Effects of Bilingualism on Novel Word Processing and Implications for Computational Sentence Processing

Calvin Li
Johns Hopkins University
cli78@jhu.edu

Simon Zeng
Johns Hopkins University
szeng7@jhu.edu

Meg Obata
Johns Hopkins University
megobata@jhu.edu

Abstract

Bilingual speakers are known to perform better in learning novel words than monolingual speakers (Kaushanskaya and Marian, 2009). This study looks to not only verify that through looking at novel word identification times for monolingual and bilingual speakers, but to also analyze the effect that the morphological richness of a language has on the syntax and morphology dependency of novel words. If the results show that bilingual speakers have a faster novel word identification time, then it suggests that a computational sentence parsing model that utilizes two languages in its training process may achieve better results. Additionally, English is known to be a morphologically poor language that only utilizes syntax in novel word processing (Kharkwal, 2014). This means that if the morphological richness of a language does have an effect on novel word dependencies, then training a computational sentence parsing model in another language other than English or other morphologically poor language may improve its accuracy. This cross language training process will be analyzed through transfer learning, where the results of a model developed for one particular task is used as the starting point for the model on a second task. In this case, a computational sentence parsing model will be first trained on a morphologically rich language. Those trained weights will then be used as initialization weights for sentence parsing on a less morphologically rich language such as English. The results of such will be compared to that of a model that does not utilize this transfer learning from a different language task to determine its effect.

1 Introduction/ Background Information

It is often an assumption that an L1 (native language) may interfere with the acquisition of novel words in L2 (non-native language) in language learning. However, researchers have found that bilingual adult speakers are better at learning novel words. Although the exact reason is unknown and there are many confounds that can affect acquisition, some general theories fall under the fact that bilinguals may have better cognitive control. This allows for things like increased memory-storage capacity or greater control of attention, which may stem from having to suppress one language depending on the context (Kaushanskaya and Marian, 2009).

However, this study does not explore the specific differences between bilingual and monolingual word processing that leads to better bilingual performance. For example, syntactic context (Fisher, Hall, Rakowitz, Gleitman, 1994) and lexical knowledge (Dahan and Brent, 1999) is known to facilitate novel word learning. If this is true, bilingual knowledge of syntax and lexicons may interfere in novel word processing.

In order to evaluate novel word processing, jabberwocky sentences are often used. Despite the fact that they have no semantic content, humans are able to parse them like regular sentences and infer meaning of individual words as well. For example, we are able to infer some meanings of pseudowords from Lewis Carroll's poem, Jabberwocky:

'Twas brillig, and the slithy toves
Did gyre and gimble in the wabe;
All mimsy were the borogoves,
And the mome raths outgrabe.

Brillig may refers to a time or day and gyre and gimble may refer to the movement of slithy toves

(Kharkwal, 2014). Kharkwal conducted 6 experiments investigating the underlying processes that derive meaning from these sentences–particularly, he investigated whether it is morphology or syntactic cues that allow us to understand the sentences. He found that in a self-paced reading test with ambiguous, unreliable syntactic information, morphological information was not used to guide sentence processing–in fact, the sentence processor waits until more robust syntactic information is available. Kharkwal suggests that this may be due to the lack of rich morphology in English.

He further explains that English's syntactic rigidity and tendency to have a fixed word order may cause English morphology to be an unreliable indicator of a word's meaning. This suggests that morphologically rich languages like Turkish, Finnish, or Romance languages may utilize morphology over syntax to guide sentence processing. This brings up our main question: does a second language affect the way in which novel words are processed? And in what way does it influence how we understand novel words?

If there are specific features within two different languages that allow a person to better acquire novel words, this can have implications for language models which are almost always monolingual. If certain features from possibly a certain combination of languages allows humans to perform better in novel word processing, this can be generalized to training computational language models for sentence processing bilingually to maximize the performance of the model.

Specifically, this can be applied to a graph-based parser with deep biaffine attention created by Dozat and Manning. It is similar to a transition-based parser, which parses sentence from left to right and keeps a "buffer", words that haven't been parsed yet, and a "stack", words whose head has not been seen or whose dependents have not been fully parsed. The parser accesses these buffers and stacks and assigns arcs to the words of a sentence. A classifier can then be trained to extract features from the stack, buffer, or previous arcs to predict the next action.

The model is a feedforward network that assigns a probability for each action according to word, tag, and label embeddings. It also assigns a weight to each possible edge and constructs a maximum spanning tree from the weighted edges. Dozat and Manning also use biaffine attention instead of bilinear or MLP mechanisms for classifiers. This is similar to a traditional affine classifier except that instead of a weight matrix, W, they use a (d x d) linear transformation of stacked LSTM output, RU, and a (d x 1) linear transformation, Ru, for the bias term, b. It is also simpler than a MLP-based approach because it only uses one bilinear layer rather than two nonlinear layers (along with nonlinearity). This allows for modeling of a prior probability of word j receiving any dependents and also the likelihood of word j receiving a specific dependent.

The parser directly models each prior probability of each class, likelihood of class (or label) given a word, the likelihood of a class given the head word, and likelihood of a class given both the word and its head (Dozat and Manning, 2017).

This model was used by Kasai and Frank in order to investigate the extent of the contribution of lexical information. Current models are trained on similar corpuses like Brown, WSJ, or Penn Treebank, which do not allow us to see how lexically dependent a model is and whether it can generalize across domains (Gildea, 2001). In order to test this, Kasai and Frank simulated jabberwocky words by setting input word vectors to 0, while retaining the POS embedding. The experiments revealed that parsers did not perform well with this type of lexical noise, indicating a strong dependency on the lexicon. They were able to improve performance by using word dropout, decreasing the reliability of lexicon in contributing to the final POS tag. However, they found that when faced with too aggressive of a word dropout rate, the parser's performance suffered, suggesting that another approach would be best suited to solve this issue, which will be explored in this study.

2 Novel Word Processing in Bilinguals

2.1 Overview

For this study, we first aim to use human data to discover whether the meaning of novel words (pseudowords) are derived from syntactic knowledge or morphological features of the word. We use English and Finnish because Kharkwal found that English speakers rely heavily on syntactic knowledge to derive meaning of pseudowords and suggested that this is because English lacks a rich morphology. Because Finnish is a morphologically rich language, comparing performance on the two languages may reveal differences in how

English speaking monolinguals or Finnish speaking monolinguals process sentences. Furthermore, we would like to see whether bilingual speakers of English and Finnish are influenced by the interaction of the two languages. The second part of this study aims to utilize the findings from human experiments and model the possible bilingual interaction to see how a language model performs with "bilingual" training.

2.2 Experiment 1: Novel Word Detection/ Processing in English

2.2.1 Overview

Given that comprehension as a whole is relatively difficult to judge, especially with respect to syntax and morphology, a better starting point is novel word detection. In order to recognize if a word is novel or not, a subset of processing must be done to determine if any meaning is associated towards it. With this novel word detection, identification time proves to be an easily accessible measure.

Previous work by Stromswold et al. (1996) and Kharkwal (2014) found that participants' performance in such tasks is influenced by sentences' syntactic structures. We first try to replicate their findings through an experiment that tests novel word detection through this measure of identification time.

2.2.2 Methods

Participants

As many monolingual English Speaking (college students) with normal vision and no language or learning disorder.

Stimuli

In this experiment, we first have two types of relative-clause sentences as our target. Namely, we have one type that is center-embedded and the other that is right-branching. Each of these sentence types are then further split into cases where a relative clause extracts the object or the subject, yielding a total of four sentence types. Sample sentences of the four different types of sentences and images of their parse trees can be seen in the Appendix. Similar to Kharkwal (2014), sentences were designed such that the embedded novel word/psuedoword is orthographically and phonologically plausible pseudowords. What that means is that each "novel" replacement word is morphologically consistent with the original lexical word in the same position of that sentence. For each sentence, we will vary the position in which the novel pseudoword is placed as seen in the dataset below that each original sentence will have 5 variants in which a noun or a verb is replaced by a pseudonoun or a pseudoverb.

For our English experiment, sentences were designed as such: for each original sentence, we swap one noun or verb with a orthographically and phonologically plausible pseudoword (denoted with the number of variants). The way we ensured this was to have the last two letters or the last syllable match with the original word.

Sample Target Sentences

Refer to Appendix A.

2.2.3 Procedure

During the experiment, we will show each participant 200 sentences similar to those shown in the sample target sentence set (shown in Appendix A). For a well-designed stimuli, we would wish the data to be perfectly balanced in metrics such as mean length, types (i.e. each type SS, SO, OS, OO will consist of 25 percent of the total sentences respectively), and frequency (each type shows up the same number of times for each participant in random order).

During actual testing, user is given a sentence from the set given (a list of example sentences). When the participant is presented each sentence, one at a time, a timer starts. The participant must read through the sentence and identify if there is a novel word. At that point, the participant has to identify the novel word (by clicking the appropriate word). The timer stops when the user has made a selection. They will be instructed to give an answer as fast as possible without sacrificing accuracy. Both the accuracy and the identification time (response time) will be recorded for further analysis.

2.3 Experiment 2: Novel Word Detection/Processing in Finnish

2.3.1 Overview

Given that the morphological rigidity of English is a defining characteristic of it and the reason for some ambiguity that exists within English, looking towards another language that is morphologically richer and seeing how it affects the morphology and syntax dependency of novel word processing and how specific languages influence how humans interpret novel words. For the purposes of this experiment, Finnish was chosen to be the morphologically rich language due to it being an agglutinative language, meaning that it has elements such as verb conjugation and declension to it, where words are spelled differently given their syntactial role within a sentence (Gerz et al, 2017). For example, one of the fifteen cases for nouns in Finnish is accusative, meaning that if a noun acts as the direct object of a transitive verb, an "n" is added on to the end of it. This means that words like "singer" in Finnish (laulaja) get altered when in sentences like "the director loved the singer" (Johtaja rakasti laulajan). In addition, due to Finnish having a similar SVO word order like English, the potential differences in results can be more directly attributed to the morphological richness versus a confounding variable of word order.

With declension for Finnish nouns and conjugations for Finnish verbs, creating psuedowords requires more structure to them, meaning that they must have particular endings in order to be orthographically similar.

2.3.2 Methods

Participants
As many monolingual Finnish Speaking (college students) with normal vision and no language or learning disorder.

Stimuli
The stimuli in this experiment are similar to that in Experiment 1, namely that there are two types of relative-clause sentences as the target. This includes center-embedded and right-branching sentences, both of which include relative clauses that are subject-extracted or object-extracted. This yields four different sentences types from which the nouns and verbs can be replaced by orthographically and phonologically plausible psuedowords.

For each sentence, the position in which the novel psuedoword will be varied such that the original sentence will have 5 variants, in which a noun or verb is replaced by a psuedonoun or a psuedoverb.

In order to ensure orthographically plausibility, these Finnish psuedowords follow the rules of declension and conjugation with respect to the rest of the sentence (except for cases in which the psuedoword was intentionally a psuedonoun instead of a psuedoverb or vice versa).

Sample Target Sentences
Refer to Appendix B.

2.3.3 Procedure

Training data containing sets of sentences such as those shown above will be presented to the monolingual Finnish speakers. When the speaker is presented each sentence, one at a time, a timer starts. The participant must read through the sentence and identify if there is a novel word and then press a button to stop the timer. At that point the participant has to identify the novel word if he/she indicated that one was present. Both the accuracy and the identification time will be recorded.

2.4 Experiment 3: Novel Word Detection/ Processing in Bilingual Speakers

2.4.1 Overview

Here we replicate Experiments 1 and 2 again but with bilingual speakers.

2.4.2 Methods

Participants
As many bilingual Finnish-English Speaking (college students) with normal vision and no language or learning disorder.

Stimuli
The stimuli is the same as presented in Experiment 1 and Experiment 2, except that these bilingual speakers will be presented both (separately, not mixed).

Sample Target Sentences
Refer to Appendix A and B.

2.4.3 Procedure

Training data containing sets of English sentences such as those shown above will be presented to the bilingual speakers. When the speaker is presented each sentence, one at a time, a timer starts. The participant must read through the sentence and identify if there is a novel word and then press a button to stop the timer. At that point the participant has to identify the novel word if he/she indicated that one was present. Both the accuracy and the identification time will be recorded.

Afterwards, training data containing sets of Finnish sentences such as those shown above will be presented to the bilingual speakers. The same procedure as before will follow and the same values will be recorded.

2.5 Discussion of Potential Results

2.5.1 Interpretation of Experiment 1 Results

The interpretation of the results of Experiment 1 is specifically related to the effect of morphological richness of a language on the morphological and syntax dependency of novel word processing. For this experiment, the language in question is English. Although Kharkwal mentioned in his study that English novel word processing relies solely on syntax, this experiment will verify that.

The collected data, in the form of accuracies and identification times, for every sentence type and psuedoword position (both of which can be captured through the unique ID provided) will first be cleaned. This means that results for which the participant guessed the wrong novel word will not be used because their corresponding identification times are no longer accurate representations of correct dependencies on morphology and syntax. After doing so, the identification time of each sentence type and psuedoword position will have the identification time of the corresponding sentence type that does not contain the psuedoword subtracted from it. For example, this would mean that the identification time of OS-1 would have the identification time of OS-orig subtracted from it.

After cleaning the data, by making key comparisons, certain conclusions can be made regarding the syntactical and morphological dependency, which are summed up in the paragraphs below.

No Syntax Dependency
In this case, our expected result would be depend on the pseudoword position. When comparing sentences of the same psuedoword position but different sentence types, identification times are similar. (i.e. OS-1 vs OO-1)

In addition, another test would be to check the identification times of sentences of the same sentence type but different psuedoword positions. If there is no syntax dependency, then the identification time should increase as the psuedoword's position is closer to the end of the sentence. (i.e. SS-1 vs SS-2 vs SS-3 vs SS-4 vs SS-5)

The reasoning is that if syntax plays a significant role in novel word processing, then identification times should differ based on the syntactical structure of the sentence (if it's right-branching or center-embedded, object-extracted or subject-extracted, etc). However, if the identification times are similar, then it suggests the opposite, that syntax plays no significant role in novel word identification and processing. The second test makes sense given that if there is no syntax dependency, then the user would be able to immediately see a psuedoword at the beginning of the sentence and recognize that it is a novel word without needing to read the rest of the sentence. Meaning that the more in the front of the sentence a pseudoword occurs, the faster the identification time as it is a linear scan.

Syntax Dependency
The expected result when comparing sentences of the same psuedoword position but different sentence types, identification times significantly differ.(i.e. OS-1 vs OO-1) This strongly indicates that syntax plays a significant role in novel word processing, given that the only variable is the sentence type (and syntax) of the sentence.

No morphological dependency
The expected result when comparing sentences of the same pseudoword position and sentence type but different psuedoword endings (noun vs verb endings), identification times are similar. (i.e. OS-1* vs OS-1) If morphology plays a significant role in novel word processing, then the identification times should differ based on the morphology of the novel word in question. If the novel word has an inappropriate ending (hence, an inappropriate morphology) and the identification time does not differ, then it indicates that the only variable (morphology) causes for no difference in identification time.

Morphological dependency
The expected result when comparing sentences of the same pseudoword position and sentence type but different psuedoword endings (noun vs verb endings), identification times significantly differ.(i.e. OS-1* vs OS-1) This strongly indicates that morphology plays a significant role in novel word processing, given that the only variable is the morphology of the psuedoword, where a psuedonoun is giving a pseudoverb ending or vice versa.

It is important to note that these conclusions are not mutually exclusive: English can be both dependent on morphology and syntax. Based on the conclusion obtained from the results, the interpretation of this experiment as well as that of Experiment 2 will yield insights into extending a bilingual approach towards computational sentence parsing, which is explained below.

2.5.2 Interpretation of Experiment 2 Results

The interpretation of the results of Experiment 2 is specifically related to the effect of morphological richness of a language on the morphological and syntax dependency of novel word processing. For this experiment, the language in question is Finnish.

The conclusions that can be made about the results from Experiment 2 can be concluded from the same paragraphs and reasonings as for Experiment 1. It is important to note again that the conclusions are not mutually exclusive, meaning that Finnish can be interpreted to be both dependent on morphology and syntax.

In fact, given that the morphological richness of Finnish causes for the morphology of the word to depend on the syntactical function that it serves within a sentence, the results are expected to show a dependency on both. In this case of this, then it strongly suggests that morphologically rich languages like Finnish have more features from which the part of speech of a word can be determined from (both morphological elements as well as syntactical elements).

If English proves to be solely dependent on only morphology or syntax whereas Finnish proves to be dependent on both, then our intention of using a morphologically rich language like Finnish to serve as a way through which pre-trained weights can be determined for a computational parsing model in English, essentially a form of transfer learning, would most likely yield positive changes in results, which will be further explored below in Section 3. In the case that a different permutation of results is obtained (both languages depend on both or Finnish depends on one of the two), then the human data does not strongly support the hypothesis that bilingualism through transfer learning will yield positive results in computational sentence parsing; however, it may still be worthwhile to attempt such a task.

2.5.3 Interpretation of Experiment 3 Results

The interpretation of the results of Experiment 3 comes from specifically comparing the identification times on the tasks between the monolingual speakers and bilingual speakers. This means that after the same type of data cleaning mentioned in the earlier part of this section (eliminating results where the participant guessed the wrong novel word and subtracting the identification time of the non-psuedoword sentences from that of the psuedoword sentences of the same sentence types), the identification times of sentences with the same psuedoword position and sentence type can be compared. If the identification times is significantly reduced for bilingual speakers in both language comparisons, then it strongly supports the idea that using parsing results from another language as pre-trained weights for a given language sentence parsing model through transfer learning would yield more accurate results, which will be further explored below in Section 3. If the identification times for bilingual speakers is higher, then the human data support for extending transfer learning of a different language to computational sentence parsing is greatly diminished, but the experiment is still worthwhile to pursue. If the identification times is mixed (higher for one language and lower for the other), then it would be worthwhile to reevaluate the experiment, looking more closely at the language chosen and the experimental methods. A clear conclusion would not be evident, but the extension of transfer learning of a different language to computational sentence parsing would yet again still potentially be worthwhile.

3 Implications for Computational Models of Sentence Parsing

A key area of study in current sentence parsers is their ability to interpret novel or OOV words. Given that humans are able to derive semantic meaning for novel words without having seen them before, an advancement in a computational model's ability to interpret novel words would bridge the gap between it and cognitive models while also allowing for the parser to be more easily generalizable across different domains.

Kasai and Frank found that current state of the art sentence parsers overly relies on lexical information and does not perform ideally when tasked with jabberwocky/pseudoword parsing. In using a BiLSTM POS tagger with character CNNs and input representations of each word being a 100-dimensional word embedding and a 25 dimensional PTB POS embedding, they found that parsing results of OOV words were unimpressive until the technique of word dropout was introduced, where a certain percentage of words embeddings were zeroed out during training. The intuition behind this is that models will be trained to depend less on lexical information due to its unreliability. When using a variety of types of word dropout, including uniform word dropout and frequency-based word dropout (frequent words are zeroed out) and open-class word dropout (non-function words are zeroed out), Kasai and Frank found that the performance of these POS taggers were significantly improved across the board, even when faced with jabberwocky/pseudowords. However, it is important to note that they found that when faced with too aggressive of a word dropout rate, the parser's performance suffered, suggesting that a solution utilizing only word dropout would not be best suited to solve this issue.

Their results further show that word dropout encourages models to make greater use of the POS information of input words; however, given that their model's input POS embeddings are randomly initialized, this gives reasons to expect that if the model were pre-trained and the weights were initialized from values from the result of another task, the model's performance would improve. Kasai and Frank argued that random initialization were used in order to encourage the model to find abstractions over the POS embeddings; while this may be true, a more specified initialization may cause for faster convergence upon an accurate result. This form of transfer learning is what this study looks to utilize to improve parsing accuracy in computational models of sentence parsing.

Transfer learning is a technique used in machine learning to combat the major assumption that in many machine learning and data mining algorithms, the training and future data must be in the same feature space and have the same distribution. In reality, this may not hold true for many datasets. Transfer learning have been hugely popular in the deep learning community, applied within many state of the art computer vision methods. There have also been work done regarding transfer learning in natural language processing tasks, including speech processing, neural machine translation and more. The core idea is to train large models with limited data by freezing previously trained model weights and randomly initializing the last couple layers of the neural network to train new parameters. In this study, we aim to train the model using more morphologically rich language datasets (in our case we use Finnish, but any other language that fulfills the criteria works) and try to use the trained model to continue training on a less morphologically rich language (like English).

Given the results of our previous section, if novel words in morphologically richer languages depend on more factors than English or if bilingual speakers have improved performance in novel word identification, then it strongly suggests that the model with pre-trained weights from a task from another language will yield the highest accuracy.

In order to test if a combination of transfer learning and word dropout will yield more accurate results, two sets of experiment will be performed.

3.1 Experiment 1: No Transfer Learning (Monolingual Benchmark)

3.1.1 Overview

The purpose of this experiment is to try and replicate the results obtained by Kasai and Frank.

3.1.2 Methods

Model Used

Graph-based Sentence Parsing Model with Biaffine Attention and BiLSTM layers. The inputs will be 100-dimensional word embeddings and 25-dimensional PTB POS embeddings.

Variables

Word Dropout Rate. The specific variations will be shown in the table below with an explanation alongside it.

Word Dropout	Explanation
No Dropout	No word dropout is performed
Unlexicalized	The head words are not annotated for phrasal nodes
Top 100	The top 100 most frequently appearing words are zeroed out
Function	All of the function words are zeroed out
Uniform (p=0.2,0.4,0.6,0.8)	Each of the words are zeroed out (p*100)% of the time
Freq (n=1,40,352,2536)	All words that appear more than n times are zeroed out
Open Class (p=0.38,0.75)	Each of the nonfunction words are zeroed out (p*100)% of the time

Dataset Used
Universal Dependency representations obtained from converting the Penn Treebank using Stanford CoreNLP. Sections 2-21 are used for training, 22 is used for development and 23 is used for testing.

Results Obtained
Parsing accuracies of each word dropout type and rate are obtained.

3.1.3 Interpretation of results

If the results differ from that of the Kasai and Frank study, then the methods used for that study must be re-evaluated. If they are similar, then we can continue with our interpretations of Experiment 2.

3.2 Experiment 2: Transfer Learning on Different Language Task

3.2.1 Overview

The purpose of this experiment is to first train the model on data from the Turku Dependency Treebank corpus using the same train-dev-test split as in Experiment 1 for the Kasai and Frank data (from the Penn Treebank). The final weights for the PTB POS embeddings of those words will then be used as the initial weights for the model that will be run on the same data and in the same exact manner as in Experiment 1. In order to match up the PTB POS embedding for a Finnish word to its corresponding English word, we will use the Finnish-English parallel corpus of fienWaC by CLARIN.SI to make correct word to word translation.

3.2.2 Methods

Model Used
Graph-based Sentence Parsing Model with Bi-affine Attention and BiLSTM layers. The inputs will be 100-dimensional word embeddings and 25-dimensional PTB POS embeddings.

Variables
Word Dropout Rate. The specific variations are the same as in the table in Experiment 1.

Dataset Used
The same dataset as in Experiment 1 is used.

3.2.3 Interpretation of results

The parsing accuracies will be compared between Experiments 1 and 2 within each type of word dropout. For example, this means that the parsing accuracy as a result of no dropout for Experiment 1 will be compared with that as a result of no dropout for Experiment 2. From this, the effect of the transfer learning can be determined and evaluated with respect to particular types of word dropout. If across the board, the parsing accuracies are higher for Experiment 2 than for Experiment 1, then it strongly suggests that transfer learning with a different language task will result in higher parsing accuracies. If there are discrepancies with respect to the type of the word dropout, then conclusions regarding what combination of transfer learning and word dropout type can yield the best results. Finally, if the parsing accuracies are lower for Experiment 2 than for Experiment 1, then it strongly suggests that transfer learning with a different language task results in lower parsing accuracies.

Ultimately, if results are obtained where the parsing accuracies are higher for Experiment 2 than for Experiment 1, it suggests that for models whose purposes are strictly monolingual, a bilingual approach with transfer learning may yield more accurate results, even for applications outside of sentence parsing.

4 Conclusion

In this paper, novel word detection and processing of bilinguals (namely of English and Finnish) is investigated to analyze the effects of both bilingualism and morphological richness of given languages. If this yields results that show that bilingual speakers or/and that speakers of morphologically richer languages have a faster identification time, then it strongly suggests that transfer learning may be an effective approach when looking towards computational model for sentence parsing. Namely, this involves training a model on a language sentence parsing task before then transferring the corresponding POS embedding weights to a new model that is used on the different language sentence parsing task. If this is the case, there is reason to believe that not only would computational sentence parsing benefit from a bilingual approach, but potentially even other monolingual tasks may benefit from a similar approach.

Acknowledgments

We would like to thank Dr. Tal Linzen for organizing this great class and Suhas Arehalli for his tremendous support throughout the course.

References

Dozat, T. & Manning, C. 2017. *Deep Biaffine Attention for Neural Dependency Parsing*. ICLR

Fisher,C.; Hall, D.G.;Rakowitz,S.;Gleitman,L. 1994. *When it is better to receive than to give: syntactic and conceptual constraints on vocabulary growth*. Lingua, 92: 333-375

Gerz, D.; Vulić, I.; Ponti, E.; Naradowsky, J.; Reichart, R.; Korhonen, A. 2018. *Language Modeling for Morphologically Rich Languages: Character-Aware Modeling for Word-Level Prediction*. Transactions of the Association for Computational Linguistics. 6: 451-465.

Gildea, D. 2001. *Corpus Variation and Parser Performance*. Proceedings of the 2001 Conference on Empirical Methods in Natural Language Processing.

Kasai, J. Frank, R. 2019. *Jabberwocky Parsing: Dependency Parsing with Lexical Noise*. Proceedings of the Society for Computation in Linguistics: Vol. 2, Article 13.

Kaushanskaya, M. Marian, V. 2009. *The bilingual advantage in novel word learning.* Psychonomic Bulletin & Review. 16(4): 705-710.

Kharkwal, G. 2014 *Taming the Jabberwocky: Examining Sentence Processing with Novel Words*.

Ljubešić, et al. *Finnish-English Parallel Corpus FienWaC 1.0*. Finnish-English Parallel Corpus FienWaC 1.0, Jožef Stefan Institute

Turku BioNLP Group. *Turku Dependency Treebank (TDT)*

Appendix A.

English Experiment

Type I: Center-embedded

Subject-extracted:

SS-orig: The director who the teacher fought with loved the singer.
SS-1: The *[launor]* who the teacher fought with loved the singer.
SS-1*: The *[tawed]* who the teacher fought with loved the singer.
SS-2: The director who the *[pagner]* fought with loved the singer.
SS-2*: The director who the *[zoonify]* fought with loved the singer.
SS-3: The director who the teacher *[helleger]* with loved the singer.
SS-3*: The director who the teacher [belding] with loved the singer.
SS-4: The director who the teacher fought with *[grawed]* the singer.
SS-4*: The director who the teacher fought with *[cerdifician]* the singer.
SS-5: The director who the teacher fought with loved the *[wrot]*.
SS-5: The director who the teacher fought with loved the *[cretianize]*.

Object-extracted:

SO-orig: The director who fought with the teacher loved the singer.
SO-1: The *[yusktor]* who fought with the teacher loved the singer.
SO-1*: The *[visified]* who fought with the teacher loved the singer.
SO-2: The director who *[brawt]* with the teacher loved the singer.
SO-2*: The director who *[calikated]* with the teacher loved the singer.
SO-3: The director who fought with the *[poulner]* loved the singer.
SO-3*: The director who fought with the *[peagify]* loved the singer.
SO-4: The director who fought with the teacher *[gomed]* the singer.
SO-4*: The director who fought with the teacher *[tissizing]* the singer.
SO-5: The director who fought with the teacher loved the *[twenger]*.
SO-5*: The director who fought with the teacher loved the *[drontigated]*.

Type II: Right- Branching

Subject-extracted:

OS-orig: The singer was loved by the director who the teacher fought with.
OS-1: The *[smonger]* was loved by the director who the teacher fought with.
OS-1*: The *[hexify]* was loved by the director who the teacher fought with.
OS-2: The singer was *[jiled]* by the director who the teacher fought with.
OS-2*: The singer was *[yuler]* by the director who the teacher fought with.
OS-3: The singer was loved by the *[trator]* who the teacher fought with.
OS-3*: The singer was loved by the *[poilize]* who the teacher fought with.

OS-4: The singer was loved by the director who the *[scamner]* fought with.
OS-4*: The singer was loved by the director who the *[wolify]* fought with.
OS-5: The singer was loved by the director who the teacher *[pright]* with.
OS-5*: The singer was loved by the director who the teacher *[bowtized]* with.

Object-extracted:

OO-orig: The singer was loved by the director who fought with the teacher.
OO-1: The *[swoncer]* was loved by the director who fought with the teacher.
OO-1*: The *[dagify]* was loved by the director who fought with the teacher.
OO-2: The singer was *[twarked]* by the director who fought with the teacher.
OO-2*: The singer was *[cinderizing]* by the director who fought with the teacher.
OO-3: The singer was loved by the *[gwaltor]* who fought with the teacher.
OO-3*: The singer was loved by the *[orded]* who fought with the teacher.
OO-4: The singer was loved by the director who *[gnot]* with the teacher.
OO-4*: The singer was loved by the director who *[bollist]* with the teacher.
OO-5: The singer was loved by the director who fought with the *[wholcher]*.
OO-5*: The singer was loved by the director who fought with the *[monded]*.

Appendix B.

Finnish Experiment

Type I: Center- embedded

Subject-extracted:

FSS-orig: Johtaja, kukan opettaja taisteli kanssa rakasti laulajan.
FSS-1: *[Eukäja]*, kukan opettaja taisteli kanssa rakasti laulajan.
FSS-1*: *[Sikällita]*, kukan opettaja taisteli kanssa rakasti laulajan.
FSS-2: Johtaja, kukan *[ojatäja]* taisteli kanssa rakasti laulajan.
FSS-2*: Johtaja, kukan *[makteta]* taisteli kanssa rakasti laulajan.
FSS-3: Johtaja, kukan opettaja *[tavastui]* kanssa rakasti laulajan.
FSS-3*: Johtaja, kukan opettaja *[jaajaja]* kanssa rakasti laulajan.
FSS-4: Johtaja, kukan opettaja taisteli kanssa *[kutuki]* laulajan.
FSS-4*: Johtaja, kukan opettaja taisteli kanssa *[rakennaja]* laulajan.
FSS-5: Johtaja, kukan opettaja taisteli kanssa rakasti *[vuansillian]*.
FSS-5*: Johtaja, kukan opettaja taisteli kanssa rakasti *[hinangi]*.

Object-extracted:

FSO-orig: Johtaja, kuka taisteli kanssa opettajan rakasti laulajan.
FSO-1: *[Gohaja]*, kuka taisteli kanssa opettajan rakasti laulajan.
FSO-1*: *[Rautta]*, kuka taisteli kanssa opettajan rakasti laulajan.
FSO-2: Johtaja, kuka *[juliesli]* kanssa opettajan rakasti laulajan.
FSO-2*: Johtaja, kuka *[woliaja]* kanssa opettajan rakasti laulajan.
FSO-3: Johtaja, kuka taisteli kanssa *[soyajan]* rakasti laulajan.
FSO-3*: Johtaja, kuka taisteli kanssa *[kouliva]* rakasti laulajan.
FSO-4: Johtaja, kuka taisteli kanssa opettajan *[pulasti]* laulajan.
FSO-4*: Johtaja, kuka taisteli kanssa opettajan *[houliaja]* laulajan.
FSO-5: Johtaja, kuka taisteli kanssa opettajan rakasti *[bouliajan]*.
FSO-5*: Johtaja, kuka taisteli kanssa opettajan rakasti *[kuwava]*.

Type II: Right- Branching

Subject-extracted:

FOS-orig: Laulajan oil rakastettu mennessä johtaja, kukan opettaja taisteli kanssa.
FOS-1: *[Talajan]* oil rakastettu mennessä johtaja, kukan opettaja taisteli kanssa.
FOS-1*: *[Teleta]* oil rakastettu mennessä johtaja, kukan opettaja taisteli kanssa.
FOS-2: Laulajan oil *[hettu]* mennessä johtaja, kukan opettaja taisteli kanssa.
FOS-2*: Laulajan oil *[talaja]* mennessä johtaja, kukan opettaja taisteli kanssa.
FOS-3: Laulajan oil rakastettu mennessä *[palaja]*, kukan opettaja taisteli kanssa.
FOS-3*: Laulajan oil rakastettu mennessä *[tuvta]*, kukan opettaja taisteli kanssa.

FOS-4: Laulajan oil rakastettu mennessä johtaja, kukan *[opetaja]* taisteli kanssa.
FOS-4*: Laulajan oil rakastettu mennessä johtaja, kukan *[tuvita]* taisteli kanssa.
FOS-5: Laulajan oil rakastettu mennessä johtaja, kukan opettaja *[tolikiti]* kanssa.
FOS-5*: Laulajan oil rakastettu mennessä johtaja, kukan opettaja *[ratjata]* kanssa.

Object-extracted:

FOO-orig: Laulajan oil rakastettu mennessä johtaja, kuka taisteli kanssa opettajan.
FOO-1: *[Touliajan]* oil rakastettu mennessä johtaja, kuka taisteli kanssa opettajan.
FOO-1*: *[Rouva]* oil rakastettu mennessä johtaja, kuka taisteli kanssa opettajan.
FOO-2: Laulajan oil *[trollosettu]* mennessä johtaja, kuka taisteli kanssa opettajan.
FOO-2*: Laulajan oil *[nouliaja]* mennessä johtaja, kuka taisteli kanssa opettajan.
FOO-3: Laulajan oil rakastettu mennessä *[foulaja]*, kuka taisteli kanssa opettajan.
FOO-3*: Laulajan oil rakastettu mennessä *[vouliva]*, kuka taisteli kanssa opettajan.
FOO-4: Laulajan oil rakastettu mennessä johtaja, kuka *[jalateli]* kanssa opettajan.
FOO-4*: Laulajan oil rakastettu mennessä johtaja, kuka *[hogataja]* kanssa opettajan.
FOO-5: Laulajan oil rakastettu mennessä johtaja, kuka taisteli kanssa *[gotorajan]*.
FOO-5*: Laulajan oil rakastettu mennessä johtaja, kuka taisteli kanssa *[haiseda]*.