Exploring Statistical and Neural Models for Noun Ellipsis Detection andResolution in English

by

Payal Khullar, Manish Shrivastava

Centre for Language Technologies Research Centre
International Institute of Information Technology
Hyderabad - 500 032, INDIA
December 2020

Exploring Statistical and Neural Models for Noun Ellipsis Detection and Resolution in English

Payal Khullar

International Institute of Information Technology Hyderabad
Gachibowli, Hyderabad
`payal.khullar@research.iiit.ac.in`

Abstract

Computational approaches to noun ellipsis resolution has been sparse, with only a naive rule-based approach that uses syntactic feature constraints for marking noun ellipsis licensors and selecting their antecedents. In this paper, we further the ellipsis research by exploring several statistical and neural models for both the subtasks involved in the ellipsis resolution process and addressing the representation and contribution of manual features proposed in previous research. Using the best performing models, we build an end-to-end supervised Machine Learning (ML) framework for this task that improves the existing F1 score by 16.55% for the detection and 14.97% for the resolution subtask. Our experiments demonstrate robust scores through pretrained BERT (Bidirectional Encoder Representations from Transformers) embeddings for word representation, and more so the importance of manual features– once again highlighting the syntactic and semantic characteristics of the ellipsis phenomenon. For the classification decision, we notice that a simple Multilayer Perceptron (MLP) works well for the detection of ellipsis; however, Recurrent Neural Networks (RNN) are a better choice for the much harder resolution step.

1 Introduction

Noun ellipsis is a linguistic phenomenon where the head noun of a noun phrase gets deleted, without making the sentence ungrammatical. For example in the sentence in (1) from (Lobeck, 1995), the noun *presentation* is elided at "[e]".

1. John's **presentation** on urban development was virtually ignored because [NP Mary's [e]] was so much more interesting.

The elided information can be retrieved from the previous context as in (1) or with the knowledge of idiomatic usage of language as in *I will be back in two [e].* where *two* means *two minutes*. It is
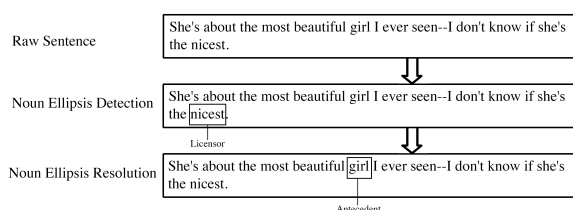

Figure 1: An example of the noun ellipsis resolution process from the dialogue L131377 of the m_44 movie of the Cornell Movie Dialogs dataset. Here, *the nicest* and *girl* denote the ellipsis licensor and antecedent respectively.

also possible that the reference of the elided information comes from extra-linguistic, situational context. For example, consider a speaker pointing towards the roses in a shop and saying an utterance as in *I will take two [e].* While human interlocutors resolve any such elided information by disambiguating from context, cognitive commonsense extension and reasoning (Chen, 2016), ellipsis resolution can be a hard task for Natural Language Processing (NLP) systems (Hardt, 1999). Resolution of ellipsis comprises two tasks - detection of the elided material and antecedent selection (Liu et al., 2016b; Nielsen, 2003). Ellipses occur in the environment of certain syntactical structures or trigger words, also known as licensors or triggers of ellipses. They are useful syntactic cues for the detection of noun ellipsis. See Figure 1 for an example of the noun ellipsis resolution process.

2 Related Work

Nominal ellipsis has been a topic of interest in theoretical linguistics for a very long time (Halliday and Hasan, 1976; Dalrymple et al., 1991; Lobeck, 1995; Lappin, 1996; Hobbs and Kehler, 1997; Hardt, 1999; Johnson, 2001; Wijnen et al., 2003; Merchant, 2004; Frazier, 2008; Chung et al., 2010; Mer-

chant, 2010; Goksun et al., 2010; Gunther, 2011; Rouveret, 2012; Lindenbergh et al., 2015; van Craenenbroeck and Merchant, 2013; Park, 2017; Hyams et al., 2017; Kim et al., 2019). Computational approaches to the ellipsis phenomenon majorly focus on the Verb Phrase Ellipsis (VPE) along with a few related phenomenon such as gapping, sluicing and do-so anaphora, for instance, the detection of VPE in the Penn Treebank using pattern match (Hardt, 1992), a transformation learning-based approach to generated patterns for VPE resolution (Hardt, 1998), the domain independent VPE detection and resolution using machine learning (Nielsen, 2003), automatically parsed text (Nielsen, 2004b), sentence trimming methods (McShane et al., 2015), linguistic principles (McShane and Babkin, 2016), improved parsing techniques that encode elided material dependencies for reconstruction of sentences containing gapping (Schuster et al., 2018), discriminative and margin infused algorithms (Dean et al., 2016), Multilayer Perceptrons (MLP) and Transformers (Zhang et al., 2019). In recent times, there has been a surge in the computational research on nominal ellipsis and closely related phenomena (Khullar et al., 2020, 2019; Lapshinova-Koltunski et al., 2018; Menzel, 2017; Menzel and Lapshinova-Koltunski, 2014). For the resolution process, we previously proposed a rule based system (Khullar et al., 2019) that detects noun ellipsis using syntactic constraints on licensors of ellipsis and resolves them by matching Part-of-Speech (POS) tag similarity between the licensor of ellipsis and the modifier of the antecedent. It later fine tunes these syntactic rules on a small curated dataset that contains 234 instances of noun ellipsis along with some negative samples (Khullar et al., 2019). For the present paper, we further the research on noun ellipses by using the NoEl corpus annotated by us previously (Khullar et al., 2020) to experiment with state-of-the-art ML models.

3 The Proposed Approach

Following the VPE resolution framework presented by (Zhang et al., 2019), we investigate a similar framework for noun ellipsis resolution in English and present alternative choices of the models at each step as shown in Figure 2. We use the NoEl corpus (Khullar et al., 2020) that marks noun ellipsis instances as a separate layer (using the stand-off annotation scheme) on the Cornell Movie Dialogs corpus (Danescu-Niculescu-Mizil and Lee, 2011).

The corpus marks a total of 946 annotations, of which 438 are described as endophoric, i.e. with a textual antecedent, and 508 exophoric, i.e. without a textual antecedent.

3.1 Noun Ellipsis Detection

From a given sentence, we first select all words belonging to the syntactic categories that can license noun ellipsis in English, i.e. cardinal and ordinal numbers, determiners and adjectives (Ross, 1967; Lobeck, 1995; Mitkov, 1999; Saito et al., 2008; Kim et al., 2019; Khullar et al., 2019) using a POS tag filter. The POS tags are obtained from state-of-the-art spaCy parser (Honnibal and Johnson, 2015). For simplicity, we refer to words with these categories as noun modifiers (although in strict linguistic terms, this might be problematic). For each of these selected noun modifiers, we follow the task specification for VPE detection used by (Nielsen, 2004a; Bos and Spenader, 2011; Liu et al., 2016a; Dean et al., 2016) and present noun ellipsis detection as a binary classification task, where given a noun modifier and the sentence in which it occurs as the input, the goal of the classifier is to predict whether the noun modifier licenses a noun ellipsis or not. Formally, for a given licensor word l_i is a licensor in a sentence s, the task is represented as follows:

$$f(l_i, s) \rightarrow \{0, 1\}$$

where 1 denotes that l_i is a licensor in s, and 0 otherwise.

We experiment with both static and contextualised word embeddings for word and context representation. For the former, we choose pretrained fastText (FT) word embeddings (Bojanowski et al., 2016) as they provide representations for rare and unknown words that might be frequent in the movie dialogues. For the latter, we use pretrained BERT embeddings from the BERT base uncased word-piece model for English (Devlin et al., 2019), as these currently offer the most powerful embeddings taking into account a large left and right context.

fastText We take pretrained FT word embeddings for the noun modifier and sentence in which it is present and sum pool to obtain a single vector that we use to train our classifiers. For the statistical models, we choose Naive Bayes and Linear Support Vector Machine (SVM), and use scikit learn (Pedregosa et al., 2011) with 5-fold cross validation for training and testing. We choose a
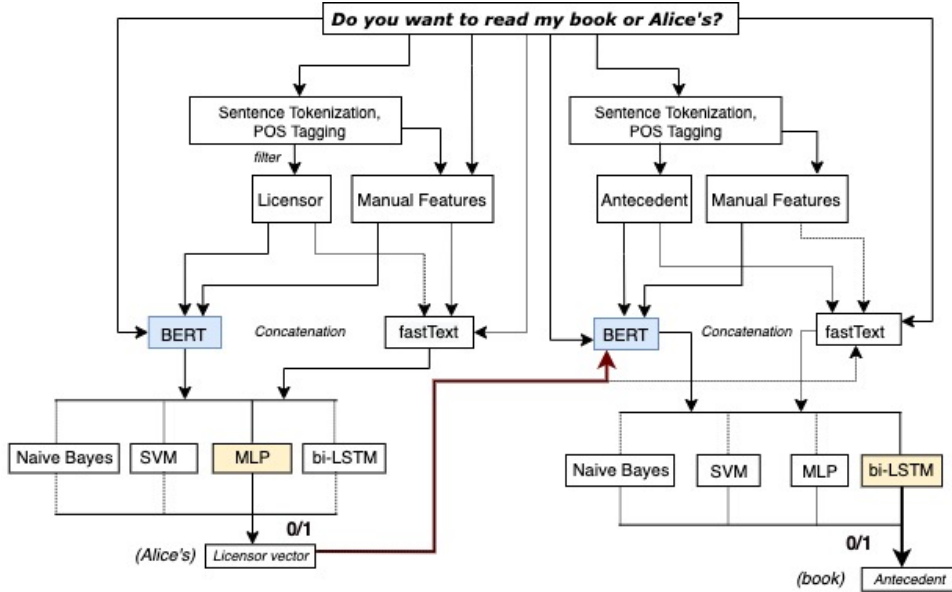
Figure 2: The proposed framework for noun ellipsis detection and resolution in English.

simple MLP and a biLSTM (Bidirectional Long Short Term Memory) as our state-of-the-art neural classifiers.

BERT We separate the sentence and the licensor with a [SEP] token and keep the sequence length to 300 as this is the maximum sentence length in the training data. After creating the concatenated set of tokens, if the number of tokens are greater than 300, we clip it to 300, otherwise we add [PAD] tokens which correspond to the embedding of 768 dimensional zero-vector. The [CLS] output of the BERT model (Devlin et al., 2019) is then fed into Naive Bayes, Linear SVM, MLP and bi-LSTM networks as above.

Manual Syntactic Features For each of these models, we additionally experiment with manual syntactic features. We use the lexical features proposed by (Dean et al., 2016) and extended lexical features by (Zhang et al., 2019), and take the five syntactic constraints on licensors of ellipsis explored by (Khullar et al., 2019) for their rule-based approach as our slot pattern features. We concatenate all these features to the embeddings from the previous step and check if they improve the classification decision.

3.2 Noun Ellipsis Resolution

We define noun ellipsis resolution as a binary classification task where given a licensor, antecedent candidate and their context, the goal of the classifier is to predict whether the antecedent candidate

is the resolution of the ellipsis licensed by the licensor. Formally, given a sentence s, the licensor l_i from the detection step, and the antecedent candidate a_j; the noun ellipsis resolution task can be defined as follows:

$$f(a_j, l_i, s) \rightarrow \{0, 1\}$$

where 1 denotes that the antecedent candidate a_j is the actual resolution of the ellipsis licensed by l_i, and 0 otherwise.

Embeddings Similar to the detection step, we take pretrained fastText word embeddings for the licensor, antecedent candidate and context, and sum pool to obtain a single vector. In case of BERT, we separate the sentence, the licensor and the antecedent candidate with a [SEP] token and follow the same steps as in the detection step.

Manual Syntactic and Semantic Features We use POS tags of the licensor and modifier of the antecedent as our syntactic features and cosine similarity between their POS tags as our semantic features, following (Khullar et al., 2019). We concatenate these features to the embeddings to explore the efficacy of adding manual features on resolution.

4 Experiments

We choose Naive Bayes and linear SVM as our statistical models, and Multilayer Perceprton (MLP) and bidirectional Long Short Term Memory (bi-LSTM) network as our neural models. For the

Subtask	Classification	Representation	Precision	Recall	F1-Score
		FT	0.4192	0.5807	0.4869
		FT+F	0.4923	0.6275	0.5517
	Naive Bayes	BERT	0.5511	0.6203	0.5837
		BERT+F	0.5990	0.6213	0.6099
		FT	0.4502	0.5560	0.4975
		FT+F	0.4824	0.6400	0.5501
	SVM	BERT	0.5766	0.6503	0.6112
		BERT+F	0.5799	0.6771	0.6247
Detection					
		FT	0.6504	0.6974	0.6731
		FT+F	0.7123	**0.8977**	0.7943
	MLP	BERT	0.6999	0.7398	0.7193
		BERT+F	**0.9116**	0.8901	**0.9007**
		FT	0.6010	0.6531	0.6260
		FT+F	0.6901	0.8745	0.7714
	bi-LSTM	BERT	0.7001	0.7503	0.7243
		BERT+F	0.8655	0.8163	0.8402

Table 1: Precision, Recall and F1-Score values of different statistical and neural models on the noun ellipsis detection task. FT stands for fastText and +F denotes concatenation of manual features to the embeddings. Values in bold depict best performance.

detection task, we take the annotated 946 positive samples (exophoric) and randomly choose 946 negative samples. Similarly, for the resolution task, we take 438 positive samples (endophoric) and 438 randomly chosen negative samples. We perform a standard 70-10-20 split to obtain the train, development and test set respectively, and follow the 5-fold cross validation procedure to capture both classes properly in each case. For MLP, we take a simple, two-layer feedforward network (FFNN) or two layers of multiple computational units interconnected in a feed-forward way without loops. We have a single hidden layer with 768 neurons and a sigmoid function. A unidirectional weight connection exists between the two successive layers. The classification decision is made by turning the input vector representations of a word with its context into a score. The network has a softmax output layer. For bi-LSTM, we have embedding layer, time-distributed translate layer, Bi-LSTM (RNN) layer, batch normalization layer, dropout layer and prediction layer. The activation used is Softmax. The loss function is calculated with cross entropy. We train in batch sizes of 16 and early stopping with max epochs of 100. In early stopping the patience is kept to be 10 and the optimizer used is

Adam. We use default values for the learning rate. We use Keras (Chollet et al., 2015) for coding the models.

5 Results

We evaluate the performance of our models in terms of F1-score, computed by taking an average F1-scores obtained from the 5-folds results. We experiment with sixteen models each for the noun ellipsis detection and resolution. The results on the testset for Precision, Recall and F1-Score values are presented in Table.2. As expected, the neural models perform significantly better than the statistical ones for both the subtasks. Our experiments show that for the detection task, BERT embeddings with a simple MLP gives best scores. This is expected because, BERT currently provides the most powerful contextual word representations, using 12 separate attention mechanism for each layer, where, at each layer, each token can focus on 12 distinct aspects of other tokens. Since Transformers (Vaswani et al., 2017) use many distinct attention heads (12*12=144 for the base BERT model), each head can focus on a different kind of constituent combinations, making BERT broadly attending over the whole sentence. In our task, the

Subtask	Classification	Representation	Precision	Recall	F1-Score
		FT	0.2992	0.5112	0.3775
		FT+F	0.3379	0.5750	0.4257
	Naive Bayes	BERT	0.3919	0.5503	0.4578
		BERT+F	0.4558	0.6230	0.5264
		FT	0.4001	0.4960	0.4429
		FT+F	0.5024	0.6040	0.5485
	SVM	BERT	0.4966	0.6003	0.5435
		BERT+F	0.5372	0.6611	0.5927
Resolution					
		FT	0.5901	0.6577	0.6221
		FT+F	0.7005	0.8122	0.7522
	MLP	BERT	0.6951	0.7798	0.7350
		BERT+F	0.8015	0.8531	0.8265
		FT	0.6009	0.6321	0.6161
		FT+F	0.7029	0.8044	0.7502
	bi-LSTM	BERT	0.7195	0.7383	0.7288
		BERT+F	**0.8652**	**0.8166**	**0.8401**

Table 2: Precision, Recall and F1-Score values of different statistical and neural models on the noun ellipsis resolution subtask. FT stands for fastText and +F denotes concatenation of manual features to the embeddings.

BERT with MLP model is robust and efficiently makes generalisations on the syntactic and semantic dependency between the elided noun and the pre-modifiers and modifiers (licensors). For the resolution task, however, bi-LSTMs work better than MLP. Unlike MLPs, Bi-LSTMs take into consideration left to right and right to left context, capturing long range dependencies in a sentence and, hence, are better suited for handling the resolution of a cohesive discourse device like ellipsis where the distance between elided phrase and antecedent can be several words. The sufficient neurons in the hidden layer with sigmoidal function ensures the MLP approximate the nonlinear relationships between the

input and output, but they are not innately designed to capture temporal relationships within a sentence. Hence, although they perform well for a task like detection that needs local information, they are outperformed by bi-LSTMs on the resolution task that requires capturing a deeper relationship between the antecedent and the elided noun. We also note that manual feature addition boosts results greatly for all models, highlighting that ellipsis is a syntactically constrained phenomenon. We finally integrate the best models for each subtask into an end-to-end pipeline, as in Figure 2. Now, instead of the gold vectors (from the annotations), the resolution model is fed the ouput licensor vector from the detection model. This obviously results into error propagation into the second model, and lowers the precision value to 82.52%, recall to 78.66% and consequently, the F1-score to 80.55% of the final system. The error in the final system comes from failing to detect actual licensors, wrongly identifying non-licensor words and correct licensor detection but failed antecedent resolution. We run our final system on the curated dataset prepared by (Khullar et al., 2019) and compare the results with their rule-based approach. As expected, this model improves the F1-score by 16.55% for noun ellipsis detection and 14.97% for noun ellipsis res-

Curated Dataset	P	R	F
Detection			
(Khullar et al., 2019)	69.15	85.53	76.47
MLP, BERT+F	91.72	94.32	93.02
Resolution			
(Khullar et al., 2019)	78.79	63.41	70.27
bi-LSTM, BERT+F	87.01	83.54	85.24

Table 3: Precision (**P**), Recall (**R**) and F1-Score (**F**) values (%) of the rule-based approach by (Khullar et al., 2019) and the neural model presented in this paper.

olution. See Table 3. The even higher accuracy on the curated dataset can be explained by the nature of the sentences in this dataset which are from textbooks, and, hence, free of grammatical errors, etc. – resulting into improved parser performance in the pre-processing step. Although, the presented models achieve high scores on both the tasks separately and in the pipeline process, the results can be further improved with hyper-parameter tuning and additional regularization.

6 Conclusion

We explored statistical and neural models for noun ellipsis detection and resolution, presenting a strong results for this task. As expected, neural classifiers perform significantly better than the statistical with the same input representation. As with several other NLP tasks, the contextual nature of BERT is useful for noun ellipsis resolution too, making robust predictions with simple neural classifiers. Finally, addition of manual features boosts the performance of all classifiers including those that use BERT, highlighting that ellipsis is a syntactically constrained phenomenon.

References

Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2016. Enriching word vectors with subword information. *arXiv preprint arXiv:1607.04606*.

Johan Bos and Jennifer Spenader. 2011. An annotated corpus for the analysis of vp ellipsis. *Language Resources and Evaluation*, 45(4):463–494.

Wei Chen. 2016. The motivation of ellipsis. *Theory and Practice in Language Studies*, 6(11):2134–2139.

François Chollet et al. 2015. Keras. https://keras.io.

Sandra Chung, William Ladusaw, and James McCloskey. 2010. Sluicing (:) between structure and inference. *In Representing language: Essays in honor of Judith Aissen*.

Jeroen van Craenenbroeck and Jason Merchant. 2013. Ellipsis phenomena. In *The Cambridge Handbook of Generative Syntax*, pages 701–745. Cambridge University Press.

Mary Dalrymple, Stuart M. Shieber, and Fernando C. N. Pereira. 1991. Ellipsis and higher-order unification. *Linguistics and Philosophy*, 14(4):399–452.

Cristian Danescu-Niculescu-Mizil and Lillian Lee. 2011. Chameleons in imagined conversations: A new approach to understanding coordination of linguistic style in dialogs. In *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics, ACL 2011*.

Kian Kenyon Dean, Jackie Chi Kit Cheung, and Doina Precup. 2016. Verb phrase ellipsis resolution using discriminative and margin-infused algorithms. *In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, page 1734–1743.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *NAACL-HLT*.

Lyn Frazier. 2008. Processing ellipsis: A processing solution to the undergeneration problem? In *Proceedings of the 26th West Coast Conference on Formal Linguistics*.

Tilbe Goksun, Tom W. Roeper, Kathy Hirsh-Pasek, and Roberta Michnick Golinkoff. 2010. From nounphrase ellipsis to verbphrase ellipsis: The acquisition path from context to abstract reconstruction.

Christine Gunther. 2011. Noun ellipsis in english: adjectival modifiers and the role of context. *The structure of the noun phrase in English: synchronic and diachronic explorations*, 15(2):279–301.

Michael Alexander Kirkwood Halliday and Ruqaiya Hasan. 1976. Cohesion in english. page 76.

Daniel Hardt. 1992. An algorithm for vp ellipsis. page 9–14. In Proceedings of the 30th annual meeting on Association for Computational Linguistics.

Daniel Hardt. 1998. Improving ellipsis resolution with transformation-based learning. AAAI fall symposium.

Daniel Hardt. 1999. Dynamic interpretation of verb phrase ellipsis. *Linguistics and Philosophy*, 22(2):187–221.

Jerry R. Hobbs and Andrew Kehler. 1997. A theory of parallelism and the case of vp ellipsis. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics*, ACL '98/EACL '98, pages 394–401, Stroudsburg, PA, USA. Association for Computational Linguistics.

Matthew Honnibal and Mark Johnson. 2015. An improved non-monotonic transition system for dependency parsing. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1373–1378, Lisbon, Portugal. Association for Computational Linguistics.

Nina Hyams, Victoria Mateu, and Lauren Winans. 2017. Ellipsis meets wh-movement: sluicing in early grammar.

Kyle Johnson. 2001. What vp ellipsis can do, and what it can't, but not why. pages 439–479.

Payal Khullar, Allen Anthony, and Manish Shrivastava. 2019. Using syntax to resolve npe in english. In *Proceedings of Recent Advances in Natural Language Processing*, pages 535–541.

Payal Khullar, Kushal Majmundar, and Manish Shrivastava. 2020. Noel: An annotated corpus for noun ellipsis in english. In *Language Resources Evaluation Conference*.

Nayoun Kim, Laurel Brehm, and Masaya Yoshida. 2019. The online processing of noun phrase ellipsis and mechanisms of antecedent retrieval. *Language, Cognition and Neuroscience*, 34(2):190–213.

Shalom Lappin. 1996. The interpretatin of ellipsis. In Shalom Lappin, editor, *The Handbook of Contemporary Semantic Theory*, pages 145–176. Blackwell.

Ekaterina Lapshinova-Koltunski, Christian Hardmeier, and Pauline Krielke. 2018. Parcorfull: a parallel corpus annotated with full coreference. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation, LREC 2018, Miyazaki, Japan, May 7-12, 2018*.

Charlotte Lindenbergh, Angeliek van Hout, and Bart Hollebrandse. 2015. Extending ellipsis research: The acquisition of sluicing in dutch. *BUCLD 39 Online Proceedings Supplement*, 39.

Chin-Ting Liu, Li-mei Chen, Yu-Ching Lin, Chia-Fang Cheng, and Hui-chen Chang. 2016a. Speech intelligibility and the production of fricative and affricate among Mandarin-speaking children with cerebral palsy. In *Proceedings of the 28th Conference on Computational Linguistics and Speech Processing (ROCLING 2016)*, pages 153–163, Tainan, Taiwan. The Association for Computational Linguistics and Chinese Language Processing (ACLCLP).

Zhengzhong Liu, Edgar Gonzalez, and Dan Gillick. 2016b. Verb phrase ellipsis detection using automatically parsed text. pages 32–40.

Anne Lobeck. 1995. *Functional Heads, Licensing, and Identification*. Oxford University Press.

Marjorie McShane and Petr Babkin. 2016. Detection and resolution of verb phrase ellipsis. *Linguistic Issues in Language Technology*, 13(1).

Marjorie McShane, Sergei Nirenburg, and Petr Babkin. 2015. Sentence trimming in service of verb phrase ellipsis resolution. In *EAPCogSci*.

Katrin Menzel. 2017. *Understanding English-German contrasts: a corpus-based comparative analysis of ellipses as cohesive devices*. Ph.D. thesis, Universitat des Saar- landes, Saarbrucken.

Katrin Menzel and Ekaterina Lapshinova-Koltunski. 2014. Kontrastive analyse deutscher und englischer kohäsionsmittel in verschiedenen diskurstypen. *tekst i dyskurs - Text und Diskurs. Zeitschrift der Abteilung für germanistische Sprachwissenschaft des Germanistischen Instituts Warschau*.

Jason Merchant. 2004. Fragments and ellipsis. *Linguistics and Philosophy*, 27(6):661–738.

Jason Merchant. 2010. *Three Kinds of Ellipsis: Syntactic, Semantic, Pragmatic?*

Ruslan Mitkov. 1999. *Anaphora Resolution*. Oxford University Press.

Leif Arda Nielsen. 2003. Using machine learning techniques for vpe detection.

Leif Arda Nielsen. 2004a. Robust VPE detection using automatically parsed text. In *Proceedings of the ACL Student Research Workshop*, pages 49–54, Barcelona, Spain. Association for Computational Linguistics.

Leif Arda Nielsen. 2004b. Verb phrase ellipsis detection using automatically parsed text.

Dongwoo Park. 2017. *When does ellipsis occur, and what is elided?* PhD dissertation, University of Maryland.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

John Robert Ross. 1967. *Constraints on variables in syntax*. Massachusetts Institute of Technology.

Alain Rouveret. 2012. Vp ellipsis, phases and the syntax of morphology. *Natural Language & Linguistic Theory*, 30(3):897–963.

Mamoru Saito, Jonah Lin, and Keiko Murasugi. 2008. Nominal-ellipsis and the structure of noun phrases in chinese and japanese. *Journal of East Asian Linguistics*, 17.

Sebastian Schuster, Joakim Nivre, and Christopher D. Manning. 2018. Sentences with gapping: Parsing and reconstructing elided predicates. *ArXiv e-prints*.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *NeurIPS*.

Frank Wijnen, Tom W. Roeper, and Hiske van der Meulen. 2003. Discourse binding: Does it begin with nominal ellipsis?

Wei-Nan Zhang, Yue Zhang, Yuanxing Liu, Donglin Di, and Ting Liu1. 2019. A neural network approach to verb phrase ellipsis resolution.