# Cross-lingual Transfer Learning for Javanese Dependency Parsing

**Fadli Aulawi Al Ghiffari, Ika Alfina,** and **Kurniawati Azizah**
Faculty of Computer Science, Universitas Indonesia, Depok, Indonesia
fadli.aulawi@ui.ac.id
{ika.alfina, kurniawati.azizah}@cs.ui.ac.id

## Abstract

While structure learning achieves remarkable performance in high-resource languages, the situation differs for under-represented languages due to the scarcity of annotated data. This study focuses on assessing the efficacy of transfer learning in enhancing dependency parsing for Javanese—a language spoken by 80 million individuals but characterized by limited representation in natural language processing. We utilized the Universal Dependencies dataset consisting of dependency treebanks from more than 100 languages, including Javanese. We propose two learning strategies to train the model: transfer learning (TL) and hierarchical transfer learning (HTL). While TL only uses a source language to pre-train the model, the HTL method uses a source language and an intermediate language in the learning process. The results show that our best model uses the HTL method, which improves performance with an increase of 10 % for both UAS and LAS evaluations compared to the baseline model.

## 1 Introduction

Despite over 80 million native speakers of Javanese (Simons et al., 2023), this language is underrepresented in NLP due to a scarcity of annotated resources. Limited works in Javanese have focused on stemmer (Soyusiawaty et al., 2020), POS tagger (Askhabi et al., 2020), sentiment analysis (Tho et al., 2021), and machine translation (Lesatari et al., 2021). However, few have explored language structure prediction, such as dependency parsing. Dependency parsing is a process that makes a structural representation of a sentence (Kübler et al., 2009) that produces a structure in the form of a dependency tree represented in a graph consisting of several connected links between words in a sentence.

Recent work, Alfina et al. (2023) created a public gold standard dataset for Javanese with 1000 sentences, published as part of the Universal Dependencies dataset (Zeman et al., 2023). This dataset covers annotation for tokenization, POS tagging, morphological features tagging, and dependency parsing tasks. The most recent parser performance (Alfina et al., 2023) using this dataset is not satisfactory, with only 77.08% on Unlabeled Attachment Score (UAS) and 71.21% on Labeled Attachment Score (LAS). The lack of training data is a typical low-resource problem considered one of the biggest NLP research problems (Ruder, 2023).

Transfer learning (TL) involves leveraging a model's knowledge from a high-resource source domain to improve performance on various NLP tasks, particularly in low-resource domains (Weiss et al., 2016), by transferring learned information to target tasks. Inspired by Maulana et al. (2022) that utilizes cross-lingual transfer learning to develop an Indonesian dependency parser, we want to try to replicate its outcome in Javanese with a limited available dataset. Moreover, we also implement hierarchical transfer learning (HTL) with two stages of transfer learning that offer increased flexibility over TL by enabling knowledge transfer between languages with a significant gap (Luo et al., 2019), as demonstrated in diverse applications, including Javanese text-to-speech (Azizah et al., 2020) and biomedical named entity recognition models (Chai et al., 2022).

We build the dependency parser model for Javanese by adopting model (Ahmad et al., 2019) that uses a self-attention encoder and a graph-based decoder. We utilize the Universal Dependency dataset v1.12 (Zeman et al., 2023) that provides dependency treebanks for more than 100 languages, including Javanese. Both TL and HTL use a selection of source languages determined by LangRank (Lin et al., 2020). Specifically, HTL employs Indonesian as an intermediary language, developing from our referenced research (Maulana et al., 2022). The empirical results show that transfer learning improves accuracy with a margin of 10% compared to



*Proceedings of the 13th International Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics: Student Research Workshop*, pages 1–9
November 1–4, 2023. 

the baseline. We also report the word embedding comparison that fastText performs better than the Javanese BERT, Javanese RoBERTa, and multilingual BERT. In summary, the main contributions of this paper are as follows:

1. Provide the first study of Javanese dependency parsing using TL and HTL strategy. We report that the HTL method can significantly improve performance compared to the training from scratch method.

2. Report the investigation of which source language and word embedding performs best for TL and HTL strategy.

## 2 Related Works

### 2.1 Dependency Parser

The dependency parser model can be developed using two methods, the transition-based and graph-based methods (Das and Sarkar, 2020). The transition-based method works by processing the word order one by one in a given sentence (Martin., 2020). Meanwhile, the graph-based method gives a score to each edge of the word relation (Martin., 2020), then looks for the best tree formed from the edges with the best scores.

Apart from these two methods, there is an approach in which the parser is built using an encoder-decoder architecture. It was first developed using a BiLSTM encoder and a deep biaffine decoder (Dozat and Manning, 2017). Encoder variations began to develop using Transformers or self-attention encoders (Vaswani et al., 2017), then subsequent studies modified it using relative positional embedding (Shaw et al., 2018). The first Javanese dependency parser (Alfina et al., 2023) uses UDPipe (Straka, 2018), which also utilizes the biaffine attention mentioned before.

In the context of transfer learning, it was found that the best combination is a self-attention encoder and a graph-based decoder (Ahmad et al., 2019), which will be used in this research. This combination has been better than other encoder-decoder combinations in cross-lingual transfer learning.

### 2.2 Transfer Learning

Transfer learning involves leveraging a pre-trained model's knowledge to enhance the performance of other models (Sarkar and Bali, 2022), addressing resource limitations in low-resource domains. Besides that, hierarchical transfer learning offers a transfer learning method in which a new layer is added before the model is transferred to the low-resource language (Luo et al., 2019). Recent work has shown that transferring multiple times could minimize the dissimilarity between the high-resource and the low-resource domain languages (Azizah et al., 2020).

Transfer learning strategy offers direct capability, which means a model is trained on a source task and then applied without any labeled data from the target task. Specifically on the parsing task, previous research already done by Kurniawan et al. (2021) and Ahmad et al. (2019) for developing an unsupervised parsing model in several languages using only English as its source language. That approach can be improved by adding fine-tuning with the available small dataset from low-resource language. Recent work (Maulana et al., 2022) shows the fine-tuning approach is better than the zero-shot one for building a parsing model in another low-resource language, Indonesian.

## 3 Method

This section concerns the model's architecture with the addition of the transfer learning method, the dataset and word embedding used to train the model, and the evaluation method of how the model is evaluated.

### 3.1 Model Architecture

This work uses an encoder-decoder architecture of Ahmad et al. (2019). No parameter modifications were made to maintain the success of the previous work. Because training and fine-tuning the model involves resources from several different languages, only language-independent labels are used where the subtype of the label is not involved.

#### 3.1.1 Encoder

We convert the words and POS tags from the sentence into their embedding form. The self-attention encoder (Vaswani et al., 2017) in this study received an embedding matrix, which concatenates the word and POS embedding matrices. The encoder produces two matrices, $M$ and $N$. $M$ matrix represents the probability of a word in column $j$ having the head of a word in row $i$. In comparison, the $N$ matrix represents the probability of a word in column $j$ having a label in row $i$.

### 3.1.2 Decoder

The decoder receives the two matrices and processes them in two following processes. First, $M$ is processed with the maximum spanning tree algorithm in the following way:

Let $G = (V, E)$ be a graph constructed using directed weighted graph $M$. In this case, a vertex is a word representation, and an edge represents the dependency score of the two words. Let $w : E \to \mathbb{R}$ be a function that assigns a weight to each edge in $E$. Then, the maximum spanning tree problem seeks to find a spanning tree $T = (V, E_T)$ of $G$ such that:

$$T = \arg\max_{T'} \sum_{e \in E_{T'}} w(e) \quad (1)$$

subject to the constraint that $T$ is a tree. Then, a list of head $H$ is generated from all the destination nodes in $E_T$. It can be denoted as:

$$H = \{d_i \mid \exists (s_i, d_i) \in E_T\}, \;\; i = 1, \ldots, n \quad (2)$$

Meanwhile, $N$ is processed to generate $L$, containing the list of labels with the highest probability for each word. Finally, the $H$ and $L$ arrays are used to build the final resulting tree from this model.

### 3.1.3 Word Embedding

This research used two types of word embedding approaches: the static type in the form of fastText and the contextual type in the form of BERT. The two types were selected to compare which type was most suitable for the Javanese parser model.

We chose fastText because of the similarity with that used in the previous research (Maulana et al., 2022). We also used BERT with two scenarios: using a different word embedding for each language (BERT and RoBERTa) and only one word embedding for all languages (multilingual BERT). The BERT and RoBERTa scenario uses all the languages involved except Croatian due to the unavailable resources.

## 3.2 Training Method

We perform two training methods: transfer learning and hierarchical transfer learning. Each method generates several models based on the number of source languages used. All models are fine-tuned with the Javanese treebank.

Standard transfer learning only uses one transfer stage from high-resource to low-resource language,

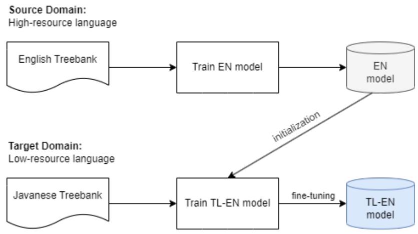


Figure 1: Illustration of standard transfer learning method

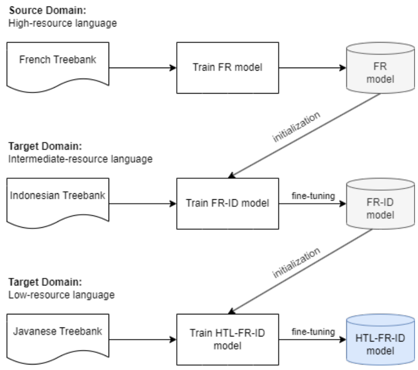


Figure 2: Illustration of hierarchical transfer learning method

as shown in Figure 1. Meanwhile, Figure 2 illustrates a hierarchical transfer learning scenario, where transferring stages are performed twice in hierarchical transfer learning. The first stage is done from a high-source language to an intermediate-resource language, and the second stage is done from an intermediate-resource language to a low-resource language.

## 3.3 Choosing Source Languages

Some languages are selected as source languages using the help of LangRank (Lin et al., 2020) and references from previous studies. This tool considers combining two main feature groups in each language pair: corpus statistics and typological information.

## 3.4 Dataset

### 3.4.1 The Javanese dataset

For the Javanese dataset, we use the only Javanese treebank available in the UD dataset v2.12, the UD_Javanese-CSUI (Alfina et al., 2023). Table 1 shows the statistics of this dataset. The set available for UD_Javanese-CSUI is only a test set because the data size is still relatively small. We do our split process by following the distribution rule of the data into train, dev, and test sets by 80%, 10%, and 10% percentages.

Table 1: The statistics of the Javanese treebank

| Description | Statistic |
|---|---|
| Sentence count | 1000 |
| Word count | 14344 |
| Unique word count | 3793 |
| Average sentence length (in words) | 14.32 |
| Universal Part-of-Speech (UPOS) tag count | 17 |
| Universal dependency relation count | 32 |
| Language-specific dependency relation count | 14 |
| Total dependency relation count | 46 |

Table 2: List of treebanks chosen for source languages, with their corresponding size in the number of sentences and words

| Treebank | Sentences | Words |
|---|---|---|
| UD_Croatian-SET (Agic and Ljubesic, 2015) | 9010 | 199409 |
| UD_English-GUM (Zeldes, 2017) | 9124 | 164396 |
| UD_French-GSD (McDonald et al., 2013) | 16341 | 400232 |
| UD_Indonesian-GSD (McDonald et al., 2013) | 5598 | 122021 |
| UD_Italian-ISDT (Bosco et al., 2022) | 14167 | 298343 |
| UD_Korean-GSD (Chun et al., 2019) | 6339 | 80322 |

#### 3.4.2 The source language dataset

Langrank recommends the top 3 languages in the following order: Indonesian, Croatian, and Korean. We also use English, one of the important languages in NLP research. These four languages are used in the standard transfer learning scenario.

For the hierarchical transfer learning scenario using Indonesian as the intermediary language, we choose English, French, and Italy as the source languages suggested by Maulana et al. (2022). In total, we use six languages as the source languages.

For each source language, we only use one treebank. If a language has more than one treebank in the UD dataset v2.12, we choose the treebank with the biggest size, as shown in Table 2.

### 3.5 Experiments Setting

#### 3.5.1 Scenarios

As explained in Section 3.2, we conducted three main scenarios:

1. Training from scratch (FS) or baseline scenario, in which the models are trained only using the target language, Javanese.

2. Standard transfer learning (TL). We construct four distinct models utilizing treebanks from each source language. Then, each model is fine-tuned using the Javanese treebank.

3. Hierarchical transfer learning (HTL). First, we train three different models using treebank from each source language. After that, the models were fine-tuned with the Indonesian treebank before being fine-tuned again with the Javanese treebank.

We also compared the performance of the four types of word embeddings for Javanese: fastText (Grave et al., 2019), Javanese BERT (Wongso et al., 2021), Javanese RoBERTa (Wongso et al., 2021), and multilingual BERT (Devlin et al., 2019).

#### 3.5.2 Environment

Implementation is done in Python environments. The training process is supported by the NVIDIA-DGX server with GPU NVIDIA A100 10GB, RAM of 64GB, and storage of 1 TB.

### 3.6 Evaluation

All models are evaluated using the unlabeled attachment score (UAS) and labeled attachment score (LAS) metrics, which are the most frequently used for evaluating the dependency parsing model (Nivre and Fang, 2017). The margin of error (MOE) with a 95% confidence level is also used to estimate the range of values within which the true population value is likely to fall.

## 4 Result and Analysis

The evaluation results for all scenarios are shown in Table 3. Scores in bold are marked as the best model in a particular word embedding type metric.

### 4.1 Models Comparison: From Scratch (FS) Model, Transfer Learning (TL) Model, and Hierarchical Transfer Learning (HTL) Model

Table 3 shows that the transfer learning model performs better than the baseline model in all word embeddings. The performance increase is quite significant, up to 13% on UAS and 14% on LAS. This verifies previous studies which explain the advantages of using transfer learning (Sarkar and Bali, 2022). The lack of resources in Javanese also indicates that transfer learning is suitable for use.

Figure 3 also shows that the hierarchical transfer learning method consistently outperforms the transfer learning method even though it is not too significant. Specifically, the comparison focused on the TL-ID and HTL models, as all models from the HTL scenario use the TL-ID model as its second base for the transferring method. The difference

Table 3: Evaluation results of all scenarios

| Word Embedding | Model | UAS | LAS |
|---|---|---|---|
| fastText | FS | 75.87 ± 2.21 | 68.97 ± 2.39 |
| | TL-ID | 84.80 ± 1.85 | 78.10 ± 2.14 |
| | TL-HR | 83.40 ± 1.92 | 76.57 ± 2.19 |
| | TL-KO | 80.68 ± 2.04 | 74.13 ± 2.26 |
| | TL-EN | 83.47 ± 1.92 | 77.27 ± 2.16 |
| | HTL-EN-ID | 84.94 ± 1.85 | **79.22 ± 2.10** |
| | HTL-FR-ID | 84.87 ± 1.85 | 77.55 ± 2.15 |
| | HTL-IT-ID | **85.84 ± 1.80** | 78.87 ± 2.11 |
| jv-BERT | FS | 74.69 ± 2.25 | 67.29 ± 2.42 |
| | TL-ID | 79.08 ± 2.10 | 72.32 ± 2.31 |
| | TL-HR | - | - |
| | TL-KO | 77.06 ± 2.17 | 70.29 ± 2.36 |
| | TL-EN | 81.73 ± 2.00 | 75.52 ± 2.22 |
| | HTL-EN-ID | **83.47 ± 1.92** | **76.64 ± 2.19** |
| | HTL-FR-ID | 81.80 ± 1.99 | 75.38 ± 2.22 |
| | HTL-IT-ID | 81.03 ± 2.02 | 73.99 ± 2.27 |
| jv-RoBERTa | FS | 69.80 ± 2.37 | 62.97 ± 2.49 |
| | TL-ID | 78.45 ± 2.12 | 72.11 ± 2.32 |
| | TL-HR | - | - |
| | TL-KO | 82.22 ± 1.97 | 76.22 ± 2.20 |
| | TL-EN | 77.13 ± 2.17 | 70.92 ± 2.35 |
| | HTL-EN-ID | 77.41 ± 2.16 | 70.85 ± 2.35 |
| | HTL-FR-ID | 83.05 ± 1.94 | 77.20 ± 2.17 |
| | HTL-IT-ID | **83.33 ± 1.92** | **77.20 ± 2.17** |
| multi-BERT | FS | 75.80 ± 2.21 | 69.04 ± 2.39 |
| | TL-ID | 82.01 ± 1.98 | 76.01 ± 2.21 |
| | TL-HR | 83.75 ± 1.90 | 77.68 ± 2.15 |
| | TL-KO | 79.78 ± 2.07 | 73.29 ± 2.28 |
| | TL-EN | 80.89 ± 2.03 | 74.13 ± 2.26 |
| | HTL-EN-ID | 82.98 ± 1.94 | 76.71 ± 2.18 |
| | HTL-FR-ID | 83.19 ± 1.93 | 77.75 ± 2.15 |
| | HTL-IT-ID | **84.45 ± 1.87** | **78.52 ± 2.12** |

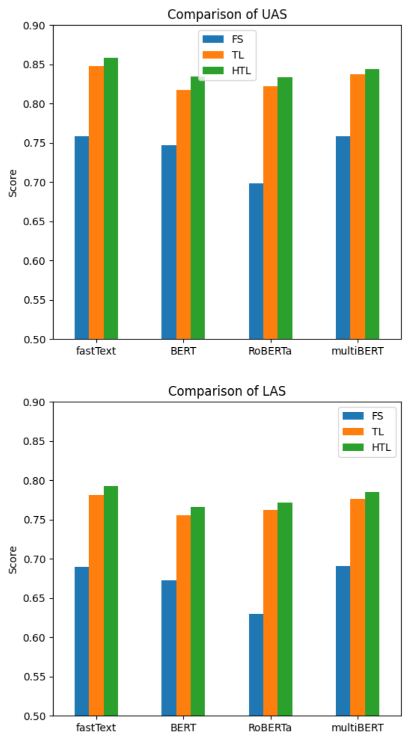


Figure 3: Comparison of the best model evaluation for each scenario

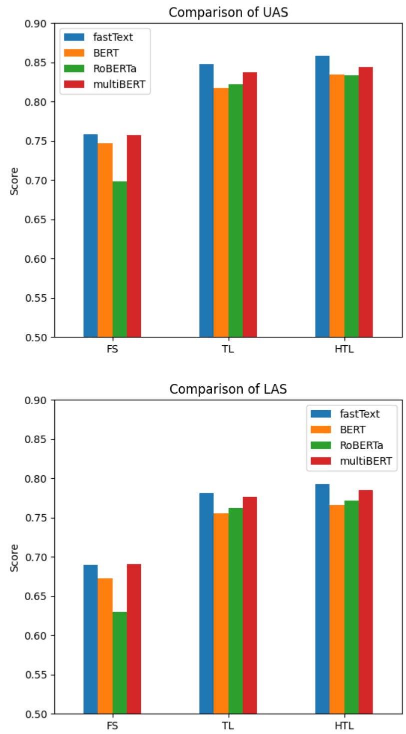


Figure 4: Comparison of the best model evaluation for each word embedding

between these two scenarios shows that adding suitable high-resource language for the initial source model can give a better performance.

### 4.2 Source Languages Comparison

Table 3 shows that two of the top three recommendations from LangRank have good results. The conclusion is that LangRank can help predict the source language in the Javanese dependency parser. However, it does not rule out the possibility that other languages also have good results. For TL, it cannot be concluded which source language achieves the best performance since different word embedding used by the model gives different results. For HTL using Indonesian as the intermediate language, Italy performs best, followed by English as the source language.

### 4.3 Word Embeddings Comparison

Figure 4 shows that the model with a higher UAS score was obtained from word embedding fastText, followed by multilingual BERT, Javanese BERT, and Javanese RoBERTa. For LAS evaluation, the sequence is fastText, multilingual BERT, Javanese RoBERTa, and Javanese BERT. Although fastText is slightly superior, the differences are insignificant when considering the models' margin of error.

Table 4: Top 10 errors of the from-scratch model and its comparison with the transfer-learning model

| Ground Truth | Prediction | FS | TL | HTL |
|---|---|---|---|---|
| obl | obj | 17 | 16 | 15 |
| obl | nsubj | 7 | 3 | 7 |
| obj | obl | 7 | 13 | 12 |
| advcl | xcomp | 5 | 5 | 6 |
| nmod | flat | 4 | 2 | 1 |
| xcomp | advcl | 4 | 5 | 5 |
| xcomp | obl | 3 | 3 | 2 |
| nmod | obl | 3 | 1 | 1 |
| nsubj | obj | 3 | 1 | 1 |
| obj | nsubj | 2 | 0 | 3 |

### 4.4 Error Analysis

Table 4 displays more detail about the performance difference. The ten labels taken are obtained from pairs with the highest errors in the from-scratch model. Some pairs significantly reduce error, but there are also pairs with no significant changes and even more errors in scenarios with transfer learning.

One noteworthy insight is the significantly increasing error of words with "obj" label that predicted with "obl". It seems contradictory that model accuracy is increasing simultaneously with the addition of transfer learning. It turns out that there are a few differences in the word labeling of both labels between the source and the target language, so the model could not predict the word label correctly.

## 5 Conclusions and Future Work

This section explains the conclusion and improvements that can be developed from this work.

### 5.1 Conclusions

This work investigates whether cross-lingual transfer learning works for dependency parsing tasks of a low-resource language, Javanese. The result shows that the cross-lingual transfer learning model is significantly better than the baseline model. Models with transfer learning can improve performance on UAS and LAS metrics by up to 10%.

The best model was obtained from the hierarchical transfer learning method using Italian and English as the source and Indonesian as the intermediary languages. Meanwhile, the standard transfer learning method achieved the best accuracy using Indonesian as the source language. However, the differences between standard transfer learning and hierarchical learning are insignificant, considering the margin of error from each scenario.

### 5.2 Future Work

We focused more on the model's learning scheme than the model's development with the highest score. We use architecture from Dozat and Manning (2017) rather than the one built by Mrini et al. (2020), the state-of-the-art dependency parsing task. So, better architecture can be used to produce a model with a higher evaluation score in the future.

Our future works also include further error analysis, especially related to the languages involved that LangRank chose. It could investigate languages with different demography and characteristics (Croatian and Korean) compared to Javanese.

## Limitations

The following are the limitations of this research:

1. There is no hyper-parameter tuning treatment in the model creation process.

2. Cross-validation is not performed in the data distribution process.

3. Only one language is used as an intermediary language in hierarchical transfer learning.


## Acknowledgements

We thank the Directorate of Research and Development, Universitas Indonesia, under Hibah PUTI 2022 (Grant No. NKB-1384/UN2.RST/HKP.05.00/2022) for funding this research. We also thank Tokopedia-UI AI Center, Faculty of Computer Science Universitas Indonesia, and Program Kompetisi Kampus Merdeka from the Ministry of Education, Culture, Research, and Technology Republic of Indonesia for the NVIDIA-DGX server we used for experiments. We also gratefully acknowledge Dr. Fajri Koto for the insightful feedback during the pre-submission mentorship.


## References


Zeljko Agic and Nikola Ljubesic. 2015. Universal Dependencies for Croatian (that work for Serbian, too). In *Proceedings of the Fifth Workshop on Balto-Slavic Natural Language Processing (BSNLP 2015)*.

Wasi Uddin Ahmad, Zhisong Zhang, Xuezhe Ma, Eduard Hovy, Kai Wei Chang, and Nanyun Peng. 2019. On difficulties of cross-lingual transfer with order differences: A case study on dependency parsing. In *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, volume 1.

Ika Alfina, Arlisa Yuliawati, Dipta Tanaya, Arawinda Dinakaramani, and Daniel Zeman. 2023. A gold standard dataset for Javanese tokenization, POS tagging, morphological features analysis, and dependency parsing.

Fa'iq Askhabi, Arie Ardiyanti Suryani, and Moch. Arif Bijaksana. 2020. Part of speech tagging in Javanese using support vector machine method. *e-Proceeding of Engineering*, 7.

Kurniawati Azizah, Mirna Adriani, and Wisnu Jatmiko. 2020. Hierarchical transfer learning for multilingual, multi-speaker, and style transfer DNN-based TTS on low-resource languages. *IEEE Access*, 8.

Cristina Bosco, Felice Dell'Orletta, and Simonetta Montemagni. 2022. *The Evalita 2014 Dependency Parsing Task*. Proceedings of the First Italian Conference on Computational Linguistics CLiC-it 2014 and of the Fourth International Workshop EVALITA 2014 9-11 December 2014, Pisa.

Zhaoying Chai, Han Jin, Shenghui Shi, Siyan Zhan, Lin Zhuo, and Yu Yang. 2022. Hierarchical shared transfer learning for biomedical named entity recognition. *BMC Bioinformatics*, 23.

Jayeol Chun, Na Rae Han, Jena D. Hwang, and Jinho D. Choi. 2019. Building universal dependency treebanks in Korean. In *LREC 2018 - 11th International Conference on Language Resources and Evaluation*.

Ayan Das and Sudeshna Sarkar. 2020. A survey of the model transfer approaches to cross-lingual dependency parsing. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 19.

Jacob Devlin, Ming Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, volume 1.

Timothy Dozat and Christopher D. Manning. 2017. Deep biaffine attention for neural dependency parsing. In *5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings*.

Edouard Grave, Piotr Bojanowski, Prakhar Gupta, Armand Joulin, and Tomas Mikolov. 2019. Learning word vectors for 157 languages. In *LREC 2018 - 11th International Conference on Language Resources and Evaluation*.

Kemal Kurniawan, Lea Frermann, Philip Schulz, and Trevor Cohn. 2021. PPT: Parsimonious parser transfer for unsupervised cross-lingual adaptation. In *EACL 2021 - 16th Conference of the European Chapter of the Association for Computational Linguistics, Proceedings of the Conference*.

Sandra Kübler, Ryan McDonald, and Joakim Nivre. 2009. *Dependency Parsing*, volume 2. Synthesis Lectures on Human Language Technologies.

Aufa Eka Putri Lesatari, Arie Ardiyanti, Arie Ardiyanti, Ibnu Asror, and Ibnu Asror. 2021. Phrase-based statistical machine translation Javanese-Indonesian. *Jurnal Media Informatika Budidarma*, 5.

Yu Hsiang Lin, Chian Yu Chen, Jean Lee, Zirui Li, Yuyan Zhang, Mengzhou Xia, Shruti Rijhwani, Junxian He, Zhisong Zhang, Xuezhe Ma, Antonios Anastasopoulos, Patrick Littell, and Graham Neubig. 2020. Choosing transfer languages for cross-lingual learning. In *ACL 2019 - 57th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*.

Gongxu Luo, Yating Yang, Yang Yuan, Zhanheng Chen, and Aizimaiti Ainiwaer. 2019. Hierarchical transfer learning architecture for low-resource neural machine translation. *IEEE Access*, 7.

Daniel Jurafsky & James H. Martin. 2020. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Prentice Hall.

Andhika Yusup Maulana, Ika Alfina, and Kurniawati Azizah. 2022. Building Indonesian dependency parser using cross-lingual transfer learning. In *2022 International Conference on Asian Language Processing (IALP)*, pages 488–493.

Ryan McDonald, Joakim Nivre, Yvonne Quirmbach-Brundage, Yoav Goldberg, Dipanjan Das, Kuzman Ganchev, Keith Hall, Slav Petrov, Hao Zhang, Oscar Täckström, Claudia Bedini, Núria Bertomeu Castelló, and Jungmee Lee. 2013. Universal dependency annotation for multilingual parsing. In *ACL 2013 - 51st Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*, volume 2.

Khalil Mrini, Franck Dernoncourt, Quan Tran, Trung Bui, Walter Chang, and Ndapa Nakashole. 2020. Rethinking self-attention: Towards interpretability in neural parsing. In *Findings of the Association for Computational Linguistics Findings of ACL: EMNLP 2020*.

Joakim Nivre and Chiao Ting Fang. 2017. Universal Dependency evaluation. In *Proceedings of the NoDaLiDa 2017 Workshop on Universal Dependencies, UDW 2017*.

Sebastian Ruder. 2023. The 4 biggest open problems in NLP.

Dipanjan Sarkar and Raghav Bali. 2022. *Transfer Learning in Action*, 1 edition. Manning Early Access Program.

Peter Shaw, Jakob Uszkoreit, and Ashish Vaswani. 2018. Self-attention with relative position representations. In *NAACL HLT 2018 - 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, volume 2.

Gary Simons, David Eberhard, and Charles Fennig. 2023. *Ethnologue: Languages of the World, 26nd Edition*. SIL International.

Dewi Soyusiawaty, Anna Hendri Soleliza Jones, and Nora Lestari Lestariw. 2020. The stemming application on affixed Javanese words by using Nazief and Adriani algorithm. In *IOP Conference Series: Materials Science and Engineering*, volume 771.

Milan Straka. 2018. UDPIPE 2.0 prototype at Conll 2018 UD Shared Task. In *CoNLL 2018 - SIGNLL Conference on Computational Natural Language Learning, Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pages 197–207.

C. Tho, Y. Heryadi, L. Lukas, and A. Wibowo. 2021. Code-mixed sentiment analysis of Indonesian language and Javanese language using lexicon-based approach. In *Journal of Physics: Conference Series*, volume 1869.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 2017-December.

Karl Weiss, Taghi M. Khoshgoftaar, and Ding Ding Wang. 2016. A survey of transfer learning. *Journal of Big Data*, 3.

Wilson Wongso, David Samuel Setiawan, and Derwin Suhartono. 2021. Causal and masked language modeling of Javanese language using transformer-based architectures. In *2021 International Conference on Advanced Computer Science and Information Systems, ICACSIS 2021*.

Amir Zeldes. 2017. The GUM corpus: creating multilayer resources in the classroom. *Language Resources and Evaluation*, 51.

Daniel Zeman, Joakim Nivre, Mitchell Abrams, Elia Ackermann, Noëmi Aepli, Hamid Aghaei, Željko Agić, Amir Ahmadi, Lars Ahrenberg, Chika Kennedy Ajede, Salih Furkan Akkurt, Gabrielė Aleksandravičiūtė, Ika Alfina, Avner Algom, Khalid Alnajjar, Chiara Alzetta, Erik Andersen, Lene Antonsen, Tatsuya Aoyama, Katya Aplonova, Angelina Aquino, Carolina Aragon, Glyd Aranes, Maria Jesus Aranzabe, Bilge Nas Arıcan, Hórunn Arnardóttir, Gashaw Arutie, Jessica Naraiswari Arwidarasti, Masayuki Asahara, Katla Ásgeirsdóttir, Deniz Baran Aslan, Cengiz Asmazoğlu, Luma Ateyah, Furkan Atmaca, Mohammed Attia, Aitziber Atutxa, Liesbeth Augustinus, Mariana Avelãs, Elena Badmaeva, Keerthana Balasubramani, Miguel Ballesteros, Esha Banerjee, Sebastian Bank, Verginica Barbu Mititelu, Starkaður Barkarson, Rodolfo Basile, Victoria Basmov, Colin Batchelor, John Bauer, Seyyit Talha Bedir, Shabnam Behzad, Kepa Bengoetxea, Ibrahim Benli, Yifat Ben Moshe, Gözde Berk, Riyaz Ahmad Bhat, Erica Biagetti, Eckhard Bick, Agnė Bielinskienė, Kristín Bjarnadóttir, Rogier Blokland, Victoria Bobicev, Loïc Boizou, Emanuel Borges Völker, Carl Börstell, Cristina Bosco, Gosse Bouma, Sam Bowman, Adriane Boyd, Anouck Braggaar, António Branco, Kristina Brokaitė, Aljoscha Burchardt, Marisa Campos, Marie Candito, Bernard Caron, Gauthier Caron, Catarina Carvalheiro, Rita Carvalho, Lauren Cassidy, Maria Clara Castro, Sérgio Castro, Tatiana Cavalcanti, Gülşen Cebiroğlu Eryiğit, Flavio Massimiliano Cecchini, Giuseppe G. A. Celano, Slavomír Čéplö, Neslihan Cesur, Savas Cetin, Özlem Çetinoğlu, Fabricio Chalub, Liyanage Chamila, Shweta Chauhan, Ethan Chi, Taishi Chika, Yongseok Cho, Jinho Choi, Jayeol Chun, Juyeon Chung, Alessandra T. Cignarella, Silvie Cinková, Aurélie Collomb, Çağrı Çöltekin, Miriam Connor, Daniela Corbetta, Francisco Costa, Marine Courtin, Mihaela Cristescu, Ingerid Løyning Dale, Philemon Daniel, Elizabeth Davidson, Leonel Figueiredo de Alencar, Mathieu Dehouck, Martina de Laurentiis, Marie-Catherine de Marneffe, Valeria de Paiva, Mehmet Oguz Derin, Elvis de Souza, Arantza Diaz de Ilarraza, Carly Dickerson, Arawinda Dinakaramani, Elisa Di Nuovo, Bamba Dione, Peter Dirix, Kaja Dobrovoljc, Adrian Doyle, Timothy Dozat, Kira Droganova, Puneet Dwivedi, Christian Ebert, Hanne Eckhoff, Masaki Eguchi, Sandra Eiche, Marhaba Eli, Ali Elkahky, Binyam Ephrem, Olga Erina, Tomaž Erjavec, Farah Essaidi, Aline Etienne, Wograine Evelyn, Sidney Facundes, Richárd Farkas, Federica Favero, Jannatul Ferdaousi, Marília Fernanda, Hector Fernandez Alcalde, Amal Fethi, Jennifer Foster, Cláudia Freitas, Kazunori Fujita, Katarína Gajdošová, Daniel Galbraith, Federica Gamba, Marcos Garcia, Moa Gärdenfors, Fabrício Ferraz Gerardi, Kim Gerdes, Luke Gessler, Filip Ginter, Gustavo Godoy, Iakes Goenaga, Koldo Gojenola, Memduh Gökırmak, Yoav Goldberg, Xavier Gómez Guinovart, Berta González Saavedra, Bernadeta Griciūtė, Matias Grioni, Loïc Grobol, Normunds Grūzītis, Bruno Guillaume, Céline Guillot-Barbance, Tunga Güngör, Nizar Habash, Hinrik Hafsteinsson, Jan Hajič, Jan Hajič jr., Mika Hämäläinen, Linh Hà Mỹ, Na-Rae Han, Muhammad Yudistira Hanifmuti, Takahiro Harada, Sam Hardwick, Kim Harris, Dag Haug, Johannes Heinecke, Oliver Hellwig, Felix Hennig, Barbora Hladká, Jaroslava Hlaváčová, Florinel Hociung, Petter Hohle, Marivel Huerta Mendez, Jena Hwang, Takumi Ikeda, Anton Karl Ingason, Radu Ion, Elena Irimia, Ọlájídé Ishola, Artan Islamaj, Kaoru Ito, Siratun Jannat, Tomáš Jelínek, Apoorva Jha, Katharine Jiang, Anders Johannsen, Hildur Jónsdóttir, Fredrik Jørgensen, Markus Juutinen, Hüner Kaşıkara, Nadezhda Kabaeva, Sylvain Kahane, Hiroshi Kanayama, Jenna Kanerva, Neslihan Kara, Ritván Karahóǧa, Andre Kåsen, Tolga Kayadelen, Sarveswaran Kengatharaiyer, Václava Kettnerová, Jesse Kirchner, Elena Klementieva, Elena Klyachko, Arne Köhn, Abdullatif Köksal, Kamil Kopacewicz, Timo Korkiakangas, Mehmet Köse, Alexey Koshevoy, Natalia Kotsyba, Jolanta Kovalevskaitė, Simon Krek, Parameswari Krishnamurthy, Sandra Kübler, Adrian Kuqi, Oğuzhan Kuyrukçu, Aslı Kuzgun, Sookyoung Kwak, Kris Kyle, Veronika Laippala, Lorenzo Lambertino, Tatiana Lando, Septina Dian Larasati, Alexei Lavrentiev, John Lee, Phuong Le Hong, Alessandro Lenci, Saran Lertpradit, Herman Leung, Maria Levina, Lauren Levine, Cheuk Ying Li, Josie Li, Keying Li, Yixuan Li, Yuan Li, KyungTae Lim, Bruna Lima Padovani, Yi-Ju Jessica Lin, Krister Lindén, Yang Janet Liu, Nikola Ljubešić, Olga Loginova, Stefano Lusito, Andry Luthfi, Mikko Luukko, Olga Lyashevskaya, Teresa Lynn, Vivien Macketanz, Menel Mahamdi, Jean Maillard, Ilya Makarchuk, Aibek Makazhanov, Michael Mandl, Christopher Manning, Ruli Manurung, Büşra Marşan, Cătălina Mărănduc, David Mareček, Katrin Marheinecke, Stella Markantonatou, Héctor Martínez Alonso, Lorena Martín Rodríguez, André Martins, Cláudia Martins, Jan Mašek, Hiroshi Matsuda, Yuji Matsumoto, Alessandro Mazzei, Ryan McDonald, Sarah McGuinness, Gustavo Mendonça, Tatiana Merzhevich, Niko Miekka, Aaron Miller, Karina Mischenkova, Anna Missilä, Cătălin Mititelu, Maria Mitrofan, Yusuke Miyao, AmirHossein Mojiri Foroushani, Judit Molnár, Amirsaeid Moloodi, Simonetta Montemagni, Amir More, Laura Moreno Romero, Giovanni Moretti, Shinsuke Mori, Tomohiko Morioka, Shigeki Moro, Bjartur Mortensen, Bohdan Moskalevskyi, Kadri Muischnek, Robert Munro, Yugo Murawaki, Kaili Müürisep, Pinkey Nainwani, Mariam Nakhlé, Juan Ignacio Navarro Horñiacek, Anna Nedoluzhko, Gunta Nešpore-Bērzkalne, Manuela Nevaci, Luong Nguyễn Thị, Huyền Nguyễn Thị Minh, Yoshihiro Nikaido, Vitaly Nikolaev, Rattima Nitisaroj, Alireza Nourian, Hanna Nurmi, Stina Ojala, Atul Kr. Ojha, Hulda Óladóttir, Adédayọ Olúòkun, Mai Omura, Emeka Onwuegbuzia, Noam Ordan, Petya Osenova, Robert Östling, Lilja Øvrelid, Şaziye Betül Özateş, Merve Özçelik, Arzucan Özgür, Balkız Öztürk Başaran, Teresa Paccosi, Alessio Palmero Aprosio, Anastasia Panova, Hyunji Hayley Park, Niko Partanen, Elena Pascual, Marco Passarotti, Agnieszka Patejuk, Guilherme Paulino-Passos, Giulia Pedonese, Angelika Peljak-Łapińska, Siyao Peng, Siyao Logan Peng, Rita Pereira, Sílvia Pereira, Cenel-Augusto Perez, Natalia Perkova, Guy Perrier, Slav Petrov, Daria Petrova, Andrea Peverelli, Jason Phelan, Jussi Piitulainen, Yuval Pinter, Clara Pinto, Tommi A Pirinen, Emily Pitler, Magdalena Plamada, Barbara Plank, Thierry Poibeau, Larisa Ponomareva, Martin Popel, Lauma Pretkalniņa, Sophie Prévost, Prokopis Prokopidis, Adam Przepiórkowski, Robert Pugh, Tiina Puolakainen, Sampo Pyysalo, Peng Qi, Andreia Querido, Andriela Rääbis, Alexandre Rademaker, Mizanur Rahoman, Taraka Rama, Loganathan Ramasamy, Joana Ramos, Fam Rashel, Mohammad Sadegh Rasooli, Vinit Ravishankar, Livy Real, Petru Rebeja, Siva Reddy, Mathilde Regnault, Georg Rehm, Arij Riabi, Ivan Riabov, Michael Rießler, Erika Rimkutė, Larissa Rinaldi, Laura Rituma, Putri Rizqiyah, Luisa Rocha, Eiríkur Rögnvaldsson, Ivan Roksandic, Mykhailo Romanenko, Rudolf Rosa, Valentin Roșca, Davide Rovati, Ben Rozonoyer, Olga Rudina, Jack Rueter, Kristján Rúnarsson, Shoval Sadde, Pegah Safari, Aleksi Sahala, Shadi Saleh, Alessio Salomoni, Tanja Samardžić, Stephanie Samson, Manuela Sanguinetti, Ezgi Sanıyar, Dage Särg, Marta Sartor, Mitsuya Sasaki, Baiba Saulīte, Yanin Sawanakunanon, Shefali Saxena, Kevin Scannell, Salvatore Scarlata, Nathan Schneider, Sebastian Schuster, Lane Schwartz, Djamé Seddah, Wolfgang Seeker, Mojgan Seraji, Syeda Shahzadi, Mo Shen, Atsuko Shimada, Hiroyuki Shirasu, Yana Shishkina, Muh Shohibussirri, Maria Shvedova, Janine Siewert, Einar Freyr Sigurdsson, João Silva, Aline Silveira, Natalia Silveira, Sara Silveira, Maria Simi, Radu Simionescu, Katalin Simkó, Mária Šimková, Haukur Barri Símonarson, Kiril Simov, Dmitri Sitchinava, Ted Sither, Maria Skachedubova, Aaron Smith, Isabela Soares-Bastos, Per Erik Solberg, Barbara Sonnenhauser, Shafi Sourov, Rachele Sprugnoli, Vivian Stamou, Steinhór Steingrímsson, Antonio Stella, Abishek Stephen, Milan Straka, Emmett Strickland, Jana Strnadová, Alane Suhr, Yogi Lesmana Sulestio, Umut Sulubacak, Shingo Suzuki, Daniel Swanson, Zsolt Szántó, Chihiro Taguchi, Dima Taji, Fabio Tamburini, Mary Ann C. Tan, Takaaki Tanaka, Dipta Tanaya, Mirko Tavoni, Samson Tella, Isabelle Tellier, Marinella Testori, Guillaume Thomas, Sara Tonelli, Liisi Torga, Marsida Toska, Trond Trosterud, Anna Trukhina, Reut Tsarfaty, Utku Türk, Francis Tyers, Sveinbjörn Hórarson, Vilhjálmur Horsteinsson, Sumire Uematsu, Roman Untilov, Zdeňka Urešová, Larraitz Uria, Hans Uszkoreit, Andrius Utka, Elena Vagnoni, Sowmya Vajjala, Socrates Vak, Rob van der Goot, Martine Vanhove, Daniel van Niekerk, Gertjan van Noord, Viktor Varga, Uliana Vedenina, Giulia Venturi, Veronika Vincze, Natalia Vlasova, Aya Wakasa, Joel C. Wallenberg, Lars Wallin, Abigail Walsh, Jonathan North Washington, Maximilan Wendt, Paul Widmer, Shira Wigderson, Sri Hartati Wijono, Seyi Williams, Mats Wirén, Christian Wittern, Tsegay Woldemariam, Tak-sum Wong, Alina Wróblewska, Mary Yako, Kayo Yamashita, Naoki Yamazaki, Chunxiao Yan, Koichi Yasuoka, Marat M. Yavrumyan, Arife Betül Yenice, Olcay Taner Yıldız, Zhuoran Yu, Arlisa Yuliawati, Zdeněk Žabokrtský, Shorouq Zahra, Amir Zeldes, He Zhou, Hanzhi Zhu, Yilun Zhu, Anna Zhuravleva, and Rayan Ziane. 2023. Universal Dependencies 2.12. LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University.