A Mention-Ranking Model for Abstract Anaphora Resolution

Ana Marasović, Leo Born, Juri Opitz and **Anette Frank**

Research Training Group AIPHES

Department of Computational Linguistics

Heidelberg University

AIPHES

ADAPTIVE PREPARATION OF INFORMATION FROM HETEROGENEOUS SOURCES

UNIVERSITÄT HEIDELBERG ZUKUNFT SEIT 1386

Abstract Anaphora Resolution

Resolution of (unrestricted) anaphors – nominal or pronominal – that refer to **abstract objects**: propositions, facts, events or properties.

The research of Iran Human Rights shows 34 people were hanged in public in Iran in 2016, and **an audience of hundreds of people, including children, were present for most of these hangings**. Human rights activists and informed members of civil society have always severely *criticized* **this issue**. / **this.**

unrestricted abstract anaphora resolution

Abstract Anaphora Resolution

ENTITY ANAPHORA RESOLUTION (COREFERENCE RESOLUTION)	ABSTRACT ANAPHORA RESOLUTION
resolving multiple ambiguous mentions of a single entity representing a person, a location or an organization	resolution of anaphoric expressions that refer to propositions, facts, events or properties
standard features: agreement, apposition, saliency, etc.	standard features for resolution of entity anaphora do not apply
considerable amounts of annotated training data	lack of sufficient amounts of annotated training data

Abstract Anaphora Resolution

ENTITY ANAPHORA RESOLUTION (COREFERENCE RESOLUTION)	ABSTRACT ANAPHORA RESOLUTION
resolving multiple ambiguous mentions of a single entity representing a person, a location or an organization	resolution of anaphoric expressions that refer to propositions, facts, events or properties
standard features: agreement, apposition, saliency, etc.	standard features for resolution of entity anaphora do not apply
considerable amounts of annotated training data	lack of sufficient amounts of annotated training data

NEURAL MODEL

GENERATE TRAINING DATA

We address resolution of **unrestricted** abstract anaphora with **artificially created training data** and a **neural model.**

Related work

task	example	data size	neural	note
event coreference (Lu and Ng, 2017)	Police said Lo Presti had hanged himself. His suicide appeared to be related to clan feuds.	9955 event coreference chains (KBP, Eng)	YES	coreference between VP and NP mentions of similar abstractness

Related work

task	example	data size	neural	note
event coreference (Lu and Ng, 2017)	Police said Lo Presti had hanged himself. His suicide appeared to be related to clan feuds.	9955 event coreference chains (Wng)	YES	coreference between VP and NP mentions of similar abstractness
sluicing (Anand and Hardt, 2016)	Harry traveled to southern Denmark to study botany. I want to know why.	4100 examples	NO	small data, not published

Related work
– shell noun resolution (Kolhatkar et al., 2013)

Environmental Defense notes that **mowing the lawn with a gas mower produces as much pollution as driving a car 172 miles**. **This fact** may explain the recent surge in the sales of old-fashioned push mowers.

Anaphoric Shell Noun (ASN)

(*fact, reason, issue, question*, …)

Related work

task	example	data size	neural	note
event coreference (Lu and Ng, 2017)	Police said Lo Presti had hanged himself. His suicide appeared to be related to clan feuds.	9955 event coreference chains (Eng)	YES	coreference between VP and NP mentions of similar abstractness
sluicing (Anand and Hardt, 2016)	Harry traveled to southern Denmark to study botany. I want to know why.	4100 examples	NO	small data, not published
anaphoric connectives (Stede and Grishina, 2016)	~~Peter was the best goal scorer.~~ Therefore he received the trophy.	140 instances (therefore)	NO	restricted in type, ambiguous and require WSD

Related work
– shell noun resolution (Kolhatkar et al., 2013)

Environmental Defense notes that **mowing the lawn with a gas mower produces as much pollution as driving a car 172 miles**. **This fact** may explain the recent surge in the sales of old-fashioned push mowers.

Anaphoric Shell Noun (ASN)

Congress has focused almost solely on **the fact** that **special education is expensive - and that it takes away money from regular education**.

Cataphoric Shell Noun (CSN)

syntactic rules: N-to, N-to-be, N-that, etc.

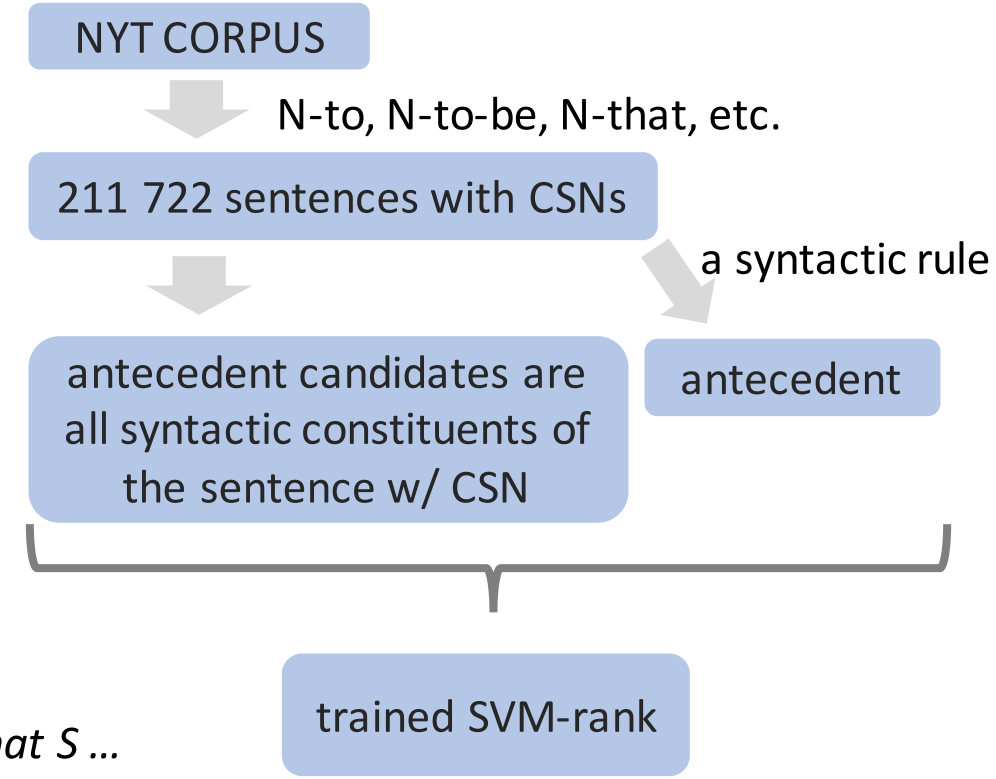
special education is expensive - and that it takes away money from regular education

antecedent

Congress has focused almost solely on **this fact**.

sentence with the shell noun

Related work
– shell noun resolution (Kolhatkar et al., 2013)

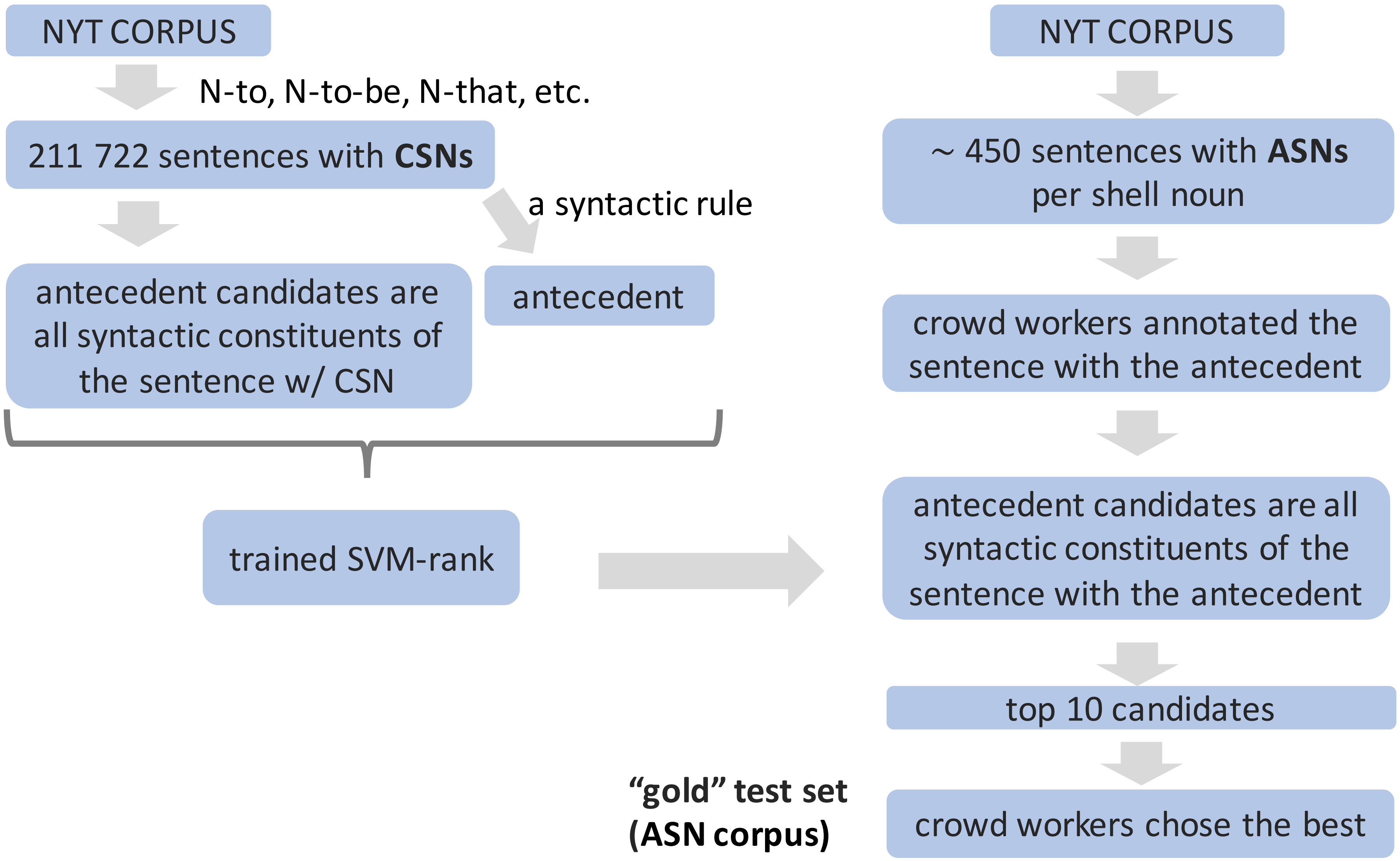


CSN:
… *the* ***fact*** *that S* …

ASN:
… *pointed to* ***this fact*** …

assumption: linguistic knowledge encoded in CSN antecedents will help in interpreting ASNs
=>
apply the SVM-rank model trained on CSN data to **predict ASN antecedents** as well

Related work
– shell noun resolution (Kolhatkar et al., 2013)



Related work

KOLHAKTAR ET AL. (2013)	OUR WORK
shell noun resolution	Unrestricted abstract anaphora (AA) resolution (nominal and pronominal)
data generation method depends on properties and categorization of shell nouns	a common syntactic construction
feature-based ranking model	neural ranking model

Intuitions – for our model

Antecedent

The research of Iran Human Rights shows 34 people were hanged in public in Iran in 2016, and **an audience of hundreds of people, including children, were present for most of these hangings**.
Human rights activists and informed members of civil society have always severely criticized **this issue**.

Anaphoric Sentence

Abstract Anaphor (AA)

Something that human right activists may criticize

How can we learn what is the correct antecedent for a given AA?

Our intuition: by ***learning the relation*** between

Anaphoric Sentence w/ AA & Antecedent

*Human rights activists … criticized **AA***

… hundreds of people, including children, were present for these hangings.

Intuitions – for training data generation

Antecedent

The research of Iran Human Rights shows 34 people were hanged in public in Iran in 2016, and **an audience of hundreds of people, including children, were present for most of these hangings**.
Human rights activists and informed members of civil society have always severely *criticized* **this issue**.

Anaphoric Sentence

Abstract Anaphor (AA)

We can extract such Antecedent – Anaphoric Sentence pairs automatically from **constructions with embedded sentences**, by a simple transformation:

Human rights activists and informed members of civil society have always severely *criticized* **[$_{S'}$ that [$_{S}$ an audience of hundreds of people, including children, were present for most of these hangings]].**

→

Human rights activists and informed members of civil society have always severely *criticized* **this / this issue.**
An audience of hundreds of people, including children, were present for most of these hangings.

Training data generation

type	head of S’	possible anaphoric phrase
empty	Ø	**this**, that
general	that, this	that, this
causal	because, as	therefore, because of this/that
temporal	while, since, etc.	during this/that
conditional	if, whether	if this/that is true

Complements

He ***doubts*** [**S’** Ø [**S** a ~~Bismarckian super state will emerge that would dominate Europe~~]], but warns of “a risk of profound change in the heart of the European Community from a Germany that is too strong, even if democratic”].

[**S** a Bismarckian super state will emerge that would dominate Europe]

Training data generation

type	head of S’	possible anaphoric phrase
empty	Ø	**this**, that
general	that, this	that, this
causal	because, as	therefore, because of this/that
temporal	while, since, etc.	during this/that
conditional	if, whether	if this/that is true

Complements

He ***doubts*** [**S’** Ø [**S** this]], but warns of “a risk of profound change in the heart of the European Community from a Germany that is too strong, even if democratic”].

[**S** a Bismarckian super state will emerge that would dominate Europe]

Training data generation

type	**head of S'**	**possible anaphoric phrase**
empty	Ø	**this, that**
general	that, this	that, this
causal	because, as	therefore, because of this/that
temporal	while, since, etc.	during this/that
conditional	if, whether	if this/that is true

Complements

He ***doubts*** **this,** but warns of "a risk of profound change in the heart of the European Community from a Germany that is too strong, even if democratic".

Anaphoric Sentence

A Bismarckian super state will emerge that would dominate Europe.

Antecedent

Training data generation

type	head of S’	possible anaphoric phrase
empty	Ø	this, that
general	that, this	that, this
causal	because, **as**	therefore, **because of this/that**
temporal	while, since, etc.	during this/that
conditional	if, whether	if this/that is true

Adjuncts

There is speculation that property casualty firms will sell even more munis [**S‘** ***as*** [~~**S** they scramble to raise cash to pay claims related to Hurrican Hugo and the Northern California earthquake~~]].

[**S** they scramble to raise cash to pay claims related to Hurrican Hugo and the Northern California earthquake]

Training data generation

type	head of S’	possible anaphoric phrase
empty	Ø	this, that
general	that, this	that, this
causal	because, **as**	therefore, **because of this/that**
temporal	while, since, etc.	during this/that
conditional	if, whether	if this/that is true

There is speculation that property casualty firms will sell even more munis [**S‘** ***because of* this**].

[**S** they scramble to raise cash to pay claims related to Hurrican Hugo and the Northern California earthquake]

Training data generation

type	head of S’	possible anaphoric phrase
empty	Ø	this, that
general	that, this	that, this
causal	because, as	therefore, **because of this/that**
temporal	while, since, etc.	during this/that
conditional	if, whether	if this/that is true

Adjuncts

There is speculation that property casualty firms will sell even more munis ***because of*** **this**.

Anaphoric Sentence

They scramble to raise cash to pay claims related to Hurrican Hugo and the Northern California earthquake.

Antecedent

Training data generation

- training data for **unrestricted** abstract anaphora resolution
- obtained using a **common** construction – a verb with an embedded sentence
- **large-scale training data**
  - 15,282 instances from the WSJ part of the PTB corpus for initial experiments
  - but much more can be extracted

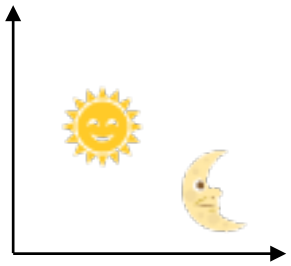
Siamese-LSTM Mention-ranking model

the highest scoring candidate is
the predicted antecedent

score that characterizes a relation between the
candidate and the anaphoric sentence

combine individual representations in a joint
representation and feed it to the FFL2

representation produced by the
bi-LSTM and the FFL1



representation produced by the
bi-LSTM and the FFL1

candidate

sentence with the anaphor

architecture trained w/
max-margin objective
(Clark and Manning, 2015)

$$score(cand, AnaphS) =$$
$$W h_{cand,\, AnaphS} + b \;\in \mathbb{R}$$

$$h_{cand,AnaphS} =$$
$$conc\begin{pmatrix} |h_{cand} - h_{AnaphS}|; \\ h_{cand} \odot h_{AnaphS} \end{pmatrix}$$

Siamese-LSTM Mention-ranking model – input

an audience of hundreds of people, including children, were present for most of these hangings

candidate

input to LSTM

emb(token)

Human rights activists and informed members of civil society have always severely criticized **this issue**.

sentence with the anaphor (AnaphS)

input to LSTM

emb(token)

Siamese-LSTM Mention-ranking model – input

an audience of hundreds of people, including children, were present for most of these hangings

candidate

input to LSTM

emb(token)	emb(AA-head)	emb(AA-ctx)

Human rights activists and informed members of civil society have always severely criticized **this issue**.

sentence with the anaphor (AnaphS)

input to LSTM

emb(token)	emb(AA-head)	emb(AA-ctx)

The anaphoric sentence may have more than one abstract anaphor (AA)

Siamese-LSTM Mention-ranking model – input

an audience of hundreds of people, including children, were present for most of these hangings

candidate

input to LSTM | emb(token) | emb(AA-head) | emb(AA-ctx) | emb(Syn-tag)

a constituent tag label is a good signal of worthy candidates

Human rights activists and informed members of civil society have always severely criticized **this issue**.

sentence with the anaphor (AnaphS)

input to LSTM | emb(token) | emb(AA-head) | emb(AA-ctx) | emb(S-tag)

The anaphoric sentence may have more than one abstract anaphor (AA)

Experiment 1: Shell noun resolution

	train	test
fact	43 809	472
reason	4 529	442
issue	2 664	303
decision	42 289	389
question	9 327	440
possibility	11 874	277

Datasets

- train data*: generated with resolution of CSNs
- test data*: anaphoric shell noun dataset annotated with crowd workers (the ASN corpus)
- dev data: a small-scaled subset of the ARRAU corpus (Uryupina et al., 2016) restricted to unconstrained abstract anaphors (ARRAU-AA)

* Obtained from Kolhatkar et al. 2013

Exp1: Shell noun resolution

	train	test	model
fact	43 809	472	MR-LSTM
	-	-	**KZH13**
	-	-	TAG-BL
reason	4 529	442	MR-LSTM
	4 529	442	MR-LSTM-tune
	-	-	**KZH13**
	-	-	TAG-BL
issue	2 664	303	MR-LSTM
	-	-	**KZH13**
	-	-	TAG-BL
decision	42 289	389	MR-LSTM
	-	-	**KZH13**
	-	-	TAG-BL
question	9 327	440	MR-LSTM
	-	-	**KZH13**
	-	-	TAG-BL
possibility	11 874	277	MR-LSTM
	-	-	**KZH13**
	-	-	TAG-BL

Baselines

- Kolhatkar et al. (2013) – **KZH13**

Exp1: Shell noun resolution

	train	test	model
fact	43 809	472	MR-LSTM
	-	-	KZH13
	-	-	**TAG-BL**
reason	4 529	442	MR-LSTM
	4 529	442	MR-LSTM-tune
	-	-	KZH13
	-	-	**TAG-BL**
issue	2 664	303	MR-LSTM
	-	-	KZH13
	-	-	**TAG-BL**
decision	42 289	389	MR-LSTM
	-	-	KZH13
	-	-	**TAG-BL**
question	9 327	440	MR-LSTM
	-	-	KZH13
	-	-	**TAG-BL**
possibility	11 874	277	MR-LSTM
	-	-	KZH13
	-	-	**TAG-BL**

Baselines

- Kolhatkar et al. (2013) – KZH13
- the **TAG baseline** randomly chooses a candidate with the tag in {S, VP, ROOT, SBAR}

Exp1: Shell noun resolution

	train	test	model	s@1	s@2	s@3	s@4
fact	43 809	472	MR-LSTM				
	-	-	KZH13				
	-	-	TAG-BL				
reason	4 529	442	MR-LSTM				
	4 529	442	MR-LSTM-tune				
	-	-	KZH13				
	-	-	TAG-BL				
issue	2 664	303	MR-LSTM				
	-	-	KZH13				
	-	-	TAG-BL				
decision	42 289	389	MR-LSTM				
	-	-	KZH13				
	-	-	TAG-BL				
question	9 327	440	MR-LSTM				
	-	-	KZH13				
	-	-	TAG-BL				
possibility	11 874	277	MR-LSTM				
	-	-	KZH13				
	-	-	TAG-BL				

Evaluation metrics:

success@n (s@n) :

the antecedent or a candidate that differs in *one word* or *one word and punctuation* is in the first n ranked candidates, $n \in \{1, 2, 3, 4\}$

Exp1: Shell noun resolution

	train	test	model	s@1	s@2	s@3	s@4
fact	43 809	472	**MR-LSTM**	**83.47**	85.38	86.44	87.08
	-	-	KZH13	70.00	**86.00**	**92.00**	**95.00**
	-	-	TAG-BL	46.99	-	-	-
reason	4 529	442	MR-LSTM	71.27	77.38	80.09	80.54
	4 529	442	**MR-LSTM-tune**	**87.78**	**91.63**	**93.44**	93.89
	-	-	KZH13	72.00	86.90	90.00	**94.00**
	-	-	TAG-BL	42.40	-	-	-
issue	2 664	303	**MR-LSTM**	**88.12**	**91.09**	**93.07**	**93.40**
	-	-	KZH13	47.00	61.00	72.00	81.00
	-	-	TAG-BL	44.92	-	-	-
decision	42 289	389	**MR-LSTM**	**76.09**	**85.86**	**91.00**	**93.06**
	-	-	KZH13	35.00	53.00	67.00	76.00
	-	-	TAG-BL	45.55	-	-	-
question	9 327	440	**MR-LSTM**	**89.77**	**94.09**	**95.00**	**95.68**
	-	-	KZH13	70.00	83.00	88.00	91.00
	-	-	TAG-BL	42.02	-	-	-
possibility	11 874	277	**MR-LSTM**	**93.14**	**94.58**	**95.31**	**95.67**
	-	-	KZH13	56.00	76.00	87.00	92.00
	-	-	TAG-BL	48.66	-	-	-

- MR-LSTM outperforms KZH13’s and TAG-BL for 5/6 shell nouns without HP tuning!

Exp1: Shell noun resolution

	train	test	model	s@1	s@2	s@3	s@4
fact	43 809	472	**MR-LSTM**	**83.47**	85.38	86.44	87.08
	-	-	KZH13	70.00	**86.00**	**92.00**	**95.00**
	-	-	TAG-BL	46.99	-	-	-
reason	4 529	442	MR-LSTM	71.27	77.38	80.09	80.54
	4 529	442	**MR-LSTM-tune**	**87.78**	**91.63**	**93.44**	93.89
	-	-	KZH13	72.00	86.90	90.00	**94.00**
	-	-	TAG-BL	42.40	-	-	-
issue	2 664	303	**MR-LSTM**	**88.12**	**91.09**	**93.07**	**93.40**
	-	-	KZH13	47.00	61.00	72.00	81.00
	-	-	TAG-BL	44.92	-	-	-
decision	42 289	389	**MR-LSTM**	**76.09**	**85.86**	**91.00**	**93.06**
	-	-	KZH13	35.00	53.00	67.00	76.00
	-	-	TAG-BL	45.55	-	-	-
question	9 327	440	**MR-LSTM**	**89.77**	**94.09**	**95.00**	**95.68**
	-	-	KZH13	70.00	83.00	88.00	91.00
	-	-	TAG-BL	42.02	-	-	-
possibility	11 874	277	**MR-LSTM**	**93.14**	**94.58**	**95.31**	**95.67**
	-	-	KZH13	56.00	76.00	87.00	92.00
	-	-	TAG-BL	48.66	-	-	-

with HP tuning

- MR-LSTM outperforms KZH13's and TAG-BL for 5/6 shell nouns without HP tuning!
- tuned MR-LSTM results for *reason* well beyond KZH13

Exp1: Shell noun resolution – tuning for *reason*

REASON

CTX-AA	AA	TAG	SHORTCUT	FFL1	FFL2	**s@1**	**s@2**	**s@3**	**s@4**
✓	✓	✓	✓	✓	✓	**87.78**	**91.63**	**93.44**	93.89
✗						85.97	87.56	89.14	89.82
	✗					86.65	88.91	91.18	91.40
		✗	✗			68.10	80.32	85.29	89.37
			✗			85.52	88.24	89.59	90.05
✗	✗	✗	✗			66.97	80.54	85.75	88.24
				✗		87.56	91.62	92.76	**94.12**
					✗	85.97	88.69	89.14	90.05

- a large performance drop when omitting syntactic info (tag, cut)
  - the model makes good use of syntactic info or it fits the bias in the tag distribution

Experiment2: unrestricted abstract anaphora VS. constrained shell noun resolution

		shell noun resolution		unrestricted AA resolution	
		CSN	ASN	ARTIFICIAL	ARRAU-AA
		train	test	train	test
# of shell nouns / anaphors		2 664 – 43 809	277 - 472	8 527	600
median # of tokens	Antec	12.75	13.87	11	20.5
	AnaphS	11.5	24	19	28
median #	Antec	2	**4.5**	2	**1**
	negatives	44.5	39	15	48
#	nominal	all	all	none	397
	pronominal	none	none	all	203

Exp2: unrestricted abstract anaphora VS. constrained shell noun resolution

		shell nouns resolution		unrestricted AA resolution	
		CSN	**ASN**	**ARTIFICIAL**	**ARRAU-AA**
		train	test	train	test
# of shell nouns / anaphors		2 664 – 43 809	277 - 472	8 527	600
median # of tokens	Antec	12.75	13.87	11	20.5
	AnaphS	11.5	24	19	28
median #	Antec	2	4.5	2	1
	negatives	44.5	39	15	48
#	nominal	all	all	none	**397**
	pronominal	none	none	all	**203**

Exp2: Unrestricted abstract anaphora resolution

	all				nominal				pronominal			
	s@1	**s@2**	**s@3**	**s@4**	**s@1**	**s@2**	**s@3**	**s@4**	**s@1**	**s@2**	**s@3**	**s@4**
PS-BL	27.67	-	-	-	30.48	-	-	-	22.17	-	-	-
TAG-BL	38.43	-	-	-	40.10	-	-	-	35.17	-	-	-

Baselines

- PS-BL: the preceding sentence baseline
- TAG-BL: randomly choses a candidate with the tag in {S, VP, ROOT, SBAR}

Exp2: Unrestricted abstract anaphora resolution

						all				nominal				pronominal			
CTX	**AA**	**TAG**	**CUT**	**FFL1**	**FFL2**	**s@1**	**s@2**	**s@3**	**s@4**	**s@1**	**s@2**	**s@3**	**s@4**	**s@1**	**s@2**	**s@3**	**s@4**
✓	✓	✓	✓	✓	✓	24.17	43.67	54.50	63.00	29.47	50.63	62.47	72.04	13.79	30.05	38.92	45.32
PS-BL						27.67	-	-	-	30.48	-	-	-	22.17	-	-	-
TAG-BL						**38.43**	-	-	-	40.10	-	-	-	**35.17**	-	-	-

- the full architecture does not outperform baselines
- MR-LSTM resolves nominal anaphors better than pronominal

Exp2: Unrestricted abstract anaphora resolution

						all				nominal				pronominal			
CTX	AA	TAG	CUT	FFL1	FFL2	s@1	s@2	s@3	s@4	s@1	s@2	s@3	s@4	s@1	s@2	s@3	s@4
✓	✓	✓	✓	✓	✓	24.17	43.67	54.50	63.00	29.47	50.63	62.47	72.04	13.79	30.05	38.92	45.32
✗						29.67	52.50	66.00	75.00	33.50	58.19	72.04	80.86	22.17	41.38	54.19	63.55
	✗					22.83	39.00	52.00	61.33	22.42	41.31	54.66	64.48	23.65	34.48	46.80	55.17
		PS-BL				27.67	-	-	-	30.48	-	-	-	22.17	-	-	-
		TAG-BL				**38.43**	-	-	-	**40.10**	-	-	-	**35.17**	-	-	-

- info about context of the anaphor lowers results,
- but the **embedding of the head of the anaphor** is important
  - and: more useful for **nominals** compared to pronouns

Exp2: Unrestricted abstract anaphora resolution

						all				nominal				pronominal			
CTX	**AA**	**TAG**	**CUT**	**FFL1**	**FFL2**	**s@1**	**s@2**	**s@3**	**s@4**	**s@1**	**s@2**	**s@3**	**s@4**	**s@1**	**s@2**	**s@3**	**s@4**
✓	✓	✓	✓	✓	✓	24.17	43.67	54.50	63.00	29.47	50.63	62.47	72.04	13.79	30.05	38.92	45.32
✗						29.67	52.50	66.00	75.00	33.50	58.19	72.04	80.86	22.17	41.38	54.19	63.55
	✗					22.83	39.00	52.00	61.33	22.42	41.31	54.66	64.48	23.65	34.48	46.80	55.17
		✗	✗			38.33	54.83	63.17	69.33	**46.60**	64.48	72.54	79.09	22.17	35.96	44.83	50.25
PS-BL						27.67	-	-	-	30.48	-	-	-	22.17	-	-	-
TAG-BL						**38.43**	-	-	**-**	40.10	-	-	-	**35.17**	-	-	-

- **omitting syntactic info** helps and **raises MR-LSTM results for nominals above both BLs!**

Exp2: Unrestricted abstract anaphora resolution

						all				nominal				pronominal			
CTX	**AA**	**TAG**	**CUT**	**FFL1**	**FFL2**	**s@1**	**s@2**	**s@3**	**s@4**	**s@1**	**s@2**	**s@3**	**s@4**	**s@1**	**s@2**	**s@3**	**s@4**
✓	✓	✓	✓	✓	✓	24.17	43.67	54.50	63.00	29.47	50.63	62.47	72.04	13.79	30.05	38.92	45.32
✗						29.67	52.50	66.00	**75.00**	33.50	58.19	72.04	80.86	22.17	41.38	54.19	63.55
	✗					22.83	39.00	52.00	61.33	22.42	41.31	54.66	64.48	23.65	34.48	46.80	55.17
		✗	✗			38.33	54.83	63.17	69.33	46.60	**64.48**	72.54	79.09	22.17	35.96	44.83	50.25
shuffling +		✗	✗			**43.83**	**56.33**	**66.33**	73.00	**51.89**	**64.48**	**73.55**	**79.85**	28.08	40.39	52.22	59.61
TAG-BL						38.43	-	-	-	40.10	-	-	-	**35.17**	-	-	-

- additional shuffling of training data boosts results significantly
- we found model variants that **surpass the baselines** for the <u>entire</u> and the <u>nominal part</u> of ARRAU-AA
- results of resolution of shell nouns were in the range 76.09 - 93.14 s@1

Exp2: Unrestricted abstract anaphora resolution

						all				nominal				pronominal			
CTX	**AA**	**TAG**	**CUT**	**FFL1**	**FFL2**	**s@1**	**s@2**	**s@3**	**s@4**	**s@1**	**s@2**	**s@3**	**s@4**	**s@1**	**s@2**	**s@3**	**s@4**
✓	✓	✓	✓	✓	✓	24.17	43.67	54.50	63.00	29.47	50.63	62.47	72.04	13.79	30.05	38.92	45.32
✗						29.67	52.50	66.00	**75.00**	33.50	58.19	72.04	80.86	22.17	41.38	54.19	63.55
	✗					22.83	39.00	52.00	61.33	22.42	41.31	54.66	64.48	23.65	34.48	46.80	55.17
		✗	✗			38.33	54.83	63.17	69.33	46.60	**64.48**	72.54	79.09	22.17	35.96	44.83	50.25
shuffling +		✗	✗			**43.83**	**56.33**	**66.33**	73.00	**51.89**	**64.48**	**73.55**	**79.85**	28.08	40.39	52.22	59.61
TAG-BL						38.43	-	-	-	40.10	-	-	-	**35.17**	-	-	-

- our model selects syntactically plausible candidates and – if disregarding syntax – discriminates candidates using deeper features

Exp2: Unrestricted abstract anaphora resolution

						all				nominal				pronominal			
CTX	**AA**	**TAG**	**CUT**	**FFL1**	**FFL2**	**s@1**	**s@2**	**s@3**	**s@4**	**s@1**	**s@2**	**s@3**	**s@4**	**s@1**	**s@2**	**s@3**	**s@4**
✓	✓	✓	✓	✓	✓	24.17	43.67	54.50	63.00	29.47	50.63	62.47	72.04	13.79	30.05	38.92	45.32
✗						29.67	52.50	66.00	**75.00**	33.50	58.19	72.04	80.86	22.17	41.38	54.19	63.55
	✗					22.83	39.00	52.00	61.33	22.42	41.31	54.66	64.48	23.65	34.48	46.80	55.17
		✗	✗			38.33	54.83	63.17	69.33	46.60	**64.48**	72.54	79.09	22.17	35.96	44.83	50.25
shuffling +		✗	✗			**43.83**	**56.33**	**66.33**	73.00	**51.89**	**64.48**	**73.55**	**79.85**	28.08	40.39	52.22	59.61
w/ the best HPs for pronominal		✗	✗			38.17	52.50	61.33	68.67	43.07	57.43	65.49	72.04	28.57	**42.86**	**53.20**	**62.07**
		PS-BL				27.67	-	-	-	30.48	-	-	-	22.17	-	-	-
		TAG-BL				38.43	-	-	**-**	40.10	-	-	-	**35.17**	-	-	-

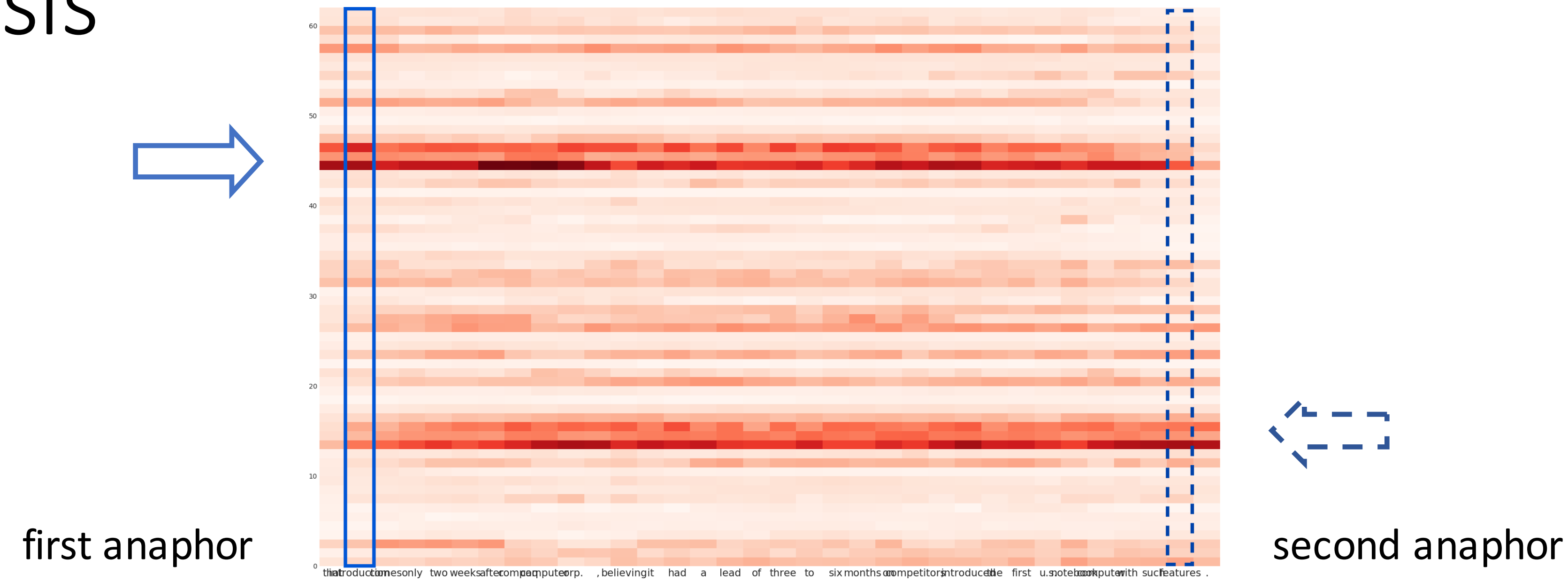
- HPs that yield good performance for nominal anaphors are not good for pronominal ones and vice versa

Exp2: Unrestricted abstract anaphora resolution

						all				nominal				pronominal			
CTX	AA	TAG	CUT	FFL1	FFL2	s@1	s@2	s@3	s@4	s@1	s@2	s@3	s@4	s@1	s@2	s@3	s@4
✓	✓	✓	✓	✓	✓	24.17	43.67	54.50	63.00	29.47	50.63	62.47	72.04	13.79	30.05	38.92	45.32
✗						29.67	52.50	66.00	**75.00**	33.50	58.19	72.04	80.86	22.17	41.38	54.19	63.55
	✗					22.83	39.00	52.00	61.33	22.42	41.31	54.66	64.48	23.65	34.48	46.80	55.17
		✗	✗			38.33	54.83	63.17	69.33	46.60	**64.48**	72.54	79.09	22.17	35.96	44.83	50.25
		✗	✗			**43.83**	**56.33**	**66.33**	73.00	**51.89**	**64.48**	**73.55**	**79.85**	28.08	40.39	52.22	59.61
		✗	✗			38.17	52.50	61.33	68.67	43.07	57.43	65.49	72.04	28.57	**42.86**	**53.20**	**62.07**
			✗			30.17	48.00	57.83	67.33	30.73	50.88	61.21	71.54	29.06	42.36	51.23	59.11
✗	✗	✗	✗			26.33	40.50	50.67	58.67	28.46	41.81	52.14	59.70	22.17	37.93	47.78	56.65
				✗		21.33	41.17	53.17	60.33	23.43	47.36	60.45	69.52	17.24	29.06	38.92	42.36
					✗	12.00	24.67	33.50	41.50	13.35	27.20	37.28	45.84	9.36	19.70	26.11	33.00

- only the head of the anaphor, the first and the second feed-forward layer contribute

Exp2: Unrestricted abstract anaphora resolution – analysis



Does a learned representation between the anaphoric sentence and an antecedent establish a relation between a specific anaphor we want to resolve and the antecedent?

The heat-maps illustrates the difference in output of the bi-LSTM for the same anaphoric sentence with two anaphors when **the first vs. second anaphor** is considered.

Clearly, the representations differ and consequently, their joint representations with the candidate as well.

What have we learned about AA resolution?

CONTRIBUTIONS

1. first **neural** mention-ranking model for resolving **unrestricted** abstract anaphora – trained on **artificially created training data**
2. **evaluation on more realistic** and **more challenging evaluation data set** (compared to KZH13)
3. **We outperform the BLs** for the nominals in ARRAU-AA **without training models for individual anaphors**

LESSONS LEARNED

1. **nominal** and **pronominal** anaphors should be learned **independently**
2. the full model **selects syntactically plausible candidates, but** the model w/o syntax info discriminates candidates using **deeper features, with better performance**
3. embedding of the anaphor ensures that **the learned relation between antecedent and anaphoric sentence is dependent on the anaphor under consideration**

Future directions

LESSONS LEARNED

- **nominal** and **pronominal** anaphors should be learned **independently**
- the full model **selects syntactically plausible candidates, but** the model w/o syntax info discriminates candidates using **deeper features, with better performance**
- embedding of the anaphor ensures that **the learned relation between ar**
  sentence is dependent on the anaphor under consideration

1. DO THIS

2. LEARN A MODEL THAT DOES BOTH

3. investigate mixtures of data from different sources (artificial + natural)

4. offer candidates from the larger context

Thank you for your attention!

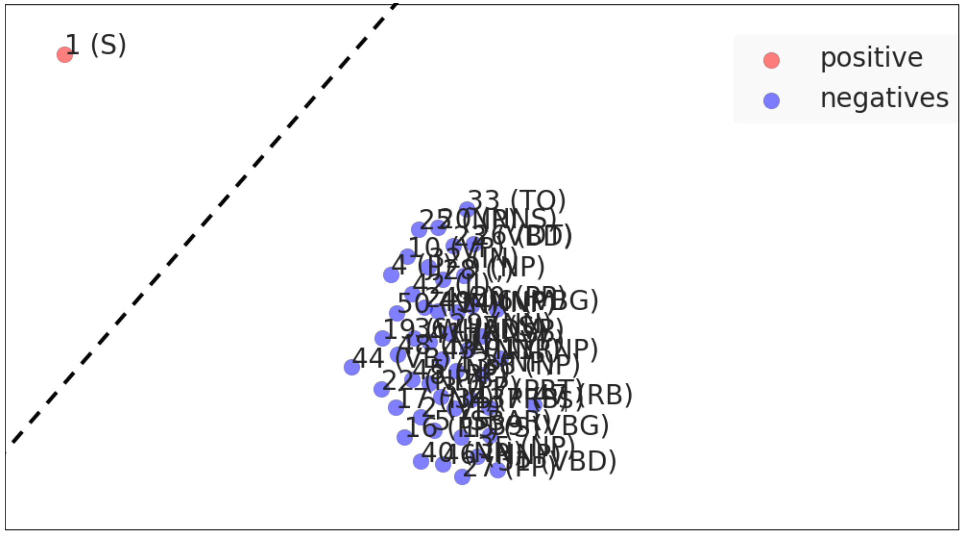
github repo with code:

https://github.com/amarasovic/neural-abstract-anaphora

References

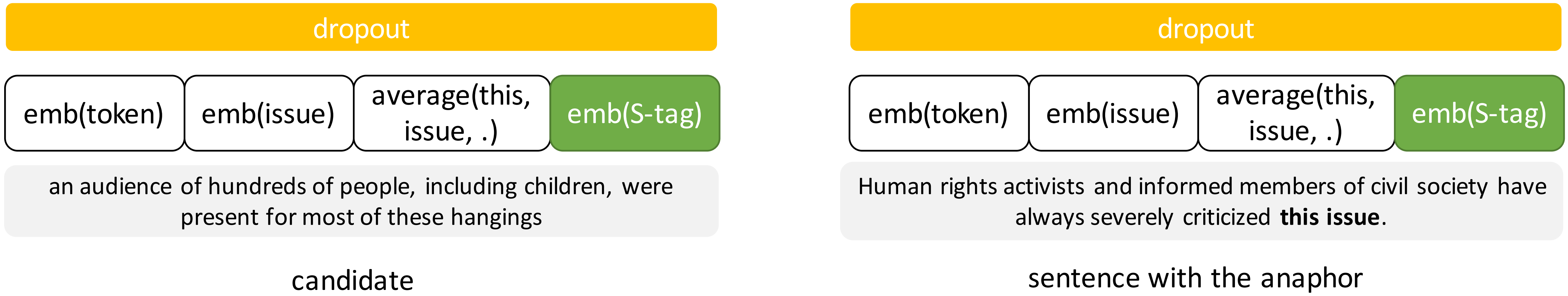
- Anand, P., and Hardt, D. (2016). Antecedent Selection for Sluicing: Structure and Content. In Proceedings of *EMNLP* (pp. 1234-1243).
- Clark, K. and Manning, C. D. (2015). Entity-centric coreference resolution with model stacking. In Proceedings of the ACL. Beijing, China.
- Kolhatkar, V., Zinsmeister, H., and Hirst, G. (2013). Interpreting Anaphoric Shell Nouns using Antecedents of Cataphoric Shell Nouns as Training Data. In Proceedings of the *EMNLP* (pp. 300-310).
- Lu, J., and Ng, V. (2017). Joint Learning for Event Coreference Resolution. In Proceedings of the ACL (pp. 90–101).
- Stede, M., and Grishina, Y. Anaphoricity in Connectives: A Case Study on German. In Proceedings of the *CORBON@HLT-NAACL* (pp. 41-46).
- Uryupina, O., Artstein, R., Bristot, A., Cavicchio, F., Rodríguez, K. J., and Poesio, M. (2016). ARRAU: Linguistically-Motivated Annotation of Anaphoric Descriptions. In Proceedings of the *LREC*.

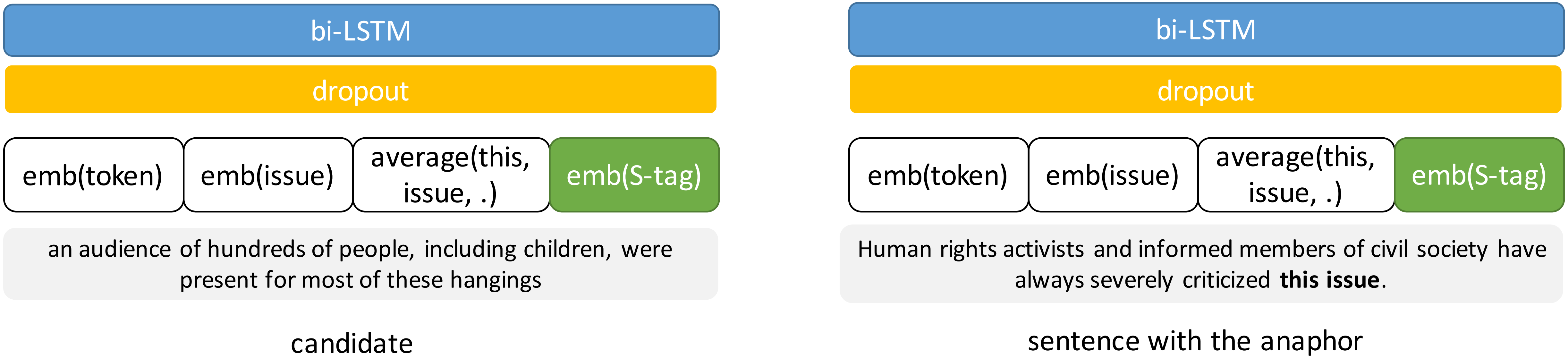
EXP2: unrestricted abstract anaphora resolution – analysis

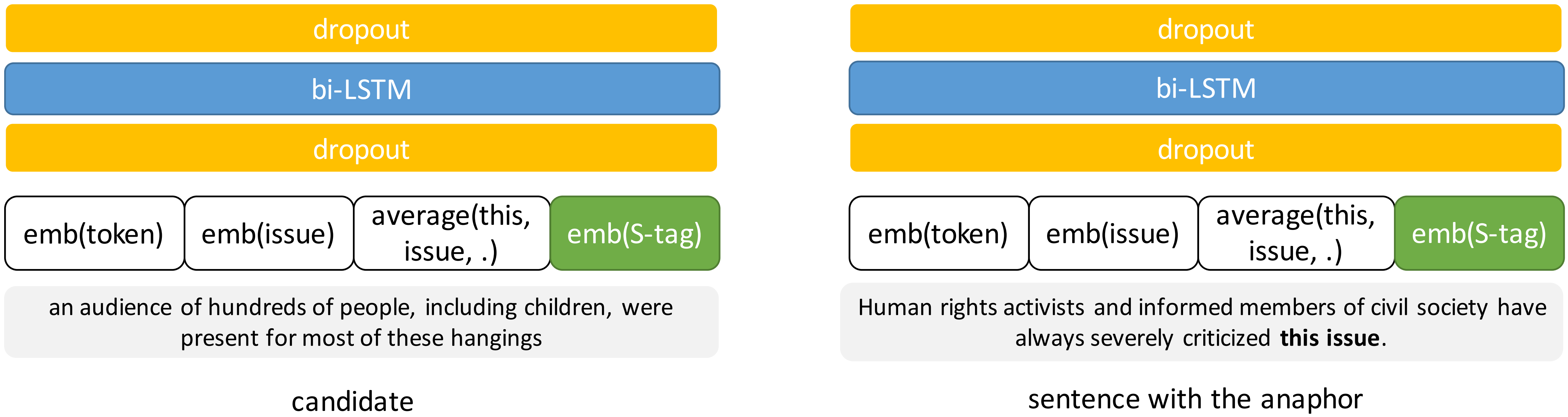


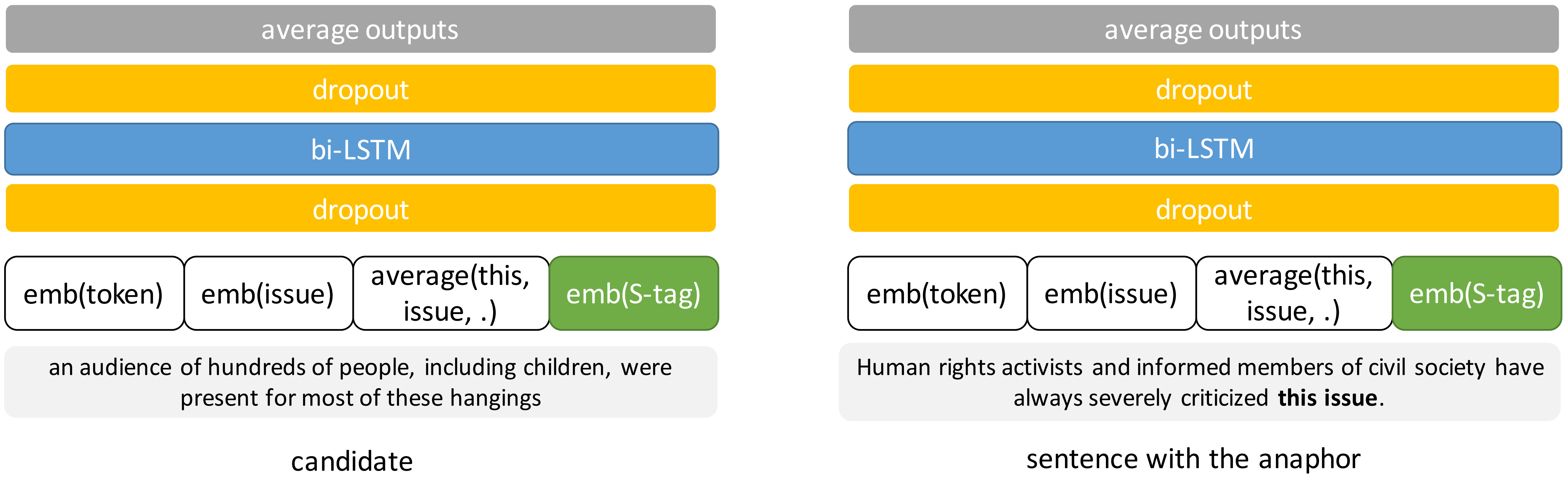
What does the max-margin objective achieve in the MR-LSTM?

It separates the best scoring antecedent from the best scoring negative candidate by separating their respective joint representations with the anaphoric sentence.

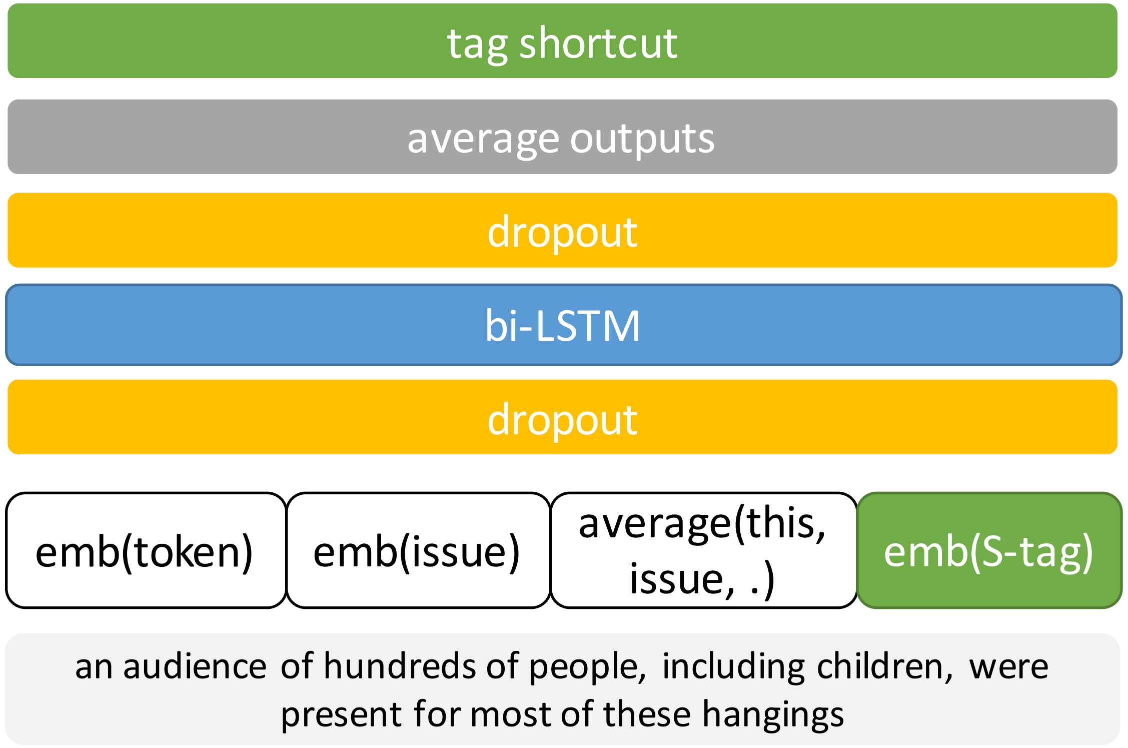

dropout
emb(token)
emb(issue)
average(this, issue, .)
emb(S-tag)
an audience of hundreds of people, including children, were present for most of these hangings
candidate
dropout
emb(token)
emb(issue)
average(this, issue, .)
emb(S-tag)
Human rights activists and informed members of civil society have always severely criticized **this issue**.
sentence with the anaphor


bi-LSTM
dropout
emb(token)
emb(issue)
average(this, issue, .)
emb(S-tag)
an audience of hundreds of people, including children, were present for most of these hangings
candidate
bi-LSTM
dropout
emb(token)
emb(issue)
average(this, issue, .)
emb(S-tag)
Human rights activists and informed members of civil society have always severely criticized **this issue**.
sentence with the anaphor

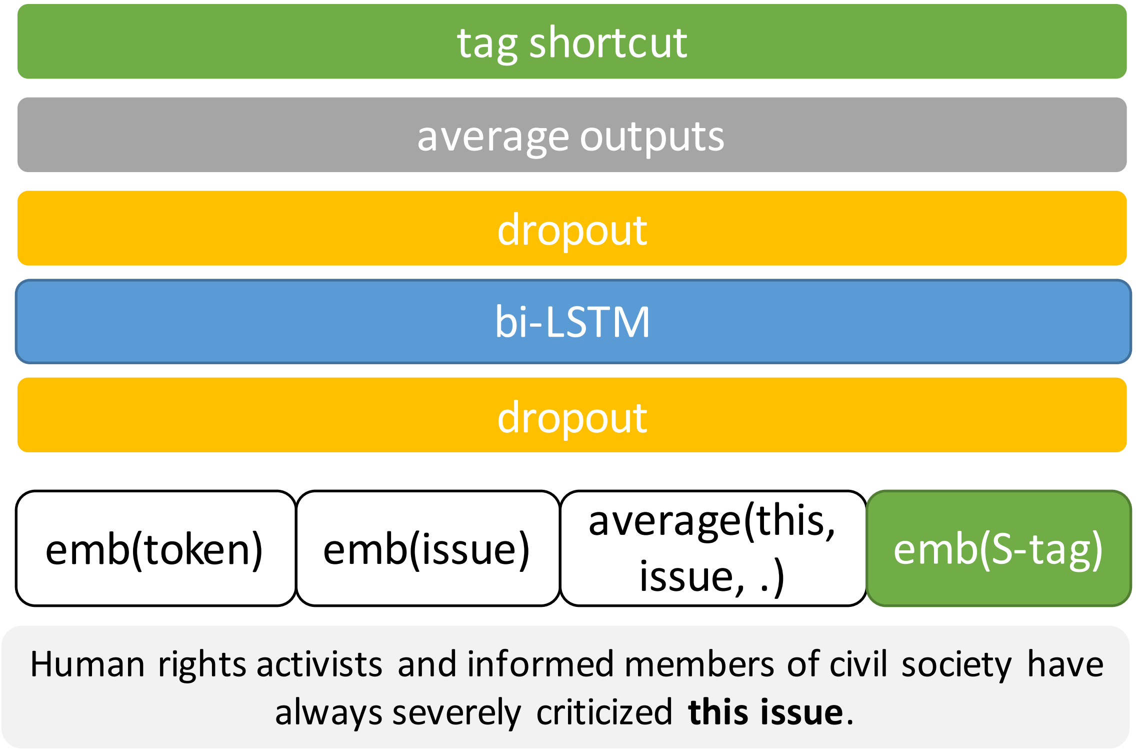

dropout
bi-LSTM
dropout
emb(token)
emb(issue)
average(this, issue, .)
emb(S-tag)
an audience of hundreds of people, including children, were present for most of these hangings
candidate
dropout
bi-LSTM
dropout
emb(token)
emb(issue)
average(this, issue, .)
emb(S-tag)
Human rights activists and informed members of civil society have always severely criticized **this issue**.
sentence with the anaphor


average outputs
dropout
bi-LSTM
dropout
emb(token)
emb(issue)
average(this, issue, .)
emb(S-tag)
an audience of hundreds of people, including children, were present for most of these hangings
candidate
average outputs
dropout
bi-LSTM
dropout
emb(token)
emb(issue)
average(this, issue, .)
emb(S-tag)
Human rights activists and informed members of civil society have always severely criticized **this issue**.
sentence with the anaphor

info about the constituent tag may be lost ⇒ add embedding of the constituent tag to the representation produced by bi-LSTM (**tag shortcut**)

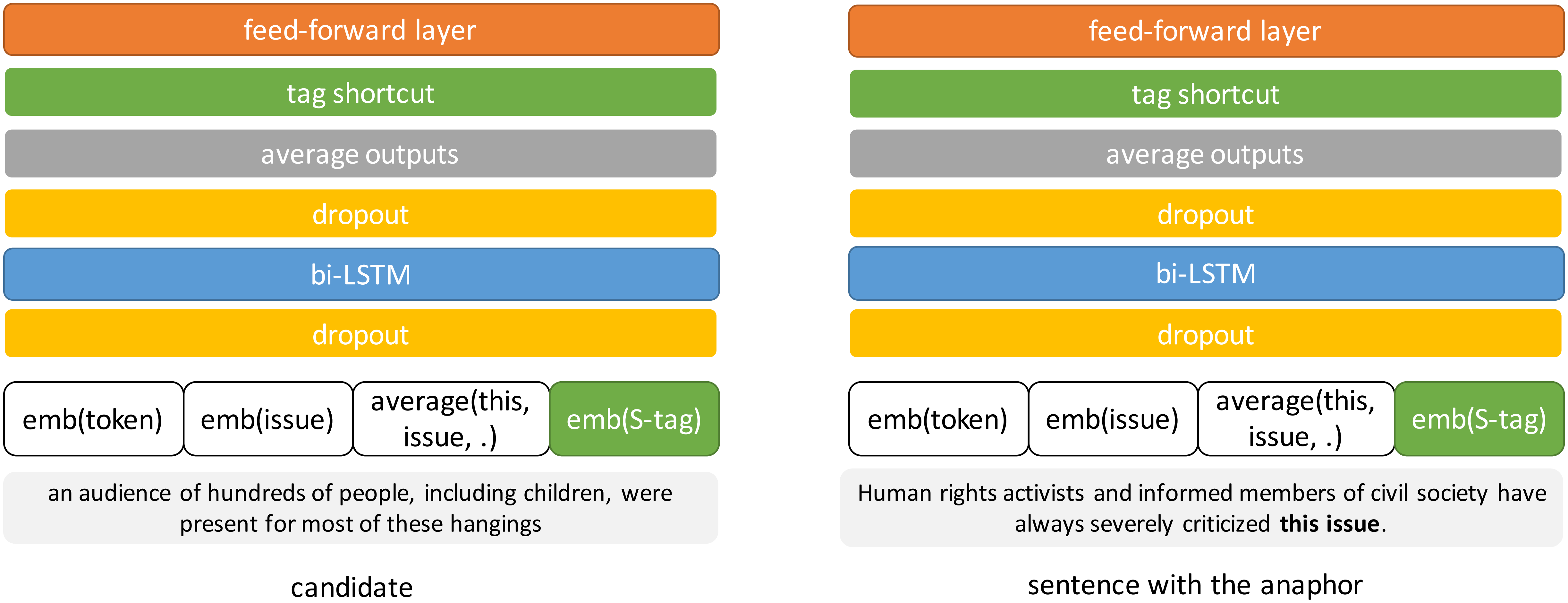


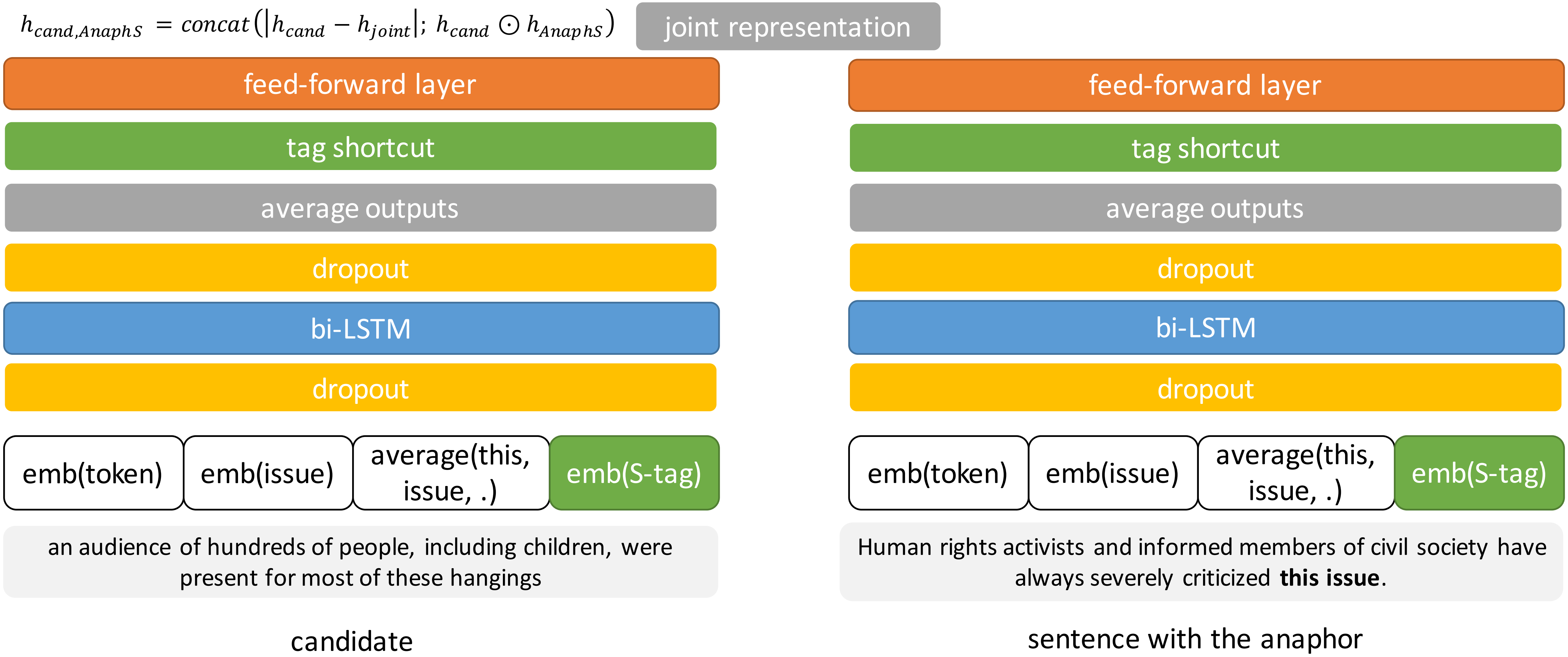
candidate

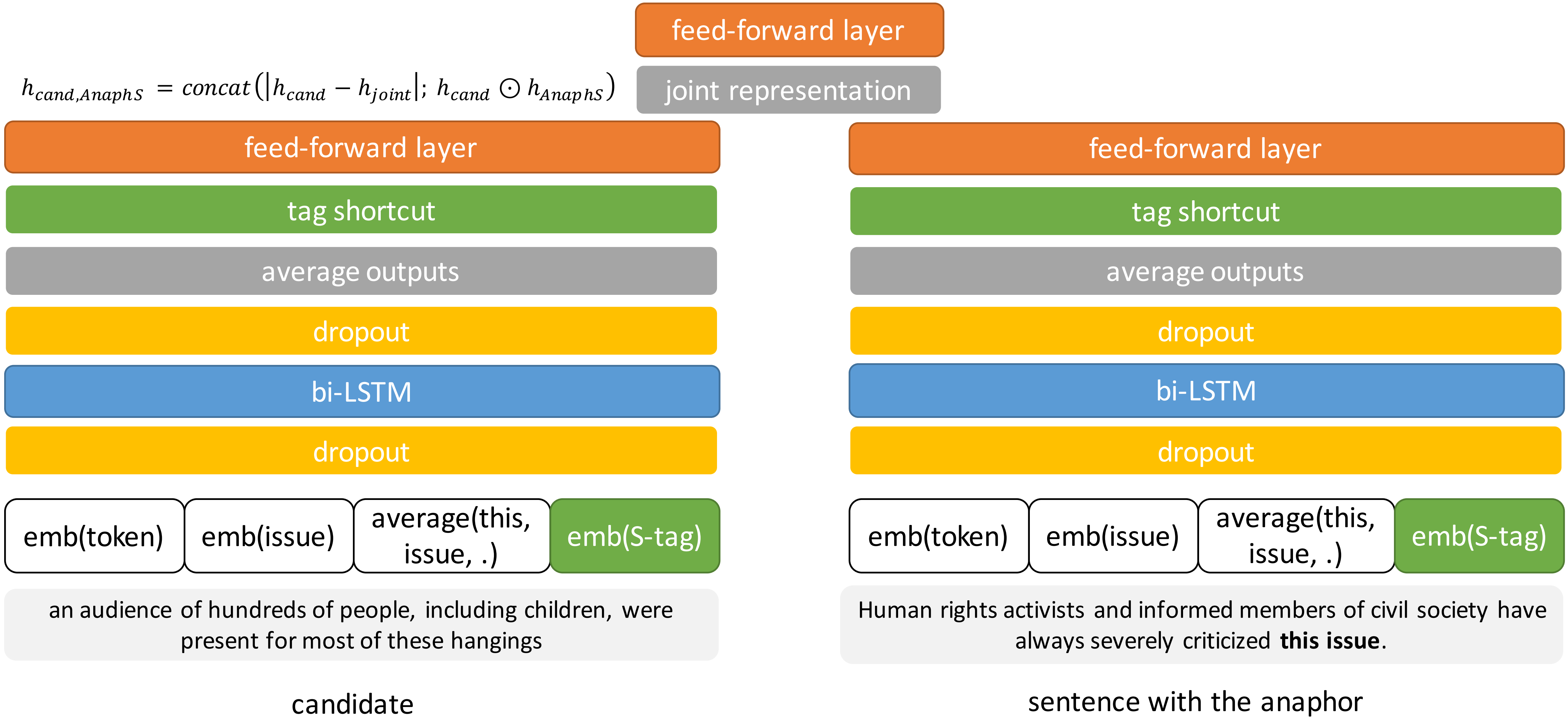


sentence with the anaphor

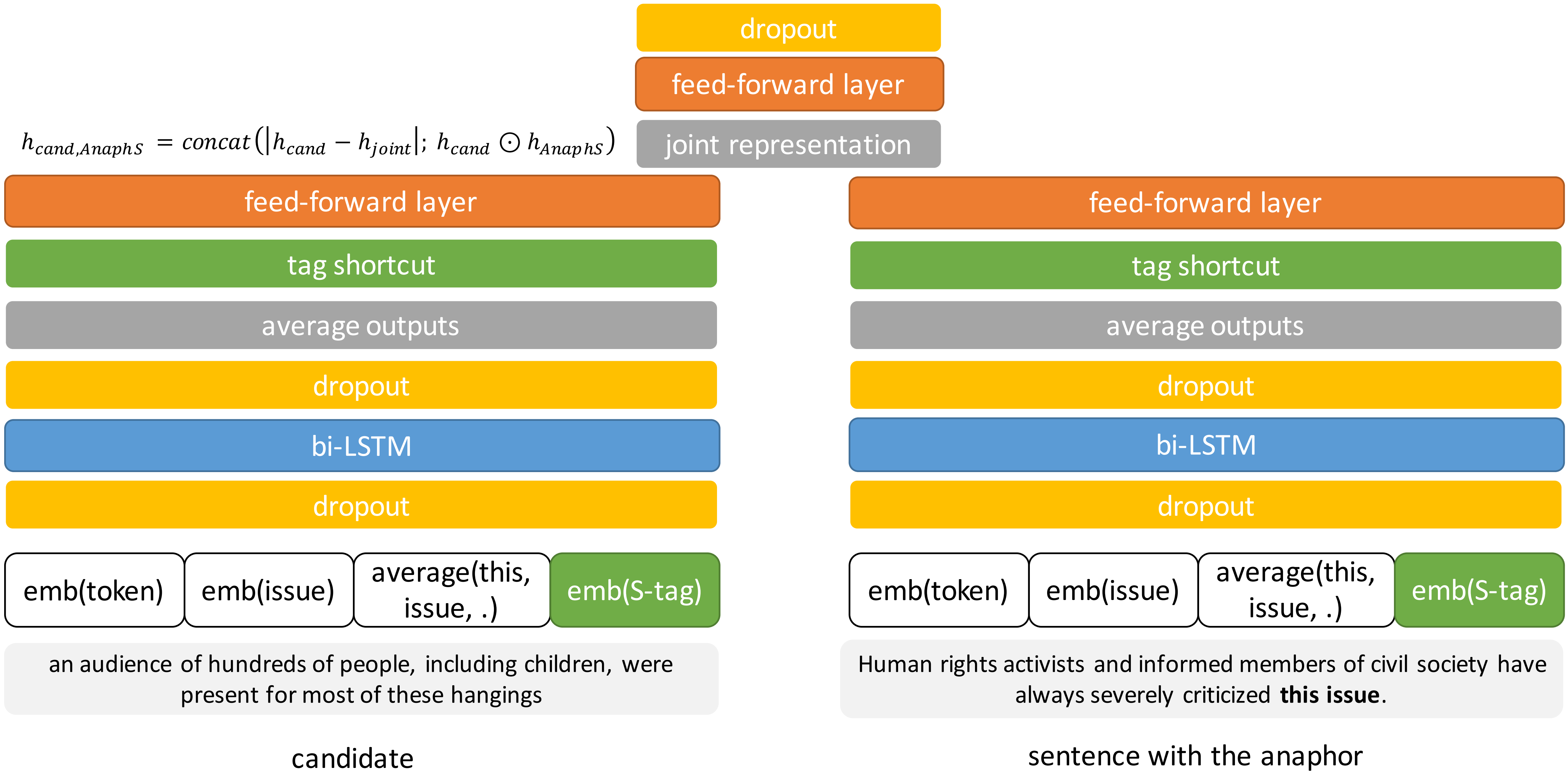
feed-forward layer for more expressivity

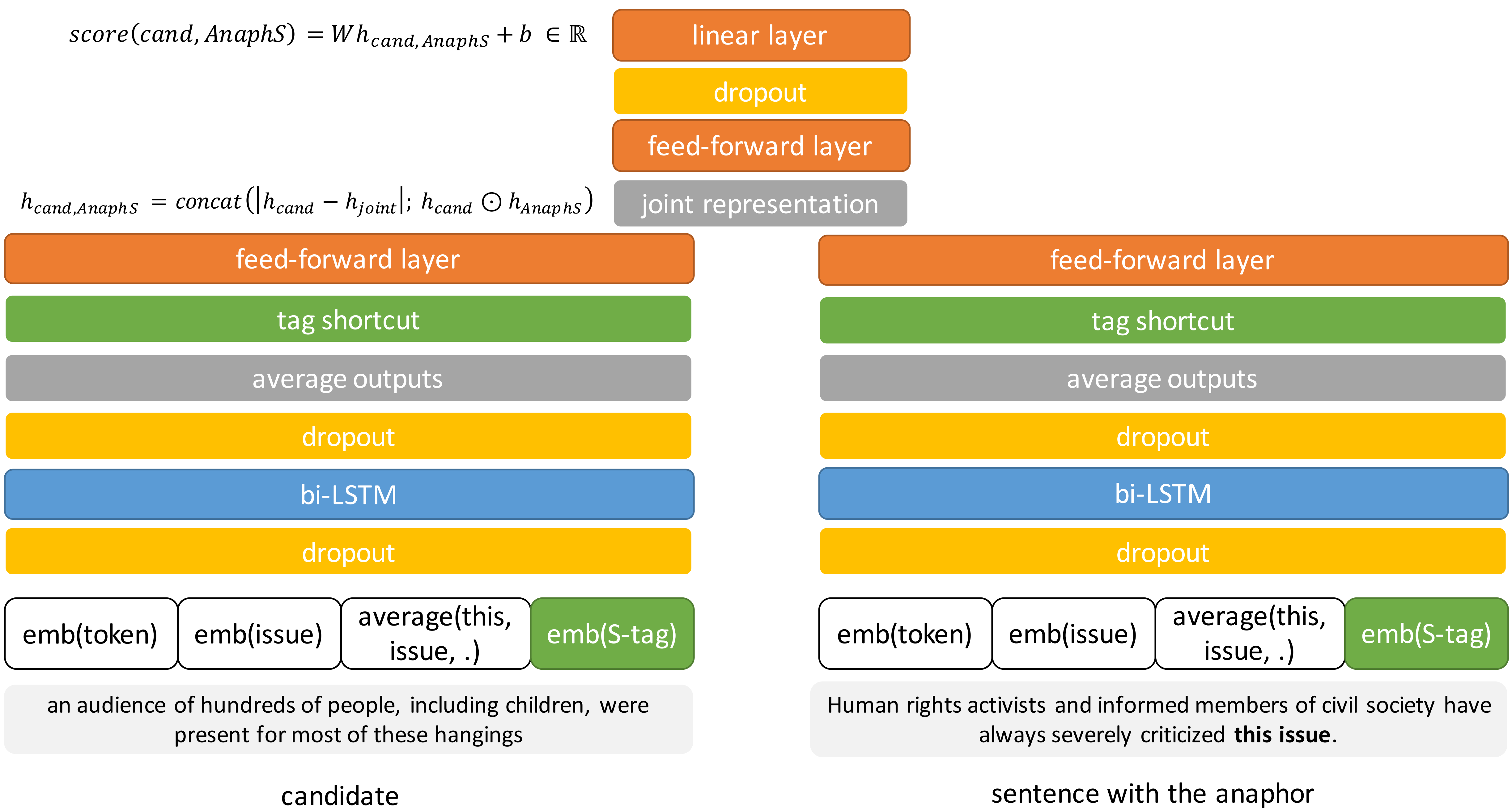


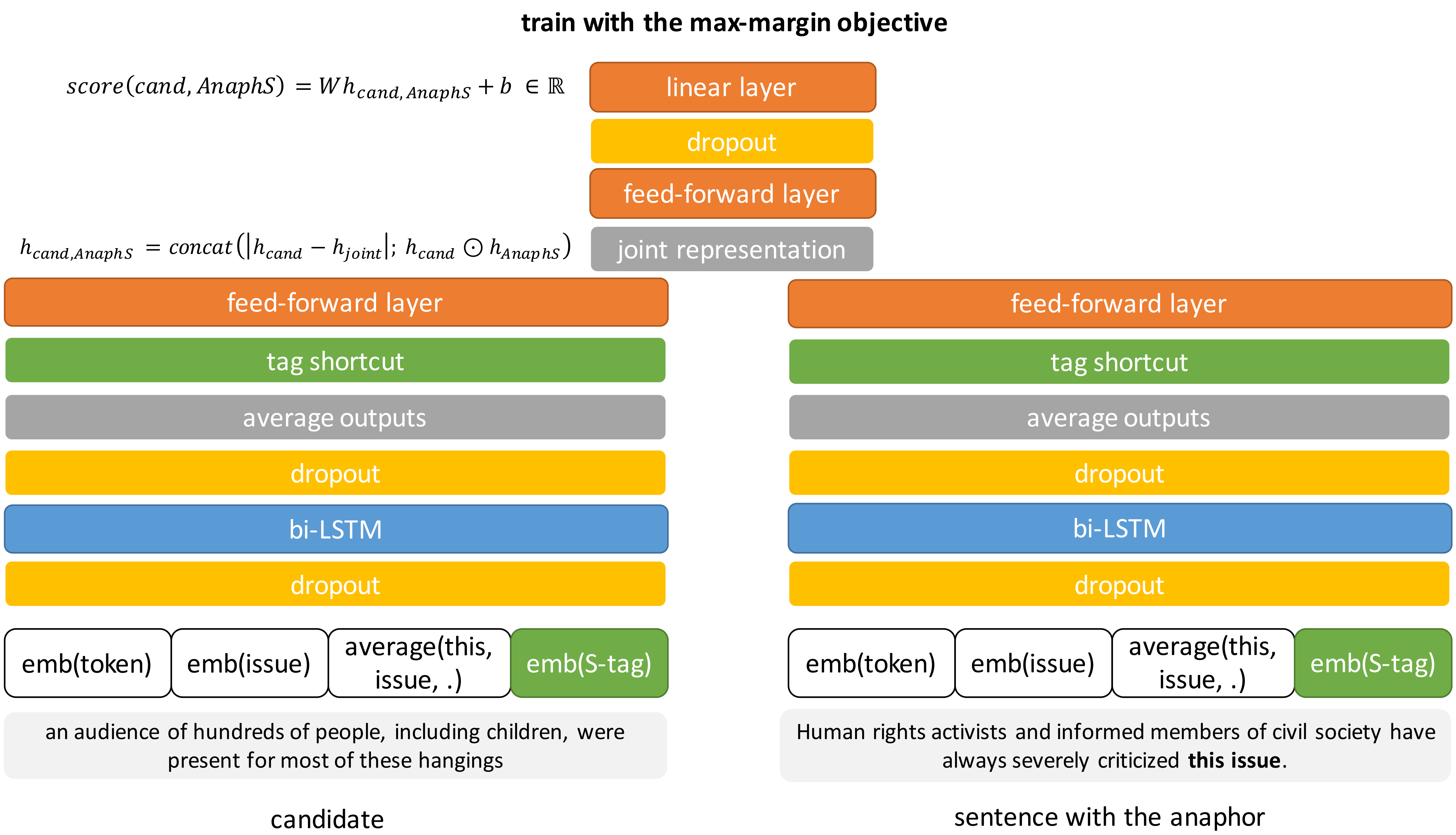

$h_{cand,AnaphS} = concat(|h_{cand} - h_{joint}|; h_{cand} \odot h_{AnaphS})$
joint representation
feed-forward layer
tag shortcut
average outputs
dropout
bi-LSTM
dropout
emb(token)
emb(issue)
average(this, issue, .)
emb(S-tag)
an audience of hundreds of people, including children, were present for most of these hangings
candidate
feed-forward layer
tag shortcut
average outputs
dropout
bi-LSTM
dropout
emb(token)
emb(issue)
average(this, issue, .)
emb(S-tag)
Human rights activists and informed members of civil society have always severely criticized **this issue**.
sentence with the anaphor



feed-forward layer
joint representation
$h_{cand,AnaphS} = concat(|h_{cand} - h_{joint}|; h_{cand} \odot h_{AnaphS})$
feed-forward layer
tag shortcut
average outputs
dropout
bi-LSTM
dropout
emb(token)
emb(issue)
average(this, issue, .)
emb(S-tag)
an audience of hundreds of people, including children, were present for most of these hangings
candidate
feed-forward layer
tag shortcut
average outputs
dropout
bi-LSTM
dropout
emb(token)
emb(issue)
average(this, issue, .)
emb(S-tag)
Human rights activists and informed members of civil society have always severely criticized **this issue**.
sentence with the anaphor


dropout
feed-forward layer
joint representation
$h_{cand,AnaphS} = concat(|h_{cand} - h_{joint}|;\ h_{cand} \odot h_{AnaphS})$
feed-forward layer
tag shortcut
average outputs
dropout
bi-LSTM
dropout
emb(token)
emb(issue)
average(this, issue, .)
emb(S-tag)
an audience of hundreds of people, including children, were present for most of these hangings
candidate
feed-forward layer
tag shortcut
average outputs
dropout
bi-LSTM
dropout
emb(token)
emb(issue)
average(this, issue, .)
emb(S-tag)
Human rights activists and informed members of civil society have always severely criticized **this issue**.
sentence with the anaphor


$score(cand, AnaphS) = Wh_{cand, AnaphS} + b \in \mathbb{R}$
linear layer
dropout
feed-forward layer
joint representation
$h_{cand,AnaphS} = concat(|h_{cand} - h_{joint}|; h_{cand} \odot h_{AnaphS})$
feed-forward layer
tag shortcut
average outputs
dropout
bi-LSTM
dropout
emb(token)
emb(issue)
average(this, issue, .)
emb(S-tag)
an audience of hundreds of people, including children, were present for most of these hangings
candidate
feed-forward layer
tag shortcut
average outputs
dropout
bi-LSTM
dropout
emb(token)
emb(issue)
average(this, issue, .)
emb(S-tag)
Human rights activists and informed members of civil society have always severely criticized **this issue**.
sentence with the anaphor


train with the max-margin objective
$score(cand, AnaphS) = Wh_{cand, AnaphS} + b \in \mathbb{R}$
linear layer
dropout
feed-forward layer
$h_{cand,AnaphS} = concat(|h_{cand} - h_{joint}|; h_{cand} \odot h_{AnaphS})$
joint representation
feed-forward layer
tag shortcut
average outputs
dropout
bi-LSTM
dropout
emb(token)
emb(issue)
average(this, issue, .)
emb(S-tag)
an audience of hundreds of people, including children, were present for most of these hangings
candidate
feed-forward layer
tag shortcut
average outputs
dropout
bi-LSTM
dropout
emb(token)
emb(issue)
average(this, issue, .)
emb(S-tag)
Human rights activists and informed members of civil society have always severely criticized **this issue**.
sentence with the anaphor