POS Driven Cross-Lingual Pronoun Prediction with Feed-Forward Neural Networks

UPPSALA UNIVERSITET

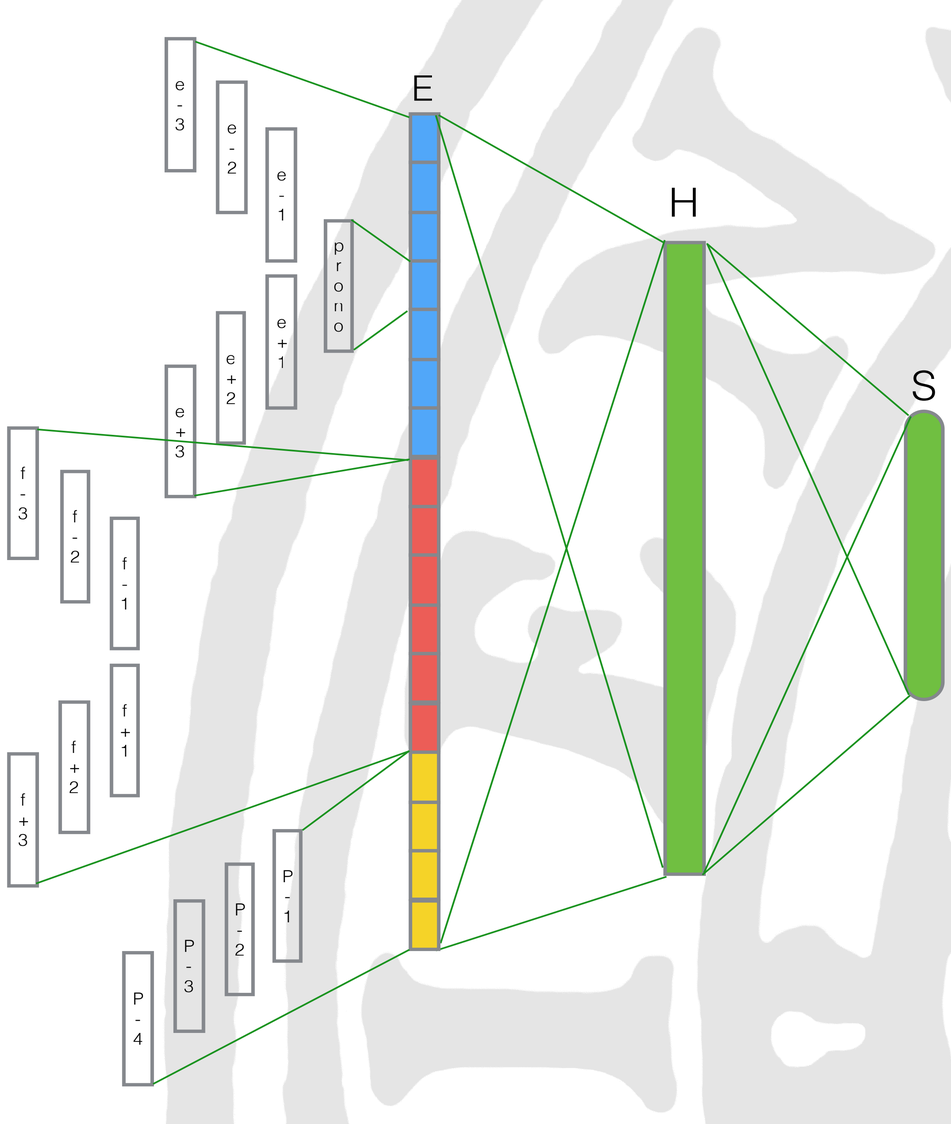
Jimmy Callin, Christian Hardmeier, Jörg Tiedemann

Introduction

- For some language pairs, pronoun translation is a discourse-driven task which requires information that lies beyond its local context.
  - The monkey ate the banana because *it* was hungry.
  - The monkey ate the banana because *it* was ripe.
  - The monkey ate the banana because *it* was tea-time.
- Inspired by (Hardmeier et. al. 2013b), we suggest a neural network-based model using preceding nouns and determiners as feature classes for suggesting antecedent candidates.
- Our model scores on par with similar models while having a simpler architecture.
- Theano implementation openly available at `https://github.com/jimmycallin/whatelles`.

Architecture

- Features extracted from source context and translation context around the missing pronoun, by encoding a number of word embeddings n words to the left and m words to the right (n+m).
- No external anaphora system.
- We finds the four closest nouns and determiners in English, and use the corresponding French nouns and articles as features.
- Additionally, we look at the French context of the missing pronoun.
- Trained using stochastic gradient descent with mini-batches and L2 regularization. Cross-entropy is used as a cost function, with a softmax output layer and embedding sizes of 50.



Context features

<S> <S> <S> it expresses our view of how we…

<S> <S> <S> __ exprime notre manière d' aborder …

POS tag features

We have this **banner** in our **off ces** in **Palo Alto**

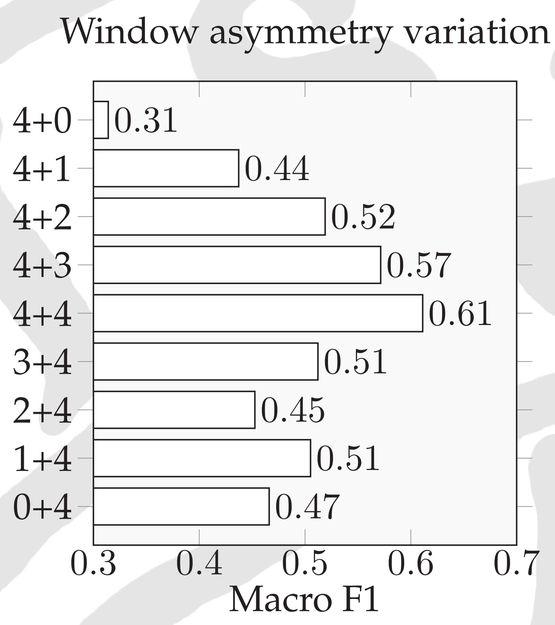
Nous avons cette **bannière** dans nos **bureaux** à **Palo** Alto

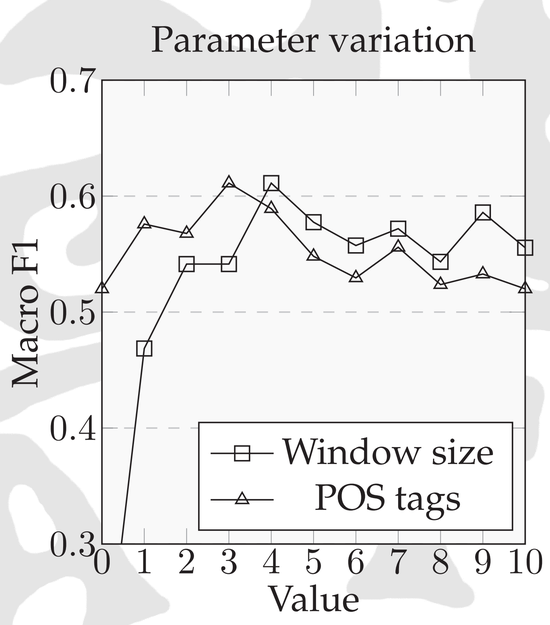
Finding optimal hyperparameters

Feature ablation

	POS	Source	Target	None
ce	**0.9236**	0.8629	*0.6405*	0.8822
cela	0.6179	**0.6324**	*0.4156*	0.6260
elle	0.2963	0.3019	*0.0930*	**0.3571**
elles	**0.2500**	0.2069	*0.1667*	0.2222
il	0.5366	0.4426	*0.3651*	**0.5620**
ils	0.8364	0.8345	*0.7050*	**0.8754**
OTHER	**0.8976**	0.8769	*0.6969*	0.8847
Macro	0.5526	0.5128	*0.3569*	**0.6299**
Micro	0.7871	0.7510	*0.5797*	**0.8019**

Window size & POS tags





Results

	ce	cela	elle	elles	il	ils	other	sum
ce	165	3	0	1	8	1	6	184
cela	5	80	4	1	21	0	18	129
elle	7	10	22	2	**22**	2	18	83
elles	0	0	0	18	0	**31**	3	51
il	11	7	9	0	64	1	12	104
ils	1	0	0	5	0	149	5	160
other	10	12	9	1	9	15	338	394
sum	199	112	44	27	124	199	400	

	Precision	Recall	F1
ce	0.8291	0.8967	0.8616
cela	0.7143	0.6202	0.6639
elle	0.5000	0.2651	0.3465
elles	0.6296	0.3333	0.4359
il	0.5161	0.6154	0.5614
ils	0.7487	0.9312	0.8301
other	0.8450	0.8579	0.8514
Macro	0.5816	0.5495	0.5530
Micro	0.7213	0.7213	0.7213

- Results indicate that the model performs on par with similar models, while being easier to train.
- There are some expected drops in performance for the less common classes heavily dependent on finding their antecedent.
- A symmetric window size is beneficial, while we are not as sure of why this is the case.
- Right context seems to be more important than left context, possibly due to the fact that pronouns in their role as subjects largely appears early in sentences.
- There is potentially too little data for creating robust embeddings, except for in a few reoccurring circumstances.