

Ellipsis Detection

PROJECT REPORT

Team Name: LingoTribe

Team Advisor: Prof. Dipti M. Sharma

Team Members: Harshita Sharma[20171099], Souvik Banerjee[20171094]

Natural Language Processing - Monsoon, 2019

International Institute of Information Technology, Hyderabad

Contents

Introduction to Ellipsis	2
Ellipsis Detection	3
Problem Description	5
Dataset preparation	5
NPE Detection	5
Output Analysis	6
Error Analysis	7
References	7
PAPER PRESENTATION	8

1. Introduction to Ellipsis

1.1 What is ellipsis?

Omission from a clause of one or more words that are nevertheless understood in the context of the remaining elements. Examples:

- a. John can play the guitar; Mary can _{play the guitar} too.
- b. I heard Mary's dog, and you heard Bill's _{dog}.

1.1.1 Types of ellipsis

There are numerous types of ellipses identified in theoretical syntax. Some of which are mentioned below:

1. **Noun-Phrase Ellipsis:** Noun ellipsis (also N-ellipsis, N'-ellipsis, NP-ellipsis, NPE, ellipsis in the DP) occurs when the noun and potentially accompanying modifiers are omitted from a noun phrase.

Examples:

- a. Fred did three onerous tasks because Susan had done two _{onerous tasks}.
- b. The first train and the second _{train} has arrived.

2. **Verb-Phrase Ellipsis:** Verb phrase ellipsis (also VP-ellipsis or VPE) is a particularly frequent form of ellipsis in English. VP-ellipsis elides a non-finite VP. The ellipsis must be introduced by an auxiliary verb or by the particle *to*. An aspect of VP-ellipsis that is unlike gapping and stripping is that it can occur forwards or backwards.

Examples:

- a. The man who wanted to order the salmon did _{order the salmon}.
- b. The man who wanted to _{order the salmon} did order the salmon.

3. **Answer-Ellipsis:** Answer ellipsis involves question-answer pairs. The question focuses on an unknown piece of information, often using an interrogative word (e.g. *who*, *what*, *when*, etc.). The corresponding answer provides the missing information and in so doing, the redundant information that appeared in the question is elided

Examples:

- a. Q: Who has been hiding the truth? A: Billy has been hiding the truth
- b. Q: What have you been trying to accomplish? A: I have been trying to accomplish
This damn crossword.

2. Ellipsis Detection

Ellipsis being a linguistic phenomenon that has received considerable attention, mostly focusing on its interpretation. In English, an instance of NPE consists of two parts: a licensor, typically a determiner or pre-modifier, that indicates the presence of a NPE; and an antecedent, which is the noun phrase to which the elided element resolves. For example:

For example, in the sentence, “*The government includes money spent on residential renovation; Dodge does not*”. The licensor “*does*” resolves to the antecedent “*includes money spent on residential renovation*”.

2.1 Importance of ellipsis detection and resolution:

Ellipsis is very common in natural language and poses many problems when dealing with data in many NLP and other computational linguistic problems. Some of them include Machine Translation, and other problem statements like Word Problem Solving - which need to simplify sentences as part of

data preprocessing before dealing with the data in the true sense of the problem.

Here, we have collected some of the striking examples which show the need to identify and resolve ellipsis for Machine Translation:

- a. When Susan brings her dog, Sam brings his [dog] too.

Manual translation: Jab susan apna kutta laati hai, Sam bhi apna lata hai.

Google Translate output: jab susaan apane kutte ko laata hai, sam use bhee laata hai.

- b. John has one hat, and Bill has five [hats].

Manual translation: John ke paas ek Topee hai, aur Bill ke pass paanch hain.

Google Translate output: John ke paas ek Topee hai, aur Bill mein paanch hain.

- c. She won't laugh but he will [laugh].

Manual translation: Vah nahi hansegi par vah.

Google Translate output: Vah nahin hansaegge lekin vah.

- d. Some school kids like syntax, and some don't [like syntax].

Manual translation: Kuch school ke bachchon ko vakya rachana pasand hai, aur kuch ko nahi.

Google Translate output: Kuch skoolee bachchon ko vaaky rachana pasand hai, aur kuch nahin.

As seen from the examples above, it is not just machine translation but also manual translation that can go wrong because of the presence of the problems posed by ellipsis. Some of these observed problems are - inflection being fused in the word in IL as seen in example c. because of which the word 'will' itself gets no translation; and the wrong understanding of the

term ‘has’ by the machine in example b. i.e. instead of ‘possession’ it interprets it as ‘to contain something inside’ and translates it to ‘mein’.

3. Problem Description

The problem of NPE detection can be broken down into three parts:

- Dataset preparation
- Ellipsis detection: Trigger and the elided phrase detection
- Analysis of the output

Constraint: Sentence boundary - NPE detection and antecedent detection is done within the sentence boundary.

4. Dataset preparation

There is no explicit datasets for nominal ellipsis. Some of the resources used to get as many as possible sentences that show NPE:

- Linguistic Resources for Ellipsis: Wikipedia, RANLP III: Selected papers and other research papers
- ParCorFull: a Parallel Corpus Annotated with Full Coreference
- The Oxford Handbook of Ellipsis
- The Syntactic Licensing of Ellipsis

Finally, a mixed dataset was created with 120 sentences with around 65 cases of NPE.

5. NPE Detection

In case of NPE in English, the trigger words are often determiners and modifiers of the elided noun. These are also known as licensors of ellipsis. For example:

- a. I caught the first train before you caught the second.

- b. Jill likes your story even though she hates Bill's.

Here, the licensors of elided noun are the cardinal number *second* and possessive proper noun *Bill's* respectively.

An interesting feature about these license of NPE in English is that they can only belong to certain syntactic categories - cardinal numbers, ordinal numbers, noun/pronoun possessives, quantifiers like 'some' etc.

5.1 Rules

Considering the patterns seen in sentences where the NPE exists the following rules have been formulated:

Rule 1: Check for ordinal numbers, if ordinal numbers are followed by punctuation or prepositions or has no noun phrase in the next three words, then it is chosen as a licensor.

Rule 2: Previous rule is applied for cardinal numbers, demonstrative determiners, possessive pronouns and quantifiers.

Rule 3: Check for interrogative determiners , if they are followed by punctuation or the previous word is a noun modifier and next three words don't add up to a noun phrase, it is chosen.

Rule 4: Check for superlative adjectives, if they are preceded by a determiner and no noun phrase exists, it is chosen.

Rule 5: Check if the selected noun modifier is immediately followed by a verb or auxiliary verb as that would indicate the end of the given noun phrase immediately after the noun modifier.

5.2 Antecedent Identification

To resolve NPE, we look for antecedents - Match POS tags of the licensor with other noun modifiers. If a POS tag matching the licensor of the NPE is found in the sentence, the system outputs the noun that the modifier with the same tag modifies as the antecedent of the NPE. If there are more than one

such modifiers found, the system selects the one nearest to the NPE as generally has a role to play in anaphora and coreference resolution tasks.

6. Output Analysis

6.1 Accuracy

The output for all input sentences were observed and analysed manually to look at the various places where the formulated rules worked and where the system failed. Initially, accuracy rate was slightly more than half owing to the fact that we had no rule for a “one-anaphora” situation. Once that rule had been added, accuracy went up to around . This is due to the fact that a lot of our training set consists of one-anaphora sentences.

6.2 Analysis of the output

Output is of the form -

When Susan brings her dog, Sam brings his too.

Licensor: ['his', 8, 'DET', 'PRP\$']

Antecedent: [['dog', 4, 'NOUN', 'NN']]

where each tuple is of the form [word , word_index, word_POS, word_tag]

It was first failing for certain sentences with one-anaphora. After a rule was added, it worked for these as well.

6. Error Analysis

8.1 Error

In this section, we discuss the kind of errors that have occurred and the pattern seen in the sentences where ellipsis detection was seen to be incorrect after the analysis of the output, so as to group similar examples and identify the type of these sentences or NPE.

The only kind of error we have gotten is wrong antecedent due to matching of the POS tag of the licenser with some closer word which is not a noun modifier.

Example - Like this sentence -

Some school kids like syntax, and some don't.

Licenser: ['some', 7, 'DET', 'DT']

Antecedent: []

Here the licenser some in some don't is classified as DET hence it is not matching with the POS tag of the some in 'Some schools' which is a quantifier(CD).

8.2 Improvements

Modifying the antecedent finding code. Instead of just looking at the immediate POS tags which match, we have a window which also checks at a little distance as well. A one-anaphora rule was also added.

7. References

- a. TAKAHASHI D. Noun Phrase Ellipsis
- b. CORVER N., KOPPEN M.V. Let's Focus on noun Ellipsis
- c. KHULLAR P., ANTONY A., SHRIVASTAVA M. Using Syntax to Resolve NPE in English
- d. SAAB A. Nominal Ellipsis
- e. LOBECK A.C. Strong Agreement and Identification: Evidence from Ellipsis in English

- f. ALEXIADOU A., GENDEL K. NP Ellipsis without Focus movement/projections: the role of Classifiers

PAPER PRESENTATION:

Name: Verb phrase ellipsis detection using automatically parsed text

Link: <https://dl.acm.org/citation.cfm?id=1220512>