Neural End-to-End Coreference Resolution using Morphological Information

Tuğba Pamay-Arslan* Kutay Acar† Gülşen Eryiğit*

İTÜ NLP Research Group
Department of [AI&Data*, Computer†] Engineering
Faculty of Computer&Informatics
Istanbul Technical University
[*pamay, †acarku18, *gulsen.cebiroglu]@itu.edu.tr

2023

Introduction

Coreference Resolution (CR)

The task of determining coreferential relations between mentions referring to the same real-world entity in a document.

For Morphologically Rich Languages

- Words consist of morphemes containing deeper information.
- May necessitate the use of *morpheme-level representations* as well as word representations.

For Pro-Dropped Languages

- Dropped-pronouns can be deducible from morphology.

Introduction

Example[1]:

Sen [benim] [anne**[m]**in] geldiğ**[i]**ni gördü**n** mü?

~~Sen~~ ~~benim~~ **annemin** geldiğ**i**ni gördü**n** mü?

You my mother came see_did

Did you see the coming of my mother?

[1] Color codes are used to indicate mentions referring to the same entity.

The Proposed Model

- Baseline-based model
- End-to-end, neural, multilingual
- `Enhanced span representation:` Incorporating morphological information explicitly in contextual embeddings obtained by BERT.
- Morphological information: UPOS and/or Morphological features
- Inspired from *Pamay-Arslan and Eryiğit (2023)*
- Our Team: `TrCr`
- Our Submitted System: `morphbase`

The Proposed Model

BASE Span Embedding

- Embeddings of a mention's first token
- Embeddings of a mention's last token
- The head-attended embedding of a mention of all tokens

$$e(s_i) = e(s_{i_{first}}) \oplus e(s_{i_{last}}) \oplus e(s_{i_{head}}) \quad (1)$$

where s_i represents the i^{th} span, and $e(s_i)$ indicates the embedding of the related span.

ITU NLP

The Proposed Model

ENHANCED Span Embedding

- Extends the first and last tokens' embeddings by incorporating one/multi-hot encoded morphological information.

$$e(s_{i_{first}}) = e(s_{i_{first}}[form]) \oplus enc(s_{i_{first}}[upos]) \oplus enc(s_{i_{first}}[feats]) \quad (2)$$

$$e(s_{i_{last}}) = e(s_{i_{last}}[form]) \oplus enc(s_{i_{last}}[upos]) \oplus enc(s_{i_{last}}[feats]) \quad (3)$$

$$e_{enh}(s_i) = e_{enh}(s_{i_{first}}) \oplus e_{enh}(s_{i_{last}}) \oplus e(s_{i_{head}}) \quad (4)$$

Experimental Setup

Hyper-parameters

- BERT (Devlin et al. (2019)) : Multilingual, base, and case sensitive.
- Default hyper-parameters[a], except maximum segment length being 256 instead of 512.

[a] https://github.com/ondfa/coref-multiling/blob/master/experiments.conf

Encoded Morphological Information

Sparse Representation

- **UPOS**: One-hot encoding
  #Unique UPOS tags:20
- **Morphological Features**: Multi-hot encoding
  #Unique Feats:210

Example

For a token:
UPOS='NOUN'
FEATS='Case=Nom|Number=Plur', then
$\mathbf{upos}_{enc} = [00100000...]$
$\mathbf{feats}_{enc} = [0100100...]$

Embedded Morphological Information

- Embedding layers with the dimension of 5 are deployed for UPOS and features, separately.
- For *Morphological Features*, multiple features are averaged out to preserve dimensionality.

Results

System	CoNLL
BASELINE	58.99
+{U,F}$_{emb}$	60.75
+{U}$_{enc}$	61.27
+{U,F}$_{enc}$ (morphbase)	**61.35**

Table: The performances of the intermediate and the proposed models evaluated on the development sets (CoNLL score in %).

– U indicates the use of universal POS tags.
– F indicates the use of morphological features.

*All models exploiting morphological information surpass the performance of the baseline model by varying amounts.

Results

System	CoNLL
BASELINE	58.99
+{U,F}$_{emb}$	60.75
+{U}$_{enc}$	61.27
+{U,F}$_{enc}$ (morphbase)	**61.35**

- {U,F}$_{emb}$ vs. BASELINE: $\uparrow$ 1.76 percentage points.
- {U}$_{enc}$ vs. {U,F}$_{emb}$: $\uparrow$ 0.52 percentage points.
- {U,F}$_{enc}$ vs. BASELINE: $\uparrow$ 2.36 percentage points.
  The sparse encoding technique performs better in capturing sparse tag combinations.

Dev Set Results on all languages

System	AVG	ca_ancora	cs_pcedt	cs_pdt	de_parcor	de_potsdam	en_gum	en_parcor	es_ancora	fr_democrat	hu_szeged	lt_lcc	pl_pcc	ru_rcr	hu_korkor	no_bokmaal	no_nynorsk	tr_itcc
BASELINE	58.99	65.60	65.72	65.66	**57.25**	56.07	**66.87**	**56.56**	67.00	57.22	58.96	66.96	64.17	**63.04**	48.38	58.44	68.78	16.15
morphbase	**61.35**	**68.85**	**67.97**	**66.05**	50.10	**63.51**	65.42	44.85	**69.98**	**59.77**	**59.19**	**72.74**	**65.61**	62.93	**53.25**	**71.02**	**69.15**	**32.63**
Diff	↑ 2.36	↑ 3.25	↑ 2.25	↑ 0.39	↓ 7.15	↑ 7.44	↓ 1.45	↓ 11.71	↑ 2.98	↑ 2.55	↑ 0.23	↑ 5.78	↑ 1.44	↓ 0.11	↑ 4.87	↑ 12.58	↑ 0.37	↑ 16.48

Table: Dev set results for individual languages in the primary metric.

- Enhanced span representation achieves 61.35% CoNLL performance on average, which is higher than 2.36 percentage points on development set.
- **morphbase** improves the performance of the following morphologically rich languages: Catalan, Czech, Hungarian, Spanish, French, Lithuanian, Polish, Norwegian, and Turkish.

Test Set Results on all languages

System	**AVG**	ca_ancora	cs_pcedt	cs_pdt	de_parcor	de_potsdam	en_gum	en_parcor	es_ancora	fr_democrat	hu_szeged	lt_lcc	pl_pcc	ru_rcr	hu_korkor	no_bokmaal	no_nynorsk	tr_itcc
BASELINE	56.96	65.26	**67.72**	**65.22**	**44.11**	57.13	**63.08**	35.19	66.93	**55.31**	55.32	63.57	66.08	**69.03**	40.71	65.10	65.78	22.75
morphbase	**59.53**	**68.23**	64.89	64.74	39.96	**64.87**	62.80	**40.81**	**69.01**	53.18	**56.41**	**64.08**	**67.88**	68.53	**52.91**	**68.17**	**66.35**	**39.22**
Diff	↑ 2.57	↑ 2.97	↓ 2.83	↓ 0.48	↓ 4.15	↑ 7.74	↓ 0.28	↑ 5.62	↑ 2.08	↓ 2.13	↑ 1.09	↑ 0.51	↑ 1.8	↓ 0.5	↑ 12.2	↑ 3.07	↑ 0.57	↑ 16.47

Table: Test set results for individual languages in the primary metric.

- Enhanced span representation achieves 59.53% CoNLL performance on average, which is higher than 2.57 percentage points on test set.
- **morphbase** improves the performance of the following morphologically rich languages: Catalan, ~~Czech~~, Hungarian, Spanish, ~~French~~, Lithuanian, Polish, Norwegian, and Turkish.

Conclusion

- **morphbase** is ranked at 7th place in the shared task.
- On individual dataset scores, our (team `TrCr`) highest rank is on Catalan (ca_ancora), which is the 5th place.
- Then, it is followed by 6th place on Turkish (tr_itcc), Hungarian (hu_korkor), German (de_potsdamcc), and English (en_parcorfull) datasets.
- For Hungarian, a significant increase is obtained on hu_korkor dataset by 4.87 and 12.2 percentage points on the development and test sets, respectively.
- For Norwegian which exhibits agglutinative characteristics on verbal suffixes, the baseline model is surpassed by 12.58% percentage points on the development set.

References I

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186. Association for Computational Linguistics.

Pamay-Arslan, T. and Eryiğit, G. (2023). Incorporating dropped pronouns into coreference resolution: The case for turkish. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics: Student Research Workshop*, pages 14–25.