

A Morphology-Aware Network for Morphological Disambiguation

Eray Yildiz, Caglar Tirkaz, H. Bahadir Sahin, Mustafa Tolga Eren, Ozan Sonmez

Huawei Turkey Research and Development Center, Umraniye, Istanbul, Turkey

{eray.yildiz, mustafa.tolga.eren}@huawei.com

{caglantirkaz, hbahadirsahin, osonmez}@gmail.com

Abstract

Agglutinative languages such as Turkish, Finnish and Hungarian require morphological disambiguation before further processing due to the complex morphology of words. A morphological disambiguator is used to select the correct morphological analysis of a word. Morphological disambiguation is important because it generally is one of the first steps of natural language processing and its performance affects subsequent analyses. In this paper, we propose a system that uses deep learning techniques for morphological disambiguation. Many of the state-of-the-art results in computer vision, speech recognition and natural language processing have been obtained through deep learning models. However, applying deep learning techniques to morphologically rich languages is not well studied. In this work, while we focus on Turkish morphological disambiguation we also present results for French and German in order to show that the proposed architecture achieves high accuracy with no language-specific feature engineering or additional resource. In the experiments, we achieve 84.12, 88.35 and 93.78 morphological disambiguation accuracy among the ambiguous words for Turkish, German and French respectively.

Introduction

Morphological analysis is generally achieved through the use of finite state transducers (FST) (Kaplan and Kay 1981; Koskenniemi 1984; Beesley and Karttunen 2003; Oflazer 1993). During morphological analysis, the surface form of the word is given as input and an FST is used to output possible morphological analyses of the input word. A morphological analysis contains a root and a set of tags so called morphemes, minimal units of meaning in a language (Oflazer 1993).

A morphological disambiguator is used to select the correct analysis among the possible analyses of a word using the context that the word appears in. The output of morphological disambiguation contains syntactic and semantic information about a word such as its POS tag, tense, polarity and it being accusative, possessive or genitive. This information is vital for some NLP tasks such as dependency parsing and semantic role labeling whereas it can be utilized in other NLP tasks such as topic modeling, named entity recognition and machine translation.

Copyright © 2016, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

While morphological disambiguation is important for natural language processing in any language, it is vital in morphologically rich languages. We specifically focus on Turkish which is an important language spoken by over 70 million people and has a complex morphology that allows construction of thousands of word forms from each root through inflectional and derivational suffixation (Hakkani-Tür, Oflazer, and Tür 2000). For instance, *yürü* (walk), *yürüdüm* (I walked), *yürüyeceksiniz* (you will walk), *yürüttüler* (they made somebody walk), *yürüyünce* (When (he/she/it) walks) and *yürüyecektiler* (They were going to walk) are some of the possible word formations of a Turkish verb root *yürü*. In all the examples “*yürü*” is the root of the words whereas the suffixes are used to change meaning.

Morphological analysis of a word might produce more than one analysis since there might be multiple interpretations of a single word. Consider the example given in Table 1 where the Turkish word “dolar” is analyzed. The output of the morphological analyzer for this word contains four possible analyses. The reason for that is each of the words “dolar”, “dola”, “dol” and “do” can be used as roots and at the same time “r”, “ar” and “lar” are all valid suffixes in Turkish. Thus, all four of the analyses are valid that lead to quite different meanings.

Table 1: Morphological analyses of the Turkish word “dolar”

Morphological Analysis	English Translation
dolar +Noun +3sg +Pnon +Nominative	dollar
dola +Verb +Positive +Aorist +3sg	he/she wraps
dol +Verb +Positive +Aorist +3sg	it fills
do +Noun +3pl +Pnon +Nominative	Multiple C (musical note)

Another reason for multiple morphological analyses is due to the fact that a morpheme might change meaning depending on the context of the word. Consider the example given in Table 2. In the first row “evi” is used in the accusative case whereas in the second row it is used as a possessive noun. The word “evi” has two morphological analyses sharing the same root. Thus, the only difference in the

analyses is at the suffix of the word. The suffix “-i” in the first sentence transforms the word into “accusative marker”, while its interpretation is “third person possessive” in the second one. In addition, some root words might have multiple meanings. For instance, the Turkish word “yüz” could be interpreted as a noun (face), a verb (swim) or a number (hundred) depending on its context.

Table 2: Multiple interpretations of the Turkish word “evi”

Turkish sentence and its translation	Morphological Analysis
Evi bulabildiniz mi? – Did you find the house?	ev +Noun +3sg +Pnon +Accusative
Evi gerçekten güzelmiş. – His/Her house is really beautiful.	ev +Noun +3sg +P3sg +Nominative

As we noted before, morphological disambiguation is important for NLP in most of the languages. For instance, although German and French do not have a morphology as rich as Turkish, NLP in these languages can still benefit from morphological disambiguation. Higher accuracies in NLP tasks such as POS tagging and dependency parsing can be obtained if the morphology of the words are taken into account (Sennrich et al. 2009), (Candito et al. 2010). We apply our general purpose morphological disambiguation method to these languages and show that high accuracies can be obtained for POS tagging and lemmatization. Possible word formations and morphological analyses of the German word “Haus” and the French word “savoir” are given in Table 3 and Table 4 respectively.

Table 3: Possible word formations and morphological analyses of the German word “haus”

Word	Morphological Analyses
haus	Noun Neuter Nominative Singular Noun Neuter Dative Singular Noun Neuter Accusative Singular
häuser	Noun Neuter Accusative Plural Noun Neuter Nominative Plural Noun Neuter Genitive Plural
häusern	Noun Neuter Dative Plural
hause	Noun Neuter Dative Singular
hauses	Noun Neuter Genitive Singular

There are various approaches proposed for morphological disambiguation based on lexical rules or statistical models. Rule based methods apply hand-crafted rules in order to select the correct morphological analyses or eliminate incorrect ones (Oflazer and Kuruöz 1994; Oflazer and Tur 1996; Daybelge and Cicekli 2007). Yüret and Türe (2006) proposed a decision list learning algorithm for extraction of Turkish morphological disambiguation rules from disambiguated training data. Tür et al. (2000) developed a statistical model which scores the probability of each analy-

Table 4: Possible word formations and morphological analyses of the French word “savoir”

Word	Morphological Analyses
savoir	Noun Masculine Singular Verb Infinitive
sais	Verb Present SecondPerson Singular Verb Present FirstPerson Singular
savons	Verb Present FirstPerson Plural
savaient	Verb Imperfect ThirdPerson Plural
saches	Verb Subjunctive SecondPerson Singular
sachant	Verb Present Participle
su	Verb Past Participle Masculine Singular

sis using trigram models of the tags and roots. Sak et al. (2007) applied a multilayer perceptron algorithm which uses n-grams of the roots and tags as features. A CRF based disambiguation model was proposed by Razieh et al. 2012. Finally, hybrid models which combine statistical and rule based approaches are also proposed (Oflazer and Tur 1996; Kutlu and Cicekli 2013).

We propose a deep neural architecture followed by the Viterbi algorithm for morphological disambiguation of words in a sentence. In this paper we focus on Turkish as an example even though the proposed model can be utilized in all morphologically rich languages. We test our approach in German and French in order to prove that the proposed method is able to work well for other languages as well. The network architecture in this paper is designed to produce a classification score for a sequence of n -words. It consists of two layers and a softmax layer. The first layer of the model builds a representation for each word using root embeddings and some syntactic and semantic features. The second layer takes as input the learned word representations and incorporates contextual information. A softmax layer uses the output of the second layer to produce a classification score. We use the neural network to produce a score for each n length sequence in a given sentence. We then employ the Viterbi algorithm to produce the morphological disambiguation for each word in the sentence by finding the most probable sequence using the output of the softmax layer.

Related Works

In a natural language processing pipeline morphological disambiguation can be considered at the same level as POS tagging. In order to perform POS tagging in English, various approaches such as rule-based models (Brill 1992), statistical models (Brill 1995), maximum entropy models (Ratnaparkhi 1997), HMMs (Cutting et al. 1992), CRFs (Lafferty, McCallum, and Pereira 2001) and decision trees (Schmid 1994) are proposed. However, morphological disambiguation is a much harder problem in general due to the fact that it requires the classification of both roots, suffixes and the corresponding labels. Moreover, compared to an agglutinative language such as Turkish, English words can take on a limited number of word forms and part-of-speech tags. Yüret and Türe (2006) observe that more than ten thousand

tag types exists in a corpus comprised of a million Turkish words. Thus, due to the high number of possible tags and the number of possible analyses in languages with productive morphology, morphological disambiguation is quite different from part-of-speech tagging in English.

The previous work on morphological disambiguation in morphologically rich languages can be summarized into three categories: rule based, statistical and hybrid approaches. In the rule-based approaches a large number of hand crafted rules are used to select the correct morphological analyses or to eliminate incorrect ones (Karlssoon et al. 1995; Oflazer and Kuruöz 1994; Oflazer and Tur 1996; Daybelge and Cicekli 2007). Statistical approaches generally utilize the statistics of root and tag sequences for selection of the best roots and tags. A statistical Turkish morphological disambiguation model which scores the probability of each tag by considering the statistics over the derivational boundaries and roots using trigrams has been proposed by Tür *et al.* (2000). They test their model on a manually disambiguated test data consisting of 2763 words and obtain 93.5% accuracy in morphological disambiguation (including non-ambiguous words). A similar morphology-aware nonparametric Bayesian model is proposed in (Chahuneau, Smith, and Dyer 2013). They integrate their generative model to NLP applications such as language modeling, word alignment and morphological disambiguation and obtain state-of-the-art results for Russian morphological disambiguation. Yüret and Türe (2006) extract Turkish morphological disambiguation rules using a decision list learner, Greedy Prepend Algorithm (GPA), and they achieve 95.8% accuracy on manually disambiguated data consisting of around 1K words. Megyesi (1999) adapt a transformation based syntactic rule learner (Brill 1995) for Hungarian and Hajic (1998) extend his work for Czech and five other languages. Sak *et al.* (2007) apply a multi-layer perceptron algorithm using a set of 23 features including tri-gram and bi-gram statistics of morphological tags and roots. They obtain 96.8% accuracy on test data consisting of 2.5K words. Ehsani *et al.* (2012) apply conditional random fields (CRFs) using several features derived from morphological and syntactic properties of words and achieve 96.8% accuracy. Görgün and Yıldız (2011) use a J48 decision tree and achieve 95.6% accuracy. There are also several studies that combine statistical and rule based approaches such as (Ezeiza et al. 1998; Oflazer and Tur 1996; Kutlu and Cicekli 2013; Orosz and Novák 2013).

Although deep learning techniques have been successfully used in various NLP tasks in English (Collobert and Weston 2008; Collobert et al. 2011; Le and Mikolov 2014; Pennington, Socher, and Manning 2014; Luong, Socher, and Manning 2013; Socher et al. 2012), this study is unique in that we create a deep learning architecture specifically suited for handling morphologically rich languages. One similar work to ours is the recent work of Luong *et al.* (2013) who introduce morphological RNNs to create word representations through composition of morphemes. However, they present their results only for English which is not morphologically as rich as languages such as Turkish or Finnish. There are also recent works that suggest integrating mor-

phological knowledge into distributed word representations such as (Cotterell and Schütze 2015) and (Cui et al. 2015). Cotterell and Schütze (2015) extend log-bilinear model (an instance of language models that make the Markov assumption as n-gram language models) in order to jointly predict the next morphological tag along with the next word, encouraging the resulting embeddings to encode morphology. On the other hand, (Cui et al. 2015) propose a method for learning embeddings which is a modified version of skip-gram algorithm (Mikolov et al. 2013) that benefits from morphological knowledge when predicting the target word. Using morphology-based word representations improves the performance for different NLP tasks such as word similarity and statistical machine translation according to empirical evaluation of Botha and Phil (2014).

Our work uses a convolutional architecture and handles any number of morphological features in order to build word representations while performing disambiguation at the same time.

Method

In this work we propose an architecture with the ability to represent morphologically rich words and model spatial dependencies among word vectors. A softmax layer that is trained on top of the layers is used to predict the likelihood of a window of words. Finally, the Viterbi algorithm is used on the outputs of the softmax layer in order to find the optimal disambiguation of the words in a sentence. We also show how unsupervised pre-training can be used to improve the performance of the designed system and achieve the state-of-the-art accuracy for Turkish morphological disambiguation.

The input to our model is a sentence where each word in the sentence needs to be disambiguated. We first tokenize the sentences and then use morphological analyzers to find possible analyses of each word in the sentence. HFST tool (Lindén, Silfverberg, and Pirinen 2009) is used to perform morphological analysis in German and French whereas (Oflazer 1993) is used for Turkish morphological analysis. NLP systems that use deep learning generally employ word embeddings in order to represent each word in a dictionary. Word embeddings are dense low dimensional real-valued vectors with each dimension corresponding to a latent feature of the word (Turian, Ratinov, and Bengio 2010). In a morphologically rich language, representing words in surface form might not be a good idea since lots of surface form words can be derived from a single root. Thus, in our design, each word in surface form is represented with a root and a set of morphological features where each root and feature has individual embeddings that are learned during training. Root and morphological feature embeddings can have varying lengths and through their concatenation surface form words are represented with fixed length embeddings.

Our architecture is illustrated in Figure 1 where individual layers are marked with (a), (b) and (c). The first layer (a) takes the root and morphological features of a single word as input and propagates to the next layer. The second layer, (b), takes a window of n words as input and propagates to the softmax layer, (c). The non-linearity in both the first and

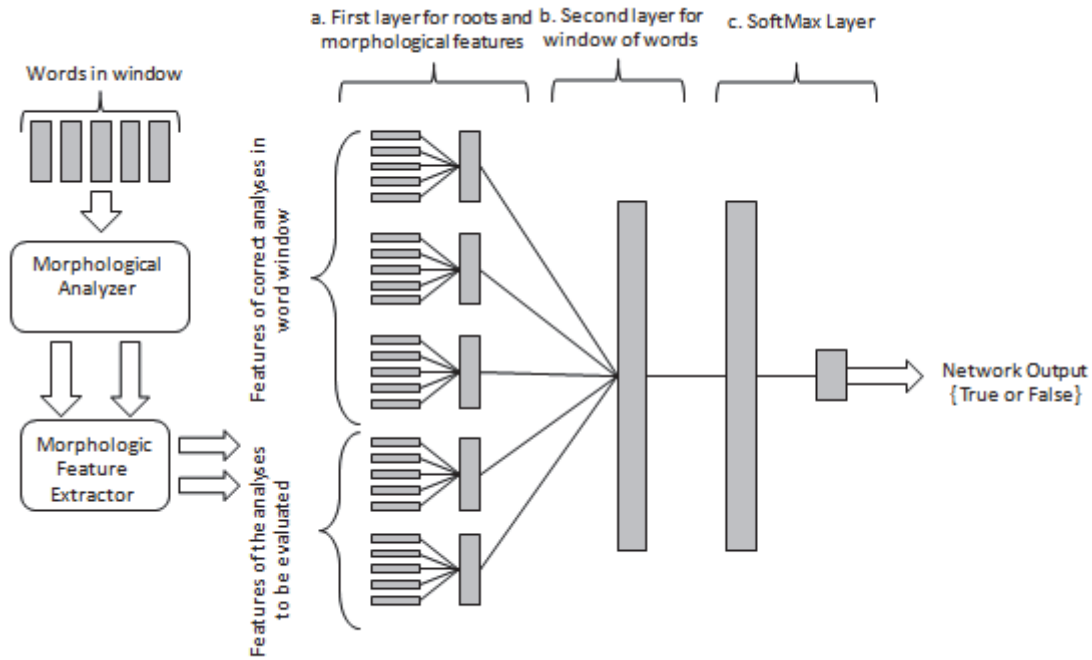


Figure 1: Structure of morphology-aware model. **Layer (a)** construct word vectors using the morphological features of the input word. **Layer (b)** allows the model to utilize contextual information by considering a window of words. **Layer (c)** is the softmax layer that makes a binary decision of whether or not the current disambiguation result is correct.

the second layers are provided through the use of *tanh* as the transfer function. The softmax layer is responsible for deciding the likelihood of the current morphological analysis of the words, i.e., a binary decision is produced with the expected result of 1 if the analysis is correct, 0 otherwise.

We train our network with the possible sequences of morphological analyses in the training data. For each sentence, and for each word, we select the $n-2$ words preceding the word and their groundtruth annotations along with the possible annotations of the last two words. We also add $n-1$ out of sentence tokens at the beginning of each sentence so that all words in the sentence are included in the training data. We label the sequences containing the correct morphological analysis as positive whereas the remaining sequences are labeled as negative. This way the model is trained to predict the correct annotation for the last two words in a sequence given that the first $n-2$ words have correct annotations. Training is performed with stochastic gradient descent and AdaGrad (Duchi, Hazan, and Singer 2011) as the optimization algorithms. At inference time, given a sentence containing words to disambiguate, we use the network to make predictions for window of words in the sentence and then use the Viterbi algorithm to select the best morphological analysis for each word.

Unsupervised pre-training of word embeddings have been employed in various NLP tasks, and their usage have improved recognition accuracies (Collobert et al. 2011; Turian, Ratniov, and Bengio 2010). In order to improve the performance of our disambiguation system we also use unsuper-

vised methods to pre-train root embeddings of words. We created a corpus comprised of 1 billion Turkish words that we collected from various sources, such as e-books and web sites. Although our corpus is rather small compared to English corpora, it is the largest text corpus in Turkish that we know of. After we trained the supervised disambiguation system as described above, we disambiguated each word in the corpus and extracted the roots of words. Next, we built representations for root forms of the words using the unsupervised skip-gram algorithm (Mikolov et al. 2013). After obtaining the pre-trained root vectors, we retrained our disambiguation system with pre-trained root embeddings. This technique allowed us to further improve the disambiguation accuracies we obtained.

As discussed earlier, the first layer takes as input the root and the morphological features of a word. The morphological features of words we use are presented in Table 5. Specifically, the set of morphological features we consider contains the root, main POS tag, minor POS tag, person and possessive agreements, plurality, gender, case marker, polarity and tense. Note that the information contained in a surface word form may differ due to morphological characteristics of a language. For instance, German and French have gender feature contrary to Turkish while Turkish words have possessive agreement and polarity. *Main POS tag*, describes the category of a word and can take on values such as noun, verb, adjective and adverb. *Minor POS tag* determines the minor morphological properties of a word such as semantic markers, causative markers and post-position. "Since",

Table 5: The morphological (morphosyntactic and morphosemantic) features we used to represent each word

Language	Morphosyntactic and Morphosemantic Features									
	Root	Main POSTag	Minor POSTag	Person Agreement	Plurality	Gender	Possessive Agreement	Case Marker	Polarity	Tense
Turkish	+	+	+	+	+	-	+	+	+	+
German	+	+	+	+	+	+	-	+	-	+
French	+	+	-	+	+	+	-	-	-	+

”While”, ”Propernoun”, ”Without” can be given as examples to this kind of morphological features in Turkish. *Person* and *possessive agreement* are used to answer the questions “who” and “whose” respectively, i.e., they are used to indicate a person or an ownership relationship. *Case marker* relate the nouns to the rest of the sentence as prepositions do in English. Nominative (none), dative(to, for), locative (at, in, on), ablative (from, out of) and genitive (of) forms are examples of the forms that can be observed in a sentence. *Polarity* of a word is positive if the word is not negated and negative otherwise. *Tense* indicates the tense of the verbs such as present, past and future tense. Additionally, we consider the moods of the verbs within tense feature. Moods express the speaker’s attitude such as indicative, imperative or subjunctive moods. In languages with grammatical gender such German and French, every noun is associated with a *gender*. The morphological analyzer we use associates each French word with one of the two genders (masculine and feminine) while it associates each German word with one of the four possible genders (masculine, feminine, neuter and no gender). Some of the suffixes in Turkish change word meaning creating derivational boundaries in the morphological analyses. The morphological features of a word given in Table 5 are extracted after the final derivational boundary. In Turkish, we add one more feature to each word named *previous tags* in order to account for the previous suffixes that the word might have. This way, our model learns the effect of suffixes that change word meaning. Some of the described morphological features exist only for certain word categories. For instance *possessive agreement* and *case marker* features can only exist in nouns, *polarity* and *tense* exist in verbs and *person agreement* exist in nouns and verbs. If a morphological feature cannot be extracted from a word, we label it as having *NULL* for the feature.

Experiments

For Turkish, we used a semi-automatically disambiguated corpus containing 1M tokens (Yüret and Türe 2006). Since this dataset is annotated semi-automatically, it also contains noise. In order to reduce the effect of noise to the recognition accuracies, we created a test set by randomly selecting sentences containing 20K of the tokens and manually annotating them. We make this test data publicly available ¹ so that Turkish morphological disambiguation algorithms can be compared more accurately in the future.

We use SPMRL 2014 dataset (Seddah and Tsarfaty 2014) for German and French. This data set is created in the Penn tree bank format and used for a shared task on statistical

parsing of morphologically rich languages. This dataset contains 1M and 500K sentences with POS tag and morphological information for German and French respectively. It provides 90% of all sentences as training set and %10 of rest of the sentences as test set. We align the features in the tree bank to the HFST outputs in order to determine the correct morphological analyses generated by the HFST tool. We use this data set for both training and testing. The development sets for each language are randomly separated from the training data and are used to optimize the embedding lengths of morphological features.

We noticed that similar parameters lead to the best performance. Thus, in the experiments, we used embedding lengths 50, 20 and 5 for roots, POS tags and the other morphological features respectively. The number of filters in the first and second layers are 30 and 40 respectively. The window length, n , that determines the number of words input to the second layer is set to 5.

Table 6: POS tagging, lemmatization and morphological disambiguation accuracies of the proposed approach for Turkish, German and French.

	Turkish(%)	German(%)	French(%)
POS Tagging	96.85	98.35	98.47
Lemma.	97.59	95.95	99.52
M. disamb.	84.12	88.35	93.78

The experiment results for POS tagging, lemmatization and morphological disambiguation in Turkish, German and French are presented in Table 6. Notice that the POS tagging and lemmatization accuracies are refer to the percentages of POS tags and lemmas predicted correctly while morphological disambiguation accuracies are refer to the percentages of the words disambiguated correctly among the ambiguous words. According to the results, we observe that even though our initial target was to be able to achieve Turkish morphological disambiguation, our model consistently obtains high accuracies in French and German as well.

In Table 7, we present the results of various models for Turkish morphological disambiguation on our hand-labeled test data. The results of the multilayer perceptron developed in (Sak, Güngör, and Saraçlar 2007) and the decision list learning algorithm developed in (Yüret and Türe 2006) are presented in lines 1 and 2 respectively. We present Turkish morphological disambiguation results obtained by our model with and without pre-training in lines 3 and 4 respectively. As we discussed before, unsupervised pre-training of the embeddings can boost accuracies of neural networks. As

¹<http://indigof:8080/Genie/disambiguationTestSet.html>

Table 7: The comparison of the disambiguation accuracy of the proposed approach with the state-of-the-art models in Turkish.

Method	Accuracy(%)
Multilayer Perceptron	82.13
Decision List	83.31
Proposed Model - w/o pre-training	84.12
Proposed Model - with pre-training	85.18

expected, morphological disambiguation accuracy increases by around 1% (around 6% reduction in error) when root embeddings are pre-trained instead of randomly initialized. We see that even without unsupervised pre-training our algorithm outperforms the current state of the art models and we are able to further improve the accuracy by pre-training of the embeddings.

Although we do not evaluate the effects of unsupervised pre-training for German and French, it is expected that higher accuracies can be achieved using unsupervised pre-training of the embeddings for these languages as well. Error analysis for Turkish morphological disambiguation shows that the root is incorrectly decided in 30% of errors. The root is correct but the POS tag is incorrectly decided in 40% of errors while 30% of errors caused by wrong decisions on other inflectional groups. When compared with the study of Sak *et al.* 2007, there is no significant difference in the distribution of mistakes. However our method performs better in root decisions due to unsupervised learning of root embeddings. As discussed before, the available data for Turkish morphological disambiguation task contains some systematic errors. Yüret and Türe (2006) report that the accuracy of the training data is below 95%. According to our observation there is a major confusion between noun and adjective POS tags in training data which affects the decision of the morphological disambiguation systems. In our experiment, we observe that 18% of the errors are caused by such confusion, whereas the ratio of these errors are reported as 22% in the experiments of Sak *et al.* (2007).

Summary and Future Work

In this paper, we present a model capable of learning word representations for languages with rich morphology. We show the utility of our approach in the task of Turkish, German and French morphological disambiguation. We also show the effect of unsupervised pre-training on recognition accuracies and improve the current state-of-the-art in Turkish morphological disambiguation. We publicly make available a manually annotated test set containing 20K tokens which we believe will benefit Turkish NLP.

This paper presents a deep learning architecture specifically aiming to handle morphologically rich languages. Nonetheless, NLP systems that work on languages such as English can also benefit from our work. Using our model, English words can be separated into morphemes so that they can be better represented. This allows creating systems that are less affected from problems such as data sparsity (Lu-

ong, Socher, and Manning 2013).

While using pre-training, we only considered the pre-trained root embeddings. It would be preferred to pre-train all the embeddings using our text corpus which we leave as future work. Another point of note is the selected embedding sizes that we used in our experiments. While we worked on a development set separated from training data for parameter selection, further investigation in parameter selection might improve the obtained accuracies.

Acknowledgments

This project is partially funded by 3140951 numbered TUBITAK-TEYDEB (The Scientific and Technological Research Council of Turkey – Technology and Innovation Funding Programs Directorate).

References

- Beesley, K. R., and Karttunen, L. 2003. *Finite state morphology*. Center for the Study of Language and Inf.
- Botha, J. A., and Blunsom, P. 2014. Compositional morphology for word representations and language modelling. *arXiv preprint arXiv:1405.4273*.
- Brill, E. 1992. A simple rule-based part of speech tagger. In *Proceedings of the Third Conference on Applied Natural Language Processing*, ANLC '92, 152–155. Stroudsburg, PA, USA: Association for Computational Linguistics.
- Brill, E. 1995. Transformation-based error-driven learning and natural language processing: A case study in part-of-speech tagging. *Comput. Linguist.* 21(4):543–565.
- Candito, M.; Nivre, J.; Denis, P.; and Anguiano, E. H. 2010. Benchmarking of statistical dependency parsers for french. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, 108–116. Association for Computational Linguistics.
- Chahuneau, V.; Smith, N. A.; and Dyer, C. 2013. Knowledge-rich morphological priors for bayesian language models. Association for Computational Linguistics.
- Collobert, R., and Weston, J. 2008. A unified architecture for natural language processing: Deep neural networks with multitask learning. In *Proceedings of the 25th International Conference on Machine Learning*, ICML '08, 160–167. New York, NY, USA: ACM.
- Collobert, R.; Weston, J.; Bottou, L.; Karlen, M.; Kavukcuoglu, K.; and Kuksa, P. 2011. Natural language processing (almost) from scratch. *J. Mach. Learn. Res.* 12:2493–2537.
- Cotterell, R., and Schütze, H. 2015. Morphological word-embeddings. In *Annual Conference of the North American Chapter of the ACL*, 1287—1292.
- Cui, Q.; Gao, B.; Bian, J.; Qiu, S.; Dai, H.; and Liu, T.-Y. 2015. Knet: A general framework for learning word embedding using morphological knowledge. *ACM Transactions on Information Systems (TOIS)* 34(1):4.
- Cutting, D.; Kupiec, J.; Pedersen, J.; and Sibun, P. 1992. A practical part-of-speech tagger. In *Proceedings of the Third Conference on Applied Natural Language Processing*, ANLC '92, 133–140. Stroudsburg, PA, USA: Association for Computational Linguistics.
- Daybelge, T., and Cicekli, I. 2007. A rule-based morphological disambiguator for turkish”, in. In *Proceedings of Recent Advances in Natural Language Processing (RANLP 2007)*, Borovets, 145–149.

- Duchi, J.; Hazan, E.; and Singer, Y. 2011. Adaptive subgradient methods for online learning and stochastic optimization. *The Journal of Machine Learning Research* 12:2121–2159.
- Ehsani, R.; Alper, M. E.; Eryigit, G.; and Adali, E. 2012. Disambiguating main POS tags for turkish. In *Proceedings of the 24th Conference on Computational Linguistics and Speech Processing, ROCLING 2012, Yuan Ze University, Chung-Li, Taiwan, September 21-22, 2012*.
- Ezeiza, N.; Alegria, I.; Arriola, J. M.; Urizar, R.; and Aduriz, I. 1998. Combining stochastic and rule-based methods for disambiguation in agglutinative languages. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics-Volume 1*, 380–384. Association for Computational Linguistics.
- Görgün, O., and Yildiz, O. T. 2011. A novel approach to morphological disambiguation for turkish. In *Computer and Information Sciences II - 26th International Symposium on Computer and Information Sciences, London, UK, 26-28 September 2011*, 77–83.
- Hajic, J., and Hladka, B. 1998. Czech language processing—pos tagging. In *Proceedings of the First International Conference on Language Resources & Evaluation*, 931–936.
- Hakkani-Tür, D. Z.; Oflazer, K.; and Tür, G. 2000. Statistical morphological disambiguation for agglutinative languages. In *Proceedings of the 18th Conference on Computational Linguistics - Volume 1, COLING '00*, 285–291. Stroudsburg, PA, USA: Association for Computational Linguistics.
- Kaplan, R. M., and Kay, M. 1981. Phonological rules and finite-state transducers. In *Linguistic Society of America Meeting Handbook, Fifty-Sixth Annual Meeting*, 27–30.
- Karlsson, F.; Voutilainen, A.; Heikkilä, J.; and Anttila, A., eds. 1995. *Constraint Grammar: A Language-Independent System for Parsing Unrestricted Text*. Mouton de Gruyter.
- Koskenniemi, K. 1984. A general computational model for word-form recognition and production. In *Proceedings of the 10th international conference on Computational Linguistics*, 178–181. Association for Computational Linguistics.
- Kutlu, M., and Cicekli, I. 2013. A hybrid morphological disambiguation system for turkish. In *Sixth International Joint Conference on Natural Language Processing, IJCNLP 2013, Nagoya, Japan, October 14-18, 2013*, 1230–1236.
- Lafferty, J.; McCallum, A.; and Pereira, F. C. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data.
- Le, Q., and Mikolov, T. 2014. Distributed representations of sentences and documents. In Jebara, T., and Xing, E. P., eds., *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*, 1188–1196. JMLR Workshop and Conference Proceedings.
- Lindén, K.; Silfverberg, M.; and Pirinen, T. 2009. Hfst tools for morphology—an efficient open-source package for construction of morphological analyzers. In *State of the Art in Computational Morphology*. Springer. 28–47.
- Luong, T.; Socher, R.; and Manning, C. D. 2013. Better word representations with recursive neural networks for morphology. In *Proceedings of the Seventeenth Conference on Computational Natural Language Learning, CoNLL 2013, Sofia, Bulgaria, August 8-9, 2013*, 104–113.
- Megyesi, B. 1999. Improving brill’s pos tagger for an agglutinative language. In *Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora*, 275–284.
- Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G. S.; and Dean, J. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, 3111–3119.
- Oflazer, K., and Kuruöz, I. 1994. Tagging and morphological disambiguation of turkish text. In *Proceedings of the Fourth Conference on Applied Natural Language Processing, ANLC '94*, 144–149. Stroudsburg, PA, USA: Association for Computational Linguistics.
- Oflazer, K., and Tur, G. 1996. Combining hand-crafted rules and unsupervised learning in constraint-based morphological disambiguation. In *Conference on Empirical Methods in Natural Language Processing*, 69–81.
- Oflazer, K. 1993. Two-level description of turkish morphology. In *Proceedings of the Sixth Conference on European Chapter of the Association for Computational Linguistics, EACL '93*, 472–472. Stroudsburg, PA, USA: Association for Computational Linguistics.
- Orosz, G., and Novák, A. 2013. Purepos 2.0: a hybrid tool for morphological disambiguation. In *RANLP*, 539–545.
- Pennington, J.; Socher, R.; and Manning, C. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543. Doha, Qatar: Association for Computational Linguistics.
- Ratnaparkhi, A. 1997. A maximum entropy model for part-of-speech tagging. In *EMNLP 1997*.
- Sak, H.; Güngör, T.; and Saraçlar, M. 2007. Morphological disambiguation of Turkish text with perceptron algorithm. In *CICLing 2007*, volume LNCS 4394, 107–118.
- Schmid, H. 1994. Probabilistic part-of-speech tagging using decision trees. In *Proceedings of the international conference on new methods in language processing. Vol. 12*.
- Seddah, D., and Tsarfaty, R. 2014. Introducing the spmrl 2014 shared task on parsing morphologically-rich languages. *SPMRL-SANCL 2014* 103.
- Sennrich, R.; Schneider, G.; Volk, M.; and Warin, M. 2009. A new hybrid dependency parser for german. *Proceedings of the German Society for Computational Linguistics and Language Technology* 115–124.
- Socher, R.; Huval, B.; Manning, C. D.; and Ng, A. Y. 2012. Semantic compositionality through recursive matrix-vector spaces. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, EMNLP-CoNLL '12*, 1201–1211. Stroudsburg, PA, USA: Association for Computational Linguistics.
- Turian, J.; Ratinov, L.; and Bengio, Y. 2010. Word representations: A simple and general method for semi-supervised learning. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, ACL '10*, 384–394. Stroudsburg, PA, USA: Association for Computational Linguistics.
- Yüret, D., and Türe, F. 2006. Learning morphological disambiguation rules for turkish. In *Proceedings of the Main Conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics, HLT-NAACL '06*, 328–334. Stroudsburg, PA, USA: Association for Computational Linguistics.