Named Entity Recognition on Twitter for Turkish using Semi-supervised Learning with Word Embeddings

Eda Okur[*†]**, Hakan Demir**[†]**, Arzucan Özgür**[†]

[*]Intel Corporation
Istanbul, Turkey
eda.okur@intel.com
[†]Department of Computer Engineering, Boğaziçi University
Istanbul, Turkey
{hakan.demir, arzucan.ozgur}@boun.edu.tr

Abstract

Recently, due to the increasing popularity of social media, the necessity for extracting information from informal text types, such as microblog texts, has gained significant attention. In this study, we focused on the Named Entity Recognition (NER) problem on informal text types for Turkish. We utilized a semi-supervised learning approach based on neural networks. We applied a fast unsupervised method for learning continuous representations of words in vector space. We made use of these obtained word embeddings, together with language independent features that are engineered to work better on informal text types, for generating a Turkish NER system on microblog texts. We evaluated our Turkish NER system on Twitter messages and achieved better F-score performances than the published results of previously proposed NER systems on Turkish tweets. Since we did not employ any language dependent features, we believe that our method can be easily adapted to microblog texts in other morphologically rich languages.

1. Introduction

Microblogging environments, which allow users to post short messages, have gained increased popularity in the last decade. Twitter, which is one of the most popular microblogging platforms, has become an interesting platform for exchanging ideas, following recent developments and trends, or discussing any possible topic. Since Twitter has an enormously wide range of users with varying interests and sharing preferences, a significant amount of content is being created rapidly. Therefore, mining such platforms can extract valuable information. As a consequence, extracting information from Twitter has become a hot topic of research. For Twitter text mining, one popular research area is opinion mining or sentiment analysis, which is surely useful for companies or political parties to gather information about their services and products (Kökciyan et al., 2013). Another popular research area is content analysis, or more specifically topic modeling, which is useful for text classification and filtering applications on Twitter (Hong and Davison, 2010). Moreover, event monitoring and trend analysis are also other examples of useful application areas on microblog texts (Kireyev et al., 2009).

In order to build successful social media analysis applications, it is necessary to employ successful processing tools for Natural Language Processing (NLP) tasks such as Named Entity Recognition (NER). NER is a critical stage for various NLP applications including machine translation, question answering and opinion mining. The aim of NER is to classify and locate atomic elements in a given text into predefined categories like the names of the persons, locations, and organizations (PLOs).

NER on well-written texts is accepted as a solved problem for well-studied languages like English. However, it still needs further work for morphologically rich languages like Turkish due to their complex structure and relatively scarce language processing tools and data sets (Şeker and Eryiğit, 2012). In addition, most of the NER systems are designed for formal texts. The performance of such systems drops significantly when applied on informal texts. To illustrate, the state-of-the-art Turkish NER system has CoNLL F-score of 91.94% on news data, but the performance drops to F-score of 19.28% when this system is adopted to Twitter data (Çelikkaya et al., 2013).

There are several challenges for NER on tweets, which are also summarized in Küçük et al. (2014), due to the very short text length and informal structure of the language used. Missing proper grammar rules and punctuation, lack of capitalization and apostrophes, usage of hashtags, abbreviations, and slang words are some of those challenges. In Twitter, using contracted forms and metonymic expressions instead of full organization or location names is very common as well. The usage of non-diacritic characters and the limited annotated data bring additional challenges for processing Turkish tweets.

Due to the dynamic language used in Twitter, heavy feature engineering is not feasible for Twitter NER. Demir and Özgür (2014) developed a semi-supervised approach for Turkish NER on formal (newswire) text using word embeddings obtained from unlabeled data. They obtained promising results without using any gazetteers and language dependent features. We adopted this approach for informal texts and evaluated it on Turkish tweets, where we achieved the state-of-the-art F-score performance. Our results show that using word embeddings for Twitter NER in Turkish can result in better F-score performance compared to using text normalization as a pre-processing step. In addition, utilizing in-domain word embeddings can be a promising approach for Twitter NER.

2. Related Work

There are various important studies of NER on Twitter for English. Ritter et al. (2011) presented a two-phase NER system for tweets, T-NER, using Conditional Random Fields (CRF) and including tweet-specific features. Liu et al. (2011) proposed a hybrid NER approach based on K-Nearest Neighbors and linear CRF. Liu et al. (2012) presented a factor graph-based method for NER on Twitter. Li et al. (2012) described an unsupervised approach for tweets, called TwiNER. Bontcheva et al. (2013) described an NLP pipeline for tweets, called TwitIE. Very recently, Cherry and Guo (2015) have shown the effectiveness of Brown clusters and word vectors on Twitter NER for English.

For Turkish NER on formal texts, Tür et al. (2003) presented the first study with a Hidden Markov Model based approach. Tatar and Çiçekli (2011) presented an automatic rule learning system. Yeniterzi (2011) used CRF for Turkish NER, and Küçük and Yazıcı (2012) proposed a hybrid approach. A CRF-based model by Şeker and Eryiğit (2012) is the state-of-the-art Turkish NER system with CoNLL F-score of 91.94%, using gazetteers. Demir and Özgür (2014) achieved a similar F-score of 91.85%, without gazetteers and language dependent features, using a semi-supervised model with word embeddings.

For Turkish NER on Twitter, Çelikkaya et al. (2013) presented the first study by adopting the CRF-based NER of Şeker and Eryiğit (2012) with a text normalizer. Küçük et al. (2014) adopted a multilingual rule-based NER by extending the resources for Turkish. Küçük and Steinberger (2014) adopted a rule-based approach for Turkish tweets, where diacritics-based expansion to lexical resources and relaxing the capitalization yielded an F-score of 48% with strict CoNLL-like metric.

3. NER for Turkish Tweets using Semi-supervised Learning

To build a NER model with a semi-supervised learning approach on Turkish tweets, we used a neural network based architecture consisting of unsupervised and supervised stages.

3.1. Unsupervised Stage

In the unsupervised stage, our aim is to learn distributed word representations, or word embeddings, in continuous vector space where semantically similar words are expected to be close to each other. Word vectors trained on large unlabeled Turkish corpus can provide additional knowledge base for NER systems trained with limited amount of labeled data in the supervised stage.

A word representation is usually a vector associated with each word, where each dimension represents a feature. The value of each dimension is defined to be representing the amount of activity for that specific feature. A distributed representation represents each word as a dense vector of continuous values. By having lower dimensional dense vectors, and by having real values at each dimension, distributed word representations are helpful to solve the sparsity problem. Distributed word representations are trained with a huge unlabeled corpus using unsupervised learning. If this unlabeled corpus is large enough, then we expect that the distributed word representations will capture the syntactic and semantic properties of each word and this will provide a mechanism to obtain similar representations for semantically and syntactically close words.

Vector space distributed representations of words are helpful for learning algorithms to reach better results in many NLP tasks, since they provide a method for grouping similar words together. The idea of using distributed word representations in vector space is applied to statistical language modeling for the first time by using a neural network based approach with a significant success by Bengio et al. (2003). The approach is based on learning a distributed representation of each word, where each dimension of such a word embedding represents a hidden feature of this word and is used to capture the word's semantic and grammatical properties. Later on, Collobert et al. (2011) proposed to use distributed word representations together with the supervised neural networks and achieved state-of-the art results in different NLP tasks, including NER for English.

We used the public tool, word2vec[1], released by Mikolov et al. (2013) to obtain the word embeddings. Their neural network approach is similar to the feed-forward neural networks (Bengio et al., 2003; Collobert et al., 2011). To be more precise, the previous words to the current word are encoded in the input layer and then projected to the projection layer with a shared projection matrix. After that, the projection is given to the non-linear hidden layer and then the output is given to softmax in order to receive a probability distribution over all the words in the vocabulary. However, as suggested by Mikolov et al. (2013), removing the non-linear hidden layer and making the projection layer shared by all words is much faster, which allowed us to use a larger unlabeled corpus and obtain better word embeddings.
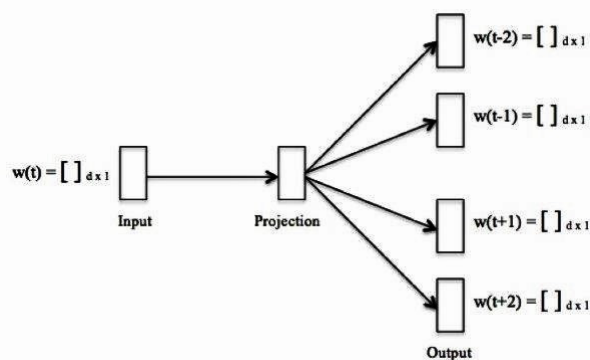

Figure 1: Skip-gram model architecture to learn continuous vector representation of words in order to predict surrounding words (Mikolov et al., 2013).

Among the methods presented in Mikolov et al. (2013), we used the continuous Skip-gram model to obtain semantic representations of Turkish words. The Skip-gram model uses the current word as an input to the projection layer

---

[1]https://code.google.com/p/word2vec/

with a log-linear classifier and attempts to predict the representation of neighboring words within a certain range. In the Skip-gram model architecture we used, we have chosen 200 as the dimension of the obtained word vectors. The range of surrounding words is chosen to be 5, so that we will predict the distributed representations of the previous 2 words and the next 2 words using the current word. Our vector size and range decisions are aligned with the choices made in the previous study for Turkish NER by Demir and Özgür (2014). The Skip-gram model architecture we used is shown in Figure 1.

3.2. Supervised Stage

In this stage, a comparably smaller amount of labeled data is used for training the final NER models. We used the publicly available neural network implementation by Turian et al. (2010)[2], which actually follows the study by Ratinov and Roth (2009), where a regularized averaged multiclass perceptron is used.

Note that although non-local features are proven to be useful for the NER task on formal text types such as news articles, their usage and benefit is questionable for informal and short text types. Due to the fact that each tweet is treated as a single document with only 140 characters, it is difficult to make use of non-local features such as context aggregation and prediction history for the NER task on tweets. On the other hand, local features are mostly related to the previous and next tokens of the current token. With this motivation, we explored both local and non-local features but observed that we achieve better results without non-local features. As a result, to construct our NER model on tweets, we used the following local features:

- Context: All tokens in the current window of size two.

- Capitalization: Boolean feature indicating whether the first character of a token is upper-case or not. This feature is generated for all the tokens in the current window.

- Previous tags: Named entity tag predictions of the previous two tokens.

- Word type information: Type information of tokens in the current window, i.e. all-capitalized, is-capitalized, all-digits, contains-apostrophe, and is-alphanumeric.

- Token prefixes: First characters with length three and four, if exists, of current token.

- Token suffixes: Last characters with length one to four, if exists, of current token.

- Word embeddings: Vector representations of words in the current window.

In addition to tailoring the features used by Ratinov and Roth (2009) for tweets, there are other Twitter-specific aspects of our NER system such as using word embeddings

trained on an unlabeled tweet corpus, applying normalization on labeled tweets, and extracting Twitter-specific keywords like hashtags, mentions, smileys, and URLs from both labeled and unlabeled Turkish tweets. For text normalization as a pre-processing step of our system, we used the Turkish normalization interface[3] developed for social media text with ill formed word detection and candidate word generation (Torunoğlu and Eryiğit, 2014).

Along with the features used, the representation scheme for named entities is also important in terms of performance for a NER system. Two popular such encoding schemes are BIO and BILOU. The BIO scheme identifies the Beginning, the Inside and the Outside of the named entities, whereas the BILOU scheme identifies the Beginning, the Inside and the Last tokens of multi-token named entities, plus the Outside if it is not a named entity and the Unit length if the entity has single token. Since it is shown by Ratinov and Roth (2009) that BILOU representation scheme significantly outperforms the BIO encoding scheme, we make use of BILOU encoding for tagging named entities in our study. Furthermore, we applied normalization to numerical expressions as described in Turian et al. (2010), which helps to achieve a degree of abstraction to numerical expressions.

4. Data Sets

4.1. Unlabeled Data

In the unsupervised stage, we used two types of unlabeled data to obtain Turkish word embeddings. The first one is a Turkish news-web corpus containing 423M words and 491M tokens, namely the BOUN Web Corpus[4] (Sak et al., 2008; Sak et al., 2011). The second one is composed of 21M Turkish tweets with 241M words and 293M tokens, where we combined 1M tweets from TS TweetS[5] by Sezer and Sezer (2013) and 20M Turkish Tweets[6] by Bolat and Amasyalı.

We applied tokenization on both Turkish news-web corpus and Turkish tweets corpus using the publicly available Zemberek[7] tool developed for Turkish. We have also applied lower-casing on both corpora in order to limit the number of unique words. Since our combined tweets corpus is composed of Twitter-specific texts, we applied what we call Twitter processing where we replaced mentions, hashtags, smileys and URLs with certain keywords.

4.2. Labeled Data

In the supervised stage, we used two types of labeled data to train and test our NER models. The first one is Turkish news data annotated with ENAMEX-type named entities, or PLOs (Tür et al., 2003). It includes 14481 person, 9409 location, and 9034 organization names in the training partition of 450K words. This data set is popularly used for performance evaluation of NER systems for Turkish, including the ones presented by Şeker and Eryiğit (2012), by Yeniterzi (2011) and by Demir and Özgür (2014).

The second type of labeled data is annotated Turkish tweets, where we used two different sets. The first set, TwitterDS-1, has around 5K tweets with 54K tokens and 1336 annotated PLOs (Çelikkaya et al., 2013). The second set, TwitterDS-2[8], which is publicly available, has 2320 tweets with around 21K tokens and 980 PLOs in total (Küçük et al., 2014). The counts for each of the ENAMEX-type named entities for these Turkish Twitter data sets are provided in Table 1.

	Twitter DS-1 (TwtDS-1)	**Twitter DS-2** (TwtDS-2)
Data Size (#tokens)	54K	21K
Person	676	457
Location	241	282
Organization	419	241
Total PLOs	1336	980

Table 1: Number of PLOs in Turkish Twitter data sets.

5. Experiments and Results

We designed a number of experimental settings to investigate their effects on Turkish Twitter NER. These settings are as follows: the text type of annotated data used for training, the text type of unlabeled data used to learn the word embeddings, using the capitalization feature or not, and applying text normalization. We evaluated all models on ENAMEX types with the CoNLL metric and reported phrase-level overall F-score performance results. To be more precise, the F-score values presented in Table 2, Table 3 and Table 4 are micro-averaged over the classes using the strict metric.

5.1. NER Models Trained on News

Most of our NER models are trained on annotated Turkish news data by Tür et al. (2003) and tested on tweets, due to the limited amount of annotated Turkish tweets.

In addition to using TwitterDS-1 and TwitterDS-2 as test sets, we detected 291 completely non-Turkish tweets out of 5040 in TwitterDS-1 and filtered them out using the isTurkish[9] tool (Şahin et al., 2013) to obtain TwitterDS-1_FT. We also used the normalized versions of these data sets. As shown in Table 2, turning off the capitalization feature is better when text normalization is not applied (bold entries), but the best results are achieved when normalization is applied and the capitalization feature is used (underlined bold entries). To observe the effects of the type of the source text used to learn the word embeddings, we have three models as Web, Twt, and Web+Twt where we used the Turkish web corpus, tweet corpus, and their combination respectively to learn the word embeddings. Including in-domain data from a relatively smaller tweet corpus together with a larger web corpus yields in better Twitter NER performance.

Test Set	Cap	Phrase-level (Overall)		
		Web	Twt	W+T
TwtDS-1	ON	36.55	35.14	38.11
	OFF	38.52	31.57	**40.18**
TwtDS-1_Norm	ON	41.82	41.54	**42.79**
	OFF	40.50	39.31	41.04
TwtDS-1_FT	ON	40.53	40.44	41.86
	OFF	41.63	36.43	**44.00**
TwtDS-1_FT_Norm	ON	45.74	46.27	**46.61**
	OFF	44.17	44.91	45.27
TwtDS-2	ON	53.14	47.72	54.01
	OFF	54.09	48.15	**55.45**
TwtDS-2_Norm	ON	55.20	52.23	**56.79**
	OFF	54.75	49.43	56.12

Table 2: Phrase-level overall F-score performance results of the NER models trained on news.

5.1.1. Word Embeddings versus Text Normalization

We examined the effects of word embeddings on the performance of our NER models, and compared them to the improvements achieved by applying normalization on Turkish tweets. The baseline NER model is built by using the features explained in section 3.2, except the capitalization and word embeddings features. Using word embeddings obtained with unsupervised learning from a large corpus of web articles and tweets results in better NER performance than applying a Twitter-specific text normalizer, as shown in Table 3. This is crucial since Turkish text normalization for unstructured data is a challenging task and requires successful morphological analysis, whereas extracting word embeddings for any language or domain is much easier, yet more effective.

NER Model	Phrase-level (Overall)		
	Twt DS-1	Twt DS-1_FT	Twt DS-2
Baseline(BL)	22.16	25.98	35.16
BL+**Norm**	33.05	39.23	37.17
BL+**WordE**	40.18	44.00	55.45
BL+**WordE+Norm**	**41.04**	**45.27**	**56.12**
Baseline(BL)+Cap	27.16	30.21	37.32
BL+Cap+**Norm**	36.70	40.78	42.18
BL+Cap+**WordE**	38.11	41.86	54.01
BL+Cap+**WordE+Norm**	**42.79**	**46.61**	**56.79**

Table 3: Phrase-level overall F-score performance results to compare word embeddings and normalization.

5.2. NER Models Trained on Tweets

Although an ideal Turkish NER model for Twitter should be trained on similar informal texts, all previous Turkish Twitter NER systems are trained on news data due to the limited amount of annotated Turkish tweets. We also experimented training NER models on relatively smaller labeled Twitter data with 10-fold cross-validation. Our best phrase-

System	Trained On	Best Settings				Test Set	Phrase-level (Overall)
		Gazet	Norm	Cap	Other		
Çelikkaya et al., (2013)	Turkish News (Tür et al., 2003)	Yes	Yes	ON	-	TwtDS-1	19.28
Küçük et al., (2014)	EMM News	Yes	No	ON	relaxed & extended gazetteer	TwtDS-1	36.11
						TwtDS-2	42.68
Küçük and Steinberger, (2014)	no training	Yes	No	OFF	diacritics expanded gazetteer	*TwtDS-1*	*38.01*
						TwtDS-2	*48.13*
Our NER Systems	Turkish News (Tür et al., 2003)	No	Yes	ON	word embeddings + filter non-Turkish	TwtDS-1	46.61
					word embeddings	**TwtDS-2**	**56.79**
	Turkish Tweets (TwtDS-2)	No	Yes	ON	word embeddings + filter non-Turkish	**TwtDS-1**	**48.96**

Table 4: Phrase-level overall F-score performance results compared to the state-of-the-art.

level F-score of 46.61% achieved on TwitterDS-1_FT is increased to 48.96% when trained on the much smaller tweets data, TwitterDS-2, instead of news data.

5.3. Comparison with the State-of-the-art

The best F-scores of the previously published Turkish Twitter NER systems (Çelikkaya et al., 2013; Küçük et al., 2014; Küçük and Steinberger, 2014) as well as our proposed NER system are shown in Table 4. We used the same training set with the first system (Çelikkaya et al., 2013) in our study, but the second NER system (Küçük et al., 2014) uses a different multilingual news data and the third system (Küçük and Steinberger, 2014), which is rule based, does not have a training phase at all. All of these previous NER systems use gazetteer lists for named entities, which are manually constructed and highly language-dependent, whereas our system does not. Note that there is no publicly available gazetteer lists in Turkish. Küçük and Steinberger (2014) achieved the state-of-the-art performance results for Turkish Twitter NER with their best model settings (shown in italic). These settings are namely using gazetteers list, with capitalization feature turned off, and with no normalization, together by expanding their gazetteer lists of named entities with diacritics variations.

Our proposed system outperforms the state-of-the-art results on both Turkish Twitter data sets, even without using gazetteers (shown in bold). We achieved our best performance results with Turkish word embeddings obtained from our Web+Tweets corpus, when we apply normalization on tweets and keep the capitalization as a feature.

6. Conclusion

We adopted a neural networks based semi-supervised approach using word embeddings for the NER task on Turkish tweets. At the first stage, we attained distributed representations of words by employing a fast unsupervised learning method on a large unlabeled corpus. At the second stage, we exploited these word embeddings together with language independent features in order to train our neural network on labeled data. We compared our results on two different Turkish Twitter data sets with the state-of-the-art NER systems proposed for Twitter data in Turkish

and showed that our system outperforms the state-of-the-art results on both data sets. Our results also show that using word embeddings from an unlabeled corpus can lead to better performance than applying Twitter-specific text normalization. We discussed the promising benefits of using in-domain data to learn word embeddings at the unsupervised stage as well. Since the only language dependent part of our Turkish Twitter NER system is text normalization, and since even without text normalization it outperforms the previous state-of-the-art results, we believe that our approach can be adapted to other morphologically rich languages. Our Turkish Twitter NER system, namely TTNER, is publicly available[10].

We believe that there is still room for improvement for NLP tasks on Turkish social media data. As a future work, we aim to construct a much larger in-domain resource, i.e., unlabeled Turkish tweets corpus, and investigate the full benefits of attaining word embeddings from in-domain data on Twitter NER.

8. Bibliographical References

Bengio, Y., Ducharme, R., Vincent, P., and Janvin, C. (2003). A neural probabilistic language model. *J. Mach. Learn. Res.*, 3:1137–1155, March.

Bontcheva, K., Derczynski, L., Funk, A., Greenwood, M. A., Maynard, D., and Aswani, N. (2013). TwitIE: An open-source information extraction pipeline for microblog text. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing*. Association for Computational Linguistics.

---

[10]http://tabilab.cmpe.boun.edu.tr/projects/ttner/

Çelikkaya, G., Torunoğlu, D., and Eryiğit, G. (2013). Named entity recognition on real data: A preliminary investigation for Turkish. In *Proceedings of the 7th International Conference on Application of Information and Communication Technologies, AICT2013*, Baku, Azarbeijan, October. IEEE.

Cherry, C. and Guo, H. (2015). The unreasonable effectiveness of word representations for twitter named entity recognition. In *NAACL HLT 2015, The 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Denver, Colorado, USA, May 31 - June 5, 2015*, pages 735–745.

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, November.

Demir, H. and Özgür, A. (2014). Improving named entity recognition for morphologically rich languages using word embeddings. In *13th International Conference on Machine Learning and Applications, ICMLA 2014, Detroit, MI, USA, December 3-6, 2014*, pages 117–122.

Hong, L. and Davison, B. D. (2010). Empirical study of topic modeling in twitter. In *Proceedings of the First Workshop on Social Media Analytics*, SOMA '10, pages 80–88, New York, NY, USA. ACM.

Kireyev, K., Palen, L., and Anderson, K. M. (2009). Applications of topics models to analysis of disaster-related twitter data. In *Proceedings of NIPS Workshop on Applications for Topic Models: Text and Beyond*.

Kökciyan, N., Çelebi, A., Özgür, A., and Üsküdarlı, S. (2013). Bounce: Sentiment classification in twitter using rich feature sets. In *Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, pages 554–561, Atlanta, Georgia, USA, June. Association for Computational Linguistics.

Küçük, D. and Steinberger, R. (2014). Experiments to improve named entity recognition on Turkish tweets. In *Proceedings of the 5th Workshop on Language Analysis for Social Media (LASM)*, pages 71–78, Gothenburg, Sweden, April. Association for Computational Linguistics.

Küçük, D. and Yazıcı, A. (2012). A hybrid named entity recognizer for Turkish. *Expert Syst. Appl.*, 39(3):2733–2742, February.

Küçük, D., Jacquet, G., and Steinberger, R. (2014). Named entity recognition on Turkish tweets. In Nicoletta Calzolari (Conference Chair), et al., editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland, may. European Language Resources Association (ELRA).

Li, C., Weng, J., He, Q., Yao, Y., Datta, A., Sun, A., and Lee, B.-S. (2012). Twiner: Named entity recognition in targeted twitter stream. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '12, pages 721–730, New York, NY, USA. ACM.

Liu, X., Zhang, S., Wei, F., and Zhou, M. (2011). Recognizing named entities in tweets. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1*, HLT '11, pages 359–367, Stroudsburg, PA, USA. Association for Computational Linguistics.

Liu, X., Zhou, M., Wei, F., Fu, Z., and Zhou, X. (2012). Joint inference of named entity recognition and normalization for tweets. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers - Volume 1*, ACL '12, pages 526–535, Stroudsburg, PA, USA. Association for Computational Linguistics.

Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient estimation of word representations in vector space. *CoRR*, abs/1301.3781.

Ratinov, L. and Roth, D. (2009). Design challenges and misconceptions in named entity recognition. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning*, CoNLL '09, pages 147–155, Stroudsburg, PA, USA. Association for Computational Linguistics.

Ritter, A., Clark, S., Mausam, and Etzioni, O. (2011). Named entity recognition in tweets: An experimental study. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP '11, pages 1524–1534, Stroudsburg, PA, USA. Association for Computational Linguistics.

Şahin, M., Sulubacak, U., and Eryiğit, G. (2013). Redefinition of Turkish morphology using flag diacritics. In *Proceedings of The Tenth Symposium on Natural Language Processing (SNLP-2013)*, Phuket, Thailand, October.

Sak, H., Güngör, T., and Saraçlar, M. (2008). Turkish language resources: Morphological parser, morphological disambiguator and web corpus. In *GoTAL 2008*, volume 5221 of *LNCS*, pages 417–427. Springer.

Sak, H., Güngör, T., and Saraçlar, M. (2011). Resources for Turkish morphological processing. *Language Resources and Evaluation*, 45(2):249–261.

Şeker, G. A. and Eryiğit, G. (2012). Initial explorations on using CRFs for Turkish named entity recognition. In *Proceedings of COLING 2012*, Mumbai, India, 8-15 December.

Sezer, T. and Sezer, B. S. (2013). TS corpus: Herkes için Türkçe derlem. In *Proceedings of the 27th National Linguistics Conference*, pages 217–255, Antalya, Turkey.

Tatar, S. and Çiçekli, I. (2011). Automatic rule learning exploiting morphological features for named entity recognition in Turkish. *J. Information Science*, 37(2):137–151.

Torunoğlu, D. and Eryiğit, G. (2014). A cascaded approach for social media text normalization of Turkish. In *5th Workshop on Language Analysis for Social Media (LASM) at EACL*, Gothenburg, Sweden, April. Association for Computational Linguistics.

Tür, G., Hakkani-Tür, D., and Oflazer, K. (2003). A statistical information extraction system for Turkish. *Natural Language Engineering*, 9(2):181–210.

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, pages 384–394, Stroudsburg, PA, USA. Association for Computational Linguistics.

Yeniterzi, R. (2011). Exploiting morphology in Turkish named entity recognition system. In *Proceedings of the ACL 2011 Student Session*, HLT-SS '11, pages 105–110, Stroudsburg, PA, USA. Association for Computational Linguistics.