# Korean Morphological Analysis with Tied Sequence-to-Sequence Multi-Task Model

**Hyun-Je Song**
Dept. of Information Technology
Chonbuk National University
Jeonju, 54896, Korea
hyunje.song@jbnu.ac.kr

**Seong-Bae Park**
Dept. of Computer Science and Engineering
Kyung Hee University
Yongin, 17104, Korea
sbpark71@khu.ac.kr

## Abstract

Korean morphological analysis has been considered as a sequence of morpheme processing and POS tagging. Thus, a pipeline model of the tasks has been adopted widely by previous studies. However, the model has a problem that it cannot utilize interactions among the tasks. This paper formulates Korean morphological analysis as a combination of the tasks and presents a tied sequence-to-sequence multi-task model for training the two tasks simultaneously without any explicit regularization. The experiments prove the proposed model achieves the state-of-the-art performance.

## 1 Introduction

Korean is an agglutinative language (Song, 2006). Thus, it is a fundamental step for understanding a sentence to analyze the grammatical structure of *eojeol*s, where an eojeol is a linguistic unit segmented by a white space. An eojeol is composed of one or more morphemes. As a result, one eojeol can be analyzed into several morpheme combinations depending on a context, which yields different part-of-speech (POS) tags of the morpheme combinations. In addition, some morphemes have a different surface form from their base form when they are derived from an eojeol. Therefore, the goal of Korean morphological analyzer is not only to decompose and recover morphemes from eojeols precisely (morpheme processing), but also to assign POS tags to the decomposed and/or recovered morphemes accurately according to a context (POS tagging).

Traditional approaches to Korean morphological analysis have adopted a pipeline model of morpheme processing and POS tagging (Lee and Rim, 2009; Na, 2015; Choi et al., 2016; Matteson et al., 2018; Song and Park, 2018). That is, they first decompose and recover morphemes from eojeols or assign so-called POSMORPH tags (Heigold et al., 2016), and then an actual POS tag sequence is determined or resolved from the POSMORPH tags using a sequential labeling algorithm. However, this pipeline model suffers from two kinds of weaknesses. One weakness is that the errors from morpheme processing are apt to be propagated to POS tagging, and the other is that it is difficult to model mutual interactions between morpheme processing and POS tagging under the pipeline model.

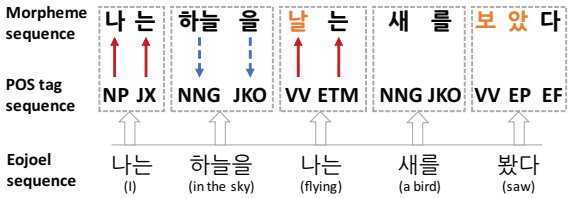


Figure 1: Korean morphological analysis of a sentence "나는 하늘을 나는 새를 봤다" of which meaning is "*I saw a bird flying in the sky*". Correct morpheme analysis helps predicting POS tags (blue dotted arrows) while POS tagging affects morpheme analysis (red arrows). Orange morphemes in the morpheme sequence are recovered morphemes. Best viewed in color.

In Korean, morpheme processing and POS tagging affect each other. That is, correct morpheme analysis helps POS tagging, and POS tags are helpful in analyzing morphemes. Figure 1 shows such an example in Korean morphological analysis. The second eojeol '하늘을 (in the sky)' is composed of two morphemes '하늘 (sky)' and '을 (*objective postposition*)'. If '하늘을' is precisely decomposed into '하늘' and '을', their POS tags can be predicted easily. This is the information flow on which the previous studies focus. On the other hand, the eojeol '나는' which appears twice in this figure is morphologically ambiguous and thus can



*Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pages 1436–1441,
Hong Kong, China, November 3–7, 2019. 

be analyzed in two ways. However, if morpheme processing obtains some information from POS tagging, it can decompose ambiguous eojeols correctly. That is, the first '나는' is decomposed into '나 (I)' and '는 (*topical postposition*)', while the second '나는' is into '날 (fly)' and '는 (*verbal ending*)'. Therefore, morpheme processing and POS tagging should be trained simultaneously.

This paper proposes a model to train morpheme processing and POS tagging simultaneously in Korean morphological analysis. The proposed model regards morpheme processing and POS tagging as individual tasks. Then, the two tasks are jointly trained under a sequence-to-sequence multi-task framework (Luong et al., 2016; Anastasopoulos and Chiang, 2018). The main characteristic of the proposed model is that there is a single decoder for both generating a morpheme sequence and assigning POS tags to the generated morphemes. As a result, the decoder shares general representations of both tasks and forces a one-to-one mapping between morphemes and POS tags without additional regularization. Morpheme processing is accomplished by a pointer-generator network (See et al., 2017), while POS tagging is done by a CRF network (Huang et al., 2015; Lample et al., 2016) utilizing the information from morpheme processing. Our experimental results show that the proposed model outperforms existing Korean morphological analyzers with its state-of-the-art performance.

## 2 Korean Morphological Analysis

### 2.1 Morpheme Processing and POS Tagging

Given an input sequence $\mathbf{x} = (w_1, ..., w_n)$ where $w_i$ is the $i$-th eojeol, Korean morphological analysis aims to produce an output sequence $\mathbf{y} = (\langle m_1, t_1 \rangle, ..., \langle m_k, t_k \rangle)$ where $m_j$ is the $j$-th morpheme and $t_j$ is its POS tag. Korean morphological analysis is different from existing NLP tasks such as English POS tagging (Toutanova et al., 2003; Manning, 2011), and joint word segmentation and POS tagging for Chinese (Zhang and Clark, 2008; Shao et al., 2017; Chen et al., 2017). In English POS tagging, a word and its tag have a one-to-one mapping so that the length of an input sequence is equal to that of an output sequence. On the other hand, they are usually different in Korean, because an eojeol consists of several morphemes. That is, $n < k$ in general. The key distinguishing factor between Chinese and Korean morphological analysis is that lemmatization that recovers a base form for every morpheme as well as morpheme segmentation is a must in Korean morphological analysis, while Chinese morphological analysis requires only word segmentation. That is, in Korean, $\bigoplus_{i=1}^{n} w_i \neq \bigoplus_{i=1}^{k} m_i$ in general, where $\bigoplus$ is a concatenate operator. According to Figure 1, the eojeol sequence '나는 하늘을 나는 새를 봤다' is different from its morpheme sequence '나는 하늘을 날는 새를 보았다' where underline marks indicate the difference.[1]

Given an annotated corpus $\mathcal{D} = (\langle \mathbf{x}_i, \mathbf{y}_i \rangle)_{i=1}^{N}$ where $N$ is the number of sentences, Korean morphological analyzer is obtained by maximizing the conditional probability $p(\mathbf{y}|\mathbf{x})$. The conditional probability can be calculated as

$$p(\mathbf{y}|\mathbf{x}) \approx \underbrace{p(\mathbf{t}|\mathbf{m}, \mathbf{x})}_{\text{POS tagging}} \quad \underbrace{p(\mathbf{m}|\mathbf{x})}_{\text{Morpheme processing}}, \tag{1}$$

where $\mathbf{m} = (m_1, \ldots, m_k)$ is a morpheme sequence, and $\mathbf{t} = (t_1, \ldots, t_k)$ is a tag sequence. Equation (1) implies that $p(\mathbf{y}|\mathbf{x})$ can be further decomposed into two another conditional probabilities. One conditional probability, $p(\mathbf{m}|\mathbf{x})$, generates a morpheme sequence $\mathbf{m}$ from the input sequence $\mathbf{x}$, and the other $p(\mathbf{t}|\mathbf{m}, \mathbf{x})$ assigns POS tags considering the morpheme sequence $\mathbf{m}$ and the input sequence $\mathbf{x}$. Therefore, $p(\mathbf{m}|\mathbf{x})$ corresponds to morpheme processing, while $p(\mathbf{t}|\mathbf{m}, \mathbf{x})$ is POS tagging.

### 2.2 Linguistic Unit in Morphological Analysis

Eojeol is the unit of spacing in Korean sentences. However, it is inappropriate to use eojeols directly in sequence-to-sequence models, because the number of eojeols is extremely huge due to the agglutinative characteristics of Korean. For instance, there exist 624,655 kinds of eojeols in the corpus used in the experiments. This large eojeol size causes a complexity problem in morphological analysis.

According to Korean orthography, an eojeol is a sequence of *syllables*. For instance, an eojeol '하늘을' is a sequence of three syllables '하', '늘', and '을'. A morpheme in Korean is also a sequence of syllables like an eojeol. The number of distinguished syllables in the corpus above is just

[1] We skip morpheme segmentation symbols in the morpheme sequence in order to emphasis the difference between the eojeol sequence and the morpheme one.

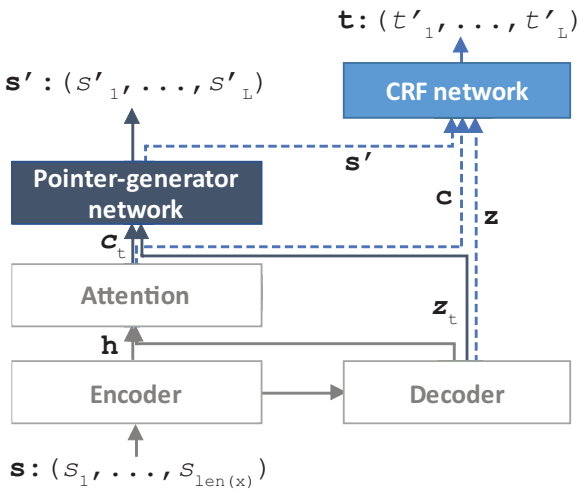


Figure 2: The proposed model for Korean morphological analysis based on a Tied sequence-to-sequence multi-task model. Since syllable is adopted as a unit for the proposed model, the input for the encoder is a syllable sequence $\mathbf{s}$. For clarity's sake, there are some dependencies not shown.

5,245 which is much smaller than that of eojeols. Therefore, this paper adopts syllable as a unit for sequence-to-sequence models for Korean morphological analysis.

## 3 Tied Sequence-to-Sequence Multi-Task Model

The proposed model is based on the sequence-to-sequence model with attention (Bahdanau et al., 2015) and extends the model for multi-task learning similarly to the work of Anastasopoulos and Chiang (2018). The proposed model consists of four parts: a recurrent encoder, a recurrent decoder, attention, and task-dependent networks. Figure 2 shows an overall structure of the proposed model. The encoder encodes an input sequence $\mathbf{x}$ into a sequence of hidden states. Since syllable is adopted as the unit for the model, the encoder (bidirectional LSTM) actually transforms an input syllable sequence $\mathbf{s} = (s_1, \ldots, s_{len(\mathbf{x})})$ into a sequence of encoder hidden states $\mathbf{h}_1, \ldots, \mathbf{h}_{len(\mathbf{x})}$, where $len(\cdot)$ is the number of syllables including white spaces. The attention transforms the encoder hidden states $\mathbf{h}$ into a sequence of context vectors $\mathbf{c}$ through attention weights. The context vectors capture relevant source-side information for both morpheme processing and POS tagging.

Unlike other multi-task models, there is only one decoder for the two tasks of morpheme processing and POS tagging in the proposed model. Note that a morpheme and its POS tag should have a one-to-one mapping relation. Since the two tasks are forced to share the same decoder states, a one-to-one mapping between a morpheme and its POS tag is guaranteed without any explicit regularization or task-specific token. In detail, the decoder (unidirectional LSTM) first computes a sequence of decoder hidden states $\mathbf{z}$. At every time $t$, the decoder hidden state $\mathbf{z}_t$ is calculated from the embedding of the previous syllable $\mathbf{s}'_{t-1}$, the decoder state $\mathbf{z}_{t-1}$, and the context vector $\mathbf{c}_t$. Then, two task-dependent networks solve their own task using the same $\mathbf{z}$.

In morpheme processing represented by solid arrows in Figure 2, a syllable $s'_t$ is produced at every time $t$. Most syllables are directly copied from the input syllables while some syllables are newly generated from a vocabulary to recover the base form of a morpheme. To do this, we adopt the pointer-generator network (See et al., 2017) which allows both copying syllables via pointing and generating syllables from a fixed vocabulary. The advantage of this approach is that we can handle non-Korean characters such as *hanja* (Chinese character) and special symbols with ease.

In POS tagging represented by dotted arrows in Figure 2, it is possible to generate a POS tag at every time $t$ by paying attention to the encoder syllable states as done in morpheme processing. However, this approach does not reflect the global dependency among adjacent tags. To solve this problem, this paper adopts a CRF network (Huang et al., 2015; Lample et al., 2016) over the pointer-generator network. When the morpheme sequence generated at the morpheme processing is given, the CRF network predicts POS tags for the morpheme sequence by considering the global dependency among tags (Lafferty et al., 2001). Furthermore, this paper uses skip-connections (He et al., 2016) in order for the CRF network to pay attention to the hidden states of the decoder. Since some POS tags can be predicted directly from input syllables, the skip-connections help the CRF network consider the information of input syllables. In detail, a sequence of syllables generated by the pointer-generator network is transformed to vectors, and the vectors are concatenated with the decoder state $\mathbf{z}$ and the context vector $\mathbf{c}$. Then, the concatenated vectors are fed to the CRF layer to predict POS tags.

| Methods | | | | F1-score | Acc (%) |
|---|---|---|---|---|---|
| Model | Morpheme processing | POS tagging | Type | | |
| Single-Task | Pointer-Generator | | | 97.36 | 95.57 |
| Multi-Task | Generator | Generator | Non-cascade | 97.10 | 95.09 |
| | Generator | CRF | Cascade | 97.25 | 95.30 |
| | Generator | CRF | Tied | 97.31 | 95.49 |
| | Pointer-Generator | Generator | Non-cascade | 97.30 | 95.45 |
| | Pointer-Generator | CRF | Cascade | 97.40 | 95.61 |
| | Pointer-Generator | CRF | Tied | **97.43** | **95.68** |
| Pipeline | Na (2015) (CRF → CRF → Post-processing) | | | 97.21 | 95.22 |
| | Song and Park (2018) (Generator → CRF) | | | 97.27 | 95.29 |
| | khaiii (CNN → Post-processing) | | | 94.88 | 91.86 |

Table 1: Performances of Korean morphological analyzers

| Information | Training | Development | Test |
|---|---|---|---|
| Sentences | 197,508 | 5,000 | 50,631 |
| Eojeols | 2,674,571 | 97,292 | 694,524 |
| Morphemes | 5,952,985 | 203,244 | 1,542,483 |
| Avg. eojeols | 13.54 | 19.46 | 13.72 |
| No. of POS tags | 42 | | |

Table 2: Simple statistics on the data set used.

Note that syllable is the unit for the model and one POS tag spans several syllables of a morpheme. Thus, to identify morpheme boundaries, morpheme processing of the proposed model generates a special symbol of morpheme ending for every explicit morpheme boundary instead of adopting the IOB tagging scheme. Such an approach reduces the number of POS tags and is helpful in segmenting eojeols into morphemes precisely.

The proposed model is trained to maximize the weighted sum of conditional log-likelihood of each task where the weights are set to be equal.

## 4 Experiments

### 4.1 Experimental Settings

The same data set with the work of Na (2015) is used for the experiments. This data set is derived from Sejong corpus.[2] Each eojeol in the Sejong corpus is annotated with pairs of a morpheme and its POS tag. The simple statistics on the data set used in the experiments is given in Table 2.

We set the syllable embedding size and the hidden size to 100. The number of LSTM layers is set to three and the batch size is 128. We use the gradient normalization with a threshold of five.

Three baselines are adopted to show the superiority of the proposed model. The first baseline is the model of Na (2015). This model consists of three sub-models (CRF segmentation, CRF tagging, and post-processing) and executes these three models in consecutive order. The second is the model of Song and Park (2018) which consists of two sub-models (generator and CRF tagger). The last is khaiii, a publicly-available CNN based morphological analyzer.[3] This analyzer assigns a POSMORPH tag for every syllable using a CNN classifier and then resolves the tag through post-processing. These baselines are all pipeline models. We also compare the proposed model with a single task non-pipeline model which generates a single sequence of morphemes and POS tags using the pointer-generator network.

The F1-measure at the morpheme level and the accuracy at the eojeol level are used as evaluation metrics, and all performances are micro-averaged.

### 4.2 Experimental Results

Table 1 shows the performances of the proposed model and the baselines. It also contains the results of the ablation study on the proposed model by changing the network of sub-tasks or the structure of multi-task models. 'Non-cascade' is similar to the standard multi-task model (Dong et al., 2015) in which each task just shares a decoder and there is no direct connection between tasks. According to Figure 2, there are no dotted arrows (marked $\mathbf{s}'$ and $\mathbf{c}$) from the Pointer-generator network and the CRF network in 'Non-cascade' model. 'Cascade'

[2] Available at https://ithub.korean.go.kr

[3] https://github.com/kakao/khaiii

| Type | Percentage |
|---|---|
| Morpheme segmentation | 5.8% |
| POS tagging | 39.3% |
| Morpheme recovery | 54.9% |

Table 3: The ratio of different error types.

is a multi-task model where the POS tagging network has a connection from the morpheme processing network but no skip connection from the decoder. In Figure 2, 'Cascade' model has no dotted arrow (marked $\mathbf{z}$) from the decoder to the CRF network. 'Tied' is the proposed model that considers both connections.

According to the table, all multi-task models that adopt Pointer-Generator and/or CRF outperform the baseline Generator-Generator, which proves their adoption is effective in improving the performance of morphological analysis. It is also noticed from the table that 'Cascade' multi-task model achieves better performance than 'Non-cascade' model, which implies that the unimpeded information from morpheme processing to POS tagging helps predict accurate POS tags. The proposed 'Tied' multi-task model yields higher performances than 'Cascade' model. When training the 'Tied' multi-task model, the feedbacks from POS tagging network influence decoder states directly. As a result, the 'Tied' model is able to learn better representation for both morpheme processing and POS tagging. All these results indicate that 'Tied' multi-task model with Pointer-Generator network and CRF is the best choice for Korean morphological analysis.

Compared to the pipeline models which showed the state-of-the-art performance previously, even simple multi-task models report a similar performance, which means that morphological analysis can be improved by training morpheme processing and POS tagging jointly. Even if the single-task non-pipeline model achieves a reasonable performance, its performance is not as high as that of the proposed model. This is because the single-task model does not consider the global tag dependency explicitly. To sum up, Korean morphological analysis should be solved by training the two tasks jointly with appropriate networks.

### 4.3 Error Analysis

Even if the proposed method achieves over 97.4% F1-score at the morpheme level, we believe that there still exists some room to improve the performance of Korean morphological analysis. To do this, we analyzed the errors by the proposed method.

Table 3 shows error types and their percentages. Morpheme segmentation is when all results are correct except some morphemes that are decomposed wrongly. This type happens mostly in Korean compound nouns where Korean compound nouns can be written as one or more eojeols. If a compound noun represented as one eojeol is given, the proposed method sometimes decomposes the compound noun into a series of nouns. 5.8% errors belong to this type. POS tagging error type is when morphemes are correctly recovered and segmented but their POS tags are predicted wrongly. Its percentage is 39.3%. In this type, 37% errors occur between noun and proper noun. The remaining but majority errors are related to morpheme recovery. That is, these errors occur when the proposed method fails in recovering morphemes from eojeols accurately. Its percentage is 54.9%. Therefore, it is inferred from the error analysis that developing a more accurate morpheme processing model is required to improve the performance of the proposed Korean morphological analyzer.

## 5 Conclusion

This paper has formulated Korean morphological analysis as a combination of morpheme processing and POS tagging. Thus, the two tasks are trained simultaneously through the tied sequence-to-sequence multi-task model with the pointer-generator network and the CRF network. According to the experiment results, the jointly trained morphological analyzer achieves higher performances than the legacy analyzers which are pipeline models of morpheme processing and POS tagging.[4]

## Acknowledgments

We would like to thank the anonymous reviewers for their helpful comments on our work. This work was supported by Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2017-0-01772).

[4]Our code is available at https://github.com/songhyunje/kma

## References


Antonios Anastasopoulos and David Chiang. 2018. Tied multitask learning for neural speech translation. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 82–91.

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural machine translation by jointly learning to align and translate. In *Proceedings of the 3rd International Conference on Learning Representations*.

Xinchi Chen, Xipeng Qiu, and Xuanjing Huang. 2017. A feature-enriched neural model for joint Chinese word segmentation and part-of-speech tagging. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, pages 3960–3966.

Jihun Choi, Jonghem Youn, and Sang-goo Lee. 2016. A grapheme-level approach for constructing a Korean morphological analyzer without linguistic knowledge. In *2016 IEEE International Conference on Big Data (Big Data)*, pages 3872–3879.

Daxiang Dong, Hua Wu, Wei He, Dianhai Yu, and Haifeng Wang. 2015. Multi-task learning for multiple language translation. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, pages 1723–1732.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778.

Georg Heigold, Guenter Neumann, and Josef van Genabith. 2016. Neural morphological tagging from characters for morphologically rich languages. *CoRR*, abs/1606.06640.

Zhiheng Huang, Wei Xu, and Kai Yu. 2015. Bidirectional LSTM-CRF models for sequence tagging. *arXiv preprint arXiv:1508.01991*.

John D. Lafferty, Andrew McCallum, and Fernando C. N. Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the 18th International Conference on Machine Learning*, pages 282–289.

Guillaume Lample, Miguel Ballesteros, Sandeep Subramanian, Kazuya Kawakami, and Chris Dyer. 2016. Neural architectures for named entity recognition. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 260–270.

Do-Gil Lee and Hae-Chang Rim. 2009. Probabilistic modeling of Korean morphology. *IEEE Transactions on Audio, Speech, and Language Processing*, 17(5):945–955.

Minh-Thang Luong, Quoc V. Le, Ilya Sutskever, Oriol Vinyals, and Lukasz Kaiser. 2016. Multi-task sequence to sequence learning. In *Proceedings of the 4th International Conference on Learning Representations*.

Christopher D. Manning. 2011. Part-of-speech tagging from 97% to 100%: Is it time for some linguistics? In *Proceedings of the 12th International Conference on Computational Linguistics and Intelligent Text Processing*, pages 171–189.

Andrew Matteson, Chanhee Lee, Youngbum Kim, and Heuiseok Lim. 2018. Rich character-level information for Korean morphological analysis and part-of-speech tagging. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 2482–2492.

Seung-Hoon Na. 2015. Conditional random fields for Korean morpheme segmentation and POS tagging. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 14(3):10:1–10:16.

Abigail See, Peter J. Liu, and Christopher D. Manning. 2017. Get to the point: Summarization with pointer-generator networks. In *Proceedings of the 55th Annual Meeting of the AssociationS for Computational Linguistics*, pages 1073–1083.

Yan Shao, Christian Hardmeier, Jörg Tiedemann, and Joakim Nivre. 2017. Character-based joint segmentation and POS tagging for Chinese using bidirectional RNN-CRF. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing*, pages 173–183.

Hyun-Je Song and Seong-Bae Park. 2018. Korean part-of-speech tagging based on morpheme generation. TR-0002-2018, Kyung Hee University, Technical report.

Jae Jung Song. 2006. *The Korean language: Structure, use and context*. Routledge.

Kristina Toutanova, Dan Klein, Christopher D. Manning, and Yoram Singer. 2003. Feature-rich part-of-speech tagging with a cyclic dependency network. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technology*, pages 173–180.

Yue Zhang and Stephen Clark. 2008. Joint word segmentation and POS tagging using a single perceptron. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 888–896.