

Finding the Right One and Resolving it

by

Payal Khullar, Arghya Bhattacharya, Manish Shrivastava

in

The SIGNLL Conference on Computational Natural Language Learning (CoNLL)

Report No: IIIT/TR/2020/-1



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

Finding the Right *One* and Resolving it

Payal Khullar Arghya Bhattacharya Manish Shrivastava

Language Technologies Research Centre
International Institute of Information Technology Hyderabad
Gachibowli, Hyderabad, Telangana-500032
{payal.khullar@research., arghya.b@research.,
m.shrivastava@}iiit.ac.in

Abstract

One-anaphora has figured prominently in theoretical linguistic literature, but computational linguistics research on the phenomenon is sparse. Not only that, the long standing linguistic controversy between the determinative and the nominal anaphoric element *one* has propagated in the limited body of computational work on one-anaphora resolution, making this task harder than it is. In the present paper, we resolve this by drawing from an adequate linguistic analysis of the word *one* in different syntactic environments - once again highlighting the significance of linguistic theory in Natural Language Processing (NLP) tasks. We prepare an annotated corpus marking *actual* instances of one-anaphora with their textual antecedents, and use the annotations to experiment with state-of-the-art neural models for one-anaphora resolution. Apart from presenting a strong neural baseline for this task, we contribute a gold-standard corpus, which is, to the best of our knowledge, the biggest resource on one-anaphora till date.

1 Introduction

One-anaphora is an anaphoric relation between a non-lexical proform (i.e. *one* or *ones*) and the head noun or the nominal group inside a noun phrase (NP). Consider the example sentence in (1) from [The British National Corpus \(2001\)](#), where the word *one* can be easily understood as *room*, from the preceding context.

1. The furniture in the lower **room**, which in every respect corresponds to the upper *one*, consists of one chair, of most antique and unsafe appearance.¹

The context from where the anaphor gets its sense and/or reference from is called the antecedent. For

¹Following the typographical conventions for one-anaphora and their antecedents by [Gardiner \(2003\)](#), we denote an antecedent noun phrase **like this** and one-anaphora *like this*.

one-anaphora, the antecedent can be a single word (head noun of the antecedent NP), as in (1), or group of nominal words – a compound noun or a head noun with its dependent, as in (2). However, the antecedent of one anaphora is never the whole NP.

2. There was much competition during the war as to who could come up with the best **bomb story**, and my mother had a great time telling this *one* to all the aunts...

One-anaphora can represent a particular case of identity-of-sense anaphora where the anaphor shares only the sense of the antecedent and not the complete reference. This category of one-anaphora is named as sense sharing one-anaphors as opposed to "contrastive anaphors" presented in (1) ([Luperfoy, 1991](#)), where the "lower" room that the anaphor refers to is in contrast with the "upper" room as antecedent. Such an interpretation can also be vague in some cases, as in (2), where the bomb story that the mother is telling might in fact be the best bomb story in the competition, but not necessarily. In other cases, it is possible that the entity that the anaphor one refers to is a subset of the entities the antecedent denotes, such as in (3), where the black car that Jack liked is actually one amongst the many cars that he saw.

3. Of all the **cars** Jack saw, he liked the black *one* the most.

This category of one-anaphora is discussed by some linguists as "member anaphora" for representative sampling ([Luperfoy, 1991](#)), or "nominal substitutes" that stand in for a meaningful head ([Halliday and Hasan, 1976](#)). The antecedent can also be the head noun with its propositional argument, such as in (4), where the anaphor resolves as *point of agreement*.

4. Even so, there are possible **points of agreement** — if not in principle, then at least in

practice. The most obvious *one* is commercial animal agriculture in its dominant form.

Sometimes, the antecedent boundary selection decision is vague, even for human evaluators. For instance, the antecedent in (5) can be *presentation on global warming* or just *presentation*, depending on the context.

5. My **presentation** on global warming was the longest *one* in the conference.

However, for a sentence like (6), there is little ambiguity that the antecedent is only the head noun *book* without its prepositional argument.

6. This **book** with yellow cover is the best *one* in the library.

It will be absurd for the anaphor to be interpreted as *book with yellow cover*, although a sloppy reading such as this is also possible.

2 Previous Work

The nominal anaphoric element *one* is extensively discussed in theoretical linguistics as one-anaphora, noun anaphora, one-insertion, one-substitution and pronominalization (Menzel, 2017, 2014; Kayne, 2015; Hankamer and Sag, 2015; Payne et al., 2013; Corver and van Koppen, 2011; Gunther, 2011; Culicover and Jackendoff, 2005; Akhtar et al., 2004; Cowper, 1992; Luperfoy, 1991; Dalrymple et al., 1991; Dahl, 1985; Radford, 1981; Baker, 1978; Halliday and Hasan, 1976; Bresnan, 1971). In computational linguistics literature, however, it has largely been ignored, despite the evident impact of one-anaphora resolution in improving the accuracy of downstream Natural Language Processing (NLP) tasks such as Machine Translation (MT) and Question Answering (QA).

To the best of our knowledge, the earliest computational approach to one-anaphora detection and resolution comes from Gardiner (2003), who presented several linguistically-motivated heuristics to distinguish one-anaphora from other non-anaphoric uses of *one* in English. For the resolution task, she used web search to select potential antecedent candidates. The second seminal work comes from Ng et al. (2005) that uses Gardiner's heuristics as features to train a Machine Learning (ML) model. The most recent work on one-anaphora comes from Recasens et al. (2016) where it has been treated as

one of the several sense anaphoric relations in English. The authors create sAnaNotes corpus where they annotate one third of the OntoNotes corpus for sense Anaphora. They use a Support Vector Machine (SVM) classifier - LIBLINEAR implementation (Fan et al., 2008) along with 31 lexical and syntactic features, to distinguish between the anaphoric and the non-anaphoric class. Trained and tested on one-third of the OntoNotes dataset annotated as the SAAnaNotes corpus, their system achieves 61.80% F1 score on the detection of all anaphoric relations, including one-anaphora. Their baseline statistical model outperforms the existing ML model for one-anaphora detection. This work, however, only limits itself to the detection part, deeming resolution of sense anaphora as a hard NLP task.

3 Getting to Know Every *One*

English has three distinct lexemes spelled as *one*—the regular third person indefinite pronoun, the indefinite cardinal numeral (determinative) and regular common count noun. There is no visible difference in their orthographic base form. However, they are totally different with respect to their morphological, syntactic, and semantic properties. On the surface, this difference can be observed in the way these forms inflect (morphology), behave in a sentence (syntax) and impart meaning (semantics) (Payne et al., 2013).

Previous efforts to classify the word *one* in English involve classification based on different functions of the word in discourse— numeric, partitive, anaphoric, generic, idiomatic, and unclassifiable; and in terms of the type of antecedent the anaphoric *one* takes— a kind, a set, an individual instance (Payne et al., 2013; Gardiner, 2003; Luperfoy, 1991; Dahl, 1985). This scheme has been extended for classification of other sense anaphoric relations as well (Recasens et al., 2013). This distinction clubs closely related types like numeric and partitive (both are determinative, roughly mean "1") to different classes. It also treats the regular count noun anaphora and determinative anaphora together as the anaphoric class. This makes the previous research miss important underlying linguistic generalisations in these forms. In syntactic literature, one-anaphora refers to an anaphoric instance of the word *one*, where its syntactic properties resemble that of a count noun (Payne et al., 2013; HuddlestonRodnry and Pullum, 2005). Like an En-

One in English		Identifying Features	Examples
Pronoun	Regular, third-person, indefinite pronoun.	Refers to an arbitrary person. As with pronouns, no plural form.	<u>One</u> must respect his elders.
Determiner	Indefinite cardinal numeral (Most common usage). Means '1'.	<ol style="list-style-type: none"> 1. When used with a head noun, it is obligatory and non-anaphoric. 2. Partitive function. It means, one entity in a set of many. Often, followed by the preposition "of". 3. When used as a noun modifier, it means 'sole', and is ommissible. 4. When used without a noun, it acts as a noun ellipsis licenser. Anaphoric to whole NP. 	<p>I will have <u>one</u> glass of water.</p> <p><u>One</u> of the keys is missing.</p> <p>You are the <u>one</u> reason I am here.</p> <p>I have two pens and my friend has <u>one</u>.</p>
Noun	Regular, common count noun.	<ol style="list-style-type: none"> 1. Means roughly 'instance thereof' - refers back to some class or type in discourse or salient in context. Anaphoric to the head noun, with or without a dependent, but never to the whole NP. Has both singular and plural forms (One Anaphora). 2. Derivative, non-anaphoric. Has both singular and plural forms. 	<p>The fictitious example being used here isn't the easiest <u>one</u> to give to an informant, but many much more difficult <u>ones</u> have been explained.</p> <p>Always take care of your loved <u>ones</u>.</p>

Table 1: The three uses of the word one in English.

glish noun, it has four inflected forms – singular (one), plural (ones), genitive singular (one's) and genitive plural (ones'). In its singular form, it can occur after a singular demonstrative determiner, a determiner followed by an adjective. It can not occur solely with an indefinite article, but a construction where an indefinite article is followed by an adjective is acceptable. With the definite article, it occurs when followed by a relative clause (Kayne, 2015).

Interestingly, this count noun instance of *one* looks very similar to the anaphoric subtype of the determinative instance of *one* on the surface. However, a close linguistic investigation clarifies that they have completely different morphological, syntactic and semantic properties (Payne et al., 2013). More importantly, they are different with respect to the kind of antecedent they take. While anaphoric noun takes noun heads as antecedents, the determinative *one* takes the whole NP. Consider the following example that Gardiner (2003) takes from

Luperfey (1991) as an instance of one-anaphora.

7. All the officers wore hats so Joe wore one too.

The problem here is that the occurrence such as in (7) is not an anaphoric noun; it is the determinative anaphor. Note that the plural form of this element is *some*, and not *ones*. Further, the constituent whose repetition this *one* word avoids is not *hats*, but the entire NP *a hat*. In ellipsis theory, this determinative *one* word here is not one-anaphora, but the licenser or trigger of an elided noun. Detection and resolution of this determinative one anaphor has actually been carried out in a part of our previous computational research on ellipsis (Khullar et al., 2020, 2019)

Right from Baker (1978), the traditional linguistic literature on one-anaphora and noun ellipsis too has confused between the noun and determiner uses of the word *one*, using them interchangeably in discussions and analysis. The faulty understanding on this phenomenon in earlier syntactic discourse,

No.	POS String Template	Example Sentences from BNC
1.	Determiner – Adjective – “one”	Her idea of the value of art criticism was <u>a simple one</u> .
2.	Determiner – (Adverb)+ – Adjective – “one”	The need for volunteers from churches, particularly in London and Scotland in the day-time, is <u>an ever constant one</u> .
3.	Determiner – “one” – Preposition	The only room available is <u>this one on</u> Friday the ninth.
4.	Determiner – “one” – Gerund/Participle	The songs contain upto eight themes, <u>each one consisting</u> of repeated phrases.
5.	Determiner – “one” –(Punct) Complementizer	Freeman wrote another clause, wrote <u>another one which</u> meant that you had to go.

Table 2: Template for fetching one-anaphora from POS tagged data.

unfortunately, propagated into the limited body of computational work on one-anaphora, and made this task harder than it really is. In the current paper, we aim to bridge this gap by drawing from a thorough linguistic investigation of anaphoric instances of the word *one* in recent linguistic studies, where clear differences between these two forms of the word have been discussed (Payne et al., 2013). Note that although Kayne (2015) prefers to give all instances of the word *one* a homogeneous internal structure, comprising a classifier merged with an indefinitive article through a variety of examples, he too identifies subtypes within this class and points out how they behave differently than one another.

The crux of the discussion on different types of *ones* in English in this section is summarised in Table 1, listing details of the classification scheme—in terms of how the word *one* behaves morphologically, syntactically and semantically in a sentence, along with identifying features and sentence examples for each type². Using this wisdom, we extend the computational research on the phenomenon.

4 Corpus Creation

In this section, we explain our efforts to build a one-anaphora corpus that contains actual instances of one-anaphora and is sizeable enough for training supervised machine learning models. We make this process easier by using linguistic theory on syntactic environment of one-anaphora. To begin with, since one-anaphora is a count noun, we select all plural *ones* as plurality is a feature of count nouns. For the singular form, we identify five POS

string sequences that capture the syntactic distribution of one-anaphora in English. The basic idea is that one-anaphora, being a regular count noun, will always occur inside of an NP. In other words, it will be preceded by a determiner or noun modifier like category and could be followed by a relative clause. All the syntactically possible combinations for one-anaphora to exist are presented in Table 1.

For our annotation purpose, we use The British National Corpus that contains over one hundred million words of British English, drawn from written and spoken sources. The text comes from a variety of sources like books, periodicals, media, letters, conversations and monologues. The text also has part of speech tags assigned by the CLAWS part-of-speech tagger (The British National Corpus, 2001). To fetch potential one-anaphora, we perform a semi-automatic search using the POS string templates discussed above. To calculate accuracy of our POS string templates, we check their output on 5000 randomly selected sentences containing the word *ones* or *ones* . Our templates retrieve 153 positive sentences. We manually check all the 5000 sentences and do not find any one-anaphora instance missed by the templates. However, of the 153 results, 18 are incorrect (false positives). Hence, we get a full recall, a precision of 88.24 and F1 score of 93.75. Although the precision is slightly low and the high F1 score is mainly contributed from the prediction of 4,847 negative instances correctly, these results show that the templates are good enough to fetch a variety of one-anaphora candidates that can be followed by manual confirmation. This is also much less expensive than previous entirely manual annotation efforts.

A simple search for the word *ones* and *ones*

²The table is an extended version of the one presented in Payne et al. (2013).

in the BNC yields 2,72,469 results. We run the templates on these sentences, which yields 15,647 unique matches. Of these, we manually check the first 1058 sentences only.³ We keep the true positive cases for the final corpus. From these 1058 sentences, we get 912 positive sentences containing 921 one-anaphora. For these 921 anaphors, we look for antecedents. Since the distance between the one-anaphor and antecedent is generally not that large (Gardiner, 2003), for finding and marking the antecedents, we only consider a context of up to three sentences, including the current sentence. If an antecedent is not present within this context or is not present at all endophorically, we leave the anaphor without its resolution marked. This decision speeds up the annotation effort.

4.1 Annotation Format

We use a stand off annotation scheme that does not modify the original text. The format of the annotation is as follows:

ANA sentence ID start index end index

ANT sentence ID start index end index

Here, ANA is short for anaphor and ANT for antecedent. Sentence ID is the unique ID given to a sentence in the BNC. We mark the boundaries with word offsets of the anaphor and antecedent in a given sentence. The simplicity of the format and stand-off annotation scheme make these annotations easy to understand and reuse.

4.2 Inter-annotator agreement

Annotation is carried out manually. Three annotators who are linguists by training and proficient in the language perform the task independently on all the sentences. For each sentence, the first annotation decision involves checking if the marked one-anaphora is correct or not. In the second step, the annotators mark antecedents for sentences they mark as correct in the first step. We calculate the inter-annotator agreement for both these steps separately. We use the Fleiss's Kappa coefficient to calculate the inter-annotator agreement between multiple annotators. For the first task, we get Fleiss's Kappa coefficient of 0.89 and for the second task, we get 0.81. These numbers confirm reliability of our annotations. Most of the disagreements occur in distinguishing between derivative

³We do not manually check all of these as it would be very arduous and expensive.

non-anaphoric and exophoric one-anaphora for the first task and boundary selection decision for the second task. All the disagreements are finally resolved at the end of the task by discussion among the three annotators and the agreed-upon cases are included in the final corpus.

4.3 Corpus Summary

In this section, we present a summary of major statistical observations of our annotated corpus along with a brief discussion.

- In the 100-million-word BNC, the word *one* occurs 2,61,093 and the word *ones* occurs 11,376. This makes their respective frequencies 0.26% and 0.01% in the corpus. Sentence wise, these frequencies are 3.97% and 0.18% respectively. From our templates, we fetch 15,647 matching sentences that contain 18,669 one-anaphora words (some sentences contain more than one one-anaphora words), both singular and plural (subject to precision error described previously). Roughly, this makes the sentence-wise frequency of one-anaphora 6.25% and word wise frequency 6.85%. We get a significantly lower frequency value as compared to that in the previous annotation efforts, which came out to be 15.2% (Ng et al., 2005) and 12.3% (Recasens et al., 2016). This is expected as most of the one-anaphora cases marked in these papers are not one-anaphoric nouns, but determinative anaphora.
- We note an interesting observation about the location of the anaphor and antecedent in the text. About 92% of the fetched one-anaphora instances come from the first and second templates alone, see Figure 1 for reference. Both these template require the anaphor to be preceded by one or more adjectives. This means that one-anaphora is most frequently followed by adjectives. This observation is in line with the analysis of one-anaphora as NP-ellipsis with adjectival remnants (Corver and van Koppen, 2011).
- In the annotated part of our corpus, we get a total of 921 one-anaphora in 912 sentences. Of these, the antecedents of 895 anaphors is present endophorically (i.e. in the text) within a context window of 3 sentences. For the remaining 26 anaphors, either the antecedent is

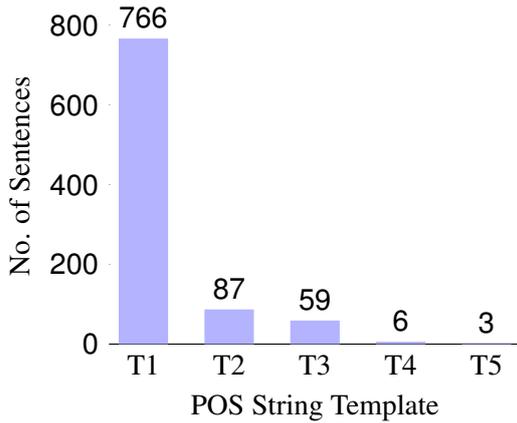


Figure 1: Number of sentences containing one-anaphora, fetched using five different templates, labelled T1 to T5.

not present in the text at all (exophoric cases), is present but not in the considered context window (and ignored for practical reasons), or the annotators are not able to agree on a single decision with certainty. This means that in our corpus, a majority one-anaphora are endophoric and, thus, can be resolved.

- We also note that a majority of the antecedents in our corpus comprise a single word only. Only 31 antecedents of out 895 are more than one word long. This implies that one-anaphora most often resolves to just the head noun of the antecedent NP. This is an important observation as antecedent boundary selection is presumably as a hard NLP task. As discussed previously, even human annotators find it difficult to make this decision in some cases. Hence, as far as one-anaphora is considered, resolving it to just the head noun of the antecedent NP is a simple and practical choice for NLP tasks.
- Finally, over 90% of the antecedents are present in the same sentence as one-anaphora, about 7% in the first previous sentence and less than 2% in the second previous sentence. The antecedent can go beyond the second previous sentence too but we do not annotate it as discussed in the annotation scheme. Although antecedents can follow one-anaphora, we do not find any such cases in our annotated corpus. Since we consider only a small part of the actual number of occurrences in the BNC, it can be safely concluded that cataphoric in-

stances are rare or very less frequent. This is in line with the observation made by [Gardiner \(2003\)](#) that the antecedent is generally located closer to one-anaphora and lies frequently in the previous context. For computational work, both these observations can be employed as manual features to improve the search for the antecedents of one-anaphora.

5 One-anaphora Resolution

In this section, we describe a framework to resolve one-anaphora in free text. We break the complete task in two subtasks – the first being the detection of the anaphor and the second the selection of the antecedent candidate from its context. See [Figure 2](#) for an overview of the framework.

5.1 Detecting One-Anaphors

Detecting instances of *one* that are one-anaphora is not a trivial task as the word *one* occurs very frequently in text and most of the times, it is not one-anaphora.⁴ To begin with, we can test the efficacy of our POS string templates on real world data, which does not come with gold tags. To do this, we use the state-of-the-art spaCy parser ([Honnibal and Johnson, 2015](#)) to automatically tag sentences from our annotated dataset and then apply the template rules to filter out matching candidates. Apart from fetching wrong candidates or missing correct ones, this template system is now also subject to parser errors. Using gold annotations, we automatically check for recall and precision value. After application of the templates on the tagged sentences from spaCy, we get a precision of 78.34%, a recall of 85.92% and F1 score of 81.96%. We now turn to supervised machine learning models to see if they offer a more accurate and robust solution.

5.1.1 Task Description

The one-anaphora detection task can be modelled as a classification problem, where, given an instance of the words *one* or *ones*, the classifier has to predict whether it is one-anaphora or not. Formally, for a given anaphor candidate ana_i in the context c , the task of one-anaphora detection is represented as follows:

$$f(ana_i, c) \rightarrow \{0, 1\}$$

where 1 denotes that ana_i is a one-anaphor in c , and 0 otherwise.

⁴The most frequent tag assigned to the word *one* in BNC is cardinal numeral ([Gardiner, 2003](#)).

5.1.2 Training/Dev/Test Data

We take the 912 sentences containing 921 one-anaphora marked in our annotated dataset as our positive set. For the negative set, we take an equal number of sentences from BNC that contain instances of *one* other than one-anaphora. Hence, our data size becomes 1824 sentences. 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.

5.2 Selecting Antecedents

There is evidence that parallelism in discourse can be applied to resolve possible readings for anaphoric entities and reference phenomenon (Hobbs and Kehler, 1997). Linguistic research also shows structural similarities between antecedent and anaphoric clauses (Luperfoy, 1991; Halliday and Hasan, 1976). An antecedent selection procedure can possibly benefit from capturing this similarity.

5.2.1 Task Description

This subtask involves selecting the right antecedent for one-anaphora, if it can be resolved. Formally, in a given context c , for an instance of one-anaphora ana_i , and the antecedent candidate ant_j ; the task of antecedent selection can be defined as follows:

$$f(ant_j, ana_i, c) \rightarrow \{0, 1\}$$

where 1 denotes that the antecedent candidate ant_j is the actual resolution of the one-anaphora ana_i , and 0 otherwise. Thus, for a given input sentence, the model can potentially select one or more antecedent candidates.

5.2.2 Training/Dev/Test Data

For antecedents, we have 895 positive samples in the annotated corpus. For the negative samples, we take all noun words other than the antecedent from the positive sentences and undersample to deal with the resulting skewed class distribution. We only take noun words since the antecedent of one-anaphora can only be a noun (optionally with dependents). As in the previous step, 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.

5.2.3 Experiments

To get representations of the word and its context, we experiment with both static and contextual types of word embeddings. For the former, we choose state-of-the-art fastText (FT) embeddings (Bojanowski et al., 2016) as they are able to provide representations of rare words and non words that might be frequent in movie dialogues. For the latter, we use BERT (Bidirectional Encoder Representations from Transformers) base uncased word-piece model for English (Devlin et al., 2019) as it currently provides the most powerful word embeddings taking into account a large left and right context. For the first subtask, we take word embeddings for the one-anaphora candidate and its context; and for the second subtask, we take word embeddings for the antecedent candidate, the gold one-anaphora vector from the annotations and their context. This way, we are able to evaluate the performance of both the subtasks separately. For fastText, we use pretrained embeddings and sumpool the embeddings of the given word and its context to obtain a single vector that we employ for training our classifiers. For both the subtasks, we experiment with a simple Multilayer Perceptron (MLP) and bidirectional Long Short Term Memory (bi-LSTM) networks. In MLP, we have 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 the 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. In case of BERT, we fine tune the pretrained BERT model. We separate the sentence and the candidate words 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. Attention mask tells the model to not focus on [PAD] tokens. The [CLS] output of the BERT model is used for classification.

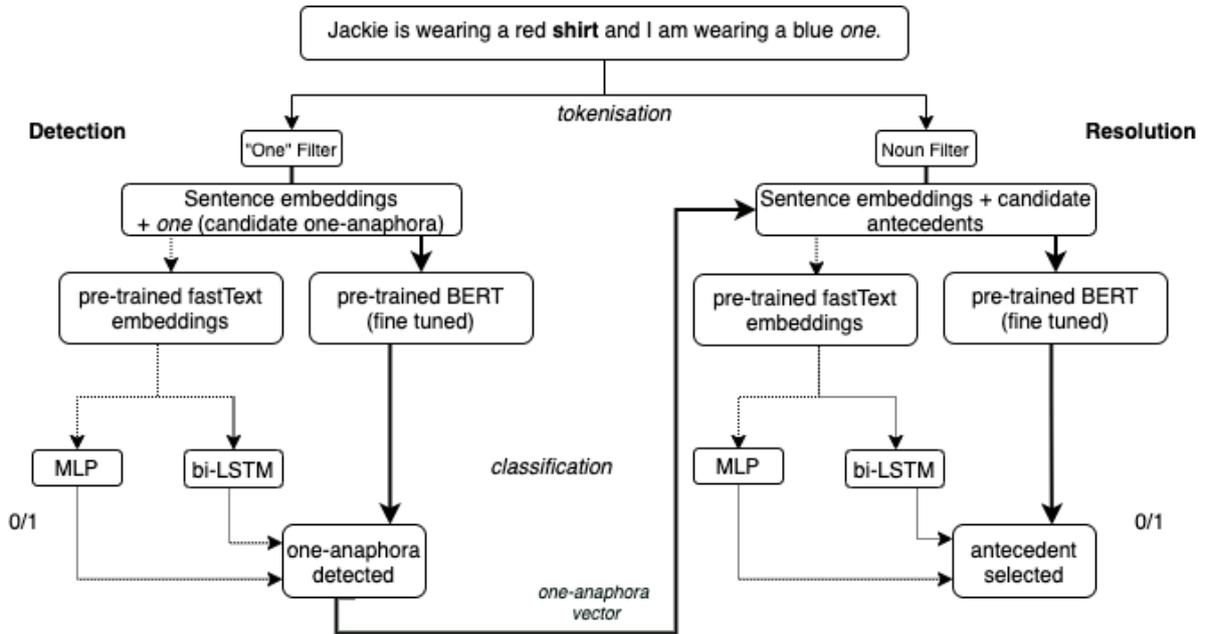


Figure 2: One anaphora detection and resolution pipeline. Classification decision is first taken on a given word one as anaphoric or non-anaphoric. The detected anaphoric one vector is passed on to the second model where the resolution decision is taken. Dotted lines represent model choices that were not included in the final pipeline.

Mathematically;

$$P(y|x) = \text{softmax}(W \cdot x + b)$$

where x denotes the input vector and y denotes the one-anaphora or antecedent for the first and second subtasks respectively. 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, 2015) for coding these models.

5.2.4 Results and Discussion

We evaluate the performance of all our detection models in terms of F1-score, computed by taking an average F1-scores obtained from the 5-folds results. The precision, recall and F1-score values of all the experiments for one-anaphora detection and antecedent selection are presented in Table 3. The majority of errors come from failing to detect actual anaphors, wrongly identifying non-anaphoric words and correct anaphor detection but failed antecedent selection. We also treat the result as incorrect when the system gives multiple antecedents for the same one-anaphora (as currently there is no way the system can make a decision in such a case).

Our experiments show that, the pre-trained fine tuned BERT model renders robust and high scores for both the subtasks. This is expected as BERT has been previously shown to give promising scores on a number of classification tasks. In our task, the model is robust and efficiently makes generalisations on the syntactic and semantic dependency between the one-anaphora with the determiners and adjectival modifiers in its context, as well as between antecedents and the anaphor. The results with the pre-trained fastText embeddings with a simple MLP are also promising. The sufficient neurons in the hidden layer with sigmoidal function ensures network approximate the nonlinear relationships between the input and output. Even though FFNNs are not designed to capture long range dependencies in a sentence that are inevitably required for handling a discourse device like one-anaphora, they can perform exceedingly well when they are infused with the contextual knowledge that they lack (Dumpala et al., 2018). This makes them suitable to resolve one-anaphora efficiently from low resource datasets like the one we use to train. This knowledge comes from the pre-trained embeddings.

We finally integrate the neural network models for each subtask into an end-to-end pipeline, see Figure 2 for an overview. Now, instead of the gold vectors, the resolution model is fed the one-

Task	P	R	F
One-Anaphora Detection			
FT, MLP	65.49	79.35	71.76
FT, Bi-LSTM	64.22	71.74	67.77
BERT (fine-tuned)	78.87	89.35	83.78
Antecedent Selection			
FT, MLP	55.24	61.29	58.11
FT, Bi-LSTM	58.97	65.74	62.17
BERT (fine-tuned)	63.07	72.33	67.38
Final Model			
See Figure 2.	59.99	70.01	64.61

Table 3: Precision (P), Recall R and F1-Score (F) values of different models for one-anaphora detection and antecedent selection tasks. Values in bold depict best performance. The two subtasks are finally integrated into a final model.

anaphora vectors from the detection model. This obviously results into error propagation into the second model, and lowers the precision value to 59.99, recall to 70.01 and consequently, the F1-score to 64.61 of the final system. Although, we achieve promising results on both the subtasks separately as well as in the pipeline process, the results can be further improved with hyperparameter tuning, additional regularization and manual feature addition.

6 Conclusion

In this paper, we used the most recent linguistic understanding of the word *one* in English to define and classify the one-anaphora phenomenon for computational linguistics research. We built a big corpus containing actual instances of one-anaphora by hand annotating sentences from BNC and used the annotations to experiment with state-of-the-art neural models for one-anaphora detection and resolution. For word and context representation, we experimented with pre-trained fastText and BERT word embeddings. We achieve promising results on a task that was deemed hard in previous NLP work, highlighting the importance of linguistic theory in NLP research. The gold standard corpus prepared for this task, containing 921 instances of one-anaphora marked in an easy-to-reuse standoff annotation scheme, will be released with this paper for future work.

References

- Nameera Akhtar, Maureen Callanan, Geoffrey K Pullum, and Barbara C Scholz. 2004. Learning antecedents for anaphoric one. *Cognition*, 4:141–145.
- Carl Lee Baker. 1978. Introduction to generative transformational syntax. *Englewood Cliffs, NJ:: Prentice-Hal*.
- Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2016. Enriching word vectors with subword information. *arXiv preprint arXiv:1607.04606*.
- Joan Bresnan. 1971. A note on the notion “identity of sense anaphora”. *Linguistic Inquiry*, 2:589–597.
- Francois Chollet. 2015. Keras. <https://keras.io>.
- Norbert Corver and Marjo van Koppen. 2011. Np-ellipsis with adjectival remnants: a micro-comparative perspective. *Natural Language & Linguistic Theory*, 29(2):371–421.
- Elizabeth A. Cowper. 1992. A concise introduction to syntactic theory. *Chicago, IL: University of Chicago Press*.
- Peter W Culicover and Ray Jackendoff. 2005. Simpler syntax. *Oxford, England: Oxford University Press*.
- Deborah Anna Dahl. 1985. The structure and function of one-anaphora in english. *Ph.D. thesis, University of Minnesota*.
- Mary Dalrymple, Stuart M. Shieber, and Fernando C.N. 1991. Ellipsis and higher order unification. *Linguistics and Philosophy*, 14:399–452.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *NAACL-HLT*.
- Sri Harsha Dumpala, Rupayan Chakraborty, and Sunil Kumar Kopparapu. 2018. Knowledge-driven feed-forward neural network for audio affective content analysis. *Workshops at the Thirty-Second AAAI Conference on Artificial Intelligence*.
- Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. 2008. Liblinear: A library for large linear classification. volume 9, page 1871–1874. *Journal of Machine Learning Research*.
- Mary Gardiner. 2003. Identifying and resolving one-anaphora. Department of Computing, Division of ICS, Macquarie University.
- 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. *Longman London*, page 76.

- Jorge Hankamer and Ivan Sag. 2015. Deep and surface anaphora. *Linguistic Inquiry*, 7:391–428.
- 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.
- Rodnry HuddlestonRodnry and Geqffry Pullum. 2005. [The cambridge grammar of the english language](#). *Zeitschrift für Anglistik und Amerikanistik*, 53.
- Richard S Kayne. 2015. English one and ones as complex determiners. *New York University*.
- 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*.
- Susann Luperfoy. 1991. Discourse pegs: A computational analysis of context-dependent referring expressions. *Ph.D. thesis, University of Texas at Austin*.
- Katrin Menzel. 2014. A corpus linguistic study of ellipsis as a cohesive device¹. *Proceedings of Corpus Linguistics*.
- Katrin Menzel. 2017. *Understanding English-German contrasts: a corpus-based comparative analysis of ellipses as cohesive devices*. Ph.D. thesis, Universität des Saarlandes, Saarbrücken.
- Hwee Tou Ng, Yu Zhou, Rober Dale, and Mary Gardiner. 2005. A machine learning approach to identification and resolution of one-anaphora. pages 1105–1110.
- John Payne, Geoffrey K. Pullum, Barbara C. Scholz, and Eva Berlage. 2013. Anaphoric one and its implications. *Language*, 4:794–829.
- Andrew Radford. 1981. *Transformational syntax: A student's guide to chomsky's extended standard theory*. Cambridge, UK: Cambridge University Press.
- Marta Recasens, Cristian Danescu-Niculescu-Mizil, and Dan Jurafsky. 2013. [Linguistic models for analyzing and detecting biased language](#). In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1650–1659, Sofia, Bulgaria. Association for Computational Linguistics.
- Marta Recasens, Zhichao Hu, and Olivia Rhinehart. 2016. Sense anaphoric pronouns: Am i one? page 1–6. *Proceedings of the Workshop on Coreference Resolution Beyond OntoNotes (CORBON 2016)*.
- The British National Corpus. 2001. *Oxford University Computing Services on behalf of the BNC Consortium*, (2).