What is ‘it’? Disambiguating the different readings of the pronoun ‘it’

Sharid Loáiciga
Uppsala University

September 13th, 2017

Pronoun Translation

Machine translation problem caused by:

- Mismatch in pronoun systems: differences in overtness, gender, number, case, formality, animacy, etc.
- Functional ambiguity: pronouns with the same surface form but different function.
- Other factors involved in the processing of reference that we do not understand well yet.

English	Facit was a fantastic company. They were born deep in the Swedish forest, and they made the best mechanical calculators in the world.
German	Facit war ein großartiges Unternehmen. Entstanden tief im schwedischen Wald, bauten sie die besten mechanischen Rechenautomaten der Welt.
French	Facit était une entreprise fantastique, fondée dans la forêt suédoise. Elle fabriquait les meilleures calculatrices mécaniques au monde.

English	And among these organisms is a bacterium by the name of Deinococcus radiodurans. It is known to be able to withstand cold, dehydration, vacuum, acid, and, most notably, radiation.
German	Unter diesen Lebewesen existiert ein Bakterium namens Deinococcus radiodurans. Seine Resistenz gegen Kälte, Dehydration, Vakuum, Säuren ist bekannt sowie insbesondere gegen Strahlung.
French	Et parmi ces organismes, il y a une bactérie appelée Deinococcus radiodurans. Elle est connue pour être capable de supporter le froid, la déshydratation, le vide, l'acide et, le plus notable, les radiations.

We use this same word, depression, to describe how a kid feels when **it** rains on his birthday, and to describe how somebody feels the minute before they commit suicide.

A sense of belonging to the European Union will develop only gradually, as the EU achieves tangible results and explains more clearly what **it** is doing for people.

So in other words, I need to tell you everything I learned at medical school. But believe me, **it** isn't going to take very long.

Pleonastic	We use this same word, depression, to describe how a kid feels when **it** rains on his birthday, and to describe how somebody feels the minute before they commit suicide.
Nominal anaphora	A sense of belonging to the European Union will develop only gradually, as the EU achieves tangible results and explains more clearly what **it** is doing for people.
Eventual anaphora	So in other words, I need to tell you everything I learned at medical school. But believe me, **it** isn't going to take very long.

Pleonastic

We use this same word, depression, to describe how a kid feels when **it** rains on his birthday, and to describe how somebody feels the minute before they commit suicide. → il

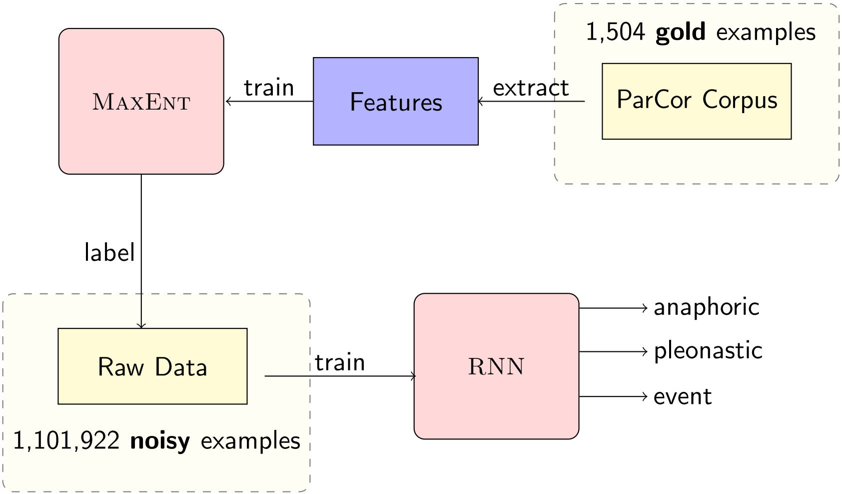
Nominal anaphora

A sense of belonging to the European Union will develop only gradually, as the EU achieves tangible results and explains more clearly what **it** is doing for people. → elle, il

Eventual anaphora

So in other words, I need to tell you everything I learned at medical school. But believe me, **it** isn't going to take very long. → cela, ça

Pipeline



Loáiciga, Guillou, and Hardmeier (2017)

Gold Data

The ParCor Corpus

Data set	Event	Anaphoric	Pleonastic	Total
Training	504	779	221	**1,504**
Dev	157	252	92	501
Test	169	270	62	501
Total	830	1,301	375	**2,506**

- ParCor is formed by TED Talks (transcribed planned speech) and EU Bookshop publications (written text).
- TED talks are particular with respect to pronoun use. Pronouns are frequent, including first and second person, but anaphoric references are not always clear.

Features

Pronoun head (hea)

1. Head word and its lemma, most of the time a verb.
2. Complement instead of head for copular verbs (*be*, *appear*, *seem*, *look*, *get*, etc).
3. Whether the head word takes a ‘that’ complement.
4. Tense of head word (verbs only).

Syntactic context (syn)

5. Whether a ‘that’ complement appears in the previous sentence.
6. Closest NP head to the left and to the right.
7. Presence or absence of extraposed sentential subjects as in ‘*So <u>it</u>’s difficult <u>to attack malaria</u> from inside malarious societies, [...]*.
8. Closest adjective to the right.

Semantic context (sem)

9 VerbNet selectional restrictions of the verb (*abstract*, *concrete* or *unknown*).

10 Likelihood of head word taking an event subject computed over the Annotated English Gigaword v.5 corpus. Two cases favouring are considered: i) when the subject is a gerund, and ii) ‘this’ pronoun subjects.

11 Non-referential probability assigned to the instance of ‘it’ by NADA.

Token context (tok)

12 Previous three tokens and next two tokens.

13 Lemmas of the next two tokens.

First System

- **Maximum Entropy - MaxEnt**
  - Trained on 1,504 gold examples
  - All features are included

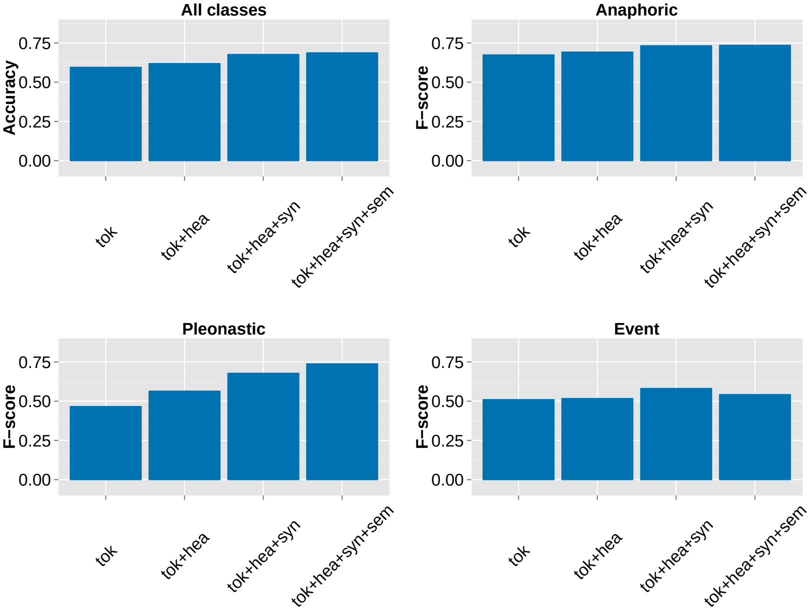
First System

- **Maximum Entropy - MaxEnt**
  - Trained on 1,504 gold examples
  - All features are included

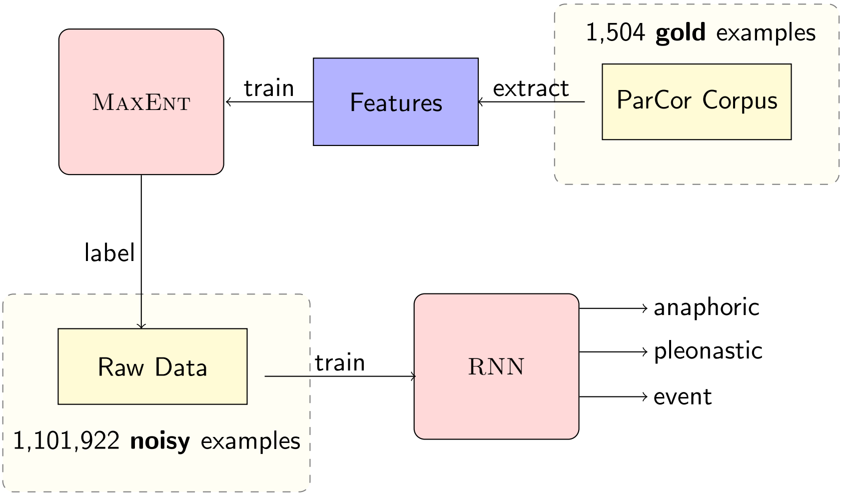
501 examples in test set

MC Baseline	**Precision**	**Recall**	**F1**	**Accuracy**
it-anaphoric	0.503	1	0.669	(270/501) 0.539
MaxEnt				
it-anaphoric	0.716	0.756	0.735	(344/501) 0.687
it-pleonastic	0.750	0.726	0.738	
it-event	**0.564**	**0.521**	**0.542**	

Ablation (MaxEnt)



Pipeline Reminder



Second System

- **Maximum Entropy - MaxEnt**
  - Trained on 1,504 gold examples
  - All features are included
- **Recurrent Neural Network - RNN**
  - Bidirectional RNN which reads the context words and then makes a decision based on the representations that it builds
  - Context window of size 50 to the left and right of the *it* to predict
  - Word-level embeddings and two GRU layers of size 90, features as one-hot vectors, softmax layer, *adam* optimizer, and categorical cross-entropy loss

Unannotated Data

	Sentences
News Commentary (from WMT)	344,805
Europarl v.7 (from WMT)	3,752,440
TED talks (from IWSLT)	380,072

Taken from the shared task on cross-lingual pronoun prediction

Systems

- **MaxEnt**

  Trained on 1,504 gold examples and all the features

- **RNNs**
  - **Gold** Trained on 1,504 gold examples
  - **Silver** Trained on 1,101,922 noisy examples annotated with the **MaxEnt** classifier
  - **Combined** Trained on gold + noisy examples

Results – 501 examples in test set

MC Baseline	**Precision**	**Recall**	**F1**	**Accuracy**
it-anaphoric	0.503	1	0.669	(270/501) 0.539
MaxEnt				
it-anaphoric	0.716	0.756	0.735	(344/501)
it-pleonastic	0.750	0.726	0.738	0.687
it-event	0.564	0.521	0.542	

Results – 501 examples in test set

MC Baseline	**Precision**	**Recall**	**F1**	**Accuracy**
it-anaphoric	0.503	1	0.669	(270/501) 0.539
MaxEnt				
it-anaphoric	0.716	0.756	0.735	(344/501)
it-pleonastic	0.750	0.726	0.738	0.687
it-event	0.564	0.521	0.542	
Rnn-Gold				
it-anaphoric	0.595	0.659	0.626	(250/501)
it-pleonastic	0.177	0.177	0.177	0.499
it-event	0.436	0.361	0.394	

Results – 501 examples in test set

MC Baseline	**Precision**	**Recall**	**F1**	**Accuracy**
it-anaphoric	0.503	1	0.669	(270/501) 0.539
MaxEnt				
it-anaphoric	0.716	0.756	0.735	(344/501)
it-pleonastic	0.750	0.726	0.738	0.687
it-event	0.564	0.521	0.542	
Rnn-Gold				
it-anaphoric	0.595	0.659	0.626	(250/501)
it-pleonastic	0.177	0.177	0.177	0.499
it-event	0.436	0.361	0.394	
Rnn-Silver				
it-anaphoric	0.706	0.552	0.620	(286/501)
it-pleonastic	0.542	0.516	0.529	0.571
it-event	0.455	0.621	0.525	

Results – 501 examples in test set

MC Baseline	**Precision**	**Recall**	**F1**	**Accuracy**
it-anaphoric	0.503	1	0.669	(270/501) 0.539
MaxEnt				
it-anaphoric	0.716	0.756	0.735	(344/501)
it-pleonastic	0.750	0.726	0.738	0.687
it-event	0.564	0.521	0.542	
Rnn-Gold				
it-anaphoric	0.595	0.659	0.626	(250/501)
it-pleonastic	0.177	0.177	0.177	0.499
it-event	0.436	0.361	0.394	
Rnn-Silver				
it-anaphoric	0.706	0.552	0.620	(286/501)
it-pleonastic	0.542	0.516	0.529	0.571
it-event	0.455	0.621	0.525	
Rnn-Combined				
it-anaphoric	0.794	0.530	0.636	(315/501)
it-pleonastic	0.582	0.742	0.652	0.629
it-event	0.520	0.746	0.613	

Results – 501 examples in test set

MC Baseline	**Precision**	**Recall**	**F1**	**Accuracy**
it-anaphoric	0.503	1	0.669	(270/501) 0.539
MaxEnt				
it-anaphoric	0.716	**0.756**	**0.735**	**(344/501)**
it-pleonastic	**0.750**	0.726	**0.738**	**0.687**
it-event	**0.564**	0.521	0.542	
Rnn-Gold				
it-anaphoric	0.595	0.659	0.626	(250/501)
it-pleonastic	0.177	0.177	0.177	0.499
it-event	0.436	0.361	0.394	
Rnn-Silver				
it-anaphoric	0.706	0.552	0.620	(286/501)
it-pleonastic	0.542	0.516	0.529	0.571
it-event	0.455	0.621	0.525	
Rnn-Combined				
it-anaphoric	**0.794**	0.530	0.636	(315/501)
it-pleonastic	0.582	**0.742**	0.652	0.629
it-event	0.520	**0.746**	**0.613**	

Error Analysis – 501 examples in test set

Is **MaxEnt** better or worse than **RNN-Combined**?

	Accuracy	
Reference relationship	MaxEnt	Rnn-Combined
(1) NP antecedent in previous 2 sentences	***(191/248)** **0.770**	(136/248) 0.548

e.g. The infectious disease that's killed more humans than any other is **malaria***.*
It's carried in the bites of infected mosquitos, and **it***'s probably our oldest scourge.*

Reference relationship	Accuracy	
	MaxEnt	Rnn-Combined
(1) NP antecedent in previous 2 sentences	***(191/248)** **0.770**	(136/248) 0.548
(2) VP antecedent in previous 2 sentences	(25/38) 0.658	**(27/38)** **0.711**

e.g. And there's hope in this next section, of this brain section of somebody else with M.S., because what it illustrates is, amazingly, the brain can **repair itself***. It just doesn't do* **it** *well enough.*

	Accuracy	
Reference relationship	MaxEnt	Rnn-Combined
(1) NP antecedent in previous 2 sentences	***(191/248)** **0.770**	(136/248) 0.548
(2) VP antecedent in previous 2 sentences	(25/38) 0.658	**(27/38)** **0.711**
(3) NP or VP antecedent not in snippet	(28/47) 0.596	(28/47) 0.596

e.g. It has spread. It has more ways to evade attack than we know. ***It****'s a shape-shifter, for one thing.*

Reference relationship	Accuracy MaxEnt	 Rnn-Combined
(1) NP antecedent in previous 2 sentences	***(191/248)** **0.770**	(136/248) 0.548
(2) VP antecedent in previous 2 sentences	(25/38) 0.658	**(27/38)** **0.711**
(3) NP or VP antecedent not in snippet	(28/47) 0.596	(28/47) 0.596
(4) Sentential or clausal antecedent	(52/88) 0.591	***(66/88)** **0.750**

e.g. **Pension systems have a hugely important economic and social role and are affected by a great variety of factors. It** *has been reflected in EU policy on pensions, which has become increasingly comprehensive over the years.*

	Accuracy	
Reference relationship	MaxEnt	Rnn-Combined
(1) NP antecedent in previous 2 sentences	***(191/248)** **0.770**	(136/248) 0.548
(2) VP antecedent in previous 2 sentences	(25/38) 0.658	**(27/38)** **0.711**
(3) NP or VP antecedent not in snippet	(28/47) 0.596	(28/47) 0.596
(4) Sentential or clausal antecedent	(52/88) 0.591	***(66/88)** **0.750**
(5) Pleonastic constructions	**(43/59)** **0.729**	(42/59) 0.728

e.g. And **it** *seemed to me that there were three levels of acceptance that needed to take place.*

Reference relationship	Accuracy MaxEnt	Accuracy Rnn-Combined
(1) NP antecedent in previous 2 sentences	***(191/248)** **0.770**	(136/248) 0.548
(2) VP antecedent in previous 2 sentences	(25/38) 0.658	**(27/38)** **0.711**
(3) NP or VP antecedent not in snippet	(28/47) 0.596	(28/47) 0.596
(4) Sentential or clausal antecedent	(52/88) 0.591	***(66/88)** **0.750**
(5) Pleonastic constructions	**(43/59)** **0.729**	(42/59) 0.728
(6) Ambiguous between event and anaphoric	(3/12) 0.250	**(7/12)** **0.583**

e.g. Today, multimedia is a desktop or living room experience, because the apparatus is so clunky . **It** *will change dramatically with small, bright, thin, high-resolution displays.*

Reference relationship	Accuracy MaxEnt	Rnn-Combined
(1) NP antecedent in previous 2 sentences	***(191/248) 0.770**	(136/248) 0.548
(2) VP antecedent in previous 2 sentences	(25/38) 0.658	**(27/38) 0.711**
(3) NP or VP antecedent not in snippet	(28/47) 0.596	(28/47) 0.596
(4) Sentential or clausal antecedent	(52/88) 0.591	***(66/88) 0.750**
(5) Pleonastic constructions	**(43/59) 0.729**	(42/59) 0.728
(6) Ambiguous between event and anaphoric	(3/12) 0.250	**(7/12) 0.583**
(7) Ambiguous between event and pleonastic	**(2/5)**	(1/5)

e.g. I did some research on how much it cost, and I just became a bit obsessed with transportation systems. And **it** *began the idea of an automated car.*

Reference relationship	Accuracy MaxEnt	Accuracy Rnn-Combined
(1) NP antecedent in previous 2 sentences	***(191/248) 0.770**	(136/248) 0.548
(2) VP antecedent in previous 2 sentences	(25/38) 0.658	**(27/38) 0.711**
(3) NP or VP antecedent not in snippet	(28/47) 0.596	(28/47) 0.596
(4) Sentential or clausal antecedent	(52/88) 0.591	***(66/88) 0.750**
(5) Pleonastic constructions	**(43/59) 0.729**	(42/59) 0.728
(6) Ambiguous between event and anaphoric	(3/12) 0.250	**(7/12) 0.583**
(7) Ambiguous between event and pleonastic	**(2/5) 0.400**	(1/5) 0.200

Putting pronoun function to work

Cross-lingual Pronoun Prediction – DiscoMT Shared Task

What is the task?

Source	If you ask for the happiness of the remembering self, **it**'s a completely different thing.
Target	si|KON vous|PRON réfléchir|VER sur|PRP le|DET bonheur|NOM du|PRP "|PUN moi|NOM du|PRP souvenir|NOM "|PUN ,|PUN **REPLACE_11** être|VER un|DET tout|PRON autre|ADJ histoire|NOM .|.
Classes	ce/c', ça/cela, elle, elles, il, ils, on, OTHER

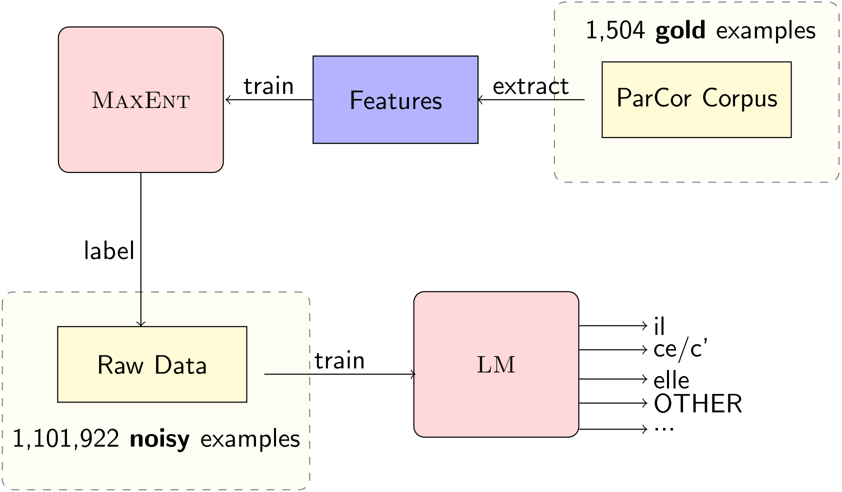
Advantages:

A set of possible translations (classes) is defined.

Controlled testing of different types of linguistic information.

Explicit anaphora or coreference resolution is not necessary.

Pipeline



Loáiciga, Guillou, and Hardmeier (2016)

Source Aware LM

Given Data

Source: Here's a jelly. It's one of my favorites, because it's got all sorts of working parts. It's got these fishing lures on the bottom.

Target: **REPLACE_0** avoir|VER ce|PRON leurre|NOM de|PRP pêche|NOM au-dessous|ADV .|.

Solution: *ils*

Source Aware LM

Training Data

Given: **REPLACE_0** avoir|VER ce|PRON leurre|NOM de|PRP pêche|NOM au-dessous|ADV .|.

–with source: **REPLACE**_*ils* avoir ce leurre de pêche au-dessous .

–with source & labels: It_*anaphoric* **REPLACE**_*ils* avoir ce leurre de pêche au-dessous .

Testing time: It **REPLACE** avoir ce leurre de pêche au-dessous .

Does it work?

Without *it*-labels

Macro-averaged Recall: **59.84%**

Pronoun	Precision	Recall
ce	**89.66**	76.47
elle	**40.00**	60.87
elles	27.27	12.00
il	**63.24**	70.49
ils	67.82	83.10
cela	76.47	41.94
on	36.36	44.44
OTHER	**88.37**	89.41

With *it*-labels

Macro-averaged Recall: **57.03%**

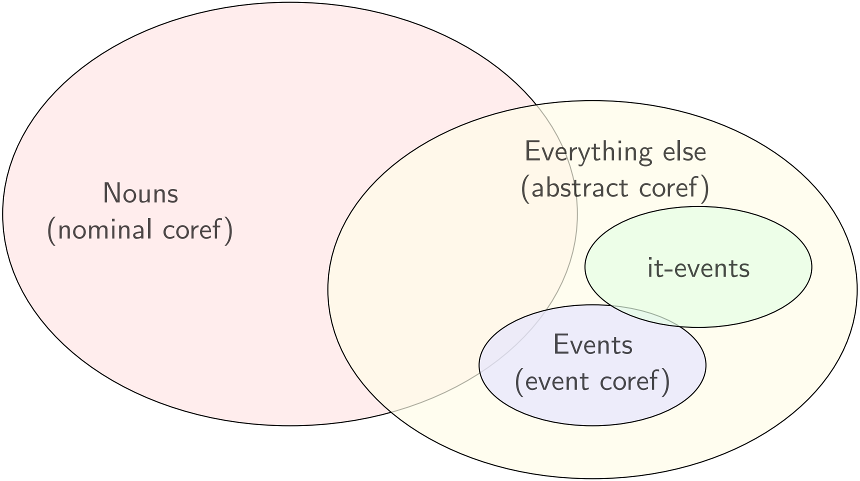
Pronoun	Precision	Recall
ce	89.09	72.06
elle	31.25	43.48
elles	**30.77**	16.00
il	54.43	70.49
ils	**69.41**	83.10
cela	**86.67**	41.94
on	**40.00**	44.44
OTHER	85.71	84.71

Conclusions

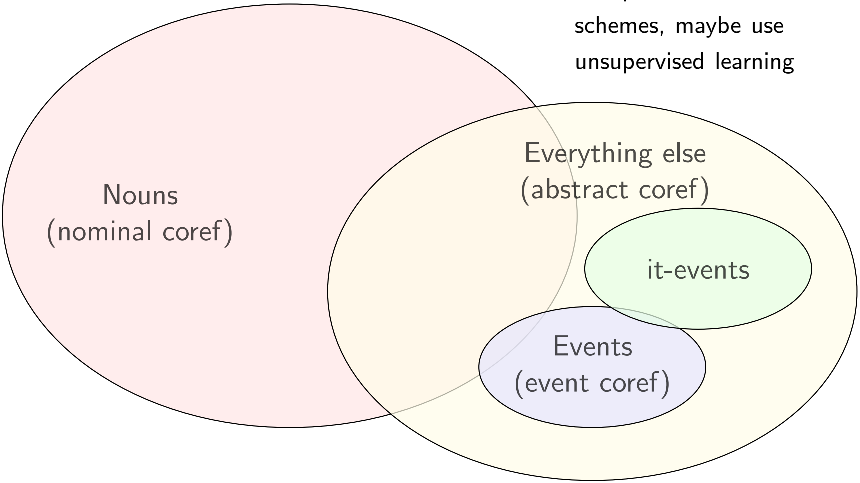
- Pronoun function is useful for the cross-lingual pronoun prediction task, indicating that it can also help the machine translation task.
- Results of training with noisy data can be improved with relatively small amounts of gold data.
- RNNs models do not seem good at identifying nominal reference but they seem rather good at identifying event reference.

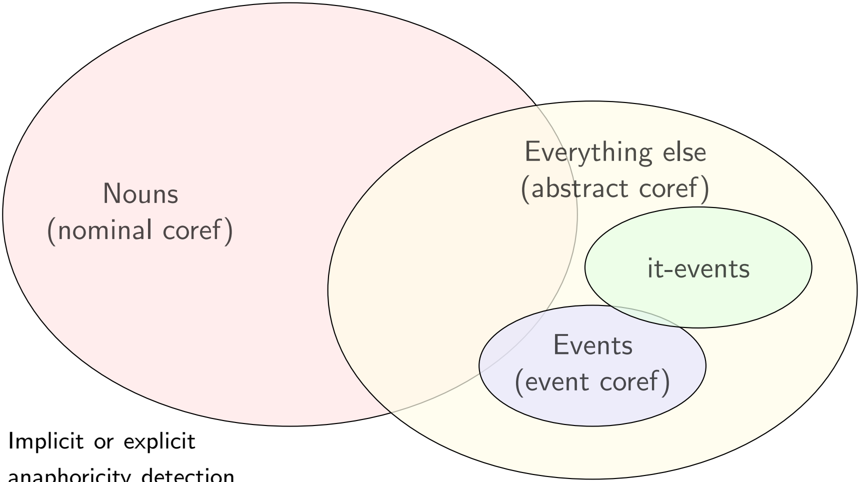
What's next?

- Improve the results - work on semi-supervised techniques
  - Re-label data for one more cycle of training with weighted version of the systems, in the spirit of Jiang, Carenini, and Ng (2016).
  - Use shared task parallel data as auxiliary training task in a multi-task learning scheme.
- Understanding of event reference and abstract reference

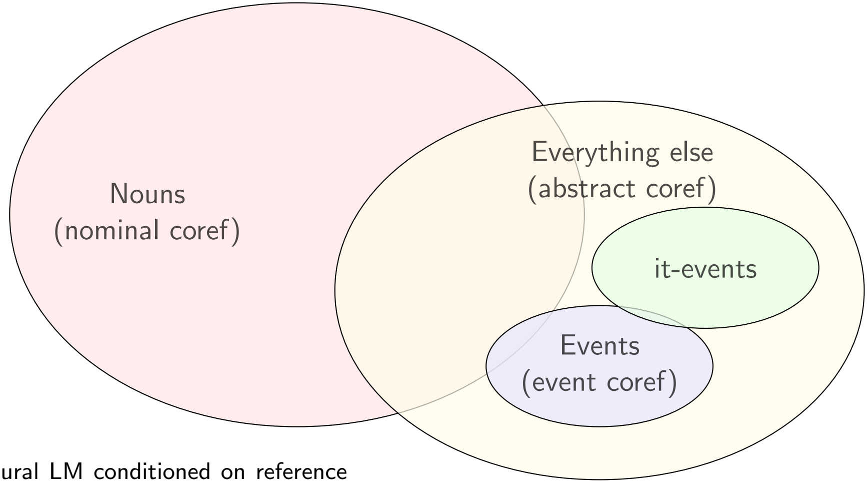

Nouns
(nominal coref)
Everything else
(abstract coref)
it-events
Events
(event coref)

Compare annotation
schemes, maybe use
unsupervised learning





Implicit or explicit anaphoricity detection for nominal coreference resolution



Neural LM conditioned on reference

(Ji et al. 2017; Yang et al. 2017)

Useful for Coref, LMs (dialogue), MT

Thank you!

References

Ji, Yangfeng et al. (2017). "Dynamic Entity Representations in Neural Language Models". In: *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. EMNLP 2017. Copenhagen, Denmark: Association for Computational Linguistics, pp. 1831–1840.

Jiang, Kailang, Giuseppe Carenini, and Raymond Ng (2016). "Training Data Enrichment for Infrequent Discourse Relations". In: *Proceedings the 26th International Conference on Computational Linguistics: Technical Papers*. COLING 2016. Osaka, Japan: The COLING 2016 Organizing Committee, pp. 2603–2614.

Loáiciga, Sharid, Liane Guillou, and Christian Hardmeier (2016). "It-disambiguation and source-aware language models for cross-lingual pronoun prediction". In: *Proceedings of the First Conference on Machine Translation*. WMT16. Berlin, Germany: Association for Computational Linguistics, pp. 581–588.

— (2017). "What is it? Disambiguating the different readings of the pronoun 'it'". In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. EMNLP 2017. Copenhagen, Denmark: Association for Computational Linguistics, pp. 1336–1342.

Yang, Zichao et al. (2017). "Reference-Aware Language Models". In: *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. EMNLP 2017. Copenhagen, Denmark: Association for Computational Linguistics, pp. 1851–1860.