Resolving Noun Ellipsis in English

TASK DESCRIPTION

NOUN Ellipsis is a linguistic phenomenon whereby certain parts of a sentence are omitted or deleted, and have to be retrieved from discourse or real-world context. For example:

⟨ L3315 m2 Jordy ⟩ Do you have coffee?
⟨ L3316 m2 Daphne ⟩ In the kitchen
⟨ L3316 m2 Jordy ⟩ I will make [*NP* some [e]] for us.

- Ellipsis RESOLUTION involves:

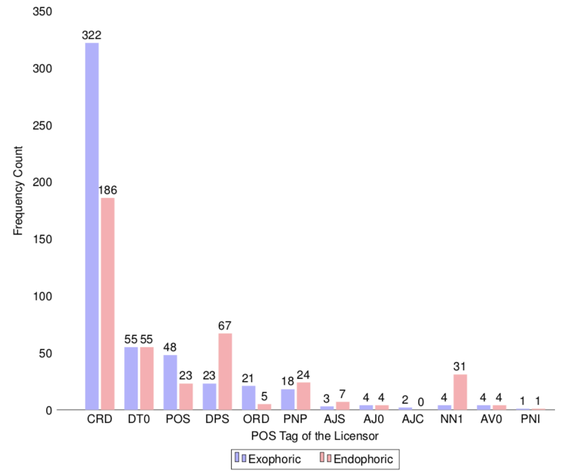
1. Detection of the ELIDED Noun: *Coffee*
2. Selection of the right ANTECEDENT Candidate: *Some*
3. Reconstruction: *I will make some coffee for us*.

ANNOTATED DATA CREATION

We annotate the first 100 movies of the Cornell Movie Dialogs dataset (16.08% of the total corpus).

Annotation guidelines described in Khullar et. al, 2020 (LREC: In Press)

- A total of 946 instances:
- 438 endophoric & 508 exophoric.
- Biggest Resource on NE
- Novel insights on distribution of licensors and antecedents.



EXPERIMENTS

• Rule Based BASELINE

1. Syntactic and semantic rules (Khullar et.al, 2019)
2. Tested on a manually curated dataset.
3. Errors mainly due to incorrect POS tags.

Task	Positive Samples	Negative Samples	Precision	Recall	F1-Score
Detection	76	132	69.15%	85.53%	76.47%
Resolution	65	29	78.79%	63.41%	70.27%

Table 2: Performance of NPE detection and resolution systems on the testset.

• Supervised ML BASELINE

1. Given [3-gram] predict whether the centre element is a licensor or not.
2. Given [licensor, antecedent candidate, context] if the antecedent candidate corresponds to an actual resolution or not.

• Neural Network pipeline
Work in Progress.

Task	Averaged Results ML Model	Precision	Recall	F1-Score
Noun Ellipsis Detection	Naive Bayes	0.6217	0.8376	0.7137
	Linear SVM	0.6407	0.8587	**0.7339**
	RBF SVM	0.6054	**0.9045**	0.7253
	Nearest Neighbors	**0.7369**	0.5949	0.6583
	Random Forest	0.1750	0.3500	0.2333
Noun Ellipsis Resolution	Naive Bayes	0.6096	0.6008	0.6052
	Linear SVM	**0.6213**	0.4258	0.5053
	RBF SVM	0.6007	0.9858	0.7465
	Nearest Neighbors	0.6061	0.3418	0.4371
	Random Forest	0.6000	**0.9989**	**0.7497**

Payal Khullar, PhD Computational Linguistics, ML & NLP Lab (LTRC, IIIT Hyderabad)