Word Expressions (MWE).

5. Bibliographical References

- Bar-haim, R., Sima'an, K., and Winter, Y. (2008). Part-ofspeech tagging of modern hebrew text. Natural Language Engineering, 14(2), 223–251.
- Bullinaria, J. A., and Levy, J. P. (2012). Extracting semantic representations from word co-occurrence statistics: stop-lists, stemming, and SVD. Behavior research methods, 44(3), 890–907.
- Caraballo, S. A. (1999). Automatic construction of a hypernym-labeled noun hierarchy from text. In Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics (pp. 120–126). Association for Computational Linguistics.
- Chen, X., Liu, Z., and Sun, M. (2014). A unified model for word sense representation and disambiguation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP) (pp. 1025–1035).
- Church, K. W., and Hanks, P. (1990). Word association norms, mutual information, and lexicography. Computational linguistics, 16(1), 22–29.
- Dunning, T. (1993). Accurate methods for the statistics of surprise and coincidence. Computational linguistics, 19(1), 61–74.
- Gasperin, C., Gamallo, P., Agustini, A., Lopes, G., and Lima, V. de. (2001). Using Syntactic Contexts for Measuring Word Similarity. In the Workshop on Semantic Knowledge Acquisition and Categorisation (ESSLI 2001).
- Gauch, S., Wang, J., and Rachakonda, S. M. (1999). A corpus analysis approach for automatic query expansion and its extension to multiple databases. ACM Transactions on Information Systems (TOIS), 17(3), 250–269.
- Harris, Z. S. (1968). Mathematical Structures of Language. New York, NY, USA: Wiley.
- Hindle, D. (1990). Noun classification from predicateargument structures. In Proceedings of the 28th annual meeting on Association for Computational Linguistics (pp. 268–275). Association for Computational Linguistics.
- Irvine, A., and Callison-Burch, C. (2013). Combining bilingual and comparable corpora for low resource machine translation. In Proceedings of the eighth workshop on statistical machine translation (pp. 262– 270).

- Itai, A., and Wintner, S. (2008). Language resources for Hebrew. Language Resources and Evaluation, 42(1), 75–98.
- Kiela, D., and Clark, S. (2014). A systematic study of semantic vector space model parameters. In Proceedings of the 2nd Workshop on Continuous Vector Space Models and their Compositionality (CVSC) at EACL (pp. 21–30).
- Kotlerman, L., Dagan, I., Szpektor, I., and Zhitomirsky-Geffet, M. (2010). Directional distributional similarity for lexical inference. Natural Language Engineering, 16(4), 359–389.
- Lapesa, G., and Evert, S. (2014). A Large Scale Evaluation of Distributional Semantic Models: Parameters, Interactions and Model Selection. Transactions of the Association for Computational Linguistics, 2, 531–545.
- Liebeskind, C., Dagan, I., and Schler, J. (2012). Statistical Thesaurus Construction for a Morphologically Rich Language. In *SEM 2012: The First Joint Conference on Lexical and Computational Semantics (pp. 59–64). Montréal, Canada.
- Lin, D. (1998). Automatic retrieval and clustering of similar words. In Proceedings of the 17th international conference on Computational Linguistics-Volume 2 (pp. 768–774). Association for Computational Linguistics.
- Lindén, K., and Piitulainen, J. (2004). Discovering Synonyms and Other Related Words. In S. Ananadiou and P. Zweigenbaum (Eds.), COLING 2004 CompuTerm 2004: 3rd International Workshop on Computational Terminology (pp. 63–70). Geneva, Switzerland.
- Melamud, O., Dagan, I., Goldberger, J., Szpektor, I., and Yuret, D. (2014). Proceedings of the Eighteenth Conference on Computational Natural Language Learning (pp. 181–190). Ann Arbor, Michigan: Association for Computational Linguistics.
- Pantel, P., and Ravichandran, D. (2004). Automatically labeling semantic classes. In Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics: HLT-NAACL 2004.
- Peirsman, Y., Heylen, K., and Speelman, D. (2008). Putting things in order. First and second order context models for the calculation of semantic similarity. In 9es Journées internationales d'Analyse statistique des Données Textuelles (JADT). Lyon, France.
- Rapp, R. (2002). The Computation of Word Associations: Comparing Syntagmatic and Paradigmatic Approaches. In COLING (Vol. 1, pp. 1–7). Stroudsburg, PA, USA.
- Rapp, R. (2008). The automatic generation of thesauri of related words for English, French, German, and Russian. International Journal of Speech Technology, 11(3–4), 147–156.
- Riedl, M., and Biemann, C. (2013). Scaling to Large3 Data: An Efficient and Effective Method to Compute Distributional Thesauri. In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, EMNLP 2013, 18-21 October 2013, Grand Hyatt Seattle, Seattle, Washington, USA, A meeting of SIGDAT, a Special Interest Group of the ACL (pp. 884–890).

- Ruge, G. (1992). Experiments on Linguistically-Based Term Associations. Information Processing and Management, 28(3), 317–332.
- Salton, G., and McGill (Eds.). (1983). Introduction to Modern Information Retrieval. McGraw-Hill.
- Schütze, H., and Pedersen, J. O. (1997). A cooccurrencebased thesaurus and two applications to information retrieval. Information Processing and Management, 33(3), 307–318.
- Smadja, F., McKeown, K. R., and Hatzivassiloglou, V. (1996). Translating collocations for bilingual lexicons: a statistical approach. Computational Linguistics, 22(1), 1–38.
- Weeds, J., and Weir, D. (2003). A general framework for distributional similarity. In Proceedings of the 2003 conference on Empirical methods in natural language processing (pp. 81–88). Association for Computational Linguistics.
- Wilcoxon, F. (1945). Individual Comparisons by Ranking Methods. Biometrics Bulletin, 1(6), 80–83.
- Yang, Y., Yih, W., and Meek, C. (2015). Wikiqa: A challenge dataset for open-domain question answering. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (pp. 2013– 2018).
- Yona, S., and Wintner, S. (2008). A finite-state morphological grammar of hebrew. Natural Language Engineering, 14(2), 173–190.

6. Language Resource References

Bentivogli, L., Pianta, E., and Pianesi, F. (2000). Coping with Lexical Gaps when Building Aligned Multilingual Wordnets. In Proceedings of the Second International Conference on Language Resources and Evaluation, LREC 2000, 31 May - June 2, 2000, Athens, Greece.

Fellbaum, C. (1998). WordNet. Wiley Online Library.

- Hayoun, A., and Elhadad, M. (2016). The Hebrew FrameNet Project. In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016). Portorož, Slovenia: European Language Resources Association (ELRA).
- Ordan, N., Ilan, B., Ordan, N., and Wintner, S. (2007). S.: Hebrew WordNet: a test case of aligning lexical databases across languages. International Journal of Translation, 19, 39–58.