Left-2-Right Dependency Parsing with Pointer Networks

Daniel Fernandez-Gonzalez, Carlos Gomez-Rodrıguez

「NAACL19」 -- Universidade da Coruña

Speaker: AntNLP(@TaoJi)

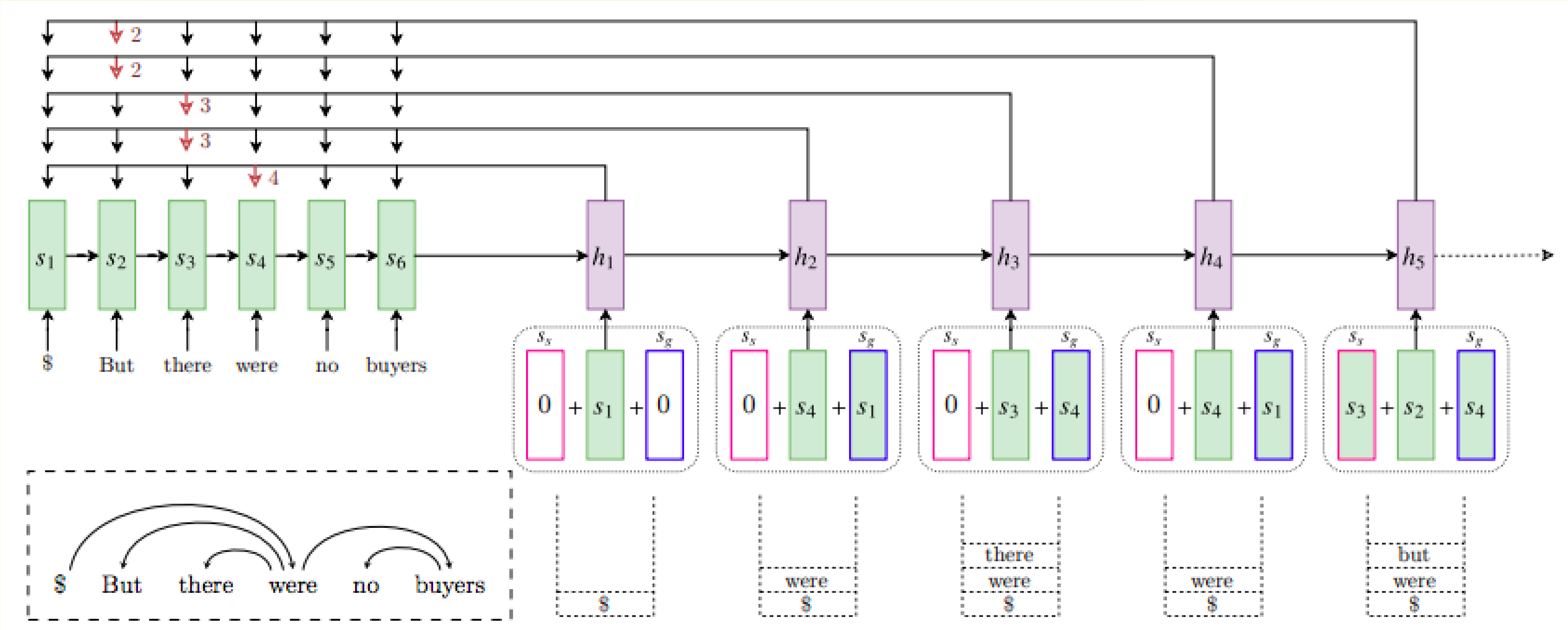
Motivation

- Their left-to-right approach is simpler than the stack-pointer parser (not requiring a stack) and reduces decoder length from $2n-1$ actions to n.

Contribution

- PTB dataset (96.04% UAS, 94.43% LAS)
- Best accuracy: $+0.17\%$ UAS, $+0.24\%$ LAS
- Ours: (95.97% UAS, 94.31% LAS)
- Speed: $10.24 \rightarrow 23.08$ sent/s

Architecture



Likelihood

$$
\begin{aligned}
P_\theta(y|x) &= \prod_{i=1}^{k} P_\theta\left(p_i|p_{<i}, x\right) \\
&= \prod_{i=1}^{k} \prod_{j=1}^{l_i} P_\theta\left(w_{i,j}|w_{i,<j}, p_{<i}, x\right)
\end{aligned}
$$

> **High-order Features:** Grandparent and Sibling.

Likelihood

$$
\begin{aligned}
P_\theta(y|x) &= \prod_{i=1}^{n} P_\theta\left(l_i | l_{<i}, x\right) \\
&= \prod_{i=1}^{n} P_\theta\left(w_h | w_i, l_{<i}, x\right)
\end{aligned}
$$

> **High-order Features:**
> Instead of using grandparent and sibling information, we just add e_{t-1}, e_{t+1} to generate d_t, which seems to be more suitable for L2R decoding.

Experiments

Parser	UAS	LAS
Chen and Manning (2014)	91.8	89.6
Dyer et al. (2015)	93.1	90.9
Weiss et al. (2015)	93.99	92.05
Ballesteros et al. (2016)	93.56	91.42
Kiperwasser and Goldberg (2016)	93.9	91.9
Alberti et al. (2015)	94.23	92.36
Qi and Manning (2017)	94.3	92.2
Fernández-G and Gómez-R (2018)	94.5	92.4
Andor et al. (2016)	94.61	92.79
Ma et al. (2018)*	95.87	94.19
This work*	**96.04**	**94.43**
Kiperwasser and Goldberg (2016)	93.1	91.0
Wang and Chang (2016)	94.08	91.82
Cheng et al. (2016)	94.10	91.49
Kuncoro et al. (2016)	94.26	92.06
Zhang et al. (2017)	94.30	91.95
Ma and Hovy (2017)	94.88	92.96
Dozat and Manning (2016)	95.74	94.08
Ma et al. (2018)*	95.84	94.21

Experiments

	Top-down		Left-to-right	
	UAS	LAS	UAS	LAS
bu	**94.42±0.02**	**90.70±0.04**	94.28±0.06	90.66±0.11
ca	93.83±0.02	91.96±0.01	**94.07±0.06**	**92.26±0.05**
cs	93.97±0.02	91.23±0.03	**94.19±0.04**	**91.45±0.05**
de	**87.28±0.07**	**82.99±0.07**	87.06±0.05	82.63±0.01
en	90.86±0.15	88.92±0.19	**90.93±0.11**	**88.99±0.11**
es	93.09±0.05	91.11±0.03	**93.23±0.03**	**91.28±0.02**
fr	**90.97±0.09**	**88.22±0.12**	90.90±0.04	88.14±0.10
it	94.08±0.04	92.24±0.06	**94.28±0.06**	**92.48±0.02**
nl	**93.23±0.09**	90.67±0.07	93.13±0.07	**90.74±0.08**
no	95.02±0.05	93.75±0.05	**95.23±0.06**	**93.99±0.07**
ro	91.44±0.11	85.80±0.14	**91.58±0.08**	**86.00±0.07**
ru	94.43±0.01	93.08±0.03	**94.71±0.07**	**93.38±0.09**

On Difficulties of Cross-Lingual Transfer with Order Differences: A Case Study on Dependency Parsing

Wasi Uddin Ahmad, Zhisong Zhang, Xuezhe Ma

Eduard Hovy, Kai-Wei Chang, Nanyun Peng

「NAACL19」 -- UC, CMU, USC

Speaker: AntNLP([@TaoJi]())

Motivation

- *order-free* vs *order-sensitive* in dep-parsing
- measure *language distance*

Conclusion

- RNN-based architectures transfer well to languages that are close to English.
- Self-attentive models have better overall cross-lingual transferability and perform especially well on distant languages.

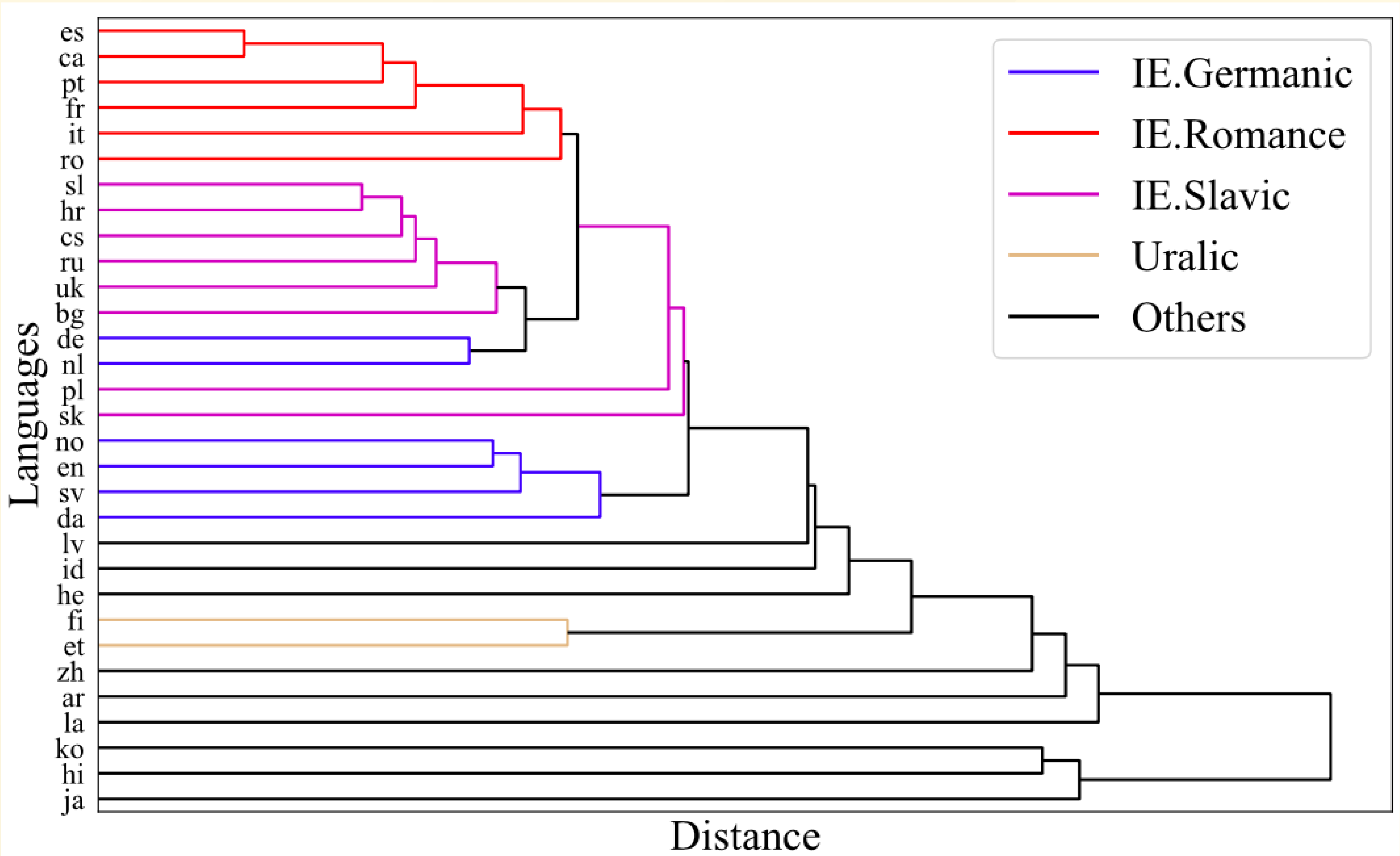
Lang. Distance

1. The World Atlas of Language Structures (WALS)
    - word order typology
2. Empirical Way
    - (modifier_upos, head_upos, dep_rel)
    - 52 selected arc types $\rightarrow$ feature vector
    - hierarchical clustering (Manhattan distance)

Lang. Distance

Language Families	Languages
Afro-Asiatic	Arabic (ar), Hebrew (he)
Austronesian	Indonesian (id)
IE.Baltic	Latvian (lv)
IE.Germanic	Danish (da), Dutch (nl), English (en), German (de), Norwegian (no), Swedish (sv)
IE.Indic	Hindi (hi)
IE.Latin	Latin (la)
IE.Romance	Catalan (ca), French (fr), Italian (it), Portuguese (pt), Romanian (ro), Spanish (es)
IE.Slavic	Bulgarian (bg), Croatian (hr), Czech (cs), Polish (pl), Russian (ru), Slovak (sk), Slovenian (sl), Ukrainian (uk)
Japanese	Japanese (ja)
Korean	Korean (ko)
Sino-Tibetan	Chinese (zh)
Uralic	Estonian (et), Finnish (fi)

Lang. Distance



Contextual Encoders

1. Deep-BiLSTM
2. Transformer
    - absolute $\rightarrow$ relative position
    - order-agnostic

Transformer

Input: $x_1, \cdots, x_n$ **Output:** $z_1, \cdots, z_n$

$$z_i = \sum_{j=1}^{n} \alpha_{ij} \left(x_j W^V + a_{ij}^V \right)$$

Weight: $\alpha_{ij} = \frac{\exp e_{ij}}{\sum_{k=1}^{n} \exp e_{ik}}$

Score: $e_{ij} = \frac{x_i W^Q \left(x_j W^K + a_{ij}^K \right)^T}{\sqrt{d_z}}$

Posi. emb: $a_{ij}^K = w_{clip(|j-i|,k)}^K$

Structured Decoders

1. Stack-Pointer Decoder
    - *order-sensitive*
2. Graph-based Decoder
    - *order-free*

Settings

- **Source:** English
- **Target:** 30 other languages
- **Pre-trained emb:** multi-lingual $300d$ FastText
- *order-free:* SelfAtt , Graph
- *order-sensitive:* RNN, Stack

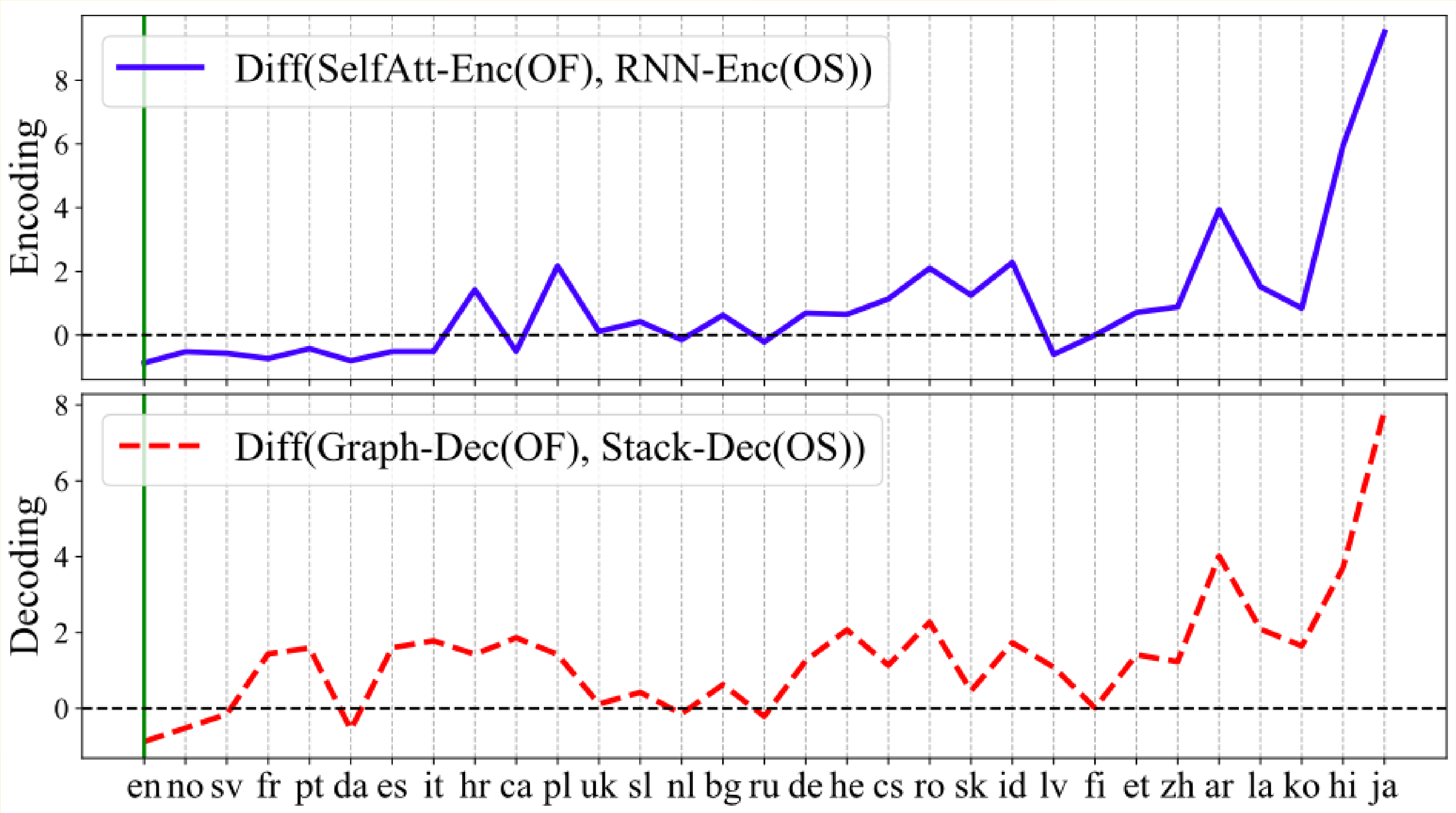
Lang	Dist. to English	SelfAtt-Graph (OF-OF)	RNN-Graph (OS-OF)	SelfAtt-Stack (OF-OS)	RNN-Stack (OS-OS)	Baseline (Guo et al., 2015)	Supervised (RNN-Graph)
en	0.00	90.35/88.40	90.44/88.31	90.18/88.06	**91.82**†/**89.89**†	87.25/85.04	90.44/88.31
no	0.06	80.80/72.81	80.67/72.83	80.25/72.07	**81.75**†/**73.30**†	74.76/65.16	94.52/92.88
sv	0.07	80.98/73.17	81.23/73.49	80.56/72.77	**82.57**†/**74.25**†	71.84/63.52	89.79/86.60
fr	0.09	77.87/72.78	**78.35**†/**73.46**†	76.79/71.77	75.46/70.49	73.02/64.67	91.90/89.14
pt	0.09	**76.61**†/67.75	76.46/**67.98**	75.39/66.67	74.64/66.11	70.36/60.11	93.14/90.82
da	0.10	76.64/67.87	77.36/68.81	76.39/67.48	**78.22**†/**68.83**	71.34/61.45	87.16/84.23
es	0.12	74.49/66.44	**74.92**†/**66.91**†	73.15/65.14	73.11/64.81	68.75/59.59	93.17/90.80
it	0.12	80.80/75.82	**81.10**/**76.23**†	79.13/74.16	80.35/75.32	75.06/67.37	94.21/92.38
hr	0.13	**61.91**†/**52.86**†	60.09/50.67	60.58/51.07	60.80/51.12	52.92/42.19	89.66/83.81
ca	0.13	73.83/65.13	**74.24**†/**65.57**†	72.39/63.72	72.03/63.02	68.23/58.15	93.98/91.64
pl	0.13	**74.56**†/**62.23**†	71.89/58.59	73.46/60.49	72.09/59.75	66.74/53.40	94.96/90.68
uk	0.13	**60.05**/**52.28**†	58.49/51.14	57.43/49.66	59.67/51.85	54.10/45.26	85.98/82.21
sl	0.13	**68.21**†/**56.54**†	66.27/54.57	66.55/54.58	67.76/55.68	60.86/48.06	86.79/82.76
nl	0.14	68.55/60.26	67.88/60.11	67.88/59.46	**69.55**†/**61.55**†	63.31/53.79	90.59/87.52
bg	0.14	**79.40**†/**68.21**†	78.05/66.68	78.16/66.95	78.83/67.57	73.08/61.23	93.74/89.61
ru	0.14	60.63/51.63	59.99/50.81	59.36/50.25	**60.87**/**51.96**	55.03/45.09	94.11/92.56
de	0.14	**71.34**†/**61.62**†	69.49/59.31	69.94/60.09	69.58/59.64	65.14/54.13	88.58/83.68
he	0.14	**55.29**/**48.00**†	54.55/46.93	53.23/45.69	54.89/40.95	46.03/26.57	89.34/84.49
cs	0.14	**63.10**†/**53.80**†	61.88/52.80	61.26/51.86	62.26/52.32	56.15/44.77	94.03/91.87
ro	0.15	**65.05**†/**54.10**†	63.23/52.11	62.54/51.46	60.98/49.79	56.01/44.04	90.07/84.50
sk	0.17	**66.65**/**58.15**†	65.41/56.98	65.34/56.68	66.56/57.48	57.75/47.73	90.19/86.38
id	0.17	**49.20**†/**43.52**†	47.05/42.09	47.32/41.70	46.77/41.28	40.84/33.67	87.19/82.60
lv	0.18	70.78/49.30	**71.43**†/**49.59**	69.04/47.80	70.56/48.53	62.33/41.42	83.67/78.13
fi	0.20	66.27/48.69	**66.36**/**48.74**	64.82/47.50	66.25/48.28	58.51/38.65	88.04/85.04
et	0.20	**65.72**†/**44.87**†	65.25/44.40	64.12/43.26	64.30/43.50	56.13/34.86	86.76/83.28
zh*	0.23	**42.48**†/**25.10**†	41.53/24.32	40.56/23.32	40.92/23.45	40.03/20.97	73.62/67.67
ar	0.26	**38.12**†/**28.04**†	32.97/25.48	32.56/23.70	32.85/24.99	32.69/22.68	86.17/81.83
la	0.28	**47.96**†/**35.21**†	45.96/33.91	45.49/33.19	43.85/31.25	39.08/26.17	81.05/76.33
ko	0.33	**34.48**†/**16.40**†	33.66/15.40	32.75/15.04	33.11/14.25	31.39/12.70	85.05/80.76
hi	0.40	**35.50**†/**26.52**†	29.32/21.41	31.38/23.09	25.91/18.07	25.74/16.77	95.63/92.93
ja*	0.49	**28.18**†/**20.91**†	18.41/11.99	20.72/13.19	15.16/9.32	15.39/08.41	89.06/78.74
Average	0.17	**64.06**†/**53.82**†	62.71/52.63	62.22/52.00	62.37/51.89	57.09/45.41	89.44/85.62

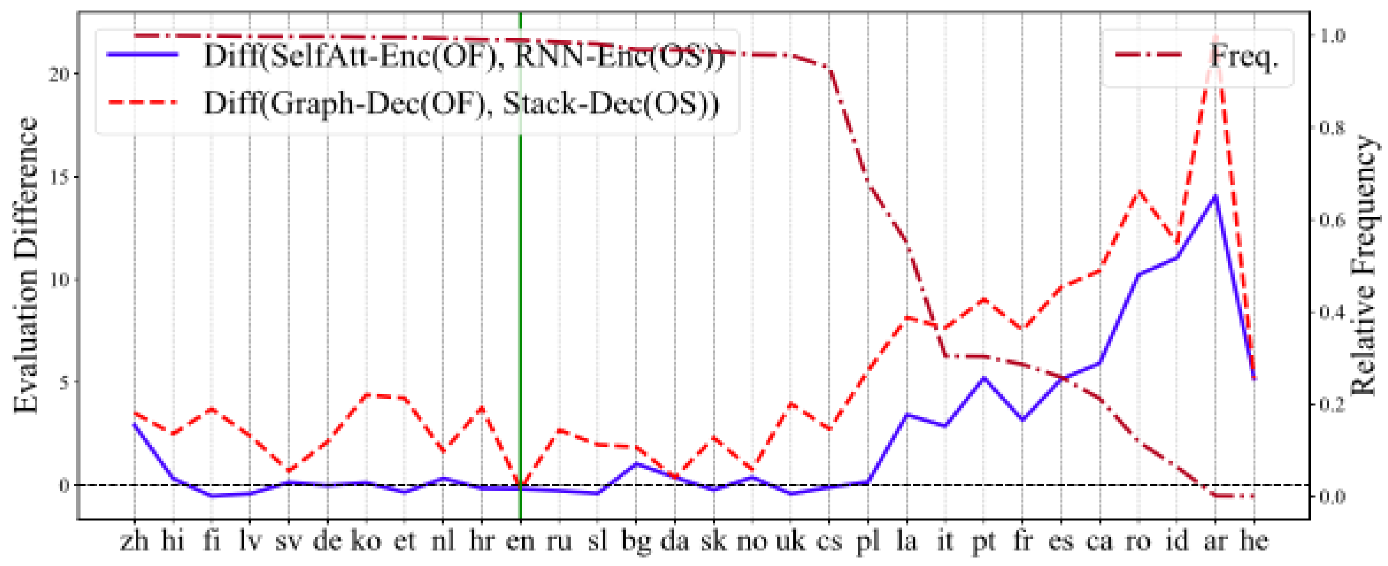
Experiments

Model	UAS%	LAS%
SelfAtt-Relative (Ours)	64.57	54.14
SelfAtt-Relative+Dir	63.93	53.62
RNN	63.25	52.94
SelfAtt-Absolute	61.76	51.71
SelfAtt-NoPosi	28.18	21.45

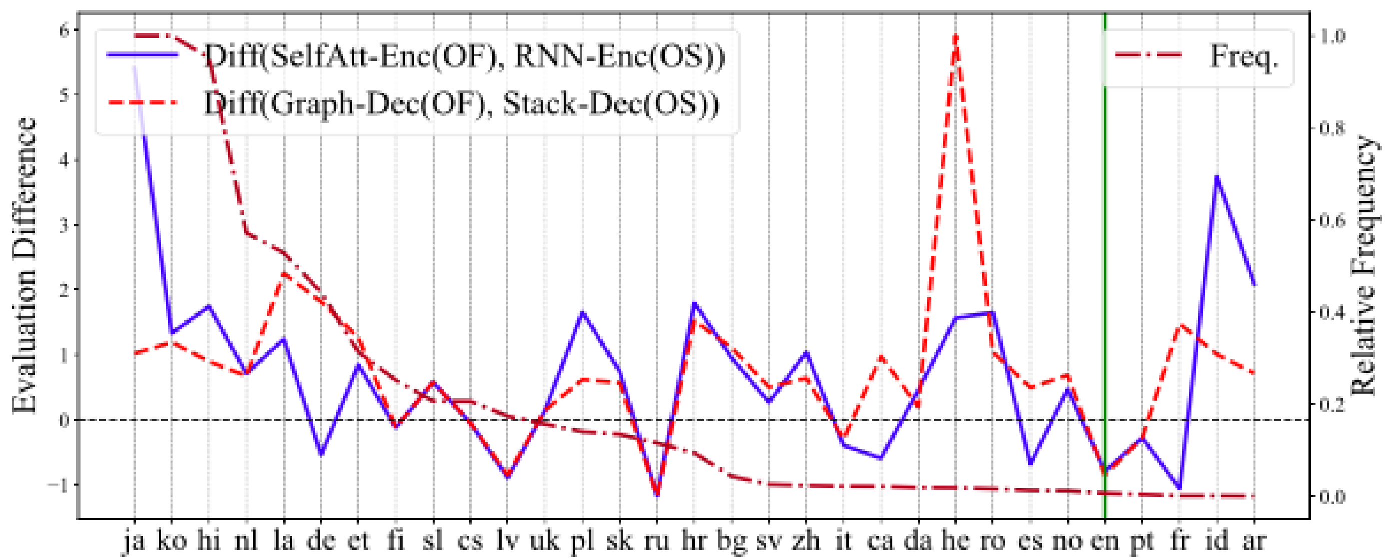
Table 3: Comparisons of different encoders (averaged results over all languages on the original training sets).

Experiments





(b) Adjective & Noun (ADJ, NOUN, amod)



(d) Object & Verb (NOUN, VERB, obj)

Thank, you!

Q&A