

Building a Hebrew Semantic Role Labeling Lexical Resource from Parallel Movie Subtitles

Ben Eyal, Michael Elhadad

Dept. of Computer Science, Ben-Gurion University of the Negev

Beer Sheva, Israel

bene@post.bgu.ac.il, elhadad@cs.bgu.ac.il

Abstract

We present a semantic role labeling resource for Hebrew built semi-automatically through annotation projection from English. This corpus is derived from the multilingual OpenSubtitles dataset and includes short informal sentences, for which reliable linguistic annotations have been computed. We provide a fully annotated version of the data including morphological analysis, dependency syntax and semantic role labeling in both FrameNet and PropBank styles. Sentences are aligned between English and Hebrew, both sides include full annotations and the explicit mapping from the English arguments to the Hebrew ones. We train a neural SRL model on this Hebrew resource exploiting the pre-trained multilingual BERT transformer model, and provide the first available baseline model for Hebrew SRL as a reference point. The code we provide is generic and can be adapted to other languages to bootstrap SRL resources.

Keywords: SRL, FrameNet, PropBank, Hebrew, Cross-lingual Linguistic Annotations

1. Introduction

Semantic role labeling (SRL) is the task that consists of annotating sentences with labels that answer questions such as “Who did what to whom, when, and where?”, the answers to these questions are called “roles.”

Two major SRL annotation schemes have been used in recent years, FrameNet (Baker et al., 1998) and PropBank (Palmer et al., 2005). Research in SRL has generated fully-annotated corpora in various languages, several CoNLL shared tasks, and many automatic SRL systems. Besides its important theoretical role, SRL has been found useful for information extraction (Christensen et al., 2010; Stanovsky et al., 2018) and question answering (FitzGerald et al., 2018).

In this work, we set out to contribute a new Hebrew corpus and lexical resources to further the advancement of the Hebrew FrameNet Project and to provide the first SRL resource covering both FrameNet and PropBank annotations (Hayoun and Elhadad, 2016). Our approach bootstraps Hebrew SRL annotations by projecting predicted annotations from English to Hebrew aligned sentences. To ensure that the resulting Hebrew annotations are reliable, we design confidence metrics on both the original English annotations and the projection method itself.

We describe previous work in multilingual FrameNet and cross-lingual SRL. We then describe the large unannotated dataset we selected, OpenSubtitles, and the method we use to align documents and sentences between English (EN) and Hebrew (HE), and to project SRL annotations from EN to HE. We present the dataset that is obtained as a result of this procedure, which includes aligned annotated sentences in EN and HE, with morphological analysis, dependency syntax and SRL in FrameNet and in PropBank formats. Finally, we present baseline automatic SRL systems in HE trained on the generated data.

Our main contribution in the paper is the description of an end to end method to bootstrap SRL datasets in low-resource settings through alignment, prediction and pro-

jection. We provide full source code and datasets for the described method https://github.com/bgunlp/hebrew_srl. The resulting dataset is the first available SRL resource for Hebrew.

1.1. FrameNet

This work is part of a larger effort to develop a Hebrew FrameNet (Hayoun and Elhadad, 2016). FrameNet is an annotation scheme and a lexical database inspired by frame semantics (Fillmore, 1976). Fillmore presented the concept of a ‘semantic frame’ as

A system of concepts related in such a way that to understand any one of them you have to understand the whole structure in which it fits.

Baker et al. (1998) introduced a linguistic resource called FrameNet, on which work is ongoing to this day. The purpose of FrameNet is to realize the idea of frame semantics in English, by building a lexical database of annotated examples of how various words are used in actual texts, grouped by semantic frame.

The project defines a formal structure for semantic frames, and various relationships among them Ruppenhofer et al. (2016). The FrameNet lexical database is comprised of the following elements, which defines useful terminology for all SRL resources:

Frames. Each frame contains a list of frame evoking words, such as “bought” in the sentence “John bought a car from Jane.” These are known as Frame Evoking Elements (FEEs) or Lexical Units (LUs). Additionally, each frame defines a list of participants and a list of constraints on and relationships between these participants. The participants are called Frame Elements (FEs).

Lexical Units. Formally, an LU is a word lemma paired with a coarse part-of-speech tag and is unique within its frame. For example, both the words **bought**

and **buying** are represented by the LU **buy.v** in the COMMERCE.BUY frame.

In the FrameNet formalism, LUs can have almost any part-of-speech tag. As an example consider the noun **purchase** (as in “a purchase was made”), which is one of the LUs in the COMMERCE.BUY frame. Verbs, however, are the most common LUs. In this work we only consider verbal LUs.

A **target** is any constituent which evokes a frame. It is the instance of an LU in a given piece of text. For example, both **buying** and **bought** could be targets which evoke the COMMERCE.BUY frame and they represent the LU **buy.v**.

Frame Elements. FrameNet classifies frame elements in terms of how central they are to the frame. The two primary levels of centrality are labeled “core” and “peripheral”.

An FE is classified as a *core* FE if it instantiates a conceptually necessary component of a frame, while making a frame unique. For example, in the COMMERCE.BUY frame, the FEs **Buyer** and **Goods** are considered core elements, while the FE **Money** (representing the thing given in exchange for the goods) is not.

In determining which FEs are core in their frame, a few formal properties are considered which may provide evidence for core status:

- When an element must be explicitly expressed, it is core. For instance, *resemble* requires two entities to compare.
- If an element receives a definite interpretation when omitted, it is core. For example, the sentence “John arrived.” is incomplete if the goal location at which John arrived cannot be inferred from context.

FEs that do not introduce additional, independent or distinct events from the main reported event are classified as *peripheral*. Peripheral FEs mark notions such as time, place, manner, degree of the main event represented by the frame.

FrameNet in English is currently at version 1.7 which consists of 1,221 frames, 13,572 LUs, and 11,428 FEs. For training purposes, there are 10,147 sentences in 107 fully-annotated texts.

1.2. FrameNet in other Languages

A great effort is made to expand FrameNet to other languages. As of today, FrameNets have been developed in Finnish (Lindén et al., 2017), Spanish (Subirats and Petruck, 2003), German (Burchardt et al., 2006), Japanese (Ohara et al., 2004), Chinese (You and Liu, 2005), Korean (Nam et al., 2014), Brazilian Portuguese (Torrent and Ellsworth, 2013), Swedish (Borin et al., 2010), French (Candito et al., 2014), Danish (Bick, 2011), Polish (Zawisławska et al., 2008), Italian (Tonelli and Pianta, 2008),

Slovenian (Lönneker-Rodman et al., 2008), and of course, Hebrew (Petrucek, 2005; Hayoun and Elhadad, 2016).

The methods to develop FrameNets presented in these papers are quite similar: assume the universality of the English frame inventory, and under that assumption, tag, either manually or semi-automatically, sentences in the desired language using this inventory. Almost every language has its specific corner cases where the English frames are insufficient. In those cases, usually new frames are created specifically for the language in question. An example of a corner case of Hebrew is multi-word lexical units like *give up* and *turn in* - in Hebrew, these LUs might not appear as contiguous words. This case was solved by allowing annotation of discontinuous units.

1.3. Annotation Projection

The idea of annotation projection is used in most research addressing cross-lingual SRL and linguistic annotation (Yarowsky et al., 2001; Padó, 2007; Padó and Lapata, 2009; van der Plas et al., 2011). In this line of work, we start with a resource-rich source language, usually English, and a parallel corpus of English and a resource-poor target language. The English sentences are annotated using an automatic tool, a semantic role labeling model in our case, and using word alignment, the annotations are projected to the target language. This process is called “direct projection.” The idea of “filtered projection” is introduced in Akbik et al. (2015): only alignments which satisfy the suggested filters are kept, while the rest are discarded. Such filters include verb filter (discard sentences where the predicate is not aligned to a verb), and translation filter (discard sentences where the aligned predicate is not a translation of the source predicate). We find in this work that filtered projection is essential to produce reliable annotations in Hebrew. In the rest of the paper, we present related work in the field of cross-lingual SRL annotation projection and the method we used to construct a Hebrew SRL dataset starting from a large parallel corpus of English/Hebrew sentence pairs. We eventually train a Hebrew SRL system on the data we produce and report on its performance on the automatically generated data.

2. Related Work

A long tradition of research has investigated how to create FrameNets and PropBanks in multiple languages using annotation projection.

For FrameNet, Padó and Lapata (2009) present an approach to automatically create a German FrameNet which formulates the search for a semantic alignment (an alignment in which each pair of aligned constituents are semantically equivalent) as an optimization problem on a bipartite graph. They report an F1 measure of 56.0 for Frame Elements prediction in German. Annesi and Basili (2010) set out to create an Italian FrameNet by applying a Hidden Markov Model to project annotations from English to Italian, with an F1 measure of 60.3 for Frame Elements prediction in Italian.

van der Plas et al. (2011) use the PropBank-style annotations for their system (PropBank resources enjoy more annotated data than FrameNet). Their target language

is French, and the pipeline consists of training a French syntactic parser, transfer the English roles to French via word alignments, and finally, train a French joint syntactic-semantic parser on the French syntactic and semantic annotations. They report argument labeling performance in French of F1=65.

Akbik et al. (2015) contributed PropBanks in seven languages (Arabic, Chinese, French, German, Hindi, Russian, Spanish) using filtered annotations projection¹. The reported argument labeling performance ranges between 65.0 (Arabic) and 82.0 (Chinese). Filtering projections significantly improves on previous work: for example, for French, the reported performance for argument exact match is F1=80.0 compared to the 65.0 reported above.

More recently, Aminian et al. (2019) presented a deep bidirectional character-level LSTM encoder-decoder model which uses annotation projection to label new languages, without using any syntactic features such as part-of-speech tags and dependency trees. This end to end model improves slightly over the baseline results presented in Akbik et al. (2015).

A recent entry to the list of automatic cross-lingual SRL systems is by Daza and Frank (2019), which does not use annotation projection in order to label semantic roles in a different language, but rather learns to simultaneously translate and label the input sentence. The reported performance F1=77.2 for German and F1=72.4 for French.

3. Data Collection

We adopt in this work the method of filtered projection of Akbik et al. (2015), and present a pipeline of filters that we introduce to control the quality of the generated dataset. We start from a noisy collection of aligned documents in large quantities (23.7M aligned sentences) in the OpenSubtitles 2016 dataset (Lison and Tiedemann, 2016).

We apply filters of different types on this data: starting with language identification, we push the data in Hebrew and English through automatic NLP annotation (POS, dependency parsing and SRL for English), we compute word-level alignment. We then design a method to rank annotated sentence pairs in terms of syntactic plausibility (that is how likely it is that the English and Hebrew parse trees will project onto each other in a clean manner). Modern Hebrew is characterized by a rich morphological system which impacts annotation schemas, especially because function words are often agglutinated together with the content words they modify. In terms of syntax, Modern Hebrew is mainly an SVO word order with no explicit case marking (except for pronouns). Personal pronouns are often skipped when the verb morphological inflection provides clear cues (e.g., “ahav-ti / I liked” instead of “ani ahav-ti”).

The overall data annotation process is summarized in Fig. 1. We describe the steps of this process in the following paragraphs.

¹<https://github.com/System-T/UniversalPropositions>

3.1. Filtering Noisy Subtitles

The OpenSubtitles 2016 dataset (Tiedemann, 2012) provides aligned data in 62 languages from movie subtitles. It specifically includes 23.7 million English-Hebrew sentence pairs. Due to the nature of the dataset, some of the pairs are very noisy. Examples of such noise include special Unicode characters unrelated to the text, e.g., musical note symbol to let hearing impaired viewers know that a song is playing, joined words (either typos or OCR artefacts) such “Amanonce” instead of “A man once,” and pairs in which the Hebrew part is actually found in English, not translated.

Given the large number of available sentence pairs, we filter noisy pairs in the dataset. The removal of noise consists of running fastText’s language detection model (Joulin et al., 2016b; Joulin et al., 2016a) on each pair, and keeping only English-Hebrew pairs. After this first filtering, we have 22.4M sentence pairs.

3.2. Hebrew Preprocessing

The importance of Hebrew preprocessing in our work is twofold: (a) we use features from the dependency parse tree later on in the pipeline, and (b) it acts as a quality gate for our data.

The preprocessing pipeline consists of morphological analysis, morphological disambiguation, and dependency parsing, all using the YAP system (More et al., 2019). In Hebrew, common words including prepositions, conjunctions and articles are written in an aggregated manner. For example, the written form for the phrase “in the house” (be ha bayit) will be written as a single token (babayit). For the goal of SRL, it is particularly relevant to segment such compound tokens so that prepositions and conjunctions can be properly aligned with their English counterparts and projected with the corresponding English sentences.

Before segmentation, the data includes 118M tokens from 2.4M word types. We predict that after segmentation, the number of tokens will rise, but the number of word types will be significantly lower. Indeed, after segmentation we are left with 188M tokens and 0.89M word types. This is a surprising number, as one would expect the number of types in Hebrew to be approximately 200-300K, not 890K. A quick check shows that 782K word types appear less than ten times in the entire dataset, leaving us with a more reasonable vocabulary of 112K word types.

Naturally, segmentation also affects sentence length, with 5.13 words before segmentation, and 8.17 after. Both of these numbers indicate that the sentences in the dataset are very short. The genre that this dataset represents is informal spoken language, with a high prevalence of personal pronouns (I, you, he, she) and modals (can, should). The three most prominent parts-of-speech are Nouns with 22M instances, personal pronouns with 13M instances, and verbs with 13M instances (including modals).

Dependency parse tree have an average depth of 2.96, which confirms that the average sentence is very short and simple, containing most often a single predicate with its arguments. We discuss below that as part of the overall preprocessing pipeline, we filter sentences that are either too short or too shallow to be of interest for SRL purposes.

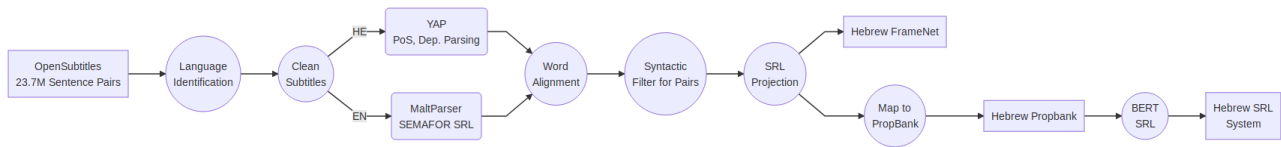


Figure 1: Overall SRL dataset production pipeline

3.3. English Preprocessing

We apply the pre-trained SEMAFOR system (Das et al., 2014) on English sentences to obtain syntax and SRL annotations. As part of SEMAFOR’s pipeline, pre-processing is done on the input sentences using MaltParser (Nivre et al., 2007) pre-trained on sections 02-21 of the WSJ section of the Penn Treebank.

On the English side, the dataset contains 148M tokens consisting of 1.5M types, with sentence length averaging at 6.45. 41M of them are nouns, 36M personal pronouns, and 23M verbs. Compare the 36M personal pronouns in English with 13M in Hebrew. This is a known syntactic aspect in Hebrew where personal pronouns are often unmarked and understood from the morphological inflection of the main verb. Such a mismatch between EN and HE challenges the methods for constituent alignment. The depth of the dependency parse tree in English is slightly smaller than in Hebrew, averaging at 2.44.

3.4. Semantic Role Labeling

The actual output from SEMAFOR consists of 55M instantiated frames across the entire dataset, but only 784 unique frames were evoked out of the 1,020 frames available in FrameNet v1.5, on which SEMAFOR is trained. On average, each sentence contains 2.4 frames and 1.17 frame elements per frame. Before filtering, the most general frame elements are also the most frequent - 3.9M instances of ENTITY, 3.3M instances of AGENT, and 3.1M instances of THEME.

After the complete filtering pipeline, however, the most frequent frames are STATEMENT, ARRIVING, BECOMING, MOTION and KILLING (reflecting the violent nature of movies) and the most frequent frame elements are THEME, SPEAKER, MESSAGE, AGENT and GOAL.

3.5. Computing Word Alignments

The original data obtained from OpenSubtitles provides aligned sentence pairs. In order to project SRL annotations from English to Hebrew, we must compute word alignment. To this end, we use the fast_align method of Dyer et al. (2013) on the set of aligned sentences. The output provides a directional mapping from English tokens in a sentence to the corresponding tokens in the associated Hebrew sentence. This alignment is not necessarily one to one.

The word alignment process identified 181M word pairs. On average, a single English token is mapped to 1.25 Hebrew after segmentation. Of these 181M pairs, 115M are one-to-one alignments, 29M are cases of 1-n alignments. In the case of a 1-n alignment, the single English token is mapped to multiple Hebrew tokens and these 29M cases cover 66M pairs.

Within the 181M collected pairs, we identify 17.6M distinct word pairs (for the full vocabulary of about 1.5M distinct words in English). Given the fact that Hebrew has rich morphology, the average rate of distinct Hebrew tokens associated to the same English token is as expected (each English form can be mapped to multiple inflected Hebrew forms in addition to the natural expected lexical diversity).

Manual inspection indicated that when one English word is mapped to many Hebrew tokens (1-3 or more), it is an indication of poor data quality. This feature was, therefore, added in the pair quality filter described below.

3.6. Span Projection

On the basis of the word alignment, it is possible to project the English SRL annotations onto the Hebrew sentence. The projection process traverses the dependency syntax tree recursively: it identifies the head word of an annotated span in English, maps it to the word aligned in Hebrew, find its subtree in the Hebrew side, and annotates it accordingly.

This simple projection method cannot work well in the presence of noisy word alignments. For example, if the head word of a span in English is aligned with zero or more than one word in Hebrew, the procedure will not produce aligned constituents. Given the abundance of data, we simply skip these problematic alignment cases. Out of the 22M sentence pairs, 6.8M annotated pairs are left after this projection filter.

3.7. Filtering Sentence Pairs

After preprocessing both English and Hebrew, we filter the 6.8M sentence pair in a way that reflects our confidence in the alignment between the annotated English and Hebrew syntactic structures.

To this end, we built a graphical user interface to visualize the word alignment, dependency tree of both English and Hebrew, the English SRL annotations, and the Hebrew SRL projections as shown in Fig. 2. This diagram summarizes all the NLP annotations available on both languages and the mapping between the SRL spans in a synthetic manner.

We manually annotated 124 sentences as one of the following six options:

- Error in sentence alignment *
- Poor translation *
- Error in word alignment
- Poor syntactic parsing
- Poor frame parsing
- Good

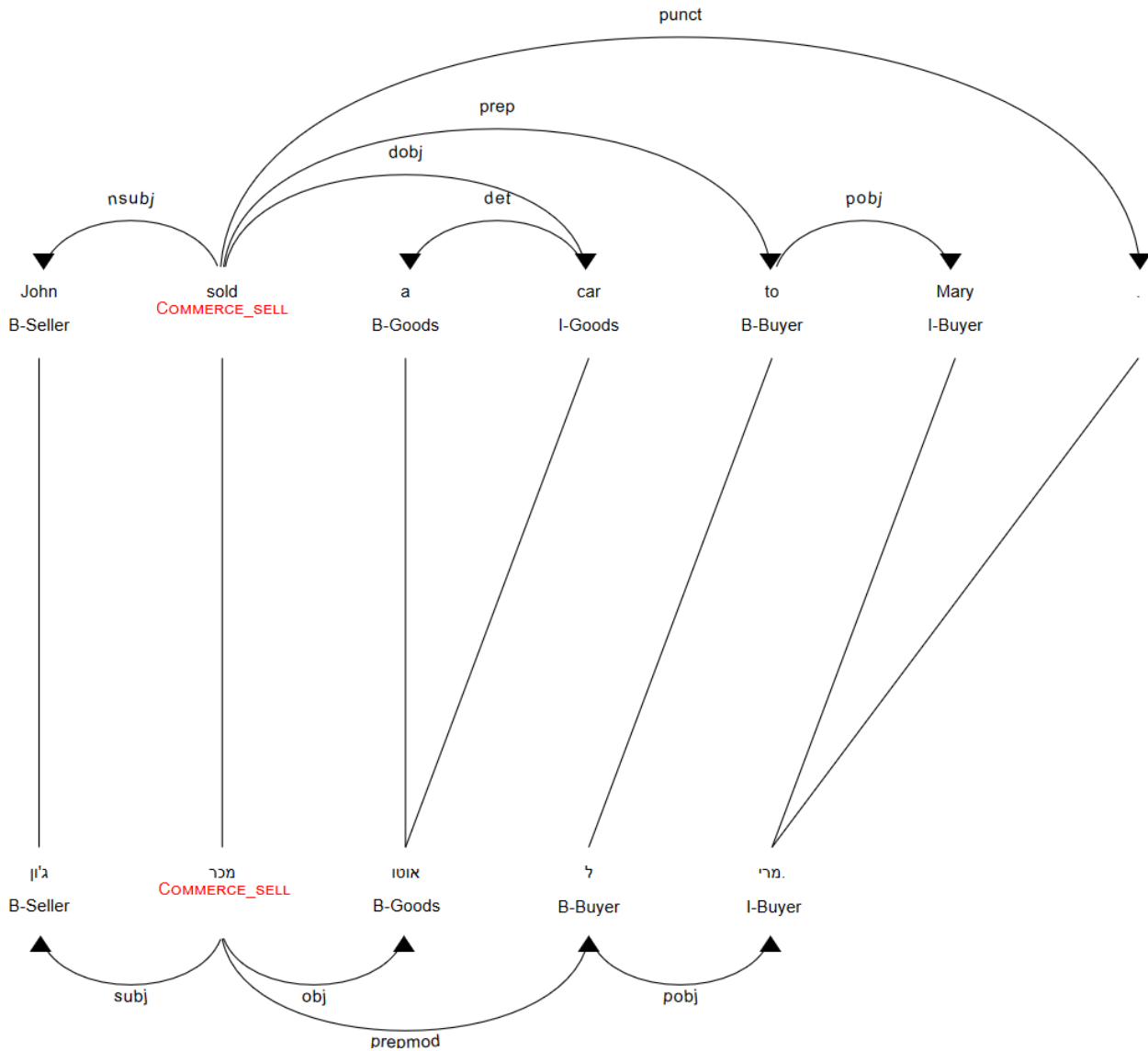


Figure 2: Visualization of the English to Hebrew SRL Projection

The items marked with an asterisk are problems in the dataset itself, *i.e.*, better tools cannot improve these sentences' quality.

Using these annotations and random resampling, we train a linear classifier to automatically annotate other sentences. We annotate a new sentence if the classifier gives it more than 80% of being "Good". For training, we use the following features:

- Sentence lengths (English/Hebrew)
- English-Hebrew sentence length ratio
- Number of frames in the English sentence
- Number of one-to-one word alignments, *i.e.*, which align one English word to one Hebrew word
- Number of one-to-many word alignments
- Dependency parse tree depth (English/Hebrew)

We interpret the prediction returned by the classifier as a score indicating the quality of the SRL match between the

English and Hebrew sentences. Higher scores are associated with longer sentences (for example, with a score over 0.80, the average sentence length is 19.74 words vs. 8.2 for the whole data) with good sentence and word alignments when verified manually.

By adjusting the threshold on the reliability score produced by this classifier, we can control how many sentence pairs we use in the training phrase of the pipeline. The distribution of sentence pairs scores and their length is shown in Fig. 3, where the numbers on top indicate for each bin, the number of sentence pairs with a score above a given threshold and the average length of the Hebrew sentences with a score above this threshold. For example, with a threshold above 0.70, there are 422.8K sentence pairs with an average length of 14.25 segmented Hebrew tokens.

3.8. Mapping to PropBank

Given a Hebrew SRL resource, we can train a Hebrew SRL system. Because there are no available FrameNet parsers for languages other than English, we decided to map the projected FrameNet annotations we collected to PropBank-

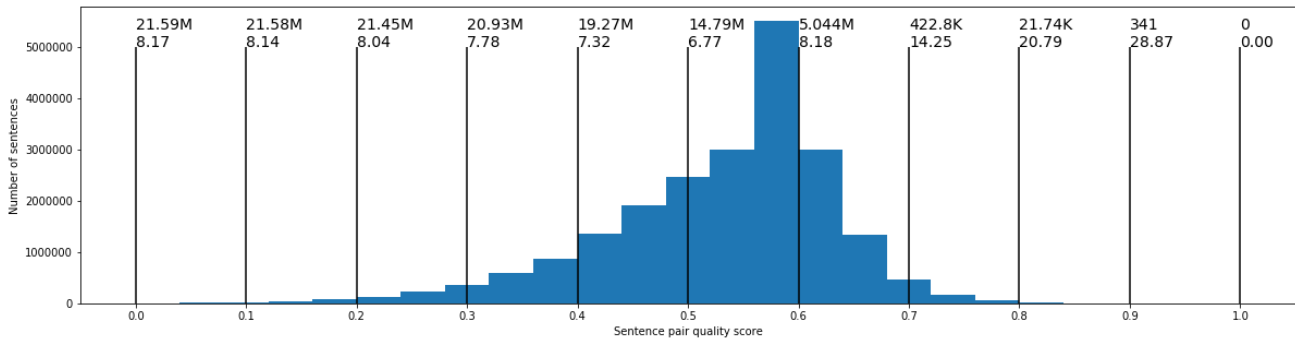


Figure 3: Histogram of sentence pairs scores and sentence length

style annotations in order to make use of existing automatic SRL systems. In this mapping, we only consider annotations with FEs marked as “Core” by FrameNet. We map these Core FEs to PropBank arguments by relying on the order in which they appear in the FrameNet frame definition, and map them accordingly to ARG0, ARG1, etc. For example, the COMMERCE_BUY frame has two core FEs: Buyer and Goods. In the sentence “Abby bought a car from Robin for \$5,000”, “Abby” is the Buyer and “a car” is the Goods. This sentence will be labeled as “[Abby_{ARG0}] bought [a car_{ARG1}] from Robin for \$5,000”. To apply this mapping to PropBank, we only keep sentences that include only Core Frame Elements². Upon further filtering with this criterion, we are left with 423K annotated sentences.

We sample from these sentences a training set of 240K sentences, 30K for development and 30K for test. The statistics of the Hebrew PropBank dataset are shown in Table 1.

4. Experimental Results

4.1. Model

We train a Hebrew PropBank semantic role labeler on the dataset we constructed to provide a baseline.

To this end, we use AllenNLP’s (Gardner et al., 2017) implementation of the model by Shi and Lin (2019), which only uses BERT (Devlin et al., 2018) for argument identification and classification. The specific BERT model we adopt is the multilingual cased version of BERT-Base which was trained, among other languages, on Hebrew. Hyper-parameters are as specified in AllenNLP’s configuration file³. This model simply encodes the whole sentence using the BERT model and predicts the BIO tags for the arguments given the predicate. A sentence is passed to BERT in the form [CLS] sentence [SEP] pred [SEP] - for example, [CLS] Barack Obama went to Paris [SEP] went [SEP]. Shi and Lin (2019) report a performance of F1=82.7 on argument identification on CoNLL 2009 in English.

²Mapping non-core elements such as space or time between FrameNet and PropBank is more interesting semantically, but we found it challenging to perform this mapping automatically. Noisy mapping would increase the risk of producing bad PropBank annotations, and we end up with a good enough number of sentences with the more stringent filter.

³https://github.com/allenai/allennlp/blob/master/training_config/bert_base_srl.jsonnet

4.2. Evaluation

As BERT works with WordPiece tokenization (Sennrich et al., 2016), we hypothesized that the segmented Hebrew words, *i.e.*, the phrase “and from you / ve min ata” is one Hebrew word which is segmented into three words, would perform worse, as BERT has not seen segmented sentences. To this end, we unsegment the sentences back to their original form (joining “and from you / ve min ata” back into one token “vemimkha”) to see if this would improve performance. Table 2 shows that this hypothesis is incorrect, as training with both segmented and unsegmented tokens provides roughly equivalent performance.

5. Conclusion

We present the first available SRL resource in Hebrew, which is constructed through annotation projection from aligned English sentences. The dataset is derived from the OpenSubtitles collection of movie subtitles. The genre is of informal spoken language. We designed a full pipeline to map a noisy collection of aligned sentences in English and Hebrew into an annotated SRL dataset in both FrameNet and PropBank styles in CoNLL 2009 format.

In order to control the quality of the generated data, we introduce a set of filters, following the approach of filtered projection of Akbik et al. (2015). We specifically verified the impact of the Hebrew rich morphology and complex word formation rules on the process.

We finally trained a neural SRL system in Hebrew as a baseline model, building on the BERT multi-lingual model. In future work, we intend to manually curate the generated data while enriching the Hebrew FrameNet database with annotated exemplar sentences.

Code, data and pre-trained models for Hebrew SRL are available at https://github.com/bgunlp/hebrew_srl.

6. Acknowledgements

This research was supported by the Lynn and William Frankel Center for Computer Science at Ben Gurion University.

7. Bibliographical References

Akbik, A., Chiticariu, L., Danilevsky, M., Li, Y., Vaithyanathan, S., and Zhu, H. (2015). Generating high quality proposition Banks for multilingual semantic role

Table 1: Dataset Statistics. (S) stands for segmented Hebrew words, (U) for unsegmented, ASL is “average sentence length”

Fold	#sentences	#tokens (S)	#types (S)	ASL (S)	#tokens (U)	#types (U)	ASL (U)
Train	240K	2.2M	60,288	9.2	1.8M	95,835	7.6
Development	30K	277K	20,633	9.2	228K	28,714	7.6
Test	30K	277K	20,664	9.2	229K	28,770	7.6

Table 2: Training Performance

	Precision	Recall	F1
Segmented	0.65	0.63	0.64
Unsegmented	0.64	0.62	0.63

- labeling. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 397–407, Beijing, China, July. Association for Computational Linguistics.
- Aminian, M., Rasooli, M. S., and Diab, M. T. (2019). Cross-lingual transfer of semantic roles: From raw text to semantic roles. *CoRR*, abs/1904.03256.
- Annesi, P. and Basili, R. (2010). Cross-lingual alignment of framenet annotations through hidden markov models. In *International Conference on Intelligent Text Processing and Computational Linguistics*, pages 12–25. Springer.
- Baker, C. F., Fillmore, C. J., and Lowe, J. B. (1998). The berkeley framenet project. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics-Volume 1*, pages 86–90. Association for Computational Linguistics.
- Bick, E. (2011). A framenet for danish. In *Proceedings of the 18th Nordic Conference of Computational Linguistics, NODALIDA 2011, May 11-13, 2011, Riga, Latvia*, pages 34–41.
- Borin, L., Dannélls, D., Forsberg, M., Gronostaj, M. T., and Kokkinakis, D. (2010). The past meets the present in swedish framenet++. In *14th EURALEX international congress*, pages 269–281.
- Burchardt, A., Erk, K., Frank, A., Kowalski, A., Padó, S., and Pinkal, M. (2006). The salsa corpus: a german corpus resource for lexical semantics. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC-2006)*, pages 969–974.
- Candito, M., Amsili, P., Barque, L., Benamara, F., De Chalendar, G., Djemaa, M., Haas, P., Huyghe, R., Mathieu, Y. Y., Muller, P., et al. (2014). Developing a french framenet: Methodology and first results. In *LREC-The 9th edition of the Language Resources and Evaluation Conference*.
- Christensen, J., Soderland, S., Etzioni, O., et al. (2010). Semantic role labeling for open information extraction. In *Proceedings of the NAACL HLT 2010 First International Workshop on Formalisms and Methodology for Learning by Reading*, pages 52–60. Association for Computational Linguistics.
- Das, D., Chen, D., Martins, A. F., Schneider, N., and Smith, N. A. (2014). Frame-semantic parsing. *Computational linguistics*, 40(1):9–56.
- Daza, A. and Frank, A. (2019). Translate and label! an encoder-decoder approach for cross-lingual semantic role labeling. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 603–615, Hong Kong, China, November. Association for Computational Linguistics.
- Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- Dyer, C., Chahuneau, V., and Smith, N. A. (2013). A simple, fast, and effective reparameterization of IBM model 2. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 644–648, Atlanta, Georgia, June. Association for Computational Linguistics.
- Fillmore, C. J. (1976). Frame semantics and the nature of language. *Annals of the New York Academy of Sciences*, 280(1):20–32.
- FitzGerald, N., Michael, J., He, L., and Zettlemoyer, L. (2018). Large-scale qa-srl parsing.
- Gardner, M., Grus, J., Neumann, M., Tafjord, O., Dasigi, P., Liu, N. F., Peters, M., Schmitz, M., and Zettlemoyer, L. S. (2017). Allennlp: A deep semantic natural language processing platform.
- Hayoun, A. and Elhadad, M. (2016). The hebrew framenet project. In *LREC*.
- Joulin, A., Grave, E., Bojanowski, P., Douze, M., Jégou, H., and Mikolov, T. (2016a). Fasttext.zip: Compressing text classification models. *arXiv preprint arXiv:1612.03651*.
- Joulin, A., Grave, E., Bojanowski, P., and Mikolov, T. (2016b). Bag of tricks for efficient text classification. *arXiv preprint arXiv:1607.01759*.
- Lindén, K., Haltia, H., Luukkonen, J., Laine, A. O., Roivainen, H., and Väisänen, N. (2017). Finnfn 1.0: The finnish frame semantic database. *Nordic Journal of Linguistics*, 40(3):287–311.
- Lison, P. and Tiedemann, J. (2016). Opensubtitles2016: Extracting large parallel corpora from movie and tv subtitles. In Nicoletta Calzolari (Conference Chair), et al., editors, *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, Paris, France, may. European Language Resources Association (ELRA).
- Lönneker-Rodman, B., Baker, C., and Hong, J. (2008). The new framenet desktop: A usage scenario for slovenian. *Programme Committee 7*, page 147.

- More, A., Seker, A., Basmova, V., and Tsarfaty, R. (2019). Joint transition-based models for morpho-syntactic parsing: Parsing strategies for MRLs and a case study from modern Hebrew. *Transactions of the Association for Computational Linguistics*, 7:33–48.
- Nam, S., Park, J., Kim, Y., Hahm, Y., Hwang, D., and Choi, K.-S. (2014). Korean framenet for semantic analysis. In *Proceedings of the 13th International Semantic Web Conference*.
- Nivre, J., Hall, J., Nilsson, J., Chanev, A., Eryigit, G., Kübler, S., Marinov, S., and Marsi, E. (2007). Malt-parser: A language-independent system for data-driven dependency parsing. *Natural Language Engineering*, 13(2):95–135.
- Ohara, K. H., Fujii, S., Otori, T., Suzuki, R., Saito, H., and Ishizaki, S. (2004). The japanese framenet project: An introduction. In *Proceedings of LREC-04 Satellite Workshop “Building Lexical Resources from Semantically Annotated Corpora” (LREC 2004)*, pages 9–11. Citeseer.
- Padó, S. and Lapata, M. (2009). Cross-lingual annotation projection for semantic roles. *Journal of Artificial Intelligence Research*, 36:307–340.
- Padó, S. (2007). *Cross-lingual annotation projection models for role-semantic information*. Saarland University.
- Palmer, M., Gildea, D., and Kingsbury, P. (2005). The proposition bank: An annotated corpus of semantic roles. *Computational linguistics*, 31(1):71–106.
- Petruck, M. R. (2005). Towards hebrew framenet. *Kernerman Dictionary News*, 13:12.
- Ruppenhofer, J., Ellsworth, M., Petruck, M. R., Johnson, C. R., and Scheffczyk, J. (2016). *FrameNet II: Extended theory and practice*. Institut für Deutsche Sprache, Bibliothek.
- Sennrich, R., Haddow, B., and Birch, A. (2016). Neural machine translation of rare words with subword units. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1715–1725, Berlin, Germany, August. Association for Computational Linguistics.
- Shi, P. and Lin, J. (2019). Simple BERT models for relation extraction and semantic role labeling. *CoRR*, abs/1904.05255.
- Stanovsky, G., Michael, J., Zettlemoyer, L. S., and Dagan, I. (2018). Supervised open information extraction. In *Proceedings of The 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, pages 885–895, New Orleans, Louisiana, june. Association for Computational Linguistics.
- Subirats, C. and Petruck, M. (2003). Surprise: Spanish framenet. In *Proceedings of CIL*, volume 17, page 188.
- Tiedemann, J. (2012). Parallel data, tools and interfaces in opus. In Nicoletta Calzolari (Conference Chair), et al., editors, *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC’12)*, Istanbul, Turkey, may. European Language Resources Association (ELRA).
- Tonelli, S. and Pianta, E. (2008). Frame information transfer from english to italian. In *LREC*.
- Torrent, T. T. and Ellsworth, M. (2013). Behind the labels: Criteria for defining analytical categories in framenet brasil. *Revista Veredas*, 17(1-).
- van der Plas, L., Merlo, P., and Henderson, J. (2011). Scaling up automatic cross-lingual semantic role annotation. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 299–304, Portland, Oregon, USA, June. Association for Computational Linguistics.
- Yarowsky, D., Ngai, G., and Wicentowski, R. (2001). Inducing multilingual text analysis tools via robust projection across aligned corpora. In *Proceedings of the First International Conference on Human Language Technology Research*.
- You, L. and Liu, K. (2005). Building chinese framenet database. In *Natural Language Processing and Knowledge Engineering, 2005. IEEE NLP-KE’05. Proceedings of 2005 IEEE International Conference on*, pages 301–306. IEEE.
- Zawisławska, M., Derwojedowa, M., and Linde-Usiekniewicz, J. (2008). A framenet for polish. In *Converging Evidence: Proceedings to the Third International Conference of the German Cognitive Linguistics Association (GCLA’08)*, pages 116–117.

Appendix

Below are the inputs to the SRL projection for the English sentence “John sold a car to Mary.”

```
{
  "frames": [
    {
      "target": {
        "name": "Commerce_sell",
        "spans": [
          {
            "start": 1,
            "end": 2,
            "text": "sold"
          }
        ]
      },
      "annotationSets": [
        {
          "rank": 0,
          "score": 81.82887993432267,
          "frameElements": [
            {
              "name": "Seller",
              "spans": [
                {
                  "start": 0,
                  "end": 1,
                  "text": "John"
                }
              ]
            },
            {
              "name": "Buyer",
              "spans": [
                {
                  "start": 4,
                  "end": 6,
                  "text": "to Mary"
                }
              ]
            },
            {
              "name": "Goods",
              "spans": [
                {
                  "start": 2,
                  "end": 4,
                  "text": "a car"
                }
              ]
            }
          ]
        }
      ]
    }
  ]
}
```

Figure 4: English SRL input

1	John	-	NNP	NNP	-	2	nsubj	-	-
2	sold	-	VBD	VBD	-	0	null	-	-
3	a	-	DT	DT	-	4	det	-	-
4	car	-	NN	NN	-	2	dobj	-	-
5	to	-	TO	TO	-	2	prep	-	-
6	Mary	-	NNP	NNP	-	5	pobj	-	-
7	.	-	.	.	-	2	punct	-	-

Figure 5: English input

1	גיון	גיון	NNP	NNP	2	subj	-	-	-
2	מכר	מכר	VB	VB	gen=M num=S per=3 tense=PAST	0	ROOT	-	-
3	אוטו	אוטו	NN	NN	gen=M num=S	2	obj	-	-
4	ל	ל	PREPOSITION	PREPOSITION		2	prepmid	-	-
5	מרי.	מרי.	NN	NN	gen=M num=P	4	pobj	-	-

Figure 6: Hebrew input

0-0 1-1 2-2 3-2 4-3 5-4 6-4

Figure 7: fast_align word alignments