Verb Phrase Anaphora: *Do(ing) so* with Heuristics

Sandhya Singh*, **Kushagra Shree**†, **Sriparna Saha**†,
Pushpak Bhattacharyya*, **Gladvin Chinnadurai**‡, **Manish Kumar Vatsa**‡
*CFILT, Indian Institute of Technology Bombay, †Indian Institute of Technology Patna,
‡LG Soft India
{sandhyasingh, pb}@cse.iitb.ac.in,
kushagra4xps@gmail.com, sriparna@iitp.ac.in,
{gladvin.durai, manish.vatsa}@lge.com

Abstract

Verb Phrase Anaphora (VPA) is a universal language phenomenon. It can occur in the form of *do so* phrase, *verb phrase ellipsis, etc.* Resolving VPA can improve the performance of Dialogue processing systems, Natural Language Generation (NLG), Question Answering (QA) and so on. In this paper, we present a novel computational approach to resolve the specific verb phrase anaphora appearing as *do so* construct and its lexical variations for the English language. The approach follows a heuristic technique using a combination of parsing from classical NLP, state-of-the-art (SOTA) Generative Pre-trained Transformer (GPT) language model and RoBERTa grammar correction model. The result indicates that our approach can resolve these specific verb phrase anaphora cases with 73.40 F1 score. The data set used for testing the specific verb phrase anaphora cases of *do so* and *doing so* is released for research purposes. This module has been used as the last module in a coreference resolution pipeline for a downstream QA task for the electronic home appliances sector.

1 Introduction

Anaphora is a linguistic construct used for maintaining coherence in the text without being repetitive. A solution to Anaphora Resolution (AR) can improve the performance of downstream tasks like Dialogue systems, Natural Language Generation (NLG), Question Answering *etc.* The process of identifying the linguist element *(anaphor)* that is referring to a preceding linguistic element *(antecedent)* in the context is known as *Anaphora Resolution (AR).*

According to Mitkov (2002), anaphoras can be classified as pronominal anaphora, lexical noun phrase anaphora, verb phrase anaphora, adverb anaphora and zero anaphora. The current state-of-the-art systems (Clark and Manning, 2016; Lee et al., 2017; Joshi et al., 2019a) have obtained high accuracy for the most prevalent type of anaphoras *i.e.* pronominal anaphora and lexical noun phrase anaphora cases. However, verb phrase anaphora, adverb anaphora and zero anaphora still remain unsolved due to the complexities involved in these language phenomena. This paper deals with one such case: **verb phrase anaphora (VPA).** The verb phrase anaphor is resolved by a preceding verb phrase plus any complement and adjunct as the antecedent. The verb phrase anaphor constructs occur as a combination of *so, this, that, it and the same thing* along with *do*. *Example 1* shows the verb phrase anaphor *doing so* referring to the verb phrase *use energy saver mode* as an antecedent. Besides antecedent identification, the grammar also enforces a syntactic modification to the antecedent as *"Using energy saver mode"* for resolution in reference to the anaphor.

> ***[Use energy saver mode]$_1$*** *in the air conditioner.*
> ***[Doing so]$_1$*** *helps reduce the load on the pocket.*
>
> *Antecedent substituted output:* **Using energy saver mode** helps reduce the load on the pocket.
> **Example 1**

In this paper, we focus on the specific case of *so anaphora* which is used in conjunction with the verb form *do*. This construct is one of the most frequent forms of verb phrase anaphora. The challenges posed by *do so* constructs are both semantic and morphosyntactic in nature. **The contributions** through this work are listed here as:

- A novel computational heuristic approach using a combination of classical NLP, transformer-based language model and a grammar correction model to resolve specific *do so* constructs as the anaphoric expression.
- A dataset of 350 data points of *do so* VPA constructs in inter-sentence and intra-sentence format is also released as a part of our research contribution.

The paper is organized as follows: Section 2 discusses the syntactic and semantic challenges

Proceedings of the 19th International Conference on Natural Language Processing (ICON), pages 92 - 98
December 15-18, 2022 ©2022 Association for Computational Linguistics

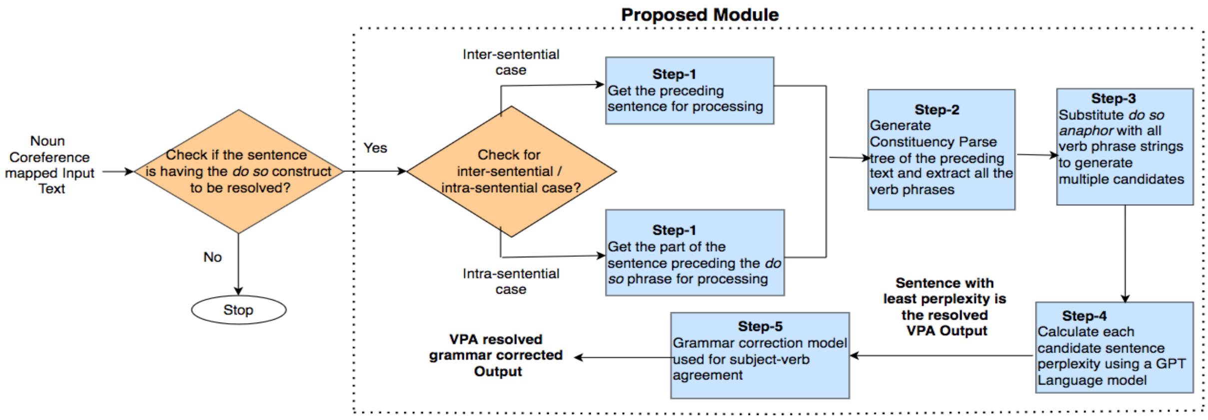

Figure 1: Proposed approach flow as the last module of a coreference resolution pipeline for a QA system.

associated with *do so* anaphor. Section 3 presents the related literature with specific focus on verb phrase anaphora. Section 4 describes our dataset and the heuristic approach used to resolve verb phrase anaphora. Section 5 analyses the results obtained followed by conclusion and future work.

2 Challenges with *do so* construct

Anaphora and ellipsis are preferred linguistic mechanisms used in conversation. In general, any anaphora resolution is a three step process: *i) Anaphor identification, ii) Antecedent candidates identification, and iii) Choosing the most likely antecedent candidate.* The general resolution technique involves eliminative constraints based on gender, number, semantic consistency and weighting preference based on proximity, centering, syntactic/semantic (role) parallelism between the anaphor and antecedent (Sayed, 2003). But the verb phrase anaphora is found to be resolved in a more complex manner.

The anaphoric verb phrases such as *do it, do that, do so, ellipsis etc.* are known to inherit the properties from their referent. As in *example 2*, the phrase *did too* not only refers to the event of eating performed by John but also to properties such as *quietly* and *from the plate*.

John ***[quietly ate the cake from the plate]***$_1$*. Jerry* ***[did too.]***$_1$
Antecedent substituted output: Jerry **quietly ate the cake from the plate** too.
Example 2

Verb phrase anaphora when resolved from a discourse perspective departs on two aspects from the standard approaches of entity resolution (Prüst et al., 1994). First, it cannot be determined by simply identifying the anaphoric verb phrase with an antecedent verb phrase. The resolution process must establish a syntactic/semantic parallelism between clauses or discourse constituent units in which the verb phrase occurs. Second, the discourse structure significantly influences the reference possibilities of verb phrase anaphora.

As we are working on a QA system for the home appliances domain, we are faced with VPA constructs in the user manual for the devices. Thus, our work focuses specifically on *do so* anaphora construct in the QA problem which we are trying to solve as a goal. It has been identified that *do so* does not refer to only the verb alone but the entire verb phrase consisting of the main verb, auxiliary verb, its complements and adjuncts as constituted in phrase structure grammar. For anaphora resolution, both terms share the load with *do* placing the semantic requirement on the antecedent whereas *so* is responsible for the anaphoric work.

My grandmother ***[knows all her grandchildren's names]***$_1$*, and she manages to* ***[do so]***$_1$ *despite her Alzheimer's.*
Example 3

The students, who ***[know French best]***$_1$*,* ***[do so]***$_1$ *because they lived in France for a year.*
Example 4

The *do so* construct also appears as the infinitive form as *"to know..."* in *example 3* and in the form where the antecedent is contained in a relative clause *"who know French ..."*, thus modifying the subject of *do so* as in *example 4*. This syntactic and semantic analysis of the construct highlights the challenges it poses in resolving it.

3 Related Work

The research on computational anaphora and coreference resolution has seen a paradigm shift from heuristic approaches to machine learning ap-

Data Points	Inter-sentential	Intra-sentential
350	210 (60%)	140 (40%)

(a) Distribution of inter-sentential & intra-sentential cases

Construct Type	Data Points
Doing so construct	130
Do so construct	149
Others (*did, does, these, this*)	71

(b) Distribution of different types of verb anaphor cases

Table 1: Data distribution statistics of our dataset

proaches in both nominal-antecedent anaphora (Ng, 2010) and non nominal-antecedent anaphora (Kolhatkar et al., 2018) categories.

Specific to VPA, considerable work is seen in the field of theoretical linguistics for different languages. Hankamer and Sag (1976) investigated verb phrase anaphora as deep or surface anaphora for the English language and Houser et al. (2006) studied the same for the Danish language. Later, Houser (2010) and Wei and hui Audrey Li (2016) studied the syntactic and semantic challenges of *do so* construct for the English and Mandarin language respectively. In dialogue systems, the problem of ellipsis has been addressed by the use of a supervised discriminative machine learning model (Kenyon-Dean et al., 2016) and joint modelling with coreference and question-answering data (Aralikatte et al., 2021). Liu et al. (2016) explored the decomposition of verb phrase ellipsis resolution into computational subtasks. Itegulov and Lebedeva (2018) experimented with identifying dependent type events for verb phrase anaphora resolution. Marasović et al. (2017) used an LSTM-Siamese Net mention-ranking model to learn abstract anaphora resolution or discourse deixis.

On the data front, datasets like OntoNotes (Pradhan et al., 2012), WikiCoref (Ghaddar and Langlais, 2016) *etc.* are the common datasets used for benchmarking the nominal-antecedent anaphora and coreference resolution models. Other datasets specifically addressing the verb phrase anaphora cases are AARAU Corpus (Poesio et al., 2018), CODI-CRAC 2021 Shared Task corpus (Khosla et al., 2021) and VP ellipsis corpus (Bos and Spenader, 2011). Though few instances of *do so* constructs are available in VPE corpus (Bos and Spenader, 2011), we could not use it for our experiment as our downstream QA system deals with a specific pattern of *do so* construct.

As the VPA constructs are consistently coming to the forefront of Dialogue systems and QA systems as a challenge, it motivated us to explore the specific case of *do so* VPA construct and create a dataset to manage our specific needs.

4 Experiment

4.1 Dataset

During the process of solving the language generation problem of a QA system, it has been noticed that the SOTA entity resolution system (Joshi et al., 2019b) is not able to resolve VPA cases. Since the SOTA model is trained on the OntoNotes (Pradhan et al., 2012) dataset, we explored its dataset and guidelines. The coreference guidelines of OntoNotes clearly state that verb is to be marked as a single-word span only if it is coreferenced with an existing noun phrase. With this guideline, it can be inferred that the deep learning model trained on this dataset will not be able to resolve the *do so* type of verb anaphoras as the dataset is not annotated to address these cases of VPA.

For our task, we constructed a targeted dataset of 350 data points with surface variations of *do so* VPA constructs, *viz. doing so, does so, did so*. The dataset contains both kinds of cases where the scope of the antecedent is either inter-sentential or intra-sentential. The data points of *do so* VPA constructs are collated from two sources: our QA system which we are automating and BNC corpus[1]. An equal number of data points were collated to balance the data for generic VPA resolution and coverage.

Two annotators helped us annotate the antecedent span for each VPA in a standoff annotation format. The antecedent span is annotated as clusters with word index based on subword tokenizer (Joshi et al., 2019b) output and stored in dictionary format.

A kappa score of 0.89 indicates a high Inter Annotator Agreement (IAA) for the antecedent spans. The detailed data stats are given in table 1. The evaluation and analysis in this paper are done on our dataset. The dataset is released as part of the contribution to further VPA research[2].

[1] BNC Consortium, The British National Corpus, XML Edition, 2007, Oxford Text Archive, http://hdl.handle.net/20.500.12024/2554.

[2] `https://github.com/Sandhya2207/VPA_dataset`

Data	MUC			B-Cube			CEAF-e			F1-avg
	R	P	F1	R	P	F1	R	P	F1	
Baseline	37.43	77.38	50.46	35.24	77.9	48.53	32.5	79.97	46.22	48.40
Our approach	68.94	66.07	**67.47**	74.67	71.91	**73.27**	80.79	78.2	**79.47**	**73.4**

Table 2: Evaluation Result of output from our approach. (R: recall, P: precision, F1: F1 score)

Input Stage	Process	Input	Output
SOTA model Output	Coreference Resolved using SOTA	Shawn turned on the killer machine to kill mosquitoes. Doing so, he says, narrows its prospects for survival.	Shawn turned on the killer machine to kill mosquitoes. Doing so, Shawn says, narrows mosquitoes’ prospects for survival.
Step-1	Identify the span of probable VPA	Shawn turned on the killer machine to kill mosquitoes. Doing so, Shawn says, narrows mosquitoes’ prospects for survival.	S1: Shawn turned on the killer machine to kill mosquitoes. S2: Doing so, Shawn says, narrows mosquitoes’ prospects for survival.
Step-2	Get the constituency parse string and consider only verb phrase	S1: Shawn turned on the killer machine to kill mosquitoes.	["turned on the killer machine to kill mosquitoes", " to kill mosquitoes", " kill mosquitoes"]
Step-3	Substitute verb anaphor with verb phrases to get candidate antecedents	["turned on the killer machine to kill mosquitoes", " to kill mosquitoes", " kill mosquitoes"]	["turned on the killer machine to kill mosquitoes, Shawn says, narrows mosquitoes prospects for survival. ", "to kill mosquitoes, Shawn says, narrows mosquitoes’ prospects for survival. ", "kill mosquitoes, Shawn says, narrows mosquitoes’ prospects for survival."]
Step-4	Get probability of correct sentence using GPT LM and pick the sentence with lowest perplexity	["turned on the killer machine to kill mosquitoes, Shawn says, narrows mosquitoes’ prospects for survival. ", "to kill mosquitoes, Shawn says, narrows mosquitoes prospects for survival. ", "kill mosquitoes, Shawn says, narrows mosquitoes’ prospects for survival."]	shawn turned on the killer machine to kill mosquitoes . turned on the killer machine to kill mosquitoes, shawn says, narrows mosquitoes’ prospects for survival .
Step-5	Correct the grammar for subject-verb agreement	shawn turned on the killer machine to kill mosquitoes . turned on the killer machine to kill mosquitoes, shawn says, narrows mosquitoes’ prospects for survival.	shawn turned on the killer machine to kill mosquitoes . turning on the killer machine to kill mosquitoes, shawn says, narrows mosquitoes’ prospects for survival .

Table 3: Input and Output of each step of our approach as shown in figure 1

4.2 Our Approach

The coreference module of our QA system works in stages. It begins with resolving all nominal antecedent cases using the SOTA coreference model (Joshi et al., 2019b). The nominal coreference clusters identified by the SOTA model are mapped to the input sentence to get noun coreference mapped output text. This text forms the input to our VPA module.

Figure 1 shows the flow of our proposed pipeline approach. In *step 1*, the nominal coreference mapped input text is checked for the presence of *do so* and all its lexical forms, *viz. doing so, does so, did so.* If the lexical text matches *do so* construct, it is a candidate for VPA resolution and is further checked for intra/inter sentential case based on the location of *do so* construct. For inter-sentential cases, the sentence preceding the sentence having the *do so* phrase is considered as the scope of its VPA antecedent. And, for the intra-sentential case, part of the sentence preceding the *do so* phrase is considered as the scope of its VPA antecedent.

In *step 2*, a constituency parse tree of the VPA antecedent text from *step 1* is generated using both Stanford CoreNLP parser (Klein et al., 2003; Manning et al., 2014) and constituency parser with ELMo embeddings (Joshi et al., 2018) for improved coverage.

From the generated parse tree, all the verb phrases are extracted as possible antecedent candidates and mapped in place of *do so* phrase to get anaphora resolved candidates as in *step 3*. This step leads to multiple candidates for identifying the correct antecedent for VPA.

At *step 4*, we get the probability of each candidate sentence using the generative pre-trained transformer (GPT) language model (Radford et al., 2018). The intuition here is that a syntactically correct candidate sentence will have higher probability as compared to incorrect candidate sentence. Using sentence probability we calculate the perplexity of all the candidate sentences. The candidate sentence with the lowest perplexity is considered as the antecedent resolved VPA output sentence.

Since subject-verb agreement is required for the correctness of the sentence in VPA, a pre-trained RoBERTa grammar correction model (Omelianchuk et al., 2020) is used to get subject-verb agreement in antecedent mapped text in *step 5*. Table 3 shows the module output after each step. Since the output of our system is to be consumed by machines, the naturality of the sentence was less of a concern.

S. No.	Input	Baseline Output	Our Approach Output	Reference Output	Remarks
1	Never put your money in a sinking company . In plainer terms , failure to do so leads to loss .	Never put your money in a sinking company . In plainer terms , failure to do so leads to loss .	Never put your money in a sinking company. In plainer terms , failure to put your money in a sinking company leads to loss .	Never put your money in a sinking company. In plainer terms, putting your money in a sinking company leads to loss .	Meaning changed for negative sentence.
2	A dolphin that watches a model place a ball in a basket might place the ball in the basket when asked to mimic t he behavior, but it may do so in a different manner.	A dolphin that watches a model place a ball in a basket might place a ball in a basket when asked to mimic the behavior, but A dolphin that watches a model place a ball in a basket may do so in a different manner.	A dolphin that watches a model place a ball in a basket might place a ball in a basket when asked to mimic the behavior, but A dolphin that watches a model place a ball in a basket may place a ball in a basket when asked to mimic the behavior , but A dolphin that watches a model place a ball in a basket may in a different manner.	A dolphin that watches a model place a ball in a basket might place the ball in the basket when asked to mimic the behavior, but dolphin may mimic the behavior in a different manner.	In case of multiple verb phrases as antecedent, not able to pick the accurate verb phrase. But the Verb phrase is available in top 3 choices.

Table 4: Error analysis of the output from our approach

5 Result

The standard evaluation metrics used for anaphora resolution is link based MUC score (Vilain et al., 1995), mention based B^3 score (Bagga and Baldwin, 1998) and optimal mapping based CEAF-E score (Luo, 2005). We evaluated our result using the standard CoNLL 2012 metric (Pradhan et al., 2012) which is calculated as an average of MUC, B-cube and CEAF metrics. Table 2 shows the precision, recall and F1 score for each metric and their average score. The baseline used for comparison is the output from the state-of-the-art BERT e2e-coreference model (Joshi et al., 2019b).

The high recall value for MUC, B-cube and CEAF metrics indicates that our approach is able to identify the antecedent and its span with higher accuracy over the baseline model. And, the overall average F1 score is showing an improvement of 25.0 value over the baseline. Table 3 shows the input and output of each step in our module.

5.1 Qualitative Analysis

Table 4 shows the error analysis of some output cases. It shows that our approach is not able to manage the following sentence formats.

- Negative sentences are not semantically correct after VPA mapping as in row 1 of the table.
- In case of multiple verb phrases in antecedents, our approach is not able to identify the boundary of prospective antecedent as in row 2 of the table.

The innovative sentence constructs at the intra-sentence level, the cataphor constructs and relative clause constructs are still an open problem to be addressed.

6 Conclusion and Future work

This paper presents a computational heuristic approach to resolve the *do so* verb phrase anaphora. The approach uses a constituency parser to get all the syntactic components of the text. From the syntactic components, all the verb phrases from preceding text is substituted in place of the verb anaphor to generate the candidate sentences. A pre-trained language model is used to select the most probable antecedent.

The result shows that our approach can identify the antecedent and its span with good accuracy on the VPA dataset developed for this experiment. The dataset used will be shared for further research. In the future, we plan to resolve the span identification issue in the intra-sentential case of *do so* construct where no conjunct is used as found in our error analysis. we also plan to investigate if our approach can be extended to other verb phrase anaphora constructs in the English language.

References

Rahul Aralikatte, Matthew Lamm, Daniel Hardt, and Anders Søgaard. 2021. Ellipsis resolution as question answering: An evaluation. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 810–817.

Amit Bagga and Breck Baldwin. 1998. Algorithms for scoring coreference chains. In *The first international conference on language resources and evaluation workshop on linguistics coreference*, volume 1, pages 563–566. Citeseer.

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

Kevin Clark and Christopher D. Manning. 2016. Deep reinforcement learning for mention-ranking coreference models. In *Empirical Methods on Natural Language Processing*.

Abbas Ghaddar and Philippe Langlais. 2016. Wikicoref: An english coreference-annotated corpus of wikipedia articles. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 136–142.

Jorge Hankamer and Ivan Sag. 1976. Deep and surface anaphora. *Linguistic inquiry*, 7(3):391–428.

Michael Houser. 2010. The syntax and semantics of do so anaphora.

Michael J Houser, Line Mikkelsen, Maziar Toosarvandani, Erin Brainbridge, and Brian Agbayani. 2006. Verb phrase pronominalization in danish: Deep or surface anaphora? *To appear in the Proceedings of WECOL*, 2007.

Daniyar Itegulov and Ekaterina Lebedeva. 2018. Handling verb phrase anaphora with dependent types and events. In *International Workshop on Logic, Language, Information, and Computation*, pages 210–222. Springer.

Mandar Joshi, Omer Levy, Daniel S. Weld, and Luke Zettlemoyer. 2019a. BERT for coreference resolution: Baselines and analysis. In *Empirical Methods in Natural Language Processing (EMNLP)*.

Mandar Joshi, Omer Levy, Daniel S Weld, and Luke Zettlemoyer. 2019b. Bert for coreference resolution: Baselines and analysis. *arXiv preprint arXiv:1908.09091*.

Vidur Joshi, Matthew Peters, and Mark Hopkins. 2018. Extending a parser to distant domains using a few dozen partially annotated examples. In *Proceedings of the Association for Computational Linguistics*, pages 1190–1199, Melbourne, Australia. Association for Computational Linguistics.

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*, pages 1734–1743.

Sopan Khosla, Juntao Yu, Ramesh Manuvinakurike, Vincent Ng, Massimo Poesio, Michael Strube, and Carolyn Rosé. 2021. The codi-crac 2021 shared task on anaphora, bridging, and discourse deixis in dialogue. In *Proceedings of the CODI-CRAC 2021 Shared Task on Anaphora, Bridging, and Discourse Deixis in Dialogue*, pages 1–15.

Dan Klein, Christopher D Manning, et al. 2003. Fast exact inference with a factored model for natural language parsing. *Advances in neural information processing systems*, pages 3–10.

Varada Kolhatkar, Adam Roussel, Stefanie Dipper, and Heike Zinsmeister. 2018. Anaphora with non-nominal antecedents in computational linguistics: A survey. *Computational Linguistics*, 44(3):547–612.

Kenton Lee, Luheng He, Mike Lewis, and Luke Zettlemoyer. 2017. End-to-end neural coreference resolution. *EMNLP*.

Zhengzhong Liu, Edgar Gonzalez, and Dan Gillick. 2016. Exploring the steps of verb phrase ellipsis.

Xiaoqiang Luo. 2005. On coreference resolution performance metrics. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 25–32.

Christopher D Manning, Mihai Surdeanu, John Bauer, Jenny Rose Finkel, Steven Bethard, and David McClosky. 2014. The stanford corenlp natural language processing toolkit. In *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations*, pages 55–60.

Ana Marasović, Leo Born, Juri Opitz, and Anette Frank. 2017. A mention-ranking model for abstract anaphora resolution. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 221–232, Copenhagen, Denmark. Association for Computational Linguistics.

Ruslan Mitkov. 2002. *Anaphora resolution*. Pearson Education.

Vincent Ng. 2010. Supervised noun phrase coreference research: The first fifteen years. In *Proceedings of the 48th annual meeting of the association for computational linguistics*, pages 1396–1411.

Kostiantyn Omelianchuk, Vitaliy Atrasevych, Artem Chernodub, and Oleksandr Skurzhanskyi. 2020. GECToR – grammatical error correction: Tag, not rewrite. In *Proceedings of the Fifteenth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 163–170, Seattle, WA, USA â†' Online. Association for Computational Linguistics.

Massimo Poesio, Yulia Grishina, Varada Kolhatkar, Nafise Sadat Moosavi, Ina Roesiger, Adam Roussel, Fabian Simonjetz, Alexandra Uma, Olga Uryupina, Juntao Yu, et al. 2018. Anaphora resolution with the arrau corpus. In *Proceedings of the First Workshop on Computational Models of Reference, Anaphora and Coreference*, pages 11–22.

Sameer Pradhan, Alessandro Moschitti, Nianwen Xue, Olga Uryupina, and Yuchen Zhang. 2012. Conll-2012 shared task: Modeling multilingual unrestricted coreference in ontonotes. In *Joint Conference on EMNLP and CoNLL-Shared Task*, pages 1–40.

Hub Prüst, Remko Scha, and Martin Van Den Berg. 1994. Discourse grammar and verb phrase anaphora. *Linguistics and Philosophy*, 17(3):261–327.

Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. 2018. Improving language understanding by generative pre-training.

Imran Q Sayed. 2003. Issues in anaphora resolution. *Stanford*.

Marc Vilain, John D Burger, John Aberdeen, Dennis Connolly, and Lynette Hirschman. 1995. A model-theoretic coreference scoring scheme. In *Sixth Message Understanding Conference (MUC-6): Proceedings of a Conference Held in Columbia, Maryland, November 6-8, 1995*.

Ting-Chi Wei and Yen hui Audrey Li. 2016. How to do so in mandarin chinese. *Journal of East Asian Linguistics*, 25(2):183–212.