UFAL Linguistic Mondays

Coreference without Borders

4 years of CorefUD and CRAC Shared Tasks

Michal Novák

Charles University
Faculty of Mathematics and Physics
Institute of Formal and Applied Linguistics

The team

Michal Novák

Martin Popel

Anna Nedoluzhko

Zdeněk Žabokrtský

Daniel Zeman

Miloslav Konopík

Ondřej Pražák

Jakub Sido

Milan Straka

Maciej Ogrodniczuk, Amir Zeldes, Barbora Dohnalová, Yilun Zhu, ...

Outline

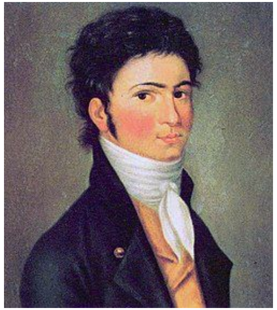
1. Introduction
2. CorefUD
3. CRAC Shared tasks
4. LLMs for coreference
5. Conclusion

Introduction

Coreference

- two or more expressions in the text (mentions) refer to the same discourse entity

Beethoven was a musical genius. The German composer began going deaf in his late twenties, yet he continued to compose masterpieces.

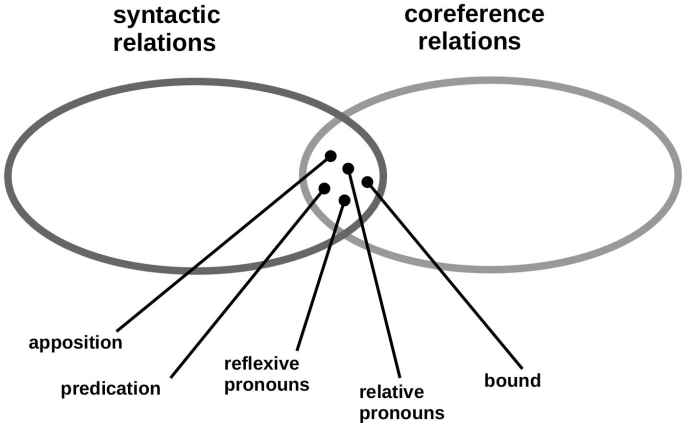


- in some languages, mentions might not be even expressed on the surface (zero mentions)

Beethoven byl hudební génius. Německý skladatel začal ztrácet sluch ve svých dvaceti letech, přesto $\varnothing_{subj}$ *dál skládal mistrovská díla.*

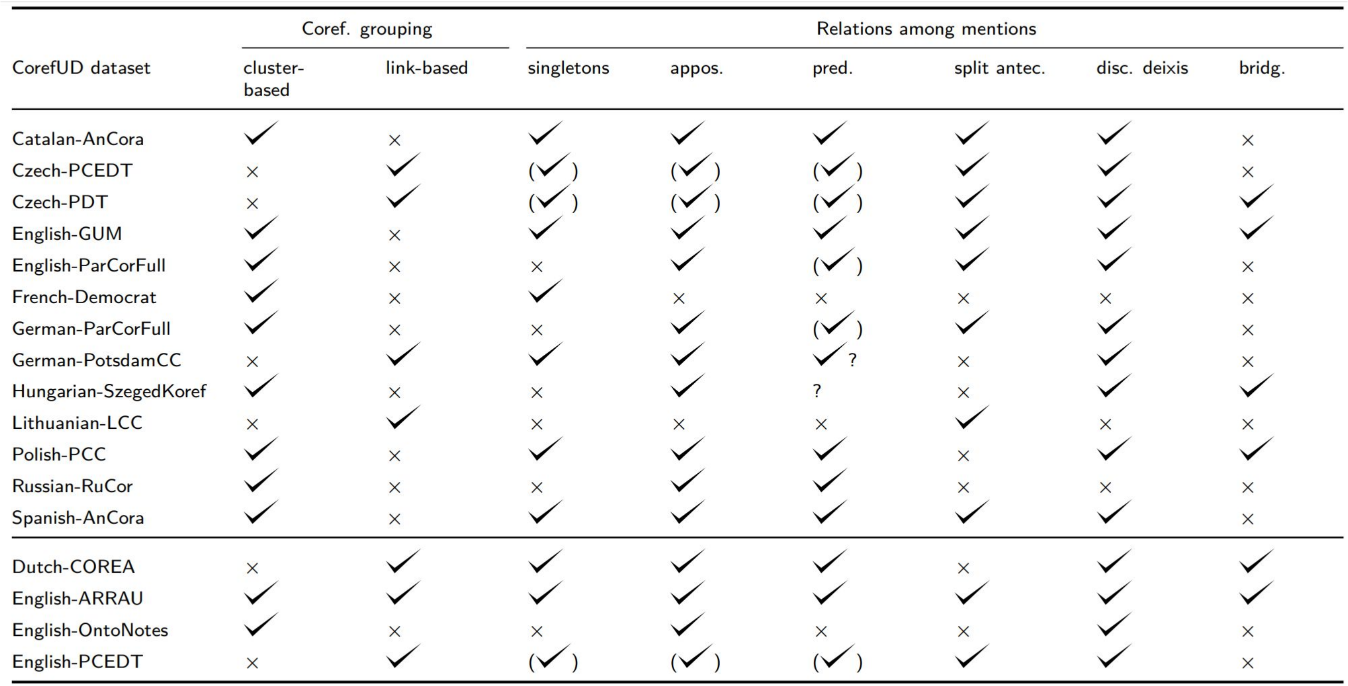
Linguistic motivations

- strong interplay of coreference and syntax
  - mentions often correspond to syntactically meaningful units (noun phrases, subject)
  - some coreference relations are expressed primarily by syntactic means (reflexive and relative constructions, apposition, predication with copula)
  - syntax is useful for the identification of zero expressions (such as pro-drop) needed for coreference
- long tradition of this approach in Prague
  - Hajičová, Panevová, Sgall:
    Coreference in the grammar and in the text (1985)
  - PDT, PCEDT



Diversity in existing resources

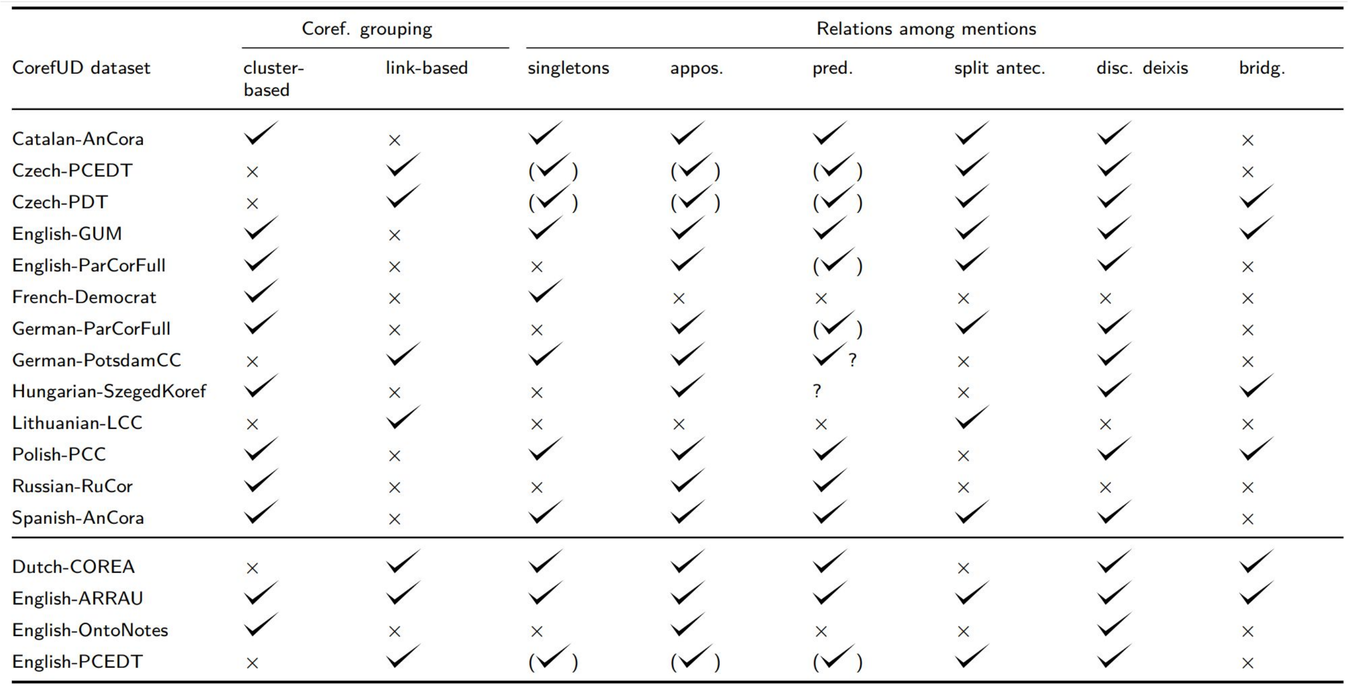
- original resources in CorefUD 0.1



	Coref. grouping		Relations among mentions					
CorefUD dataset	cluster-based	link-based	singletons	appos.	pred.	split antec.	disc. deixis	bridg.
Catalan-AnCora	✓	×	✓	✓	✓	✓	✓	×
Czech-PCEDT	×	✓	(✓)	(✓)	(✓)	✓	✓	×
Czech-PDT	×	✓	(✓)	(✓)	(✓)	✓	✓	✓
English-GUM	✓	×	✓	✓	✓	✓	✓	✓
English-ParCorFull	✓	×	×	✓	(✓)	✓	✓	×
French-Democrat	✓	×	✓	×	×	×	×	×
German-ParCorFull	✓	×	×	✓	(✓)	✓	✓	×
German-PotsdamCC	×	✓	✓	✓	✓ ?	×	✓	×
Hungarian-SzegedKoref	✓	×	×	✓	?	×	✓	✓
Lithuanian-LCC	×	✓	×	×	×	✓	×	×
Polish-PCC	✓	×	✓	✓	✓	×	✓	✓
Russian-RuCor	✓	×	×	✓	✓	×	×	×
Spanish-AnCora	✓	×	✓	✓	✓	✓	✓	×
Dutch-COREA	×	✓	✓	✓	✓	×	✓	✓
English-ARRAU	✓	✓	✓	✓	✓	✓	✓	✓
English-OntoNotes	✓	×	×	✓	×	×	✓	×
English-PCEDT	×	✓	(✓)	(✓)	(✓)	✓	✓	×

Diversity in existing resources

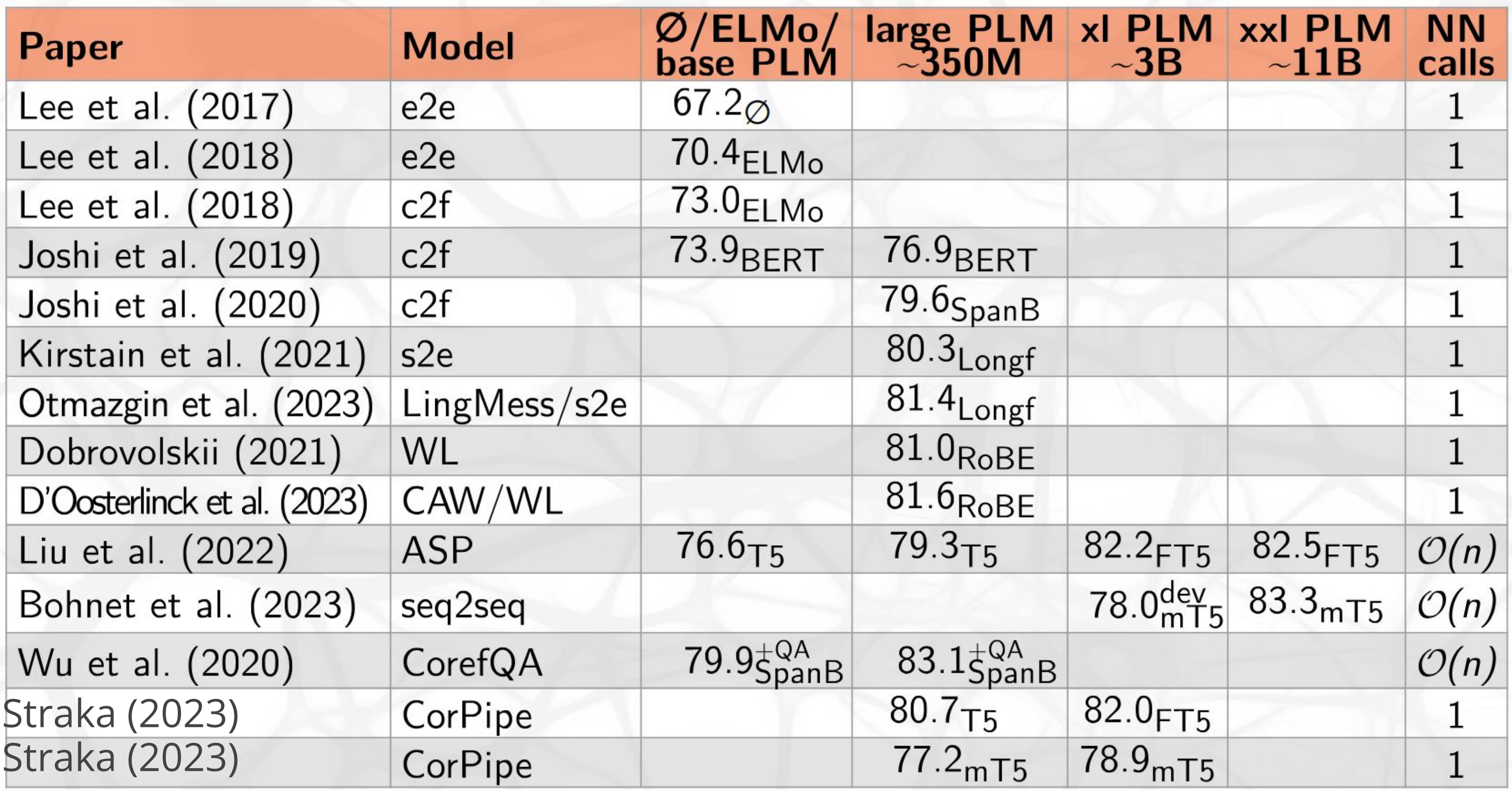
- original resources in CorefUD 0.1



	Coref. grouping		Relations among mentions					
CorefUD dataset	cluster-based	link-based	singletons	appos.	pred.	split antec.	disc. deixis	bridg.
Catalan-AnCora	✓	×	✓	✓	✓	✓	✓	×
Czech-PCEDT	×	✓	(✓)	(✓)	(✓)	✓	✓	×
Czech-PDT	×	✓	(✓)	(✓)	(✓)	✓	✓	✓
English-GUM	✓	×	✓	✓	✓	✓	✓	✓
English-ParCorFull	✓	×	×	✓	(✓)	✓	✓	×
French-Democrat	✓	×	✓	×	×	×	×	×
German-ParCorFull	✓	×	×	✓	(✓)	✓	✓	×
German-PotsdamCC	×	✓	✓	✓	✓ ?	×	✓	×
Hungarian-SzegedKoref	✓	×	×	✓	?	×	✓	✓
Lithuanian-LCC	×	✓	×	×	×	✓	×	×
Polish-PCC	✓	×	✓	✓	✓	×	✓	✓
Russian-RuCor	✓	×	×	✓	✓	×	×	×
Spanish-AnCora	✓	×	✓	✓	✓	✓	✓	×
Dutch-COREA	×	✓	✓	✓	✓	×	✓	✓
English-ARRAU	✓	✓	✓	✓	✓	✓	✓	✓
English-OntoNotes	✓	×	×	✓	×	×	✓	×
English-PCEDT	×	✓	(✓)	(✓)	(✓)	✓	✓	×

- different formats

Coreference Resolution systems



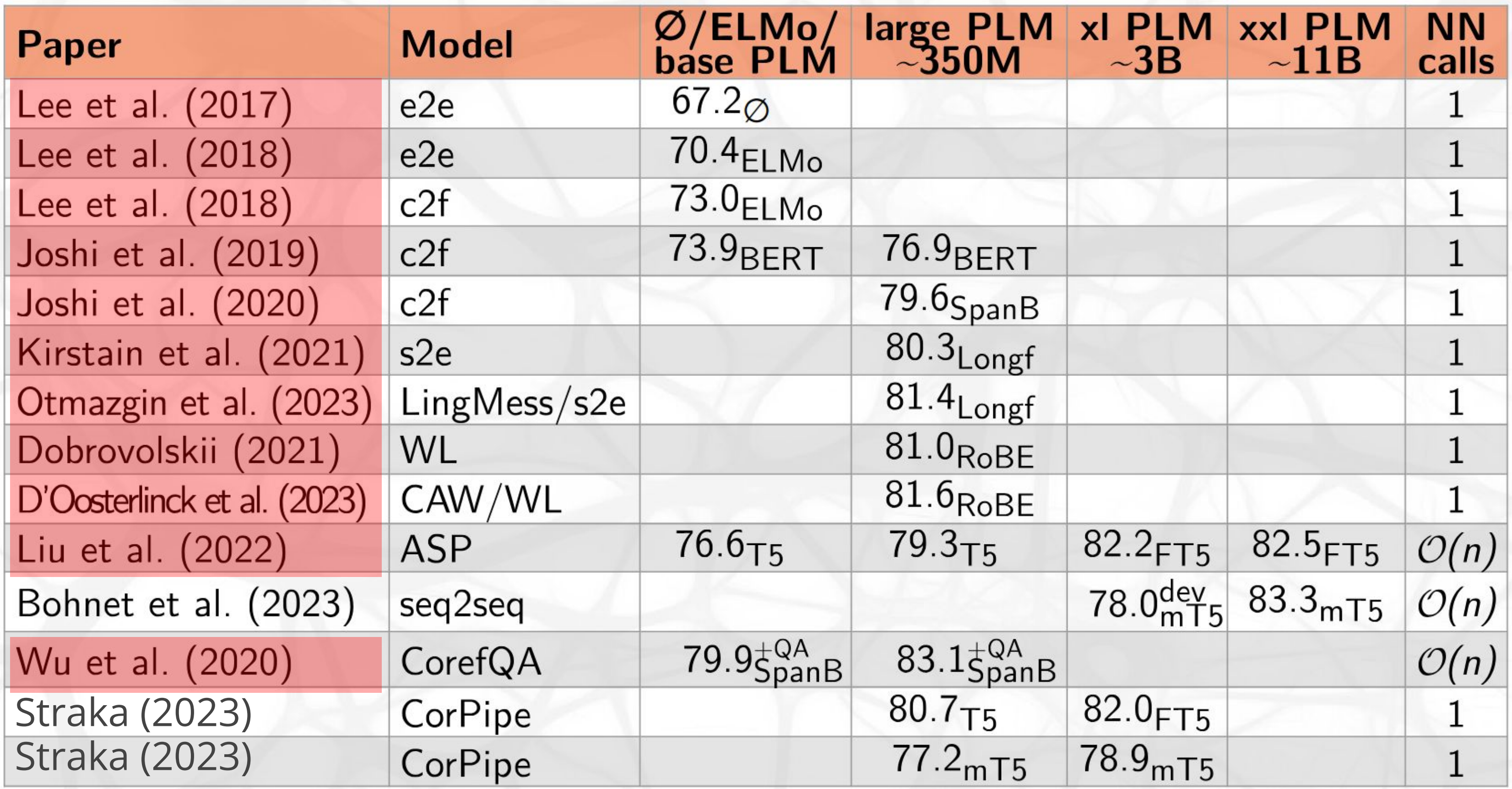
Paper	Model	Ø/ELMo/ base PLM	large PLM ∼350M	xl PLM ∼3B	xxl PLM ∼11B	NN calls
Lee et al. (2017)	e2e	$67.2_{\varnothing}$				1
Lee et al. (2018)	e2e	70.4_{ELMo}				1
Lee et al. (2018)	c2f	73.0_{ELMo}				1
Joshi et al. (2019)	c2f	73.9_{BERT}	76.9_{BERT}			1
Joshi et al. (2020)	c2f		79.6_{SpanB}			1
Kirstain et al. (2021)	s2e		80.3_{Longf}			1
Otmazgin et al. (2023)	LingMess/s2e		81.4_{Longf}			1
Dobrovolskii (2021)	WL		81.0_{RoBE}			1
D'Oosterlinck et al. (2023)	CAW/WL		81.6_{RoBE}			1
Liu et al. (2022)	ASP	76.6_{T5}	79.3_{T5}	82.2_{FT5}	82.5_{FT5}	$\mathcal{O}(n)$
Bohnet et al. (2023)	seq2seq			$78.0^{\text{dev}}_{\text{mT5}}$	83.3_{mT5}	$\mathcal{O}(n)$
Wu et al. (2020)	CorefQA	$79.9^{\text{+QA}}_{\text{SpanB}}$	$83.1^{\text{+QA}}_{\text{SpanB}}$			$\mathcal{O}(n)$
Straka (2023)	CorPipe		80.7_{T5}	82.0_{FT5}		1
Straka (2023)	CorPipe		77.2_{mT5}	78.9_{mT5}		1

from Straka (2023)

Coreference Resolution systems

evaluated only on English OntoNotes

(and GAP in some cases)



Paper	Model	Ø/ELMo/ base PLM	large PLM ~350M	xl PLM ~3B	xxl PLM ~11B	NN calls
Lee et al. (2017)	e2e	$67.2_{\varnothing}$				1
Lee et al. (2018)	e2e	70.4_{ELMo}				1
Lee et al. (2018)	c2f	73.0_{ELMo}				1
Joshi et al. (2019)	c2f	73.9_{BERT}	76.9_{BERT}			1
Joshi et al. (2020)	c2f		79.6_{SpanB}			1
Kirstain et al. (2021)	s2e		80.3_{Longf}			1
Otmazgin et al. (2023)	LingMess/s2e		81.4_{Longf}			1
Dobrovolskii (2021)	WL		81.0_{RoBE}			1
D'Oosterlinck et al. (2023)	CAW/WL		81.6_{RoBE}			1
Liu et al. (2022)	ASP	76.6_{T5}	79.3_{T5}	82.2_{FT5}	82.5_{FT5}	$\mathcal{O}(n)$
Bohnet et al. (2023)	seq2seq			$78.0^{\text{dev}}_{\text{mT5}}$	83.3_{mT5}	$\mathcal{O}(n)$
Wu et al. (2020)	CorefQA	$79.9^{+\text{QA}}_{\text{SpanB}}$	$83.1^{+\text{QA}}_{\text{SpanB}}$			$\mathcal{O}(n)$
Straka (2023)	CorPipe		80.7_{T5}	82.0_{FT5}		1
Straka (2023)	CorPipe		77.2_{mT5}	78.9_{mT5}		1

from Straka (2023)

Universal Dependencies

- framework for consistent annotation of grammar (morphology and dependency syntax) across different languages
- successful in:
  - establishing an annotation standard
  - facilitating NLP research
  - serving as a resource for linguistic studies

Universal Dependencies

- framework for consistent annotation of grammar (morphology and dependency syntax) across different languages
- successful in:
  - establishing an annotation standard
  - facilitating NLP research
  - serving as a resource for linguistic studies
- Dan Zeman in both teams

CorefUD

CorefUD timeline

CorefUD

- a multi-lingual collection of corpora annotated with coreference and anaphora
- harmonized using the same annotation scheme
- combines annotation of coreference/anaphora (always manual) with annotation of morphology and dependency syntax (manual if available, otherwise automatic)
- Universal Dependencies (Nivre et al., 2017)
  - source of inspiration
  - keeping the maximum compliance with its standards
- CorefUD 1.3 (Novák et al., 2025)
  - http://hdl.handle.net/11234/1-5896
  - 6th release since 2021
  - 28 datasets for 18 languages

Current languages and datasets

CorefUD 1.3

Public

- Ancient Greek-PROIEL (Haug and Jøhndal, 2008)
- Ancient Hebrew-PTNK (Swanson et al., 2024)
- Catalan-AnCora (Recasens and Martí, 2010)
- Czech-PCEDT (Nedoluzhko et al., 2016)
- Czech-PDT (Hajič et al., 2020)
- English-GUM (Zeldes, 2017)
- English-LitBank (Bamman et al., 2019)
- English-ParCorFull (Lapshinova-Koltunski et al., 2018)
- French-ANCOR (Muzerelle et al., 2014)
- French-Democrat (Landragin, 2021)
- German-ParCorFull (Lapshinova-Koltunski et al., 2018)
- German-PotsdamCC (Bourgonje and Stede, 2020)
- Hindi-HDTB (Mujadia et al., 2016)
- Hungarian-KorKor (Vadász, 2022)
- Hungarian-SzegedKoref (Vincze et al., 2018)
- Korean-ECMT (Nam et al., 2020)
- Lithuanian-LCC (Žitkus and Butkienė, 2018)
- Norwegian-BokmaalNARC (Mæhlum et al., 2022)
- Norwegian-NynorskNARC (Mæhlum et al., 2022)
- Old Church Slavonic-PROIEL (Haug and Jøhndal, 2008)
- Polish-PCC (Ogrodniczuk et al., 2013)
- Russian-RuCor (Toldova et al., 2014)
- Spanish-AnCora (Recasens and Martí, 2010)
- Turkish-ITCC (Pamay and Eryiğit, 2018)

Non-public

- Dutch-COREA (Hendrickx et al., 2008)
- English-ARRAU (Uryupina et al., 2020)
- English-OntoNotes (Weischedel et al., 2011)
- English-PCEDT (Nedoluzhko et al., 2016)

Current languages and datasets

CorefUD 1.3

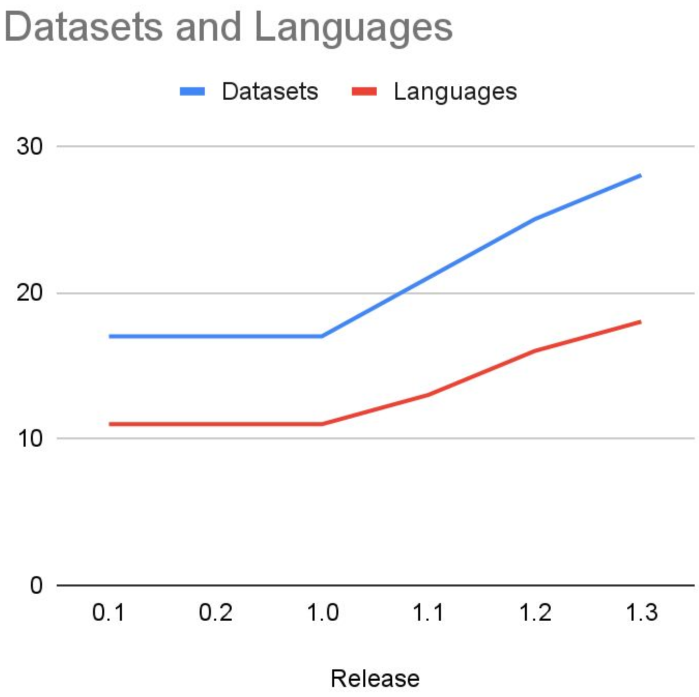
Public

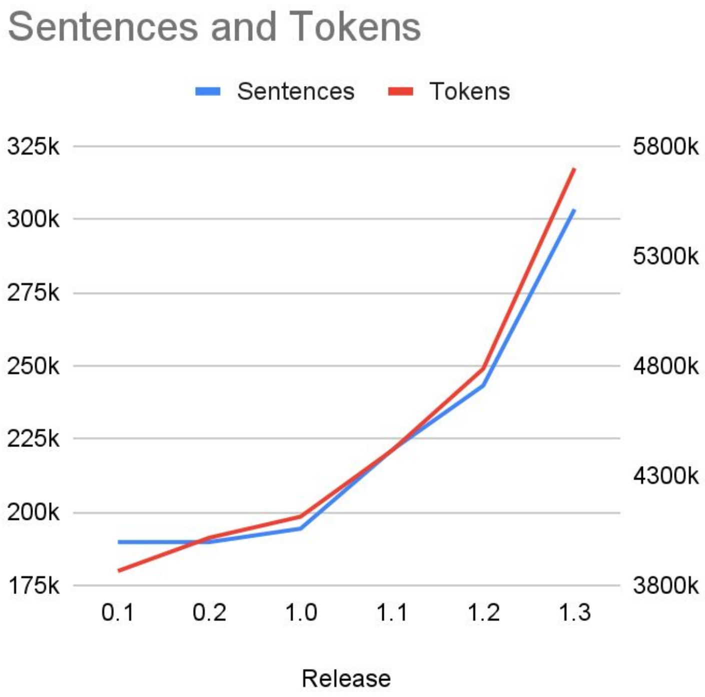
- Ancient Greek-PROIEL (Haug and Jøhndal, 2008)
- Ancient Hebrew-PTNK (Swanson et al., 2024)
- Catalan-AnCora (Recasens and Martí, 2010)
- Czech-PCEDT (Nedoluzhko et al., 2016)
- Czech-PDT (Hajič et al., 2020)
- English-GUM (Zeldes, 2017)
- English-LitBank (Bamman et al., 2019)
- English-ParCorFull (Lapshinova-Koltunski et al., 2018)
- French-ANCOR (Muzerelle et al., 2014)
- French-Democrat (Landragin, 2021)
- German-ParCorFull (Lapshinova-Koltunski et al., 2018)
- German-PotsdamCC (Bourgonje and Stede, 2020)
- Hindi-HDTB (Mujadia et al., 2016)
- Hungarian-KorKor (Vadász, 2022)
- Hungarian-SzegedKoref (Vincze et al., 2018)
- Korean-ECMT (Nam et al., 2020)
- Lithuanian-LCC (Žitkus and Butkienė, 2018)
- Norwegian-BokmaalNARC (Mæhlum et al., 2022)
- Norwegian-NynorskNARC (Mæhlum et al., 2022)
- Old Church Slavonic-PROIEL (Haug and Jøhndal, 2008)
- Polish-PCC (Ogrodniczuk et al., 2013)
- Russian-RuCor (Toldova et al., 2014)
- Spanish-AnCora (Recasens and Martí, 2010)
- Turkish-ITCC (Pamay and Eryiğit, 2018)

Non-public

- Dutch-COREA (Hendrickx et al., 2008)
- English-ARRAU (Uryupina et al., 2020)
- English-OntoNotes (Weischedel et al., 2011)
- English-PCEDT (Nedoluzhko et al., 2016)

Size over time





Format

- CorefUD 1.0 format
    - since CorefUD 1.0
    - fully compliant with the CoNLL-U format
    - may be included in UD releases
    - represented in the MISC field

```
# sent_id = ln94200-149-p2s3
# text = Pokud nebude jeho smlouva zrušena, je rozhodnut zanechat vůbec fotbalu, uvedl manažer brazilského hráče Ricardo Fuica.
# orig_file_sentence ln94200_149#5
1   Pokud       pokud      SCONJ J,------------- _          5  mark        5:mark                _
2   nebude      být        AUX   VB-S---3F-NAI-- Aspect=... 5  aux:pass    5:aux:pass            _
3   jeho        jeho       DET   P9XXXZS3------- Gender[... 4  det         4:det                 Entity=(e62601--1-gstype:spec)|Functor=4:APP
4   smlouva     smlouva    NOUN  NNFS1-----A---- Case=No... 5  nsubj:pass  5:nsubj:pass          Functor=5:PAT
5   zrušena     zrušený    ADJ   VsQW----X-APP-- Aspect=... 8  advcl       8:advcl:pokud         SpaceAfter=No|LDeriv=zrušit|Functor=8:COND
6   ,           ,          PUNCT Z:------------- _          5  punct       5:punct               _
7   je          být        AUX   VB-S---3P-AAI-- Aspect=... 8  cop         8:cop                 Functor=8:EFF
7.1 on          #PersPron  PRON  _               Case=No... _  _           8:nsubj|9:nsubj:xsubj Entity=(e62601--1-gstype:spec)|Functor=8:ACT.cop,9:ACT
8   rozhodnut   rozhodnutý ADJ   VsYS----X-APP-- Aspect=... 13 ccomp       13:ccomp              LDeriv=rozhodnout|Functor=13:PAT
9   zanechat    zanechat   VERB  Vf--------A-P-- Aspect=... 8  xcomp       8:xcomp               Functor=8:MANN
10  vůbec       vůbec      PART  TT------------- _          11 advmod:emph 11:advmod:emph        Functor=11:EXT
11  fotbalu     fotbal     NOUN  NNIS2-----A---- Animacy... 9  obl:arg     9:obl:arg:gen         SpaceAfter=No|Functor=9:PAT
12  ,           ,          PUNCT Z:------------- _          8  punct       8:punct               _
13  uvedl       uvést      VERB  VpYS----R-AAP-- Aspect=... 0  root        0:root                Functor=0:PRED
14  manažer     manažer    NOUN  NNMS1-----A---- Animacy... 13 nsubj       13:nsubj              Entity=(e62605--1-gstype:spec|Functor=13:RSTR
15  brazilského brazilský  ADJ   AAMS2----1A---- Animacy... 16 amod        16:amod               Entity=(e62601--2-gstype:spec|Functor=16:RSTR
16  hráče       hráč       NOUN  NNMS2-----A---- Animacy... 14 nmod        14:nmod:gen           Entity=e62601)|Functor=14:APP
17  Ricardo     Ricardo    PROPN NNMS1-----A---- Animacy... 14 flat        14:flat               Functor=14:RSTR
18  Fuica       Fuica      PROPN NNMS1-----A---- Animacy... 14 flat        14:flat               Entity=e62605)|Functor=14:ACT|SpaceAfter=No
19  .           .          PUNCT Z:------------- _          13 punct       13:punct              _
```

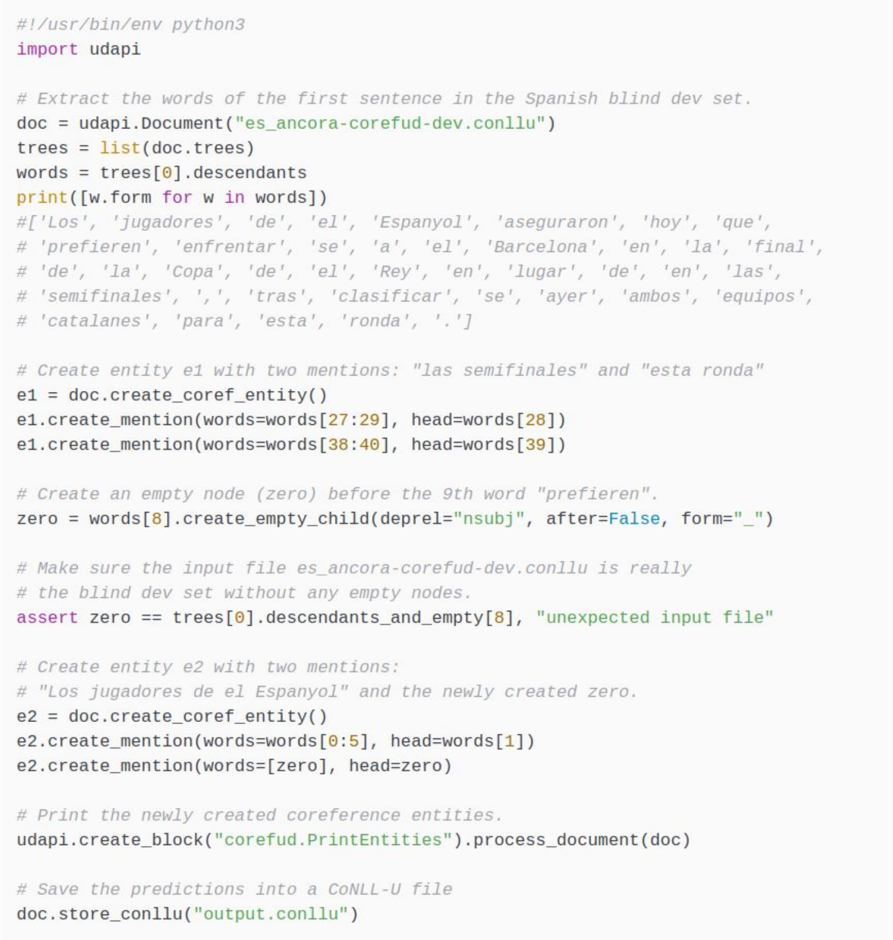
Format

- CorefUD 1.0 format
    - since CorefUD 1.0
    - fully compliant with the CoNLL-U format
    - may be included in UD releases
    - represented in the MISC field

```
# sent_id = ln94200-149-p2s3
# text = Pokud nebude jeho smlouva zrušena, je rozhodnut zanechat vůbec fotbalu, uvedl manažer brazilského hráče Ricardo Fuica.
# orig_file_sentence ln94200_149#5
1   Pokud       pokud      SCONJ J,------------- _          5  mark        5:mark                _
2   nebude      být        AUX   VB-S---3F-NAI-- Aspect=... 5  aux:pass    5:aux:pass            _
3   jeho        jeho       DET   P9XXXZS3------- Gender[... 4  det         4:det                 Entity=(e62601--1-gstype:spec)|Functor=4:APP
4   smlouva     smlouva    NOUN  NNFS1-----A---- Case=No... 5  nsubj:pass  5:nsubj:pass          Functor=5:PAT
5   zrušena     zrušený    ADJ   VsQW----X-APP-- Aspect=... 8  advcl       8:advcl:pokud         SpaceAfter=No|LDeriv=zrušit|Functor=8:COND
6   ,           ,          PUNCT Z:------------- _          5  punct       5:punct               _
7   je          být        AUX   VB-S---3P-AAI-- Aspect=... 8  cop         8:cop                 Functor=8:EFF
7.1 on          #PersPron  PRON  _               Case=No... _  _           8:nsubj|9:nsubj:xsubj Entity=(e62601--1-gstype:spec)|Functor=8:ACT.cop,9:ACT
8   rozhodnut   rozhodnutý ADJ   VsYS----X-APP-- Aspect=... 13 ccomp       13:ccomp              LDeriv=rozhodnout|Functor=13:PAT
9   zanechat    zanechat   VERB  Vf--------A-P-- Aspect=... 8  xcomp       8:xcomp               Functor=8:MANN
10  vůbec       vůbec      PART  TT------------- _          11 advmod:emph 11:advmod:emph        Functor=11:EXT
11  fotbalu     fotbal     NOUN  NNIS2-----A---- Animacy... 9  obl:arg     9:obl:arg:gen         SpaceAfter=No|Functor=9:PAT
12  ,           ,          PUNCT Z:------------- _          8  punct       8:punct               _
13  uvedl       uvést      VERB  VpYS----R-AAP-- Aspect=... 0  root        0:root                Functor=0:PRED
14  manažer     manažer    NOUN  NNMS1-----A---- Animacy... 13 nsubj       13:nsubj              Entity=(e62605--1-gstype:spec|Functor=13:RSTR
15  brazilského brazilský  ADJ   AAMS2----1A---- Animacy... 16 amod        16:amod               Entity=(e62601--2-gstype:spec|Functor=16:RSTR
16  hráče       hráč       NOUN  NNMS2-----A---- Animacy... 14 nmod        14:nmod:gen           Entity=e62601)|Functor=14:APP
17  Ricardo     Ricardo    PROPN NNMS1-----A---- Animacy... 14 flat        14:flat               Functor=14:RSTR
18  Fuica       Fuica      PROPN NNMS1-----A---- Animacy... 14 flat        14:flat               Entity=e62605)|Functor=14:ACT|SpaceAfter=No
19  .           .          PUNCT Z:------------- _          13 punct       13:punct              _
```

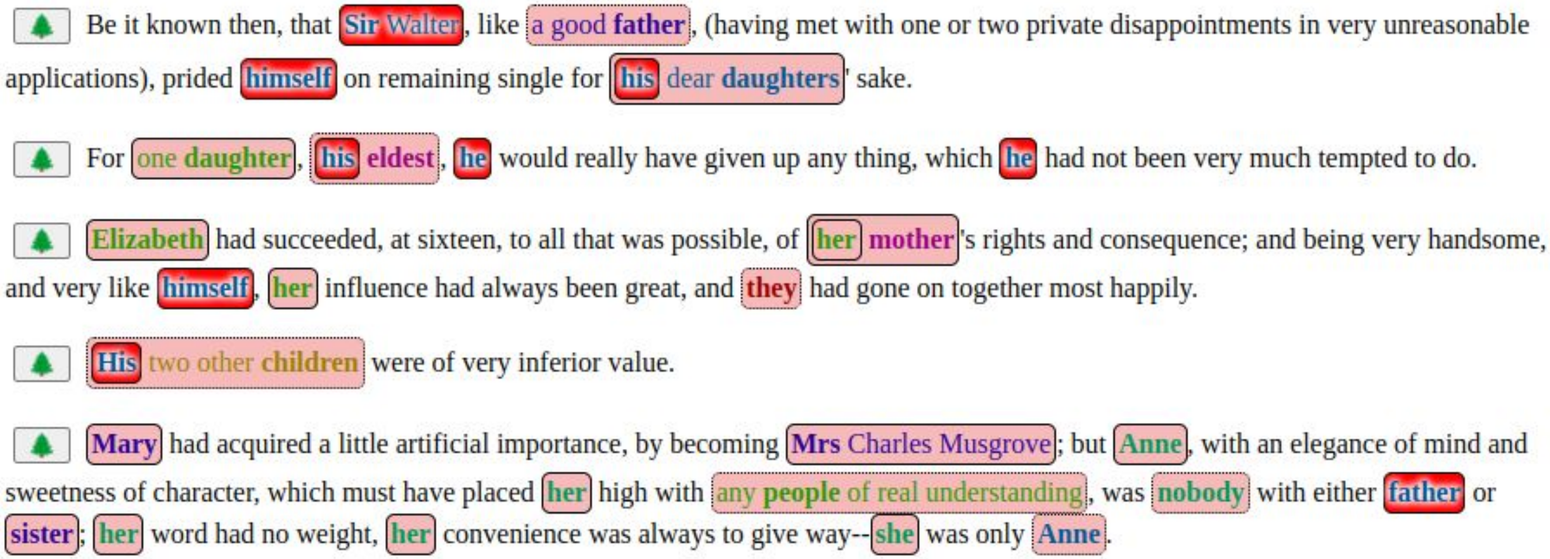
API

- Udapi (Popel et al., 2017)
  - toolkit for text and UD data manipulation
  - querying, statistics
  - format conversions
  - visualization
- coreference object model
  - mention
  - entity
  - bridging links



```
#!/usr/bin/env python3
import udapi

# Extract the words of the first sentence in the Spanish blind dev set.
doc = udapi.Document("es_ancora-corefud-dev.conllu")
trees = list(doc.trees)
words = trees[0].descendants
print([w.form for w in words])
#['Los', 'jugadores', 'de', 'el', 'Espanyol', 'aseguraron', 'hoy', 'que',
# 'prefieren', 'enfrentar', 'se', 'a', 'el', 'Barcelona', 'en', 'la', 'final',
# 'de', 'la', 'Copa', 'de', 'el', 'Rey', 'en', 'lugar', 'de', 'en', 'las',
# 'semifinales', ',', 'tras', 'clasificar', 'se', 'ayer', 'ambos', 'equipos',
# 'catalanes', 'para', 'esta', 'ronda', '.']

# Create entity e1 with two mentions: "las semifinales" and "esta ronda"
e1 = doc.create_coref_entity()
e1.create_mention(words=words[27:29], head=words[28])
e1.create_mention(words=words[38:40], head=words[39])

# Create an empty node (zero) before the 9th word "prefieren".
zero = words[8].create_empty_child(deprel="nsubj", after=False, form="_")

# Make sure the input file es_ancora-corefud-dev.conllu is really
# the blind dev set without any empty nodes.
assert zero == trees[0].descendants_and_empty[8], "unexpected input file"

# Create entity e2 with two mentions:
# "Los jugadores de el Espanyol" and the newly created zero.
e2 = doc.create_coref_entity()
e2.create_mention(words=words[0:5], head=words[1])
e2.create_mention(words=[zero], head=zero)

# Print the newly created coreference entities.
udapi.create_block("corefud.PrintEntities").process_document(doc)

# Save the predictions into a CoNLL-U file
doc.store_conllu("output.conllu")
```

Visualization



🌲 Be it known then, that Sir Walter, like a good father, (having met with one or two private disappointments in very unreasonable applications), prided himself on remaining single for his dear daughters' sake.

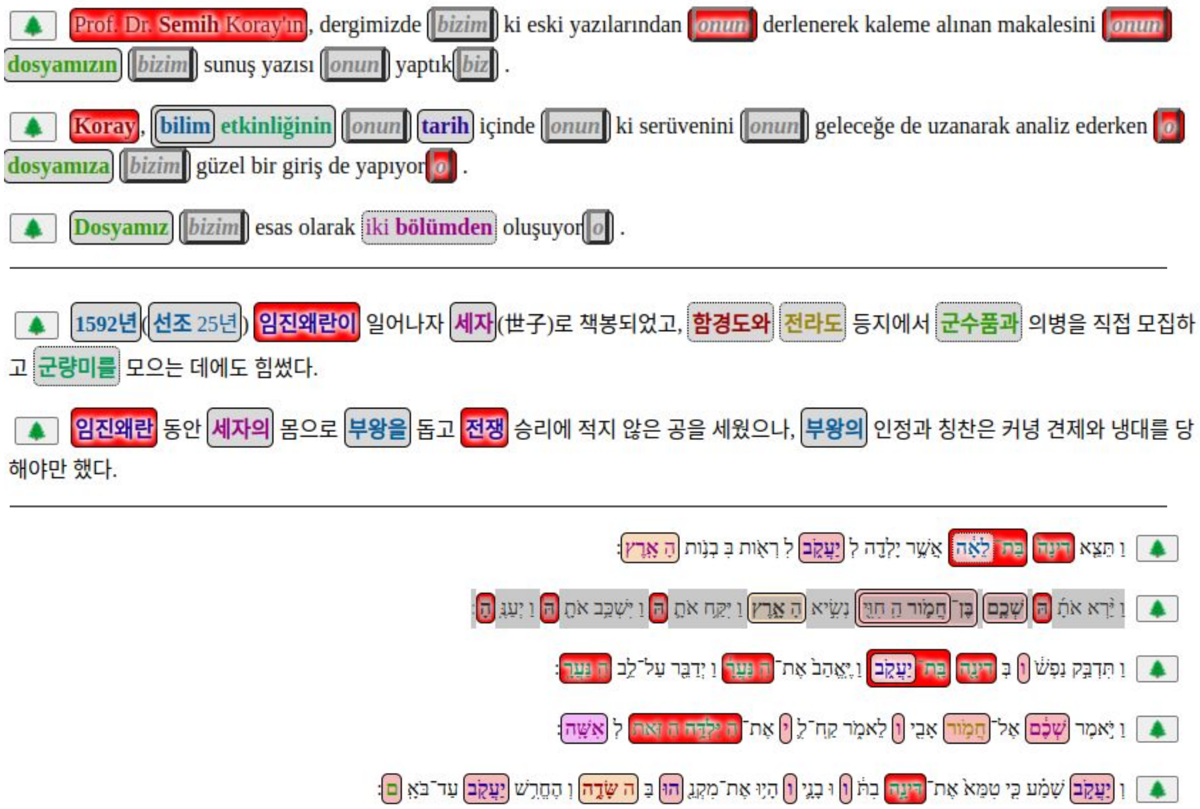
🌲 For one daughter, his eldest, he would really have given up any thing, which he had not been very much tempted to do.

🌲 Elizabeth had succeeded, at sixteen, to all that was possible, of her mother's rights and consequence; and being very handsome, and very like himself, her influence had always been great, and they had gone on together most happily.

🌲 His two other children were of very inferior value.

🌲 Mary had acquired a little artificial importance, by becoming Mrs Charles Musgrove; but Anne, with an elegance of mind and sweetness of character, which must have placed her high with any people of real understanding, was nobody with either father or sister; her word had no weight, her convenience was always to give way--she was only Anne.

Visualization



Prof. Dr. Semih Koray'ın, dergimizde bizim ki eski yazılarından onun derlenerek kaleme alınan makalesini onun dosyamızın bizim sunuş yazısı onun yaptık biz .

Koray, bilim etkinliğinin onun tarih içinde onun ki serüvenini onun geleceğe de uzanarak analiz ederken o dosyamıza bizim güzel bir giriş de yapıyor o .

Dosyamız bizim esas olarak iki bölümden oluşuyor o .

1592년(선조 25년) 임진왜란이 일어나자 세자(世子)로 책봉되었고, 함경도와 전라도 등지에서 군수품과 의병을 직접 모집하고 군량미를 모으는 데에도 힘썼다.

임진왜란 동안 세자의 몸으로 부왕을 돕고 전쟁 승리에 적지 않은 공을 세웠으나, 부왕의 인정과 칭찬은 커녕 견제와 냉대를 당해야만 했다.

וַ תֵּצֵא דִינָה בַּת־לֵאָה אֲשֶׁר יָלְדָה לְ יַעֲקֹב לִ רְאוֹת בִּ בְנוֹת הָ אָרֶץ׃

וַ יַּרְא אֹתָ הּ שְׁכֶם בֶּן־חֲמוֹר הַ חִוִּי נְשִׂיא הָ אָרֶץ וַ יִּקַּח אֹתָ הּ וַ יִּשְׁכַּב אֹתָ הּ וַ יְעַנֶּ הָ׃

וַ תִּדְבַּק נַפְשׁ וֹ בְּ דִינָה בַּת־יַעֲקֹב וַ יֶּאֱהַב אֶת־הַ נַּעֲרָ וַ יְדַבֵּר עַל־לֵב הַ נַּעֲרָ׃

וַ יֹּאמֶר שְׁכֶם אֶל־חֲמוֹר אָבִי ו לֵאמֹר קַח־לִ י אֶת־הַ יַּלְדָּה הַ זֹּאת לְ אִשָּׁה׃

וְ יַעֲקֹב שָׁמַע כִּי טִמֵּא אֶת־דִּינָה בִתּ וֹ וּ בָנָי ו הָיוּ אֶת־מִקְנֵ הוּ בַּ שָּׂדֶה וְ הֶחֱרִשׁ יַעֲקֹב עַד־בֹּא ם׃

Adding new datasets

- only datasets with free licenses
  - Chinese and Arabic OntoNotes
- datasets enhancing diversity preferred
  - language
    - non-European (e.g. Korean, Hindi)
    - ancient (Ancient Greek, Ancient Hebrew, Old Church Slavonic)
  - domain
    - LitBank: English but fiction, longer contexts
    - Ancor: French but spoken

Adding new datasets

- implementing the conversion pipeline
  - by the core team members
    - majority of datasets
    - time-consuming
  - cooperation of the core team with external volunteers
    - outsourcing it with no or weak supervision
    - initialize and outsource
  - ideally by the authors of the datasets
    - to promote their dataset

Design decisions

- mention
  - no ID
  - full span specified
  - may be discontinuous
- mention head
  - determined from the dependency tree
  - using the Udapi block `corefud.MoveHead`
- coreference entities / clusters
  - grouping by co-indexing (not by links)
  - singletons allowed
- zero mentions
  - represented by empty nodes using enhanced UD graphs
  - empty nodes may have multiple parents
- bridging relations
  - cluster-to-cluster
  - mention-to-mention would require mention IDs
- morphology and syntax
  - UD-based
  - gold if easy to convert
  - otherwise using UDPipe 2

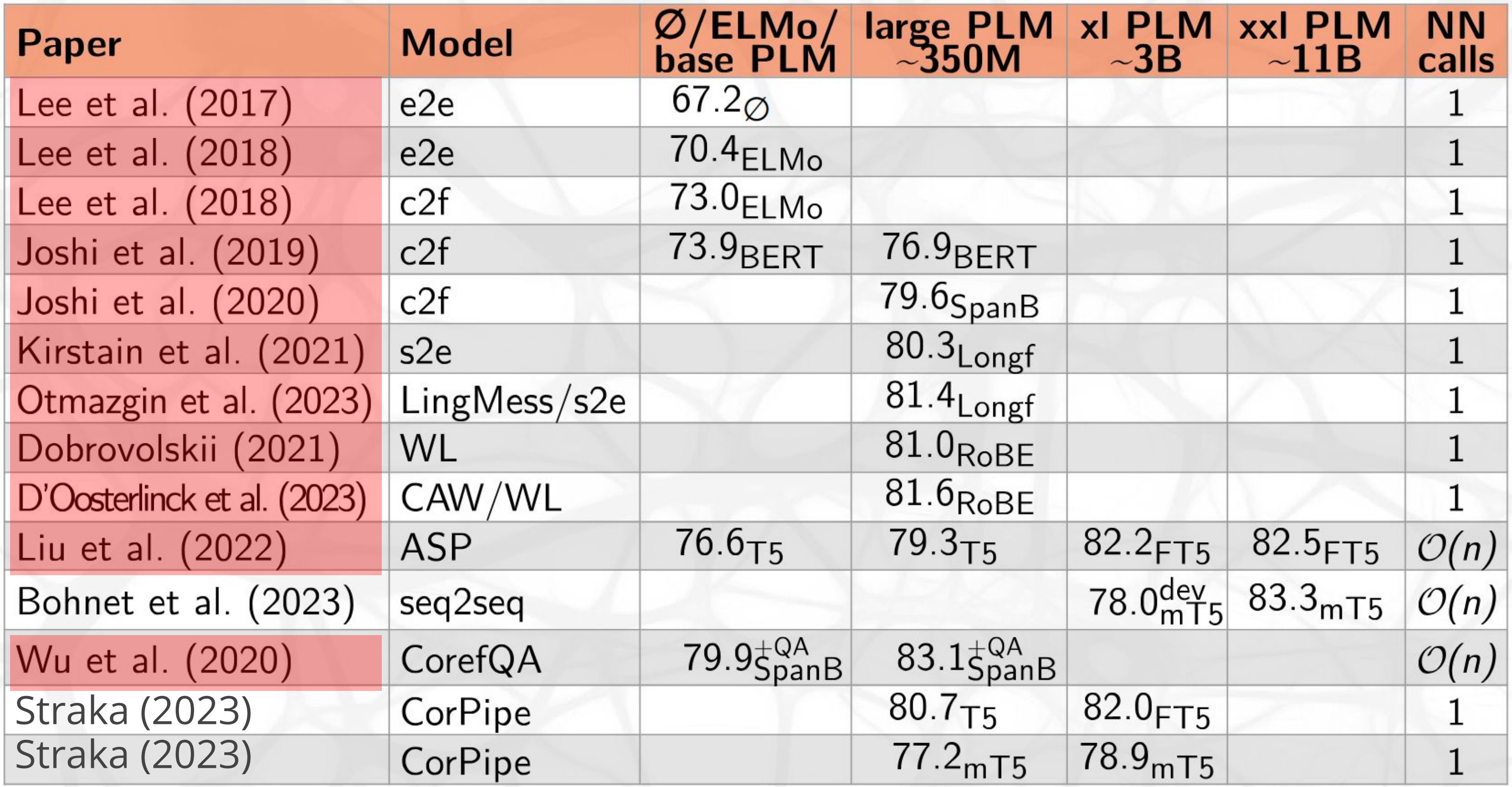
CorefUD vs. Universal Anaphora

CRAC Shared Tasks

Coreference Resolution systems

evaluated only on English OntoNotes

(and GAP in some cases)



Paper	Model	Ø/ELMo/ base PLM	large PLM $\sim$350M	xl PLM $\sim$3B	xxl PLM $\sim$11B	NN calls
Lee et al. (2017)	e2e	$67.2_{\varnothing}$				1
Lee et al. (2018)	e2e	70.4_{ELMo}				1
Lee et al. (2018)	c2f	73.0_{ELMo}				1
Joshi et al. (2019)	c2f	73.9_{BERT}	76.9_{BERT}			1
Joshi et al. (2020)	c2f		79.6_{SpanB}			1
Kirstain et al. (2021)	s2e		80.3_{Longf}			1
Otmazgin et al. (2023)	LingMess/s2e		81.4_{Longf}			1
Dobrovolskii (2021)	WL		81.0_{RoBE}			1
D'Oosterlinck et al. (2023)	CAW/WL		81.6_{RoBE}			1
Liu et al. (2022)	ASP	76.6_{T5}	79.3_{T5}	82.2_{FT5}	82.5_{FT5}	$\mathcal{O}(n)$
Bohnet et al. (2023)	seq2seq			$78.0^{\text{dev}}_{\text{mT5}}$	83.3_{mT5}	$\mathcal{O}(n)$
Wu et al. (2020)	CorefQA	$79.9^{+\text{QA}}_{\text{SpanB}}$	$83.1^{+\text{QA}}_{\text{SpanB}}$			$\mathcal{O}(n)$
Straka (2023)	CorPipe		80.7_{T5}	82.0_{FT5}		1
Straka (2023)	CorPipe		77.2_{mT5}	78.9_{mT5}		1

from Straka (2023)

Shared task in general

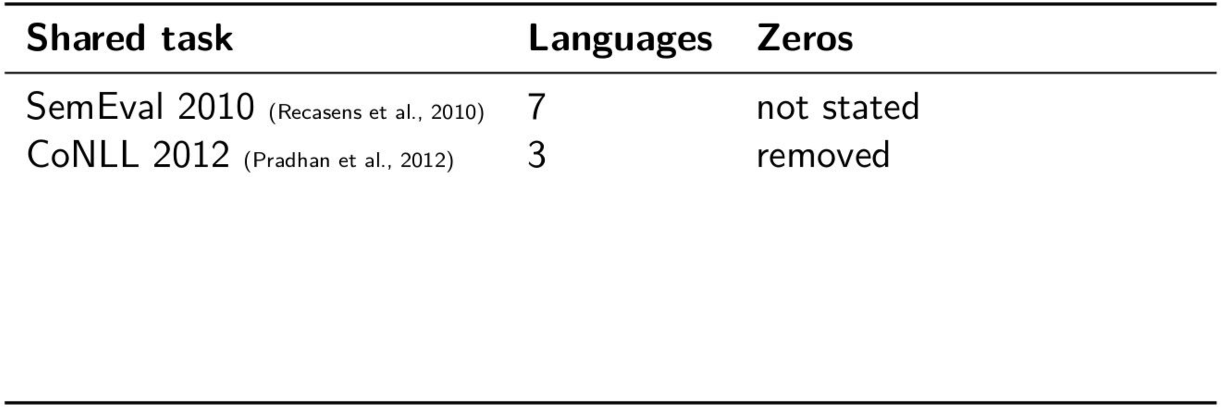
- collaborative evaluation campaign organized within the research community
  - popular in the fields such as Natural Language Processing, Machine Learning
- provides a common problem, shared dataset and evaluation framework so that multiple teams can develop and compare their systems under the same conditions
- purposes:
  - to encourage research progress on a specific task
  - to increase reproducibility and comparability of the methods
  - to provide benchmark datasets, evaluation tools and baseline models
- examples:
  - CoNLL Shared Tasks
  - SemEval
  - WMT

Shared task details

- CRAC Shared Task on Multilingual Coreference Resolution
  - collocated with the Workshop on Computational Models of Reference, Anaphora and Coreference (CRAC)
  - based on CorefUD

Shared task details

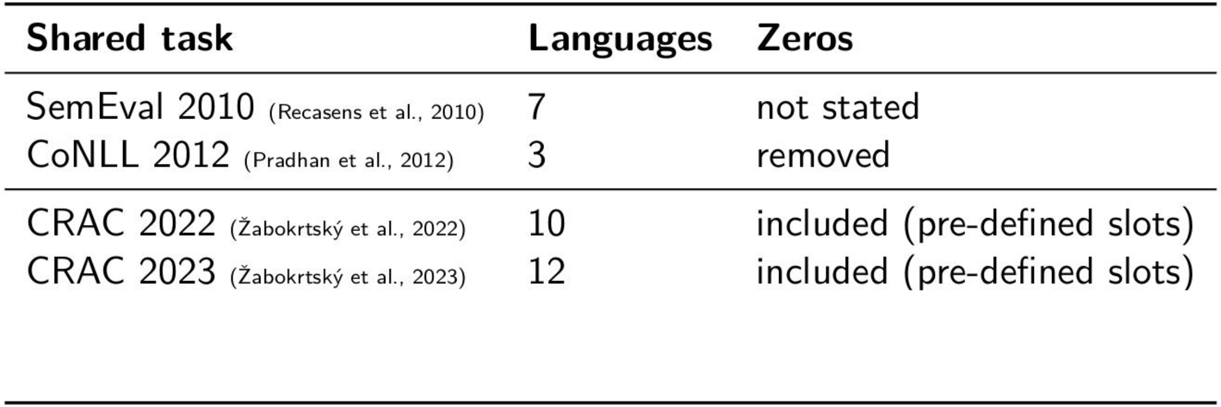
- CRAC Shared Task on Multilingual Coreference Resolution
  - collocated with the Workshop on Computational Models of Reference, Anaphora and Coreference (CRAC)
  - based on CorefUD
- task
  - identify mentions in texts and predict which mentions belong to the same coref. cluster



Shared task	**Languages**	**Zeros**
SemEval 2010 (Recasens et al., 2010)	7	not stated
CoNLL 2012 (Pradhan et al., 2012)	3	removed

Shared task details

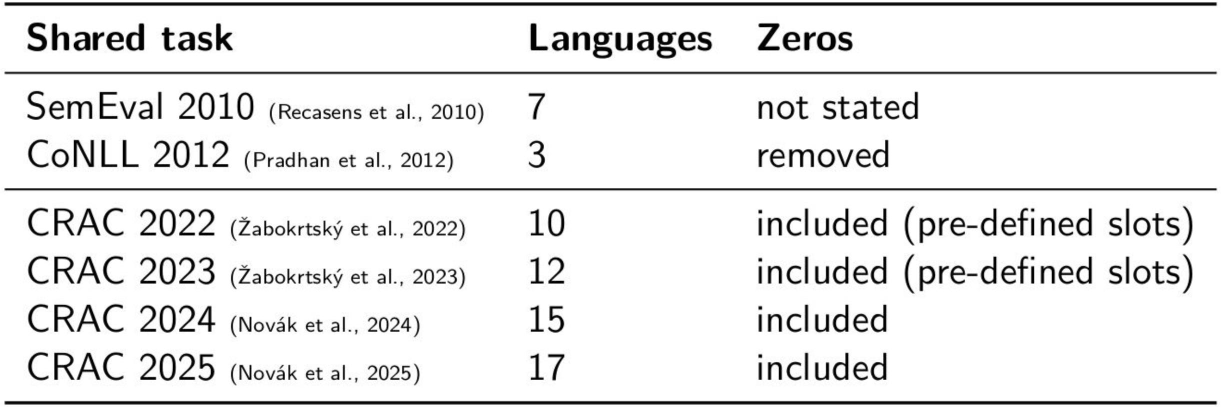
- CRAC Shared Task on Multilingual Coreference Resolution
  - collocated with the Workshop on Computational Models of Reference, Anaphora and Coreference (CRAC)
  - based on CorefUD
- task
  - identify mentions in texts and predict which mentions belong to the same coref. cluster



Shared task	Languages	Zeros
SemEval 2010 (Recasens et al., 2010)	7	not stated
CoNLL 2012 (Pradhan et al., 2012)	3	removed
CRAC 2022 (Žabokrtský et al., 2022)	10	included (pre-defined slots)
CRAC 2023 (Žabokrtský et al., 2023)	12	included (pre-defined slots)

Shared task details

- CRAC Shared Task on Multilingual Coreference Resolution
  - collocated with the Workshop on Computational Models of Reference, Anaphora and Coreference (CRAC)
  - based on CorefUD
- task
  - predict empty nodes
  - identify mentions in texts and predict which mentions belong to the same coref. cluster



Shared task	**Languages**	**Zeros**
SemEval 2010 (Recasens et al., 2010)	7	not stated
CoNLL 2012 (Pradhan et al., 2012)	3	removed
CRAC 2022 (Žabokrtský et al., 2022)	10	included (pre-defined slots)
CRAC 2023 (Žabokrtský et al., 2023)	12	included (pre-defined slots)
CRAC 2024 (Novák et al., 2024)	15	included
CRAC 2025 (Novák et al., 2025)	17	included

Evaluation

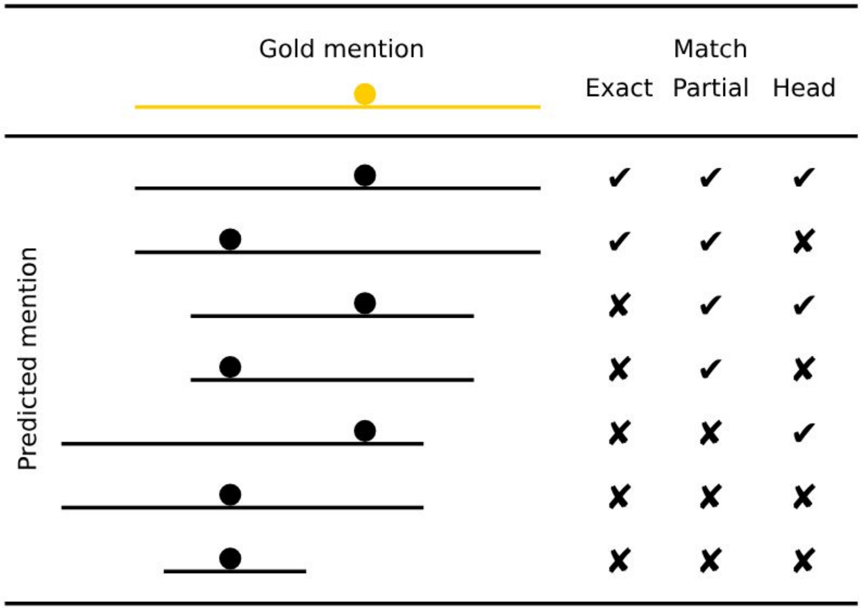
- scoring is complex
    - comparing clusters of mentions
    - MUC (Vilain et al., 1995), B3 (Bagga and Baldwin, 1998), CEAF-e, CEAF-m (Luo, 2005), BLANC (Recasens and Hovy, 2011), LEA (Moosavi and Strube, 2016)
    - each one has P/R/F1

Evaluation

- scoring is complex
    - comparing clusters of mentions
    - **MUC** (Vilain et al., 1995), **B3** (Bagga and Baldwin, 1998), **CEAF-e**, CEAF-m (Luo, 2005), BLANC (Recasens and Hovy, 2011), LEA (Moosavi and Strube, 2016)
    - each one has P/R/**F1**
    - CoNLL F1 score
        - **average**
        - macro-averaged over all datasets

Evaluation

- scoring is complex
  - comparing clusters of mentions
  - **MUC** (Vilain et al., 1995), **B3** (Bagga and Baldwin, 1998), **CEAF-e**, CEAF-m (Luo, 2005), BLANC (Recasens and Hovy, 2011), LEA (Moosavi and Strube, 2016)
  - each one has P/R/**F1**
  - CoNLL F1 score
    - **average**
    - macro-averaged over all datasets
- gold and predicted mentions must be matched
  - exact match: used traditionally for English
  - partial match: CRAC'22
  - head match: since CRAC'23



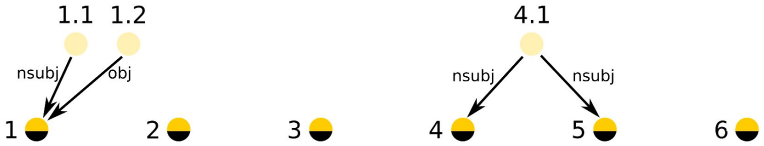
Exact	Partial	Head
✓	✓	✓
✓	✓	✗
✗	✓	✓
✗	✓	✗
✗	✗	✓
✗	✗	✗
✗	✗	✗

Evaluation

- zero matching
  - no special treatment until CRAC 2024
    - all empty nodes were already reconstructed in the input
  - predicted empty nodes no longer guaranteed to align 1:1 with the gold empty nodes
  - dependency-based matching
  - priority to the accurate assignment of both parents and dep. types, but parents are enough

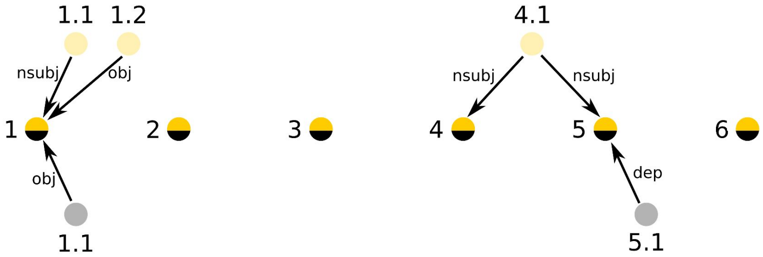
Evaluation

- zero matching
  - no special treatment until CRAC 2024
    - all empty nodes were already reconstructed in the input
  - predicted empty nodes no longer guaranteed to align 1:1 with the gold empty nodes
  - dependency-based matching
  - priority to the accurate assignment of both parents and dep. types, but parents are enough



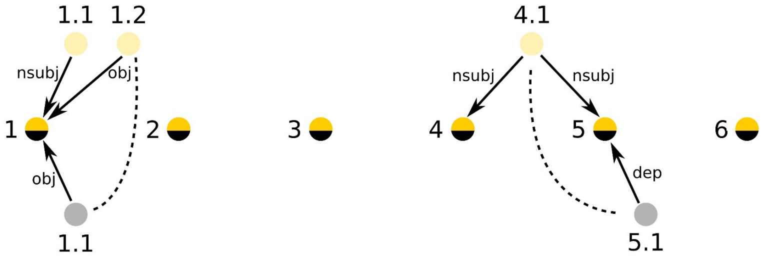
Evaluation

- zero matching
  - no special treatment until CRAC 2024
    - all empty nodes were already reconstructed in the input
  - predicted empty nodes no longer guaranteed to align 1:1 with the gold empty nodes
  - dependency-based matching
  - priority to the accurate assignment of both parents and dep. types, but parents are enough



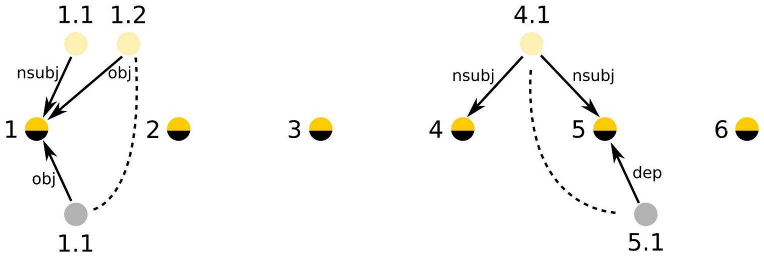
Evaluation

- zero matching
  - no special treatment until CRAC 2024
    - all empty nodes were already reconstructed in the input
  - predicted empty nodes no longer guaranteed to align 1:1 with the gold empty nodes
  - dependency-based matching
  - priority to the accurate assignment of both parents and dep. types, but parents are enough



Evaluation

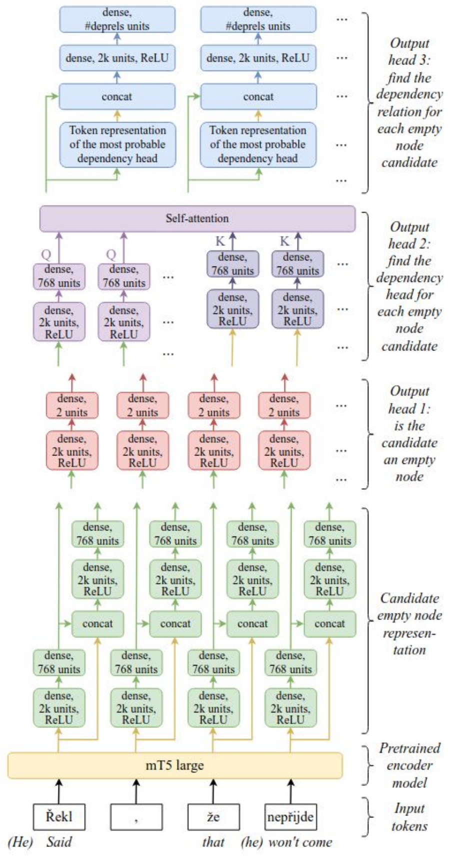
- zero matching
  - no special treatment until CRAC 2024
    - all empty nodes were already reconstructed in the input
  - predicted empty nodes no longer guaranteed to align 1:1 with the gold empty nodes
  - dependency-based matching
  - priority to the accurate assignment of both parents and dep. types, but parents are enough



- mention-decomposable score for zeros
  - difficult to compute traditional scores only on a specific type of mentions
  - allows it but due to oversimplification may lead to inaccuracies

Baseline systems

- empty nodes prediction (by Milan Straka)
  - based on XLM-RoBERTa large (Conneau et al., 2020)
  - two empty-node candidates for each word
  - its representation processed by three prediction heads:
    - empty node
    - word order
    - dependecy relation
  - trained on a combination of all CorefUD datasets with zeros
  - macro-avg F1 = 82.9 (CRAC’24)
- coreference resolution (by Ondřej Pražák)
  - same each year
  - based on the system by (Pražák et al., 2021), originally proposed by (Lee et al., 2018)
  - built on multi-lingual BERT
  - same system for all languages



dense, #deprels units
dense, 2k units, ReLU
concat
Token representation of the most probable dependency head
...
Output head 3: find the dependency relation for each empty node candidate
Self-attention
Q
K
dense, 768 units
dense, 2k units, ReLU
Output head 2: find the dependency head for each empty node candidate
dense, 2 units
dense, 2k units, ReLU
Output head 1: is the candidate an empty node
dense, 768 units
dense, 2k units, ReLU
concat
Candidate empty node represen-tation
mT5 large
Pretrained encoder model
Řekl
,
že
nepřijde
Input tokens
(He) Said
that
(he) won't come

from Straka (2024)

Shared Task Data

CorefUD 1.3

Public

- Ancient Greek-PROIEL (Haug and Jøhndal, 2008)
- Ancient Hebrew-PTNK (Swanson et al., 2024)
- Catalan-AnCora (Recasens and Martí, 2010)
- Czech-PCEDT (Nedoluzhko et al., 2016)
- Czech-PDT (Hajič et al., 2020)
- English-GUM (Zeldes, 2017)
- English-LitBank (Bamman et al., 2019)
- English-ParCorFull (Lapshinova-Koltunski et al., 2018)
- French-ANCOR (Muzerelle et al., 2014)
- French-Democrat (Landragin, 2021)
- German-ParCorFull (Lapshinova-Koltunski et al., 2018)
- German-PotsdamCC (Bourgonje and Stede, 2020)
- Hindi-HDTB (Mujadia et al., 2016)
- Hungarian-KorKor (Vadász, 2022)
- Hungarian-SzegedKoref (Vincze et al., 2018)
- Korean-ECMT (Nam et al., 2020)
- Lithuanian-LCC (Žitkus and Butkienė, 2018)
- Norwegian-BokmaalNARC (Mæhlum et al., 2022)
- Norwegian-NynorskNARC (Mæhlum et al., 2022)
- Old Church Slavonic-PROIEL (Haug and Jøhndal, 2008)
- Polish-PCC (Ogrodniczuk et al., 2013)
- Russian-RuCor (Toldova et al., 2014)
- Spanish-AnCora (Recasens and Martí, 2010)
- Turkish-ITCC (Pamay and Eryiğit, 2018)

Non-public

- Dutch-COREA (Hendrickx et al., 2008)
- English-ARRAU (Uryupina et al., 2020)
- English-OntoNotes (Weischedel et al., 2011)
- English-PCEDT (Nedoluzhko et al., 2016)

Shared Task Data: content differences

- non-public datasets excluded
- ParCorFull excluded in 2025
  - the smallest dataset
  - the largest variance across training runs

Shared Task Data: content differences

CRAC’25 Shared Task Data

Public

- Ancient Greek-PROIEL (Haug and Jøhndal, 2008)
- Ancient Hebrew-PTNK (Swanson et al., 2024)
- Catalan-AnCora (Recasens and Martí, 2010)
- Czech-PCEDT (Nedoluzhko et al., 2016)
- Czech-PDT (Hajič et al., 2020)
- English-GUM (Zeldes, 2017)
- English-LitBank (Bamman et al., 2019)
- ~~English-ParCorFull (Lapshinova-Koltunski et al., 2018)~~
- French-ANCOR (Muzerelle et al., 2014)
- French-Democrat (Landragin, 2021)
- ~~German-ParCorFull (Lapshinova-Koltunski et al., 2018)~~
- German-PotsdamCC (Bourgonje and Stede, 2020)
- Hindi-HDTB (Mujadia et al., 2016)
- Hungarian-KorKor (Vadász, 2022)
- Hungarian-SzegedKoref (Vincze et al., 2018)
- Korean-ECMT (Nam et al., 2020)
- Lithuanian-LCC (Žitkus and Butkienė, 2018)
- Norwegian-BokmaalNARC (Mæhlum et al., 2022)
- Norwegian-NynorskNARC (Mæhlum et al., 2022)
- Old Church Slavonic-PROIEL (Haug and Jøhndal, 2008)
- Polish-PCC (Ogrodniczuk et al., 2013)
- Russian-RuCor (Toldova et al., 2014)
- Spanish-AnCora (Recasens and Martí, 2010)
- Turkish-ITCC (Pamay and Eryiğit, 2018)

~~Non-public~~

- ~~Dutch-COREA (Hendrickx et al., 2008)~~
- ~~English-ARRAU (Uryupina et al., 2020)~~
- ~~English-OntoNotes (Weischedel et al., 2011)~~
- ~~English-PCEDT (Nedoluzhko et al., 2016)~~

Shared Task Data: annotation differences

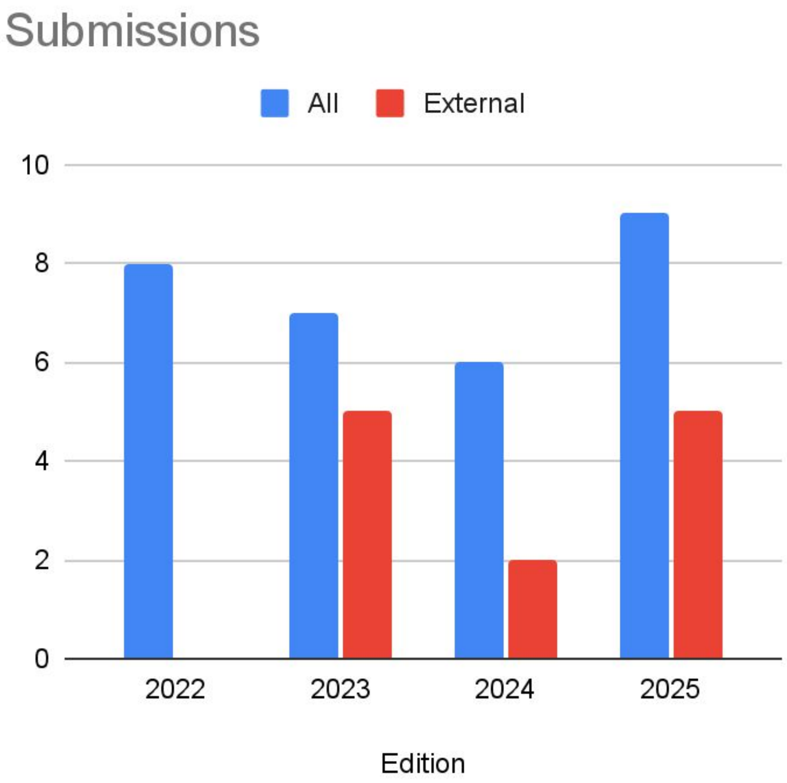
Data type	Starting point	Empty nodes	Coreference	Morpho-syntax	Forms of empty nodes
Gold (train / dev)	All	manual	manual	original (manual if available, otherwise automatic)	deleted
Input (dev / test)	Coref. and zeros from scratch	deleted	deleted	automatic UDPipe 2	deleted
	Coref. from scratch	automatic baseline	deleted	automatic UDPipe 2	deleted
	Refine the baseline	automatic baseline	automatic baseline	automatic UDPipe 2	deleted

Participants

- advertising
  - among colleagues and our students
  - ACL Portal mailing list
  - direct emails to authors of coreference-related papers (by Anja Nedoluzhko)

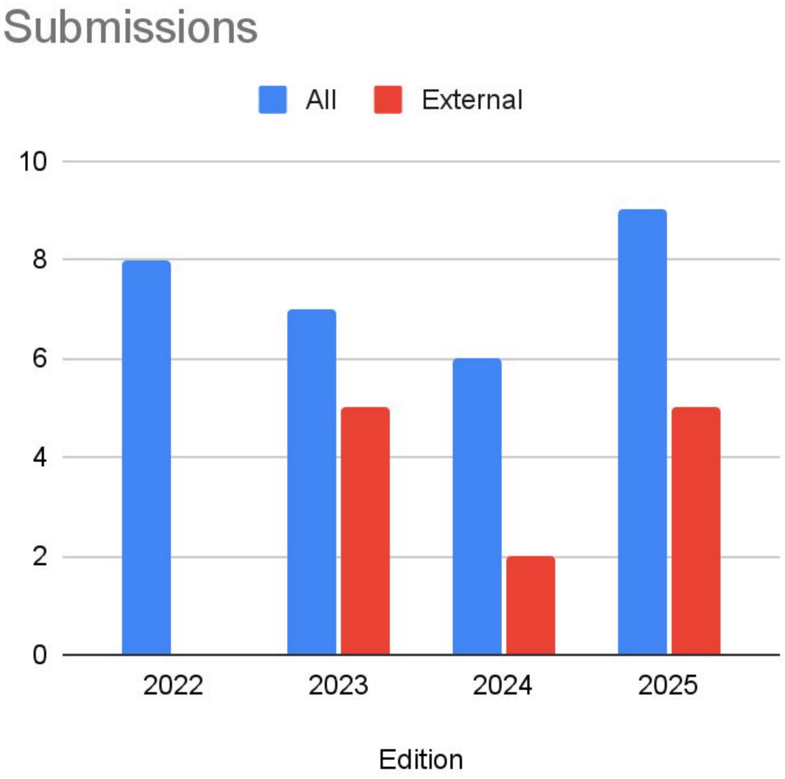
Participants

- advertising
  - among colleagues and our students
  - ACL Portal mailing list
  - direct emails to authors of coref-related papers (by Anja Nedoluzhko)
  - personally at conferences



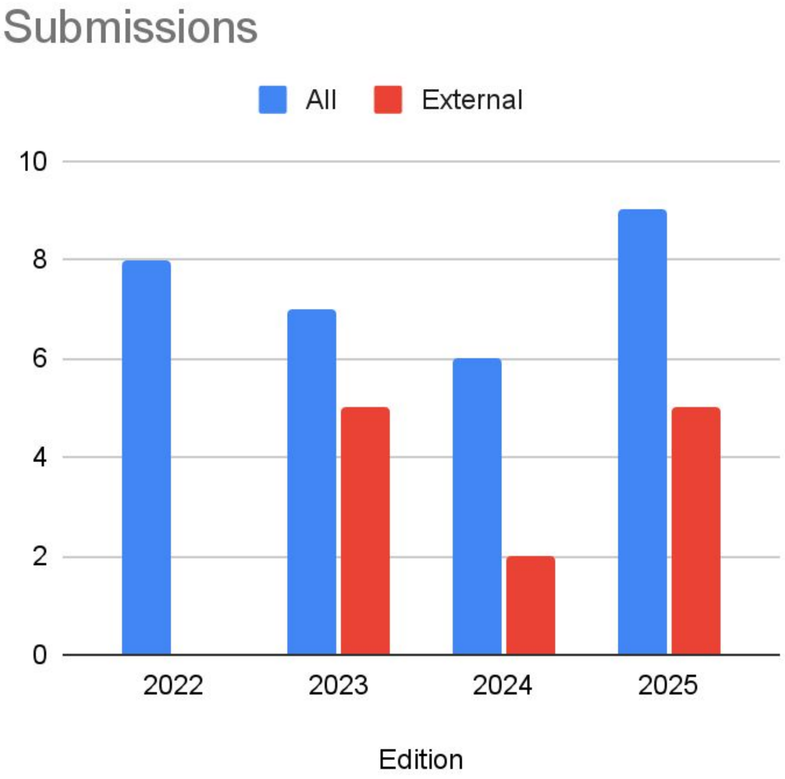
Participants

- advertising
    - among colleagues and our students
    - ACL Portal mailing list
    - direct emails to authors of coref-related papers (by Anja Nedoluzhko)
    - personally at conferences
- not easy to attract
    - CR shifted to the fringes?
    - only workshop?
    - our organization and dissemination mistakes?



Participants

- advertising
    - among colleagues and our students
    - ACL Portal mailing list
    - direct emails to authors of coref-related papers (by Anja Nedoluzhko)
    - personally at conferences
- not easy to attract
    - CR shifted to the fringes?
    - only workshop?
    - our organization and dissemination mistakes?
- still thankful to all the participants
    - Milan Straka (the winner of each edition)
    - Natalia Skachkova from DFKI
    - Ondřej Pražák from UWB



Multilingual CR Performance

- winner’s CoNLL F1

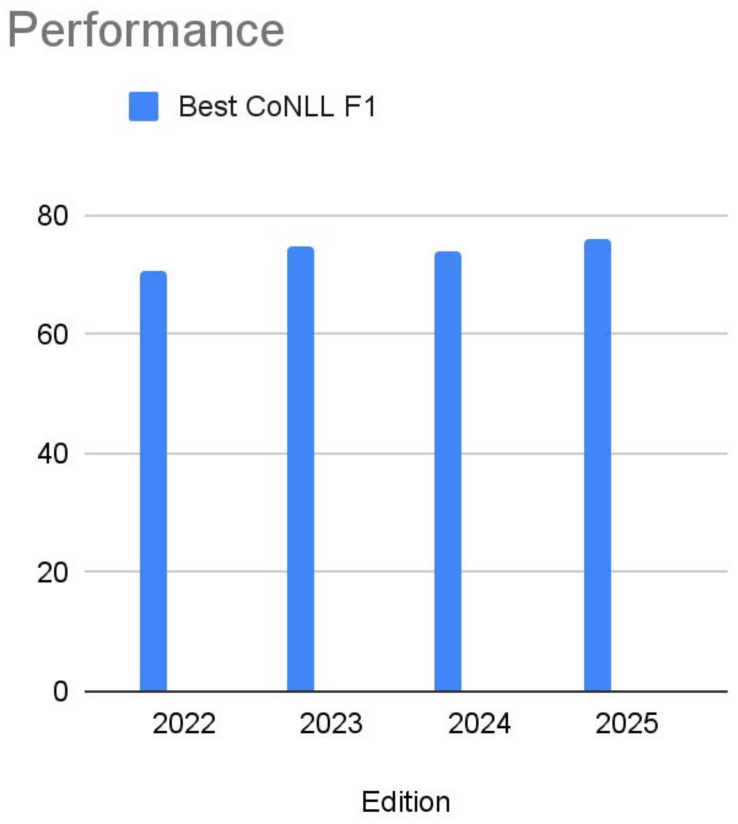
Multilingual CR Performance

- winner's CoNLL F1

CorPipe system
by Milan Straka

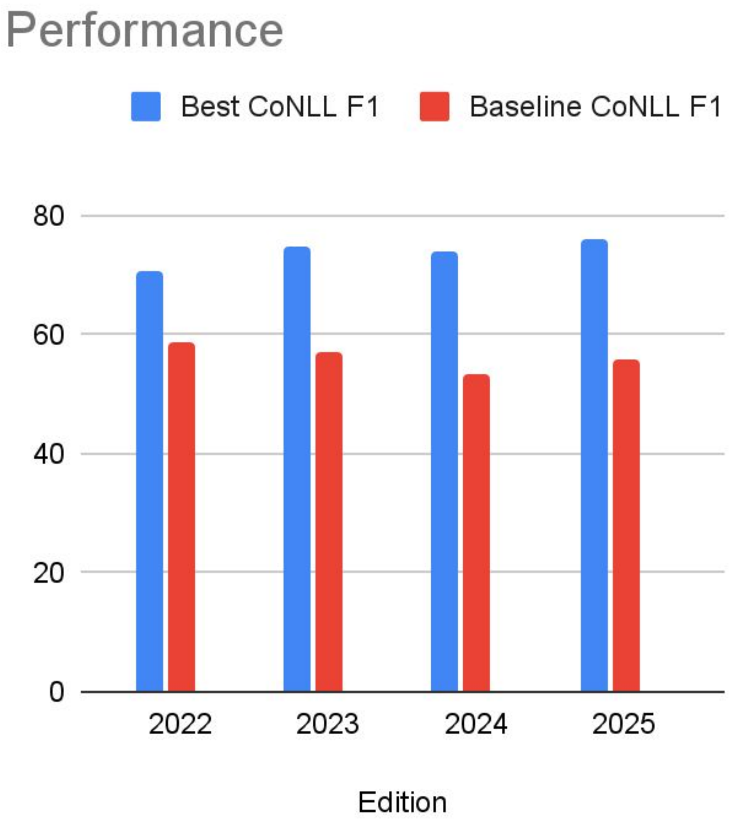
Multilingual CR Performance

- winner's CoNLL F1
    - around 70-75
    - changes in data and evaluation setup across years
    - not directly comparable



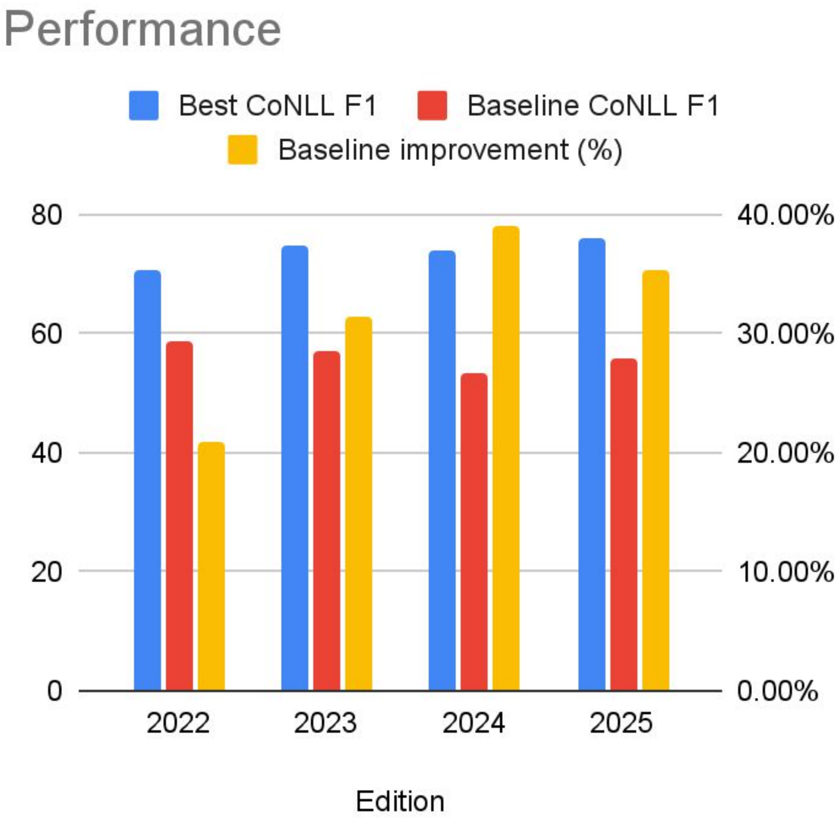
Multilingual CR Performance

- winner's CoNLL F1
  - around 70-75
  - changes in data and evaluation setup across years
  - not directly comparable
- baseline
  - still the same system, just retrained
  - the task gets more difficult



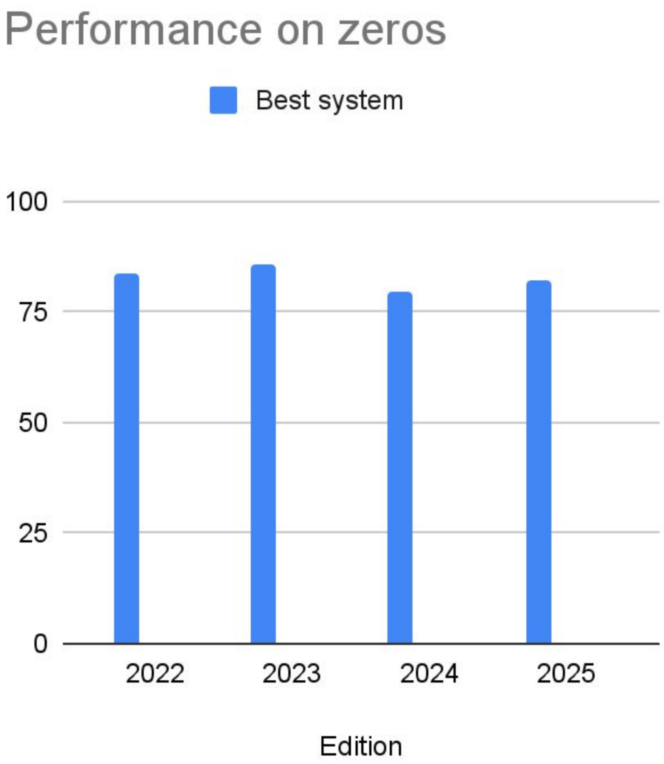
Multilingual CR Performance

- winner's CoNLL F1
  - around 70-75
  - changes in data and evaluation setup across years
  - not directly comparable
- baseline
  - still the same system, just retrained
  - the task gets more difficult
- improvement of the winner over the baseline (%)
  - fair comparison
  - +10% / year
  - 2025: decrease caused by removal of ParCorFull from the shared task data



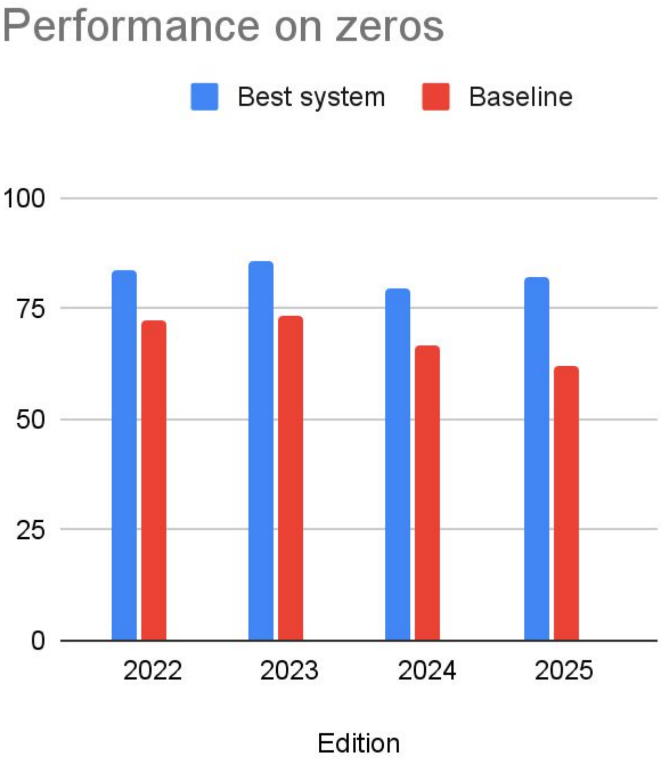
Performance on zeros

- mention-decomposable score for zeros
  - averaged over all datasets with zeros (2022: 5, 2023: 6, 2024-25: 10)
- moving to more realistic setup
  - 2024
  - drop in absolute performance



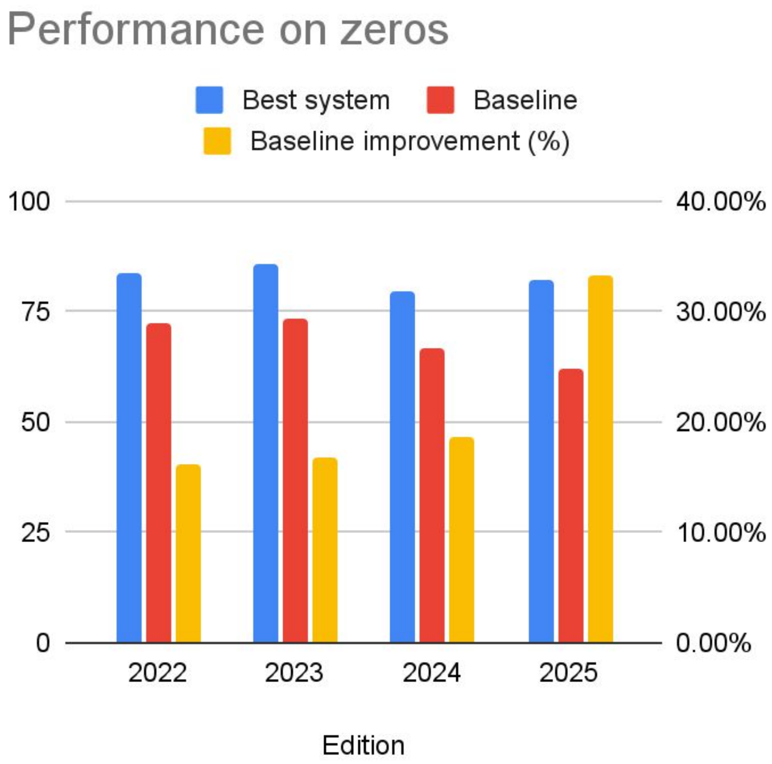
Performance on zeros

- mention-decomposable score for zeros
  - averaged over all datasets with zeros (2022: 5, 2023: 6, 2024-25: 10)
- moving to more realistic setup
  - 2024
  - drop in absolute performance



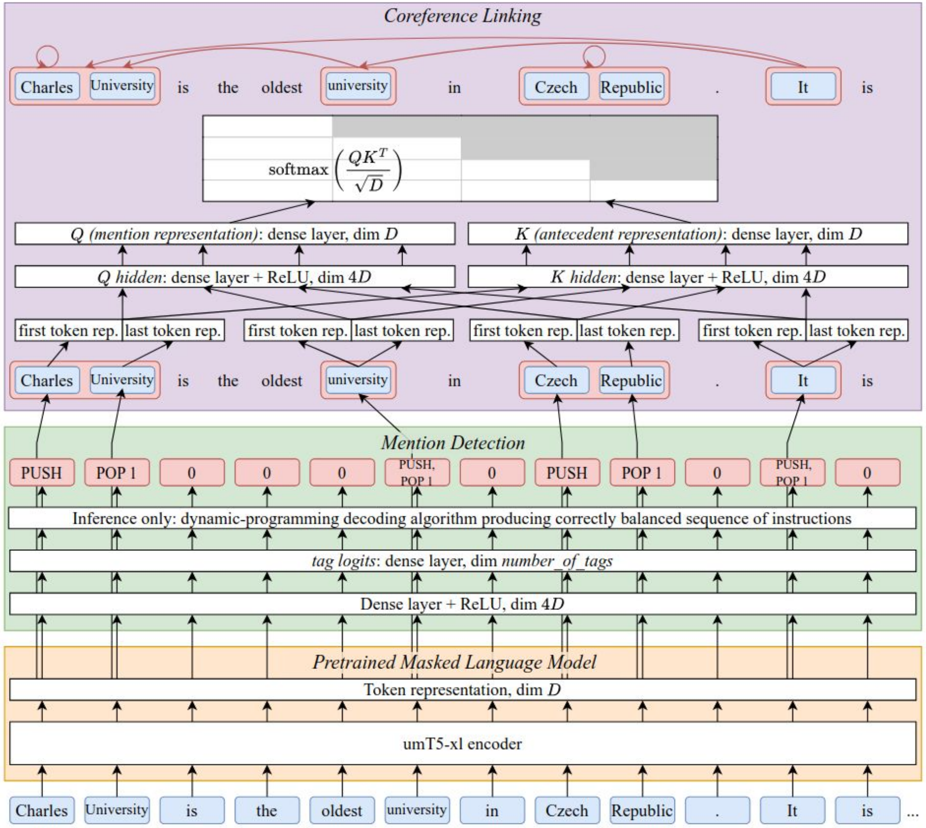
Performance on zeros

- mention-decomposable score for zeros
  - averaged over all datasets with zeros (2022: 5, 2023: 6, 2024-25: 10)
- moving to more realistic setup
  - 2024
  - drop in absolute performance
- improvement of the winner over the baseline
  - 2025: large drops in baseline performance on Czech data



Traditional approach to CR

- CorPipe by Milan Straka
- built on top of encoder-based language models
- neural architecture on top generating the annotation



Coreference Linking
Charles
University
is
the
oldest
university
in
Czech
Republic
.
It
is
softmax $\left(\frac{QK^T}{\sqrt{D}}\right)$
Q (mention representation): dense layer, dim D
K (antecedent representation): dense layer, dim D
Q hidden: dense layer + ReLU, dim 4D
K hidden: dense layer + ReLU, dim 4D
first token rep.
last token rep.
Mention Detection
PUSH
POP 1
0
PUSH, POP 1
Inference only: dynamic-programming decoding algorithm producing correctly balanced sequence of instructions
tag logits: dense layer, dim number_of_tags
Dense layer + ReLU, dim 4D
Pretrained Masked Language Model
Token representation, dim D
umT5-xl encoder
...

LLMs for Coreference

LLMs for CR

- using LLMs for coreference resolution
  - LLMs are decoder-based models generating text
- previous attempts (Gan et al., 2024, Hicke and Mimno, 2024, Le and Ritter, 2023, Saputa et al., 2024, Vadász, 2023)
  - failed to reach state-of-the-art performance
  - not tested in highly multilingual setup
  - ignoring zeros

LLM track in CRAC’25 Shared Task

- two tracks in CRAC’25 Shared task
    - LLM Track
        - LLM-based submissions
        - plaintext format with zeros (empty nodes) support
        - import/export tools
    - Unconstrained Track
        - for traditional approaches
        - same setup as in 2024
        - baseline systems provided

Data modifications due to the LLM Track

- dev and test sets capped to 25k words
  - to lower the computational cost of evaluation
  - half of the original size but affecting only a few datasets
  - variance across training runs increased only marginally

Data modifications due to the LLM Track

Public

- Ancient Greek-PROIEL (Haug and Jøhndal, 2008)
- Ancient Hebrew-PTNK (Swanson et al., 2024)
- Catalan-AnCora (Recasens and Martí, 2010)
- Czech-PCEDT (Nedoluzhko et al., 2016)
- Czech-PDT (Hajič et al., 2020)
- English-GUM (Zeldes, 2017)
- English-LitBank (Bamman et al., 2019)
- ~~English-ParCorFull (Lapshinova-Koltunski et al., 2018)~~
- French-ANCOR (Muzerelle et al., 2014)
- French-Democrat (Landragin, 2021)
- ~~German-ParCorFull (Lapshinova-Koltunski et al., 2018)~~
- German-PotsdamCC (Bourgonje and Stede, 2020)

~~Non-public~~

- ~~Dutch-COREA (Hendrickx et al., 2008)~~
- ~~English-ARRAU (Uryupina et al., 2020)~~

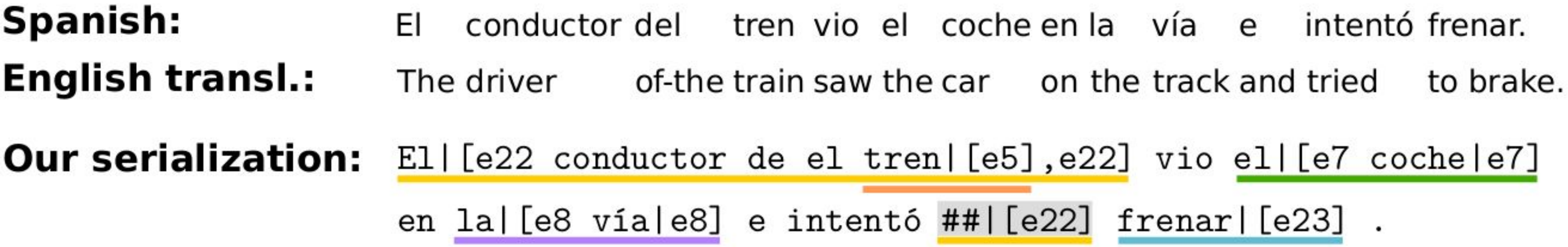
CRAC'25 Shared Task Data, capped

- Hindi-HDTB (Mujadia et al., 2016)
- Hungarian-KorKor (Vadász, 2022)
- Hungarian-SzegedKoref (Vincze et al., 2018)
- Korean-ECMT (Nam et al., 2020)
- Lithuanian-LCC (Žitkus and Butkienė, 2018)
- Norwegian-BokmaalNARC (Mæhlum et al., 2022)
- Norwegian-NynorskNARC (Mæhlum et al., 2022)
- Old Church Slavonic-PROIEL (Haug and Jøhndal, 2008)
- Polish-PCC (Ogrodniczuk et al., 2013)
- Russian-RuCor (Toldova et al., 2014)
- Spanish-AnCora (Recasens and Martí, 2010)
- Turkish-ITCC (Pamay and Eryiğit, 2018)

- ~~English-OntoNotes (Weischedel et al., 2011)~~
- ~~English-PCEDT (Nedoluzhko et al., 2016)~~

Data modifications due to the LLM Track

- plaintext format
  - mention bracketing
  - with support for zeros
  - CoNLL-U still needed for evaluation
  - import / export script available

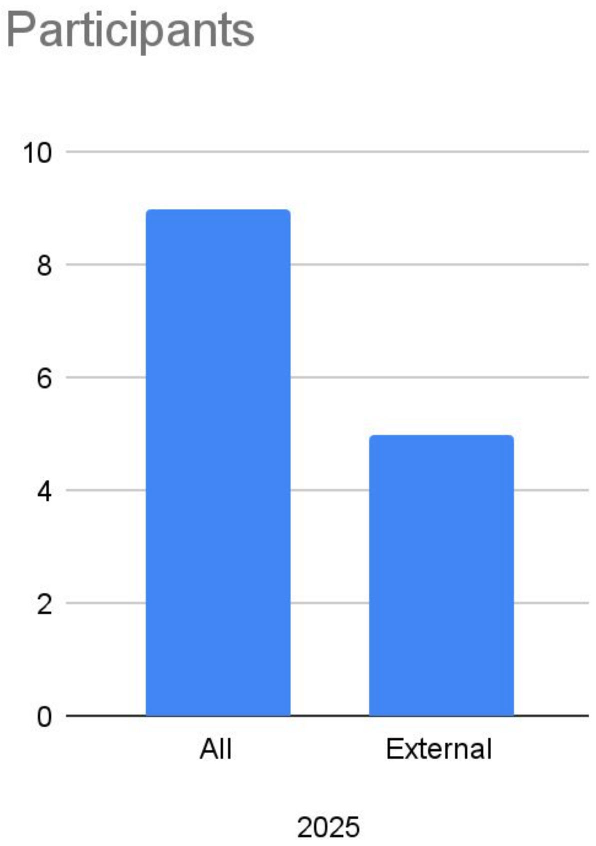


Spanish: El conductor del tren vio el coche en la vía e intentó frenar.

English transl.: The driver of-the train saw the car on the track and tried to brake.

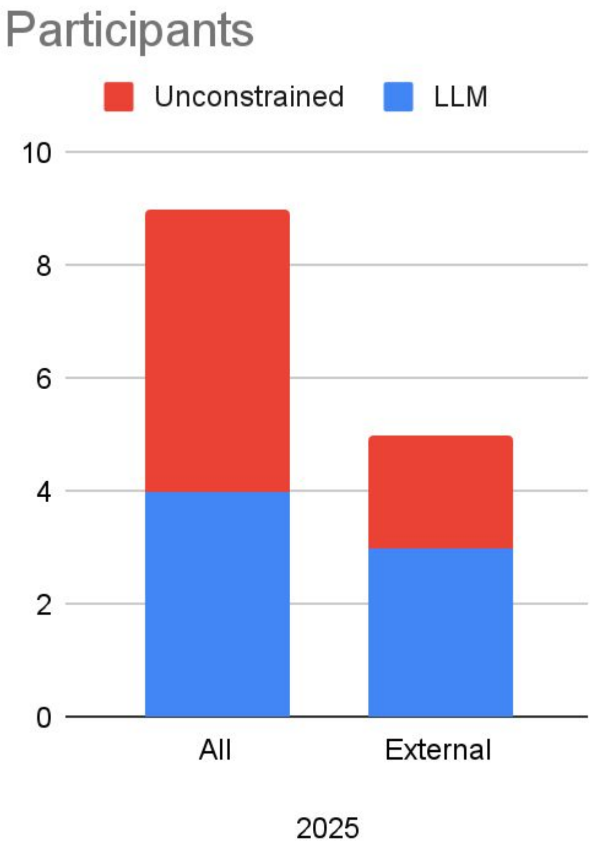
Our serialization: `El|[e22 conductor de el tren|[e5],e22] vio el|[e7 coche|e7] en la|[e8 vía|e8] e intentó ##|[e22] frenar|[e23] .`

Participants across tracks

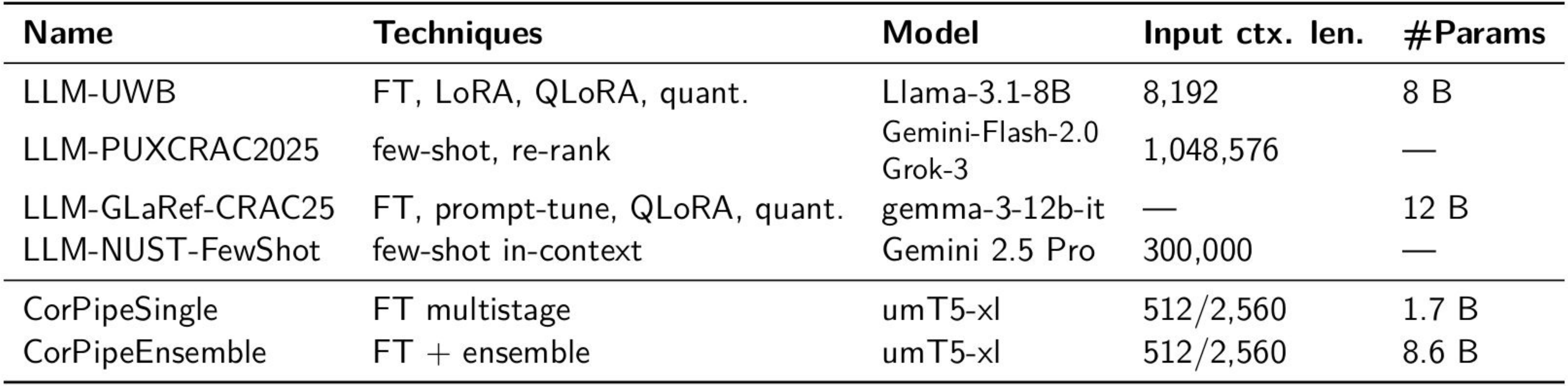


Participants across tracks

- half and half
- external participants slightly preferred the LLM track



Systems in the LLM track

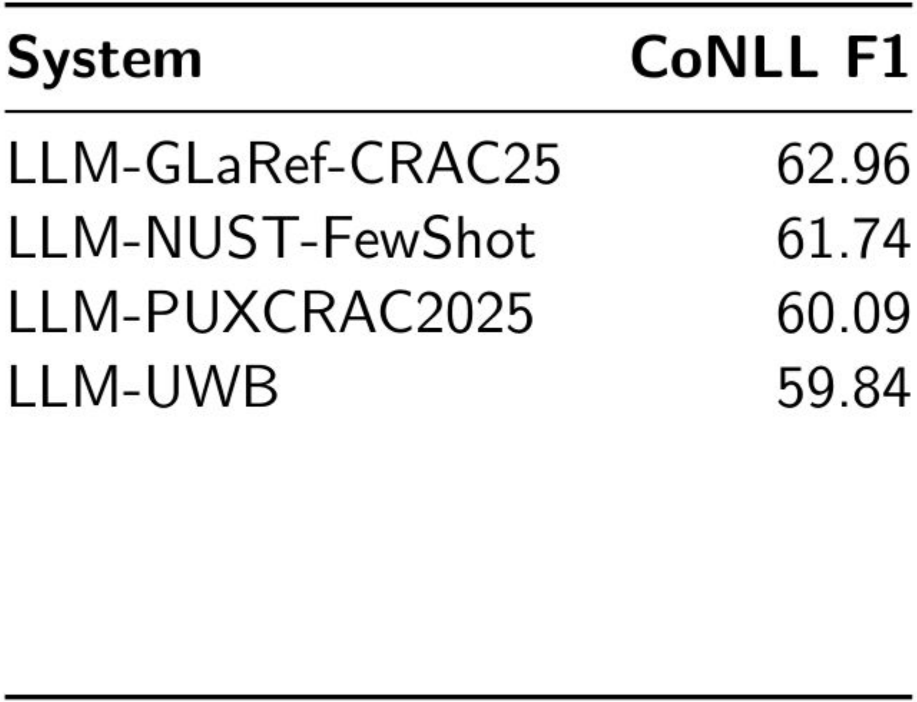


Name	Techniques	Model	Input ctx. len.	#Params
LLM-UWB	FT, LoRA, QLoRA, quant.	Llama-3.1-8B	8,192	8 B
LLM-PUXCRAC2025	few-shot, re-rank	Gemini-Flash-2.0 Grok-3	1,048,576	—
LLM-GLaRef-CRAC25	FT, prompt-tune, QLoRA, quant.	gemma-3-12b-it	—	12 B
LLM-NUST-FewShot	few-shot in-context	Gemini 2.5 Pro	300,000	—
CorPipeSingle	FT multistage	umT5-xl	512/2,560	1.7 B
CorPipeEnsemble	FT + ensemble	umT5-xl	512/2,560	8.6 B

- based on both closed-source and open-source models
- approaches:
  - few-shot prompting
  - QLoRA fine-tuning
- able to process much longer contexts at once
- models are bigger

Systems in the LLM track: performance

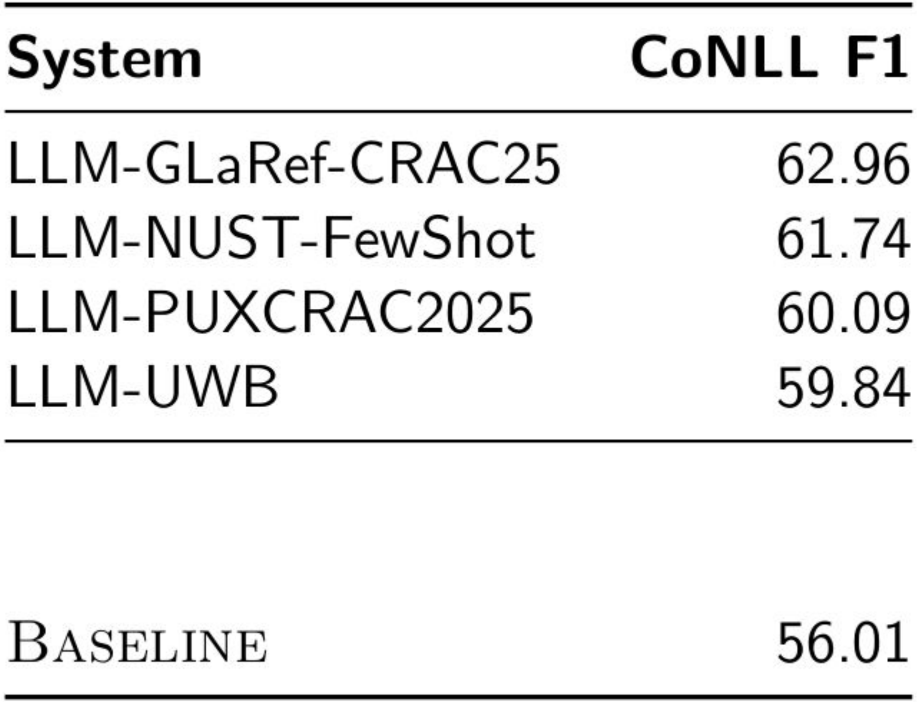
- within the LLM track: comparable performance



System	CoNLL F1
LLM-GLaRef-CRAC25	62.96
LLM-NUST-FewShot	61.74
LLM-PUXCRAC2025	60.09
LLM-UWB	59.84

Systems in the LLM track: performance

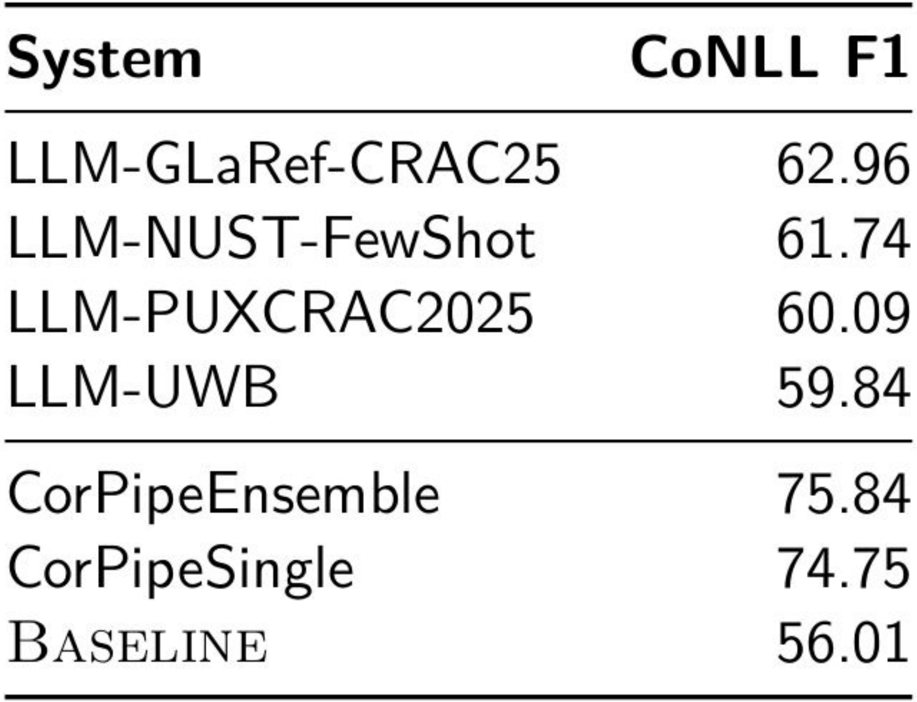
- within the LLM track: comparable performance
- outperforming baseline



System	CoNLL F1
LLM-GLaRef-CRAC25	62.96
LLM-NUST-FewShot	61.74
LLM-PUXCRAC2025	60.09
LLM-UWB	59.84
BASELINE	56.01

Systems in the LLM track: performance

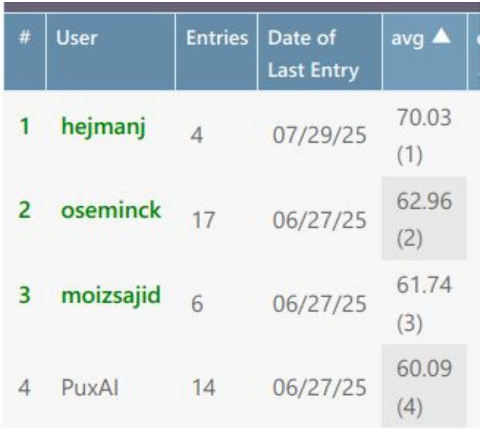
- within the LLM track: comparable performance
- outperforming baseline
- much worse than the best-performing traditional systems



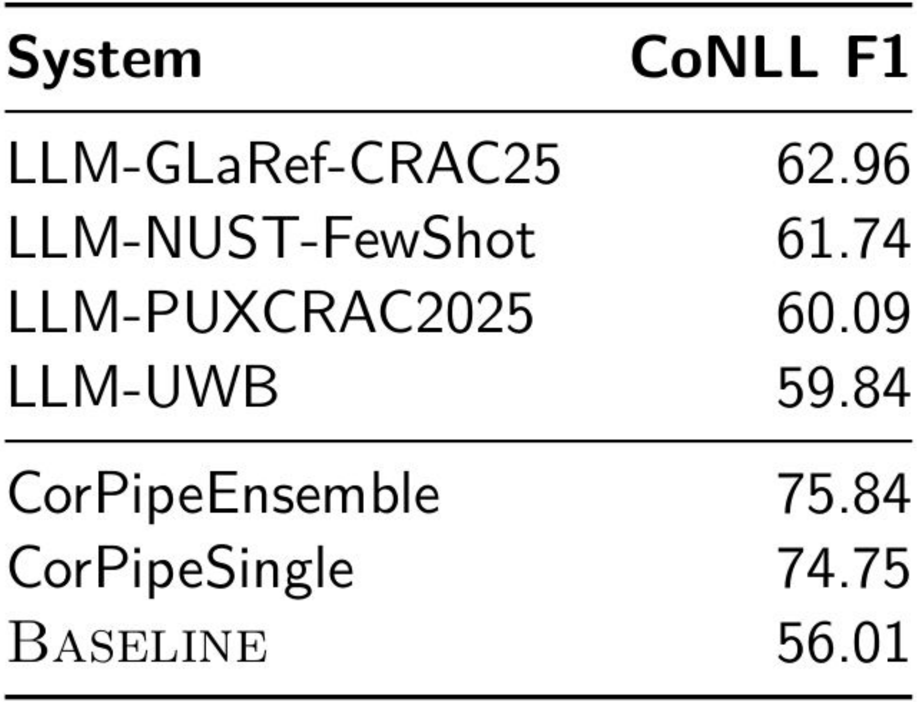
System	CoNLL F1
LLM-GLaRef-CRAC25	62.96
LLM-NUST-FewShot	61.74
LLM-PUXCRAC2025	60.09
LLM-UWB	59.84
CorPipeEnsemble	75.84
CorPipeSingle	74.75
BASELINE	56.01

Systems in the LLM track: performance

- within the LLM track: comparable performance
- outperforming baseline
- much worse than the best-performing traditional systems



#	User	Entries	Date of Last Entry	avg ▲
1	hejmanj	4	07/29/25	70.03 (1)
2	oseminck	17	06/27/25	62.96 (2)
3	moizsajid	6	06/27/25	61.74 (3)
4	PuxAI	14	06/27/25	60.09 (4)



System	CoNLL F1
LLM-GLaRef-CRAC25	62.96
LLM-NUST-FewShot	61.74
LLM-PUXCRAC2025	60.09
LLM-UWB	59.84
CorPipeEnsemble	75.84
CorPipeSingle	74.75
BASELINE	56.01

Conclusion

Conclusion

- CorefUD collection and CRAC Shared Tasks
  - standardizing multilingual coreference data
  - standardizing evaluation framework
  - turning attention to some more challenging aspects of coreference resolution, e.g. zeros
- LLMs for coreference resolution
  - encouraging experiments and research in this direction

Future Work

- CRAC 2026 Shared Task
  - will LLMs dethrone traditional approaches?
  - LLMs should be better in long-document coreference resolution
  - preparation period shortened by 4 months
- LLM-driven annotation guideline harmonization
  - with Anja Nedoluzhko and Katja Lapshinova-Koltunski

The team

Michal Novák

Martin Popel

Anna Nedoluzhko

Zdeněk Žabokrtský

Daniel Zeman

Miloslav Konopík

Ondřej Pražák

Jakub Sido

Milan Straka

Maciej Ogrodniczuk, Amir Zeldes, Barbora Dohnalová, Yilun Zhu, ...

Summary

- standardizing multilingual coreference data
- standardizing evaluation framework
- turning attention to some more challenging aspects of coreference resolution
- encouraging research on using LLMs for coreference resolution

https://ufal.cz/corefud

https://ufal.cz/corefud/crac25