Resolving Abstract Anaphors in Discourse — Uphill Battles with Neural Ranking Models and Automatic Data Extraction

Ana Marasović
Heidelberg University

Joint work with: Anette Frank, Juri Opitz, Leo Born

AIPHES
ADAPTIVE INFORMATIONSAUFBEREITUNG AUS HETEROGENEN QUELLEN

RUPRECHT-KARLS-
UNIVERSITÄT
HEIDELBERG

Heidelberger Institut für
Theoretische Studien

TECHNISCHE
UNIVERSITÄT
DARMSTADT

Analyzing Sentiment in Discourse

U.S. car dealers had an average of 59 days supply of cars in their lots at the end of September, according to Ward's Automotive Reports. But Mr. Tonkin said dealers should slash stocks to between 15 and 30 days to reduce the costs of financing inventory. His message is getting a chilly reception in Detroit, where the Big Three auto makers are already being forced to close plants because of soft sales and reduced dealer orders. Even before Mr. Tonkin's broadside, some large dealers said they were cutting inventories.

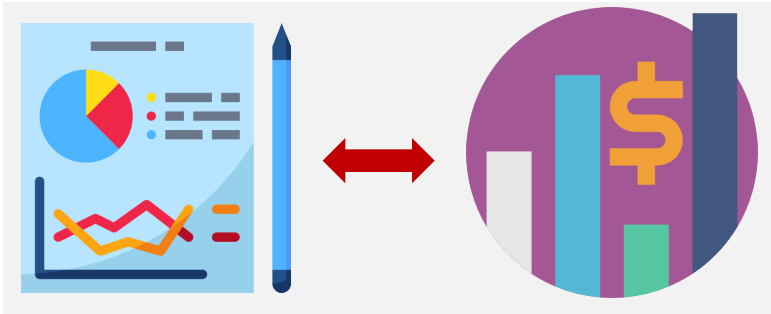
Ford Motor Co. and **Chrysler Crp. representatives** criticized **Mr. Tonkin's plan** as unworkable.

NAACL'18

SENTENCE-LEVEL OPINION ANALYSIS MODEL

FCA
FIAT CHRYSLER AUTOMOBILES

HOLDERS



TARGET

DISCOURSE-LEVEL OPINION ANALYSIS MODEL

Abstract Anaphora Resolution (AAR)

U.S. car dealers had an average of 59 days supply of cars in their lots at the end of September, according to Ward's Automotive Reports. But Mr. Tonkin said dealers should slash stocks to between 15 and 30 days to reduce the costs of financing inventory. His message is getting a chilly reception in Detroit, where the Big Three auto makers are already being forced to close plants because of soft sales and reduced dealer orders. Even before Mr. Tonkin's broadside, some large dealers said they were cutting inventories. Ford Motor Co. and Chrysler Crp. representatives criticized **Mr. Tonkin's plan** as unworkable.

ABSTRACT OBJECT ANTECEDENT

ABSTRACT ANAPHOR

AAR aims to resolve

- **nominal** expressions (e.g. Mr. Tonkin's plan, this issue, those two actions)
- **pronominal** expressions (e.g. this, that, it)

that refer to *abstract-object-antecedents* such as:

- facts
- events
- plans
- actions
- situations.

Entity Coreference Resolution vs. AAR

U.S. car dealers had an average of 59 days supply of cars in their lots at the end of September, according to Ward's Automotive Reports. But Mr. Tonkin said dealers should slash stocks to between 15 and 30 days to reduce the costs of financing inventory. His message is getting a chilly reception in Detroit, where the Big Three auto makers are already being forced to close plants because of soft sales and reduced dealer orders. Even before Mr. Tonkin's broadside, some large dealers said they were cutting inventories. Ford Motor Co. and Chrysler Crp. representatives criticized **Mr. Tonkin's plan** as unworkable.

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

Entity Coreference Resolution vs. AAR

U.S. car dealers had an average of 59 days supply of cars in their lots at the end of September, according to Ward's Automotive Reports. But Mr. Tonkin said dealers should slash stocks to between 15 and 30 days to reduce the costs of financing inventory. His message is getting a chilly reception in Detroit, where the Big Three auto makers are already being forced to close plants because of soft sales and reduced dealer orders. Even before Mr. Tonkin's broadside, some large dealers said they were cutting inventories. Ford Motor Co. and Chrysler Crp. representatives criticized **Mr. Tonkin's plan** as unworkable.

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

Entity Coreference Resolution vs. AAR

U.S. car dealers had an average of 59 days supply of cars in their lots at the end of September, according to Ward's Automotive Reports. But Mr. Tonkin said dealers should slash stocks to between 15 and 30 days to reduce the costs of financing inventory. His message is getting a chilly reception in Detroit, where the Big Three auto makers are already being forced to close plants because of soft sales and reduced dealer orders. Even before Mr. Tonkin's broadside, some large dealers said they were cutting inventories. Ford Motor Co. and Chrysler Crp. representatives criticized **Mr. Tonkin's plan** as unworkable.

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**
- ✓ **EXTRACT TRAINING DATA**

EMNLP'17

Difficulties of Unrestricted AAR

Unrestricted AAR is a difficult task

- the antecedents vary in size (Vieira et al., 2005)
- the antecedents vary in syntactic type (Vieira et al., 2005; Dipper and Zinsmeister, 2002)
- the antecedents vary in distance from the anaphor (Dipper and Zinsmeister, 2002)
- no 1:1 relation between the antecedent's syntactic and the anaphor's semantic type (Webber, 1991).

Even **annotators often disagree on the exact boundaries of antecedents**, resulting in agreement of around Krippendorff's $\alpha = 0.55$ (Artstein and Poesio, 2006).

Varieties of Abstract Anaphors and Resources

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 & NP mentions of similar abstractness

Varieties of Abstract Anaphors and Resources

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 & NP mentions of similar abstractness
abstract event resolution (Jauhar et al., 2015)	Real-estate market indicators *plummet much further than a local economy in recession*. This was seen in the late 1960s in LA.	1729 train / 180 dev / 243 test	NO	*this, that, it* from CoNLL-12 ST (Pradhan et al., 2012) w/ verbal preceding mention

Varieties of Abstract Anaphors and Resources

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 & NP mentions of similar abstractness
abstract event resolution (Jauhar et al., 2015)	Real-estate market indicators *plummet much further than a local economy in recession*. This was seen in the late 1960s in LA.	1729 train / 180 dev / 243 test	NO	*this, that, it* from CoNLL-12 ST (Pradhan et al., 2012) w/ verbal preceding mention
sluicing (Anand and Hardt, 2016)	He resorted to that. I don't know why.	1159 train / 453 dev / 173 test	NO	

Varieties of Abstract Anaphors and Resources

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 & NP mentions of similar abstractness
abstract event resolution (Jauhar et al., 2015)	Real-estate market indicators *plummet much further than a local economy in recession*. This was seen in the late 1960s in LA.	1729 train / 180 dev / 243 test	NO	*this, that, it* from CoNLL-12 ST (Pradhan et al., 2012) w/ verbal preceding mention
sluicing (Anand and Hardt, 2016)	He resorted to that. I don't know why.	1159 train / 453 dev / 173 test	NO	
anaphoric connectives (Stede and Grishina, 2016)	*Peter was the best goal scorer*. Therefore he received the trophy.	140 instances	NO	restricted in type, ambiguous & require WSD

Varieties of Abstract Anaphors and Resources

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 & NP mentions of similar abstractness
abstract event resolution (Jauhar et al., 2015)	Real-estate market indicators *plummet much further than a local economy in recession*. This was seen in the late 1960s in LA.	1729 train / 180 dev / 243 test	NO	*this, that, it* from CoNLL-12 ST (Pradhan et al., 2012) w/ verbal preceding mention
sluicing (Anand and Hardt, 2016)	He resorted to that. I don't know why.	1159 train / 453 dev / 173 test	NO	
anaphoric connectives (Stede and Grishina, 2016)	*Peter was the best goal scorer*. Therefore he received the trophy.	140 instances	NO	restricted in type, ambiguous & require WSD
shell nouns (Kolhatkar et al., 2013)	*Hundreds of people were present for these hangings*. Human rights activists have always criticized this issue.	2664–43809 train / Ø dev/ 303–472 test	NO	6 shell nouns, use the specific properties and categorization of shell nouns

Varieties of Abstract Anaphors and Resources

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 & NP mentions of similar abstractness
abstract event resolution (Jauhar et al., 2015)	Real-estate market indicators *plummet much further than a local economy in recession*. This was seen in the late 1960s in LA.	1729 train / 180 dev / 243 test	NO	*this, that, it* from CoNLL-12 ST (Pradhan et al., 2012) w/ verbal preceding mention
sluicing (Anand and Hardt, 2016)	He resorted to that. I don't know why.	1159 train / 453 dev / 173 test	NO	
anaphoric connectives (Stede and Grishina, 2016)	*Peter was the best goal scorer*. Therefore he received the trophy.	140 instances	NO	restricted in type, ambiguous & require WSD
shell nouns (Kolhatkar et al., 2013)	*Hundreds of people were present for these hangings*. Human rights activists have always criticized this issue.	2664–43809 train / Ø dev/ 303–472 test	NO	6 shell nouns, use the specific properties and categorization of shell nouns

pronominal or nominal but not both

very scarce

Varieties of Abstract Anaphors and Resources

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 & NP mentions of similar abstractness
abstract event resolution (Jauhar et al., 2015)	Real-estate market indicators *plummet much further than a local economy in recession*. This was seen in the late 1960s in LA.	1729 train / 180 dev / 243 test	NO	*this, that, it* from CoNLL-12 ST (Pradhan et al., 2012) w/ verbal preceding mention
sluicing (Anand and Hardt, 2016)	He resorted to that. I don't know why.	1159 train / 453 dev / 173 test	NO	
anaphoric connectives (Stede and Grishina, 2016)	*Peter was the best goal scorer*. Therefore he received the trophy.	140 instances	NO	restricted in type, ambiguous & require WSD
shell nouns (Kolhatkar et al., 2013)	*Hundreds of people were present for these hangings*. Human rights activists have always criticized this issue.	2664–43809 train / Ø dev/ 303–472 test	NO	6 shell nouns, use the specific properties and categorization of shell nouns
unrestricted AAR (Marasović et al., 2017)	Revenue from tourism this year is projected to total $1.3 billion, down from $2.2 billion last year. Because of this and the huge trade gap, the deficit in China's current account is expected to widen sharply from the $3.8 billion deficit last year.	we extract training data, 600 test instances	YES	*abstract* and *plan* type anaphors from WSJ part of ARRAU (Uryupina et al., 2016)

AAR – An Uphill Battle

GOAL

> an unified approach for **unrestricted** AAR

AAR – An Uphill Battle

GOAL

> an unified approach for **unrestricted** AAR

APPROACH

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

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

ANAPHORIC SENTENCE

Ford Motor Co. and Chrysler Crp. representatives criticized Mr. Tonkin's plan as unworkable.

ANTECEDENT

dealers should slash stocks to between 15 and 30 days to reduce the costs of financing inventory

something Ford and Chrysler may criticize

AAR – An Uphill Battle

GOAL

> an unified approach for **unrestricted** AAR

APPROACH

> learning what is the correct antecedent for a given AA ≈
learning the relation between the sentence with AA and the antecedent

OBSTACLE

> **scarce** training and evaluation data

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

Ford Motor Co. and Chrysler Crp. representatives criticized [$_{S'}$ **that** [$_{S}$ dealers should slash stocks to between 15 and 30 days to reduce the costs of financing inventory]].

→

Ford Motor Co. and Chrysler Crp. representatives criticized this / this issue. Dealers should slash stocks to between 15 and 30 days to reduce the costs of financing inventory.

AAR with a Mention-Ranking Siamese NN and Extracted Training Data

EMNLP'17

1. Solving the resource bottleneck – by harvesting training data
2. Resolving Abstract Anaphors in a Relational Model
   – A Mention-Ranking Siamese Network Model

ONGOING

3. A closer look at training data extraction for AAR
4. Ongoing Challenges

AAR with a Mention-Ranking Siamese NN and Extracted Training Data

EMNLP'17

1. Solving the resource bottleneck – by harvesting training data
2. Resolving Abstract Anaphors in a Relational Model
   – A Mention-Ranking Siamese Network Model

ONGOING

3. A closer look at training data extraction for AAR
4. Ongoing Challenges

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)

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)

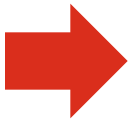
Congress has focused almost solely on **the fact** that special education is expensive - and that it takes away money from regular education.

Cataphoric Shell Noun (CSN)

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)

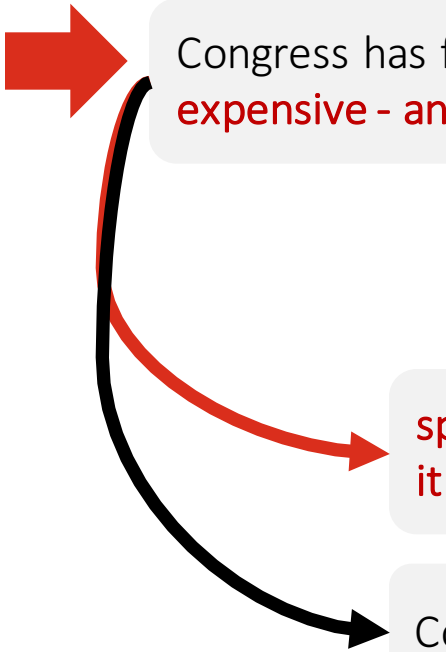
Related Work
– Shell Noun Resolution (Kolhatkar et al., 2013)

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.

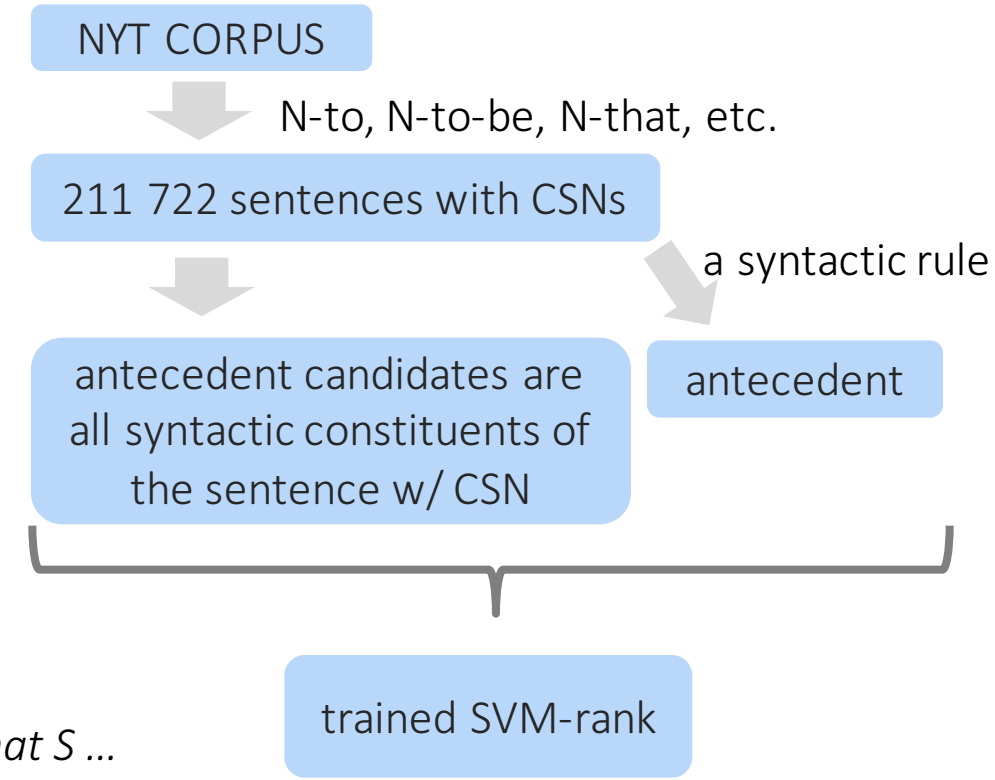
special education is expensive - and that it takes away money from regular education

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

We extend this approach to unrestricted AAR

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

Training Data Acquisition
–Silver Data

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 Acquisition
–Silver Data

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 Acquisition
–Silver Data

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 Acquisition
–Silver Data

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 Acquisition
–Silver Data

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 because of this]].

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

Training Data Acquisition
–Silver Data

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 Acquisition
–Silver Data

✓harvesting 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

AAR with a Mention-Ranking Siamese NN and Extracted Training Data

EMNLP'17

1. Solving the resource bottleneck – by harvesting training data
2. Resolving Abstract Anaphors in a Relational Model
   – A Mention-Ranking Siamese Network Model

ONGOING

3. A closer look at training data extraction for AAR
4. Ongoing Challenges

Reminder on the intuitions for our model

GOAL

> an unified approach for **unrestricted** AAR

APPROACH

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

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

ANAPHORIC SENTENCE

Ford Motor Co. and Chrysler Crp. representatives criticized Mr. Tonkin's plan as unworkable.

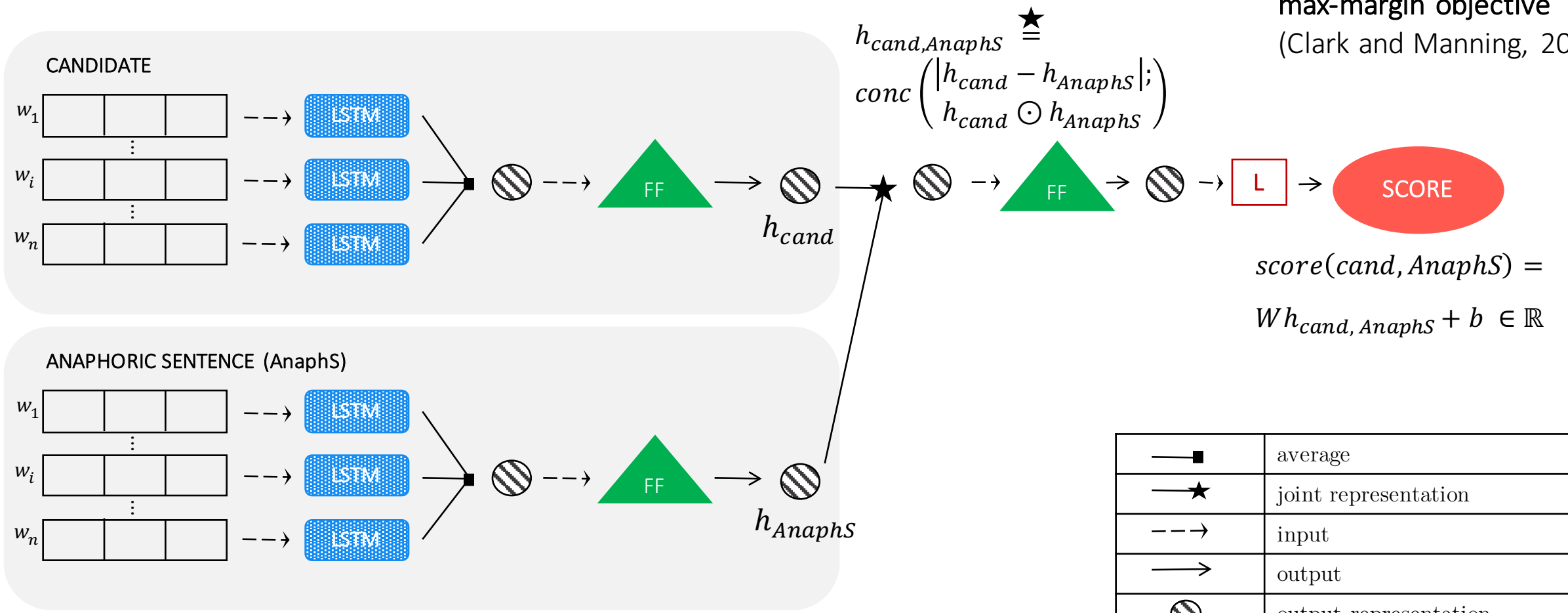
ANTECEDENT

dealers should slash stocks to between 15 and 30 days to reduce the costs of financing inventory

something Ford and Chrysler may criticize

Siamese-LSTM Mention-Ranking Model

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



——■	average
——★	joint representation
- - →	input
——→	output
⊘	output representation
L	linear layer
▲	ELU feed-forward layer

Siamese-LSTM Mention-Ranking Model – Input

dealers should slash stocks to between 15 and 30 days to reduce the costs of financing inventory

candidate

input to LSTM

emb(token)

Ford Motor Co. and Chrysler Crp. representatives criticized Mr. Tonkin's **plan** as unworkable.

sentence with the anaphor (AnaphS)

input to LSTM

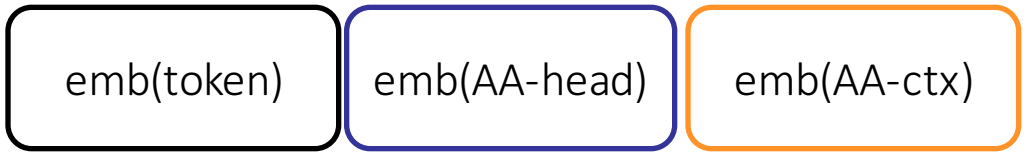
emb(token)

Siamese-LSTM Mention-Ranking Model – Input

dealers should slash stocks to between 15 and 30 days to reduce the costs of financing inventory

candidate

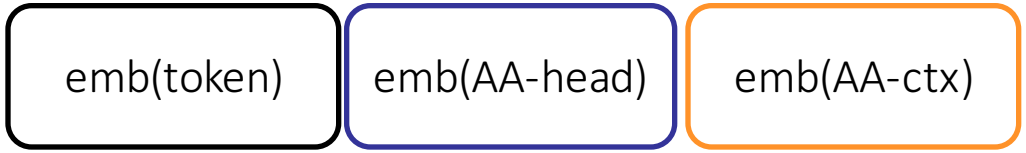
input to LSTM



Ford Motor Co. and Chrysler Crp. representatives criticized Mr. Tonkin's **plan** as unworkable.

sentence with the anaphor (AnaphS)

input to LSTM



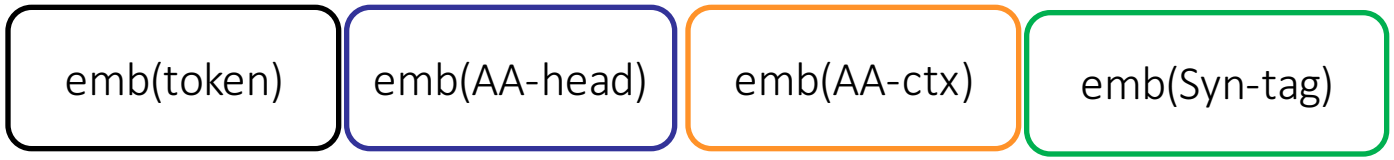
The anaphoric sentence may have more than one abstract anaphor (AA)

Siamese-LSTM Mention-Ranking Model – Input

dealers should slash stocks to between 15 and 30 days to reduce the costs of financing inventory

candidate

input to LSTM

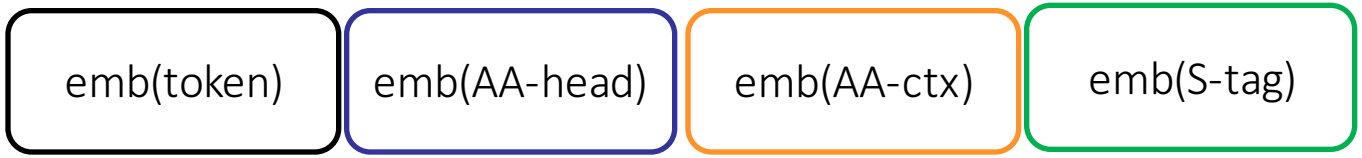


a constituent tag label is a good signal of worthy candidates

Ford Motor Co. and Chrysler Crp. representatives criticized Mr. Tonkin's **plan** as unworkable.

sentence with the anaphor (AnaphS)

input to LSTM



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

Experiment 1: Shell Noun Resolution

Datasets

- train data: extracted 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)

	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

Restricted task setup

Candidates for the antecedent are all constituents from the sentence that contains the antecedent.

U.S. car dealers had an average of 59 days supply of cars in their lots at the end of September, according to Ward's Automotive Reports. But Mr. Tonkin said dealers should slash stocks to between 15 and 30 days to reduce the costs of financing inventory. His message is getting a chilly reception in Detroit, where the Big Three auto makers are already being forced to close plants because of soft sales and reduced dealer orders. Even before Mr. Tonkin's broadside, some large dealers said they were cutting inventories. Ford Motor Co. and Chrysler Crp. representatives criticized **Mr. Tonkin's plan** as unworkable.

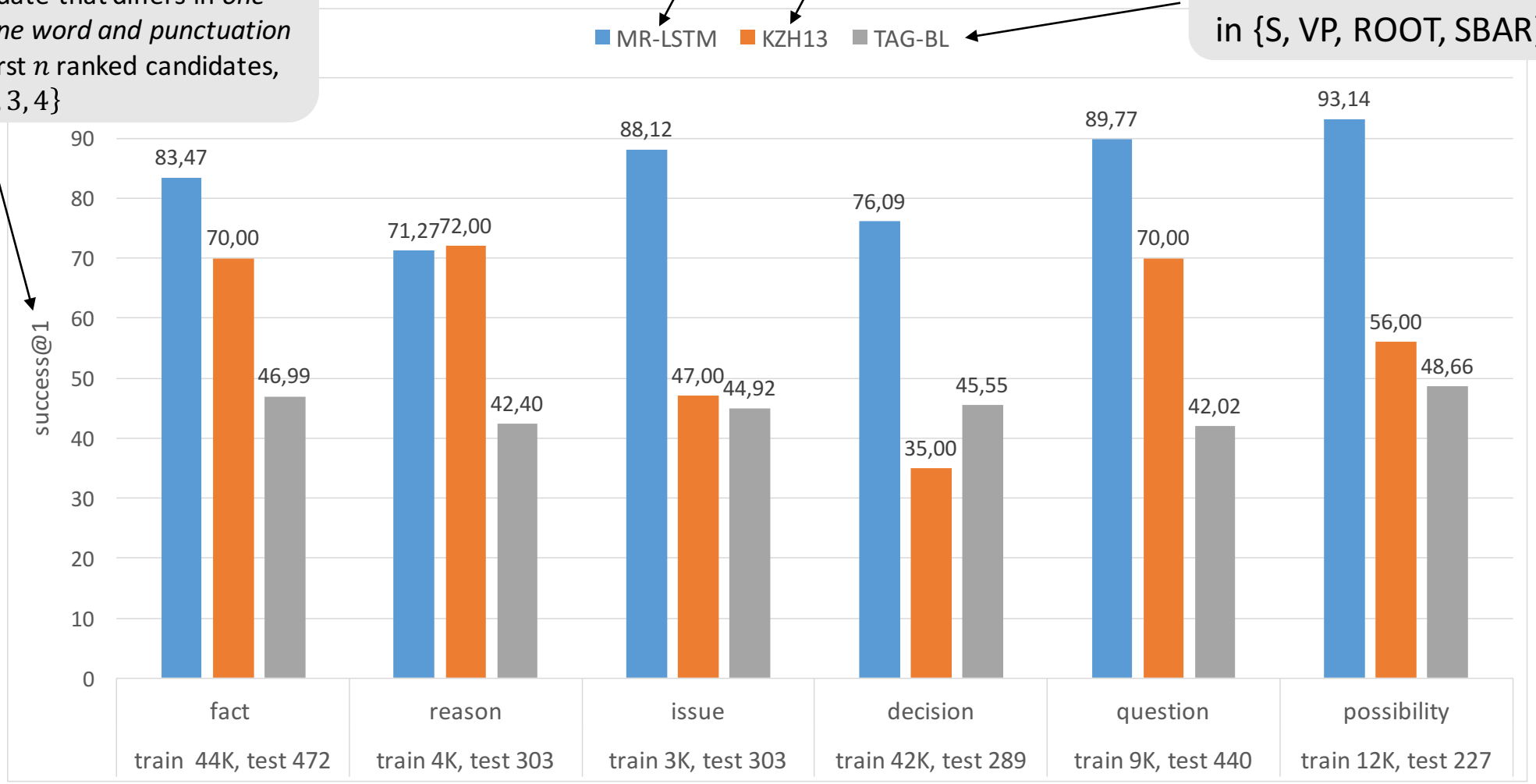
Experiment 1: Shell Noun Resolution

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, $n \in \{1, 2, 3, 4\}$

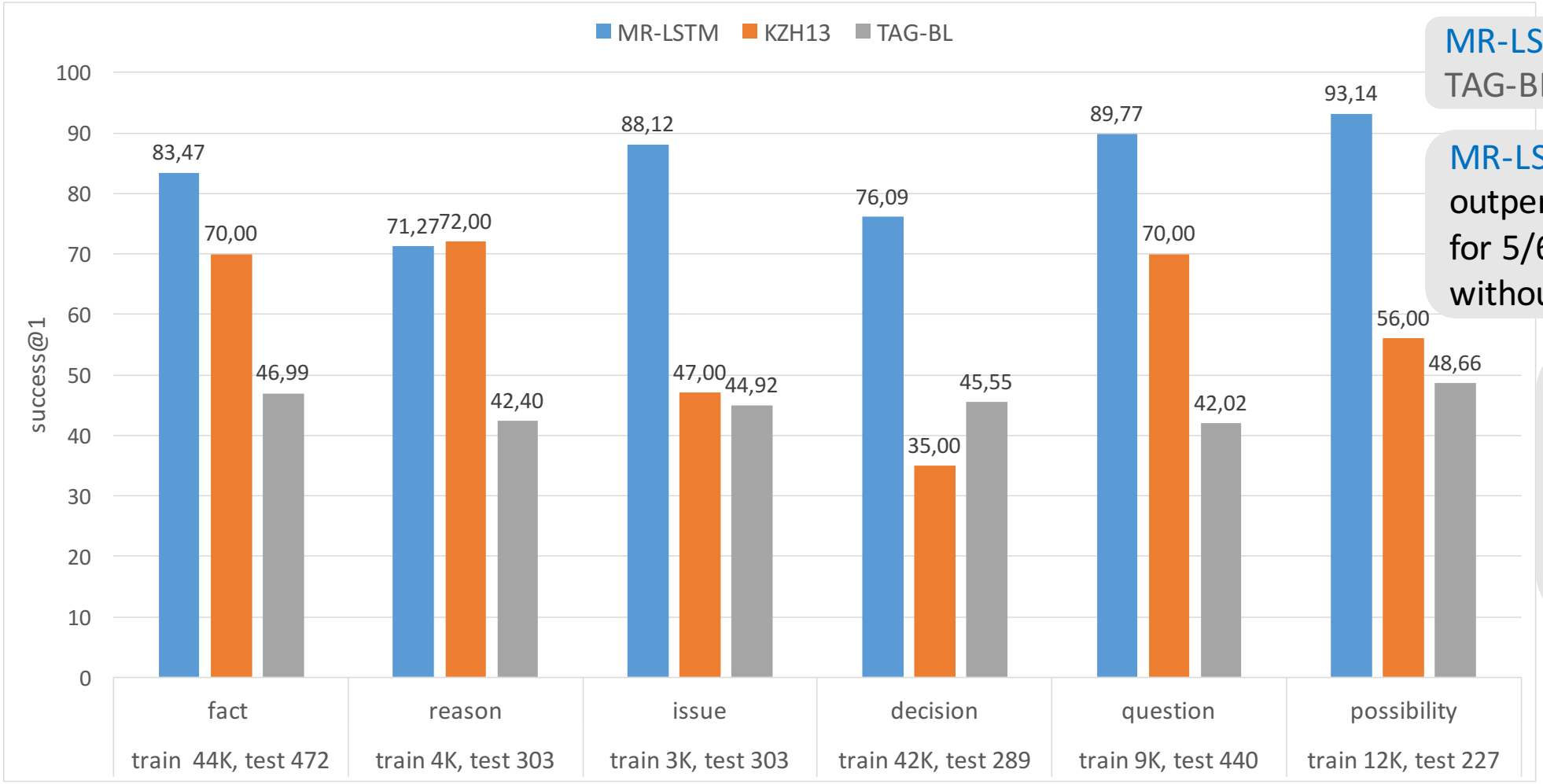
OUR

Kolhatkar et al. (2013)

randomly chooses a candidate with the tag in {S, VP, ROOT, SBAR}



Experiment 1: Shell Noun Resolution



MR-LSTM beats the TAG-BL

MR-LSTM outperforms KZH13's for 5/6 shell nouns without HP tuning

MR-LSTM tuned for *reason* well beyond KZH13

Experiment 2: Unrestricted Abstract Anaphora Resolution

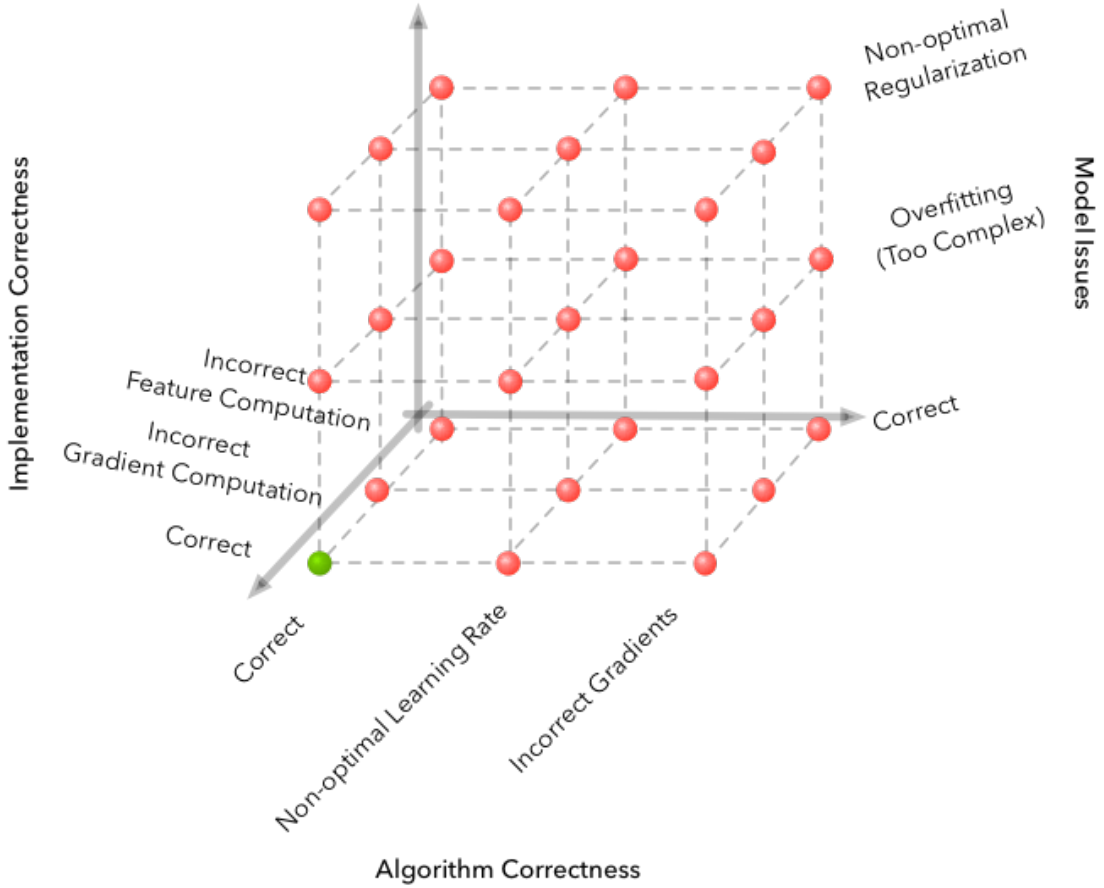
Datasets

- train data: extracted silver data ← pronominal, silver
- dev data: anaphoric shell noun dataset annotated with crowd workers (the ASN corpus) ← nominal, gold
- test data: a *small-scaled subset* of the ARRAU corpus (Uryupina et al., 2016) restricted to unrestricted abstract anaphors (ARRAU-AA) ← pronominal & nominal, gold

Restricted task setup

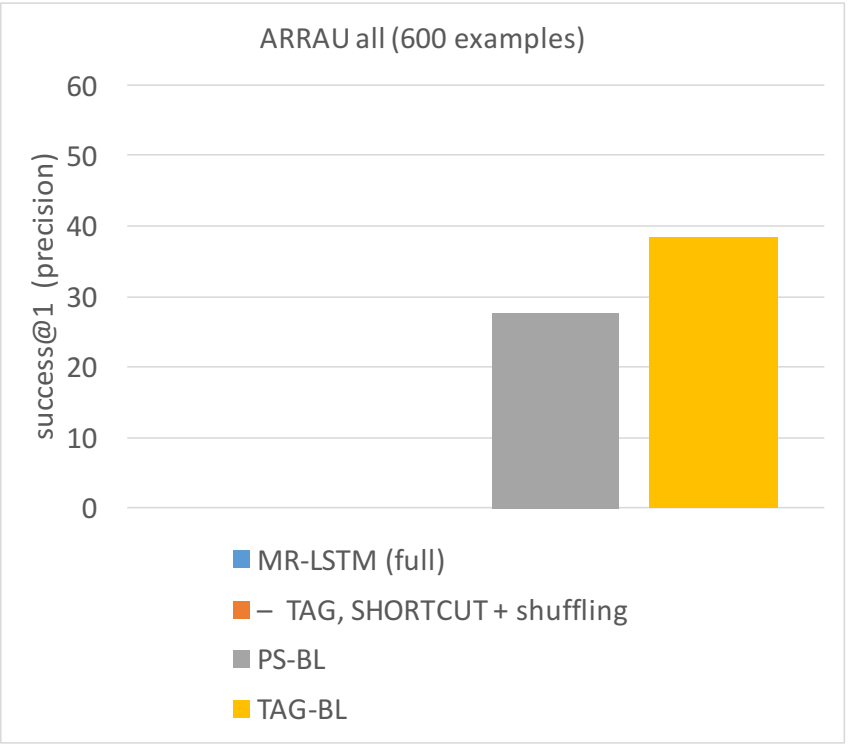
Candidates for the antecedent are all constituents from sentence that contains the antecedent.

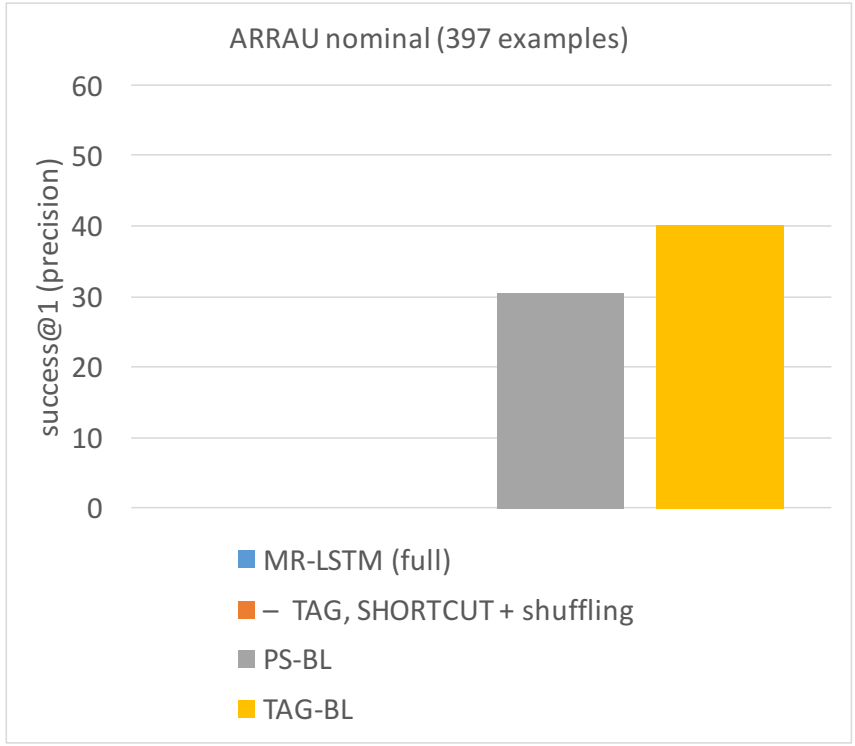
Experiment 2: Unrestricted Abstract Anaphora Resolution
– Hard Task & Exponentially Difficult Debugging

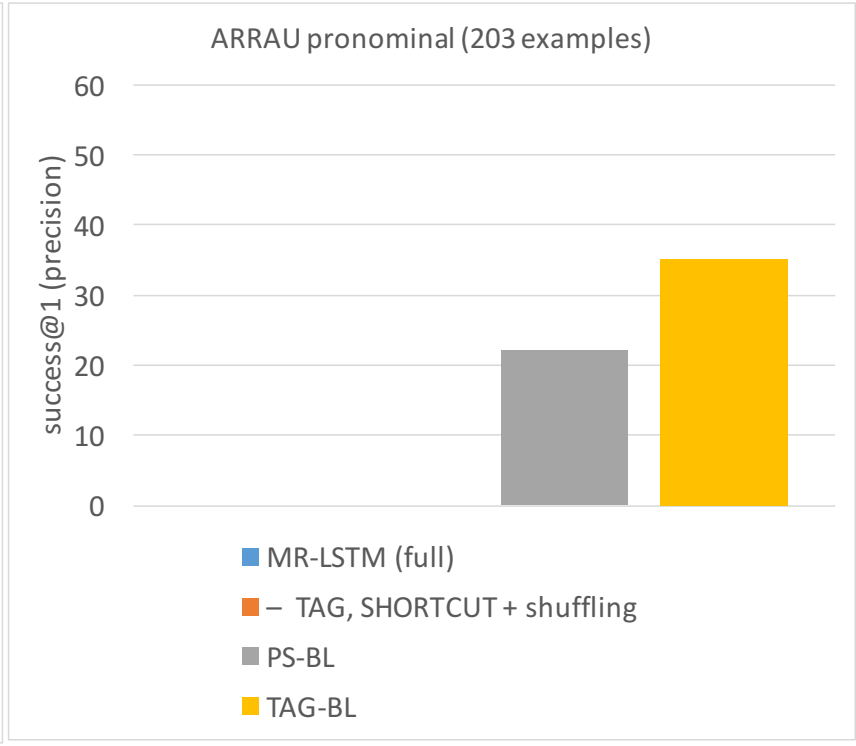


©Photo courtesy of S. Zayd Enam.

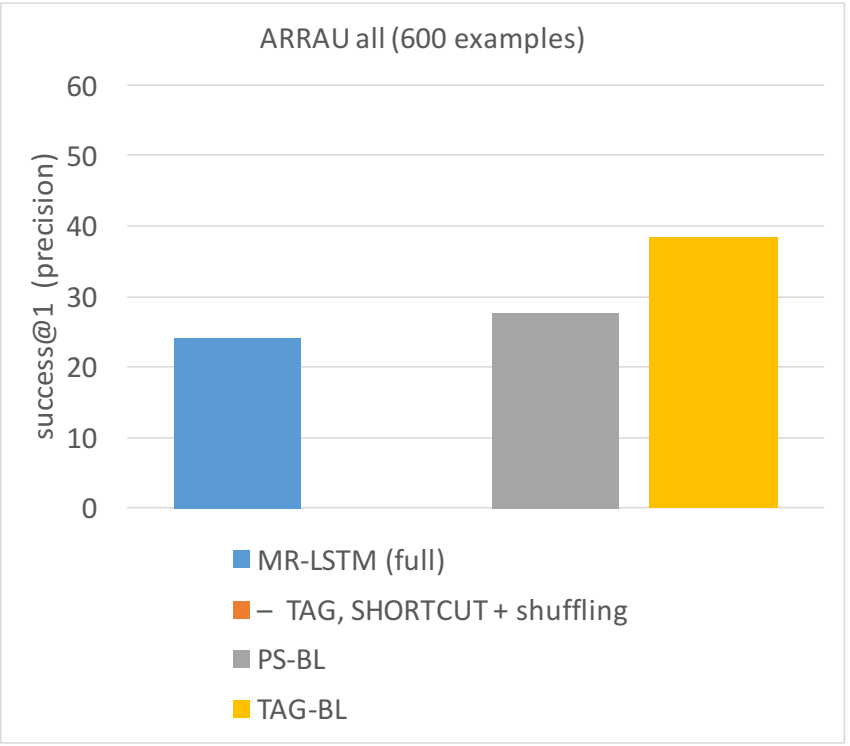
Experiment 2: Unrestricted AAR – all vs. nominal. vs. pronominal – Baselines

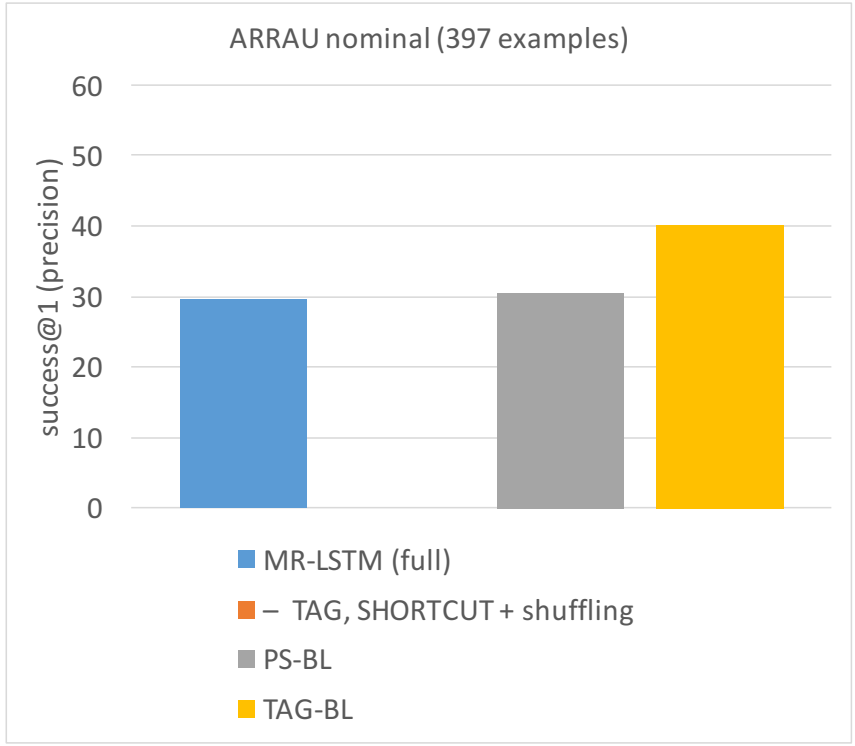


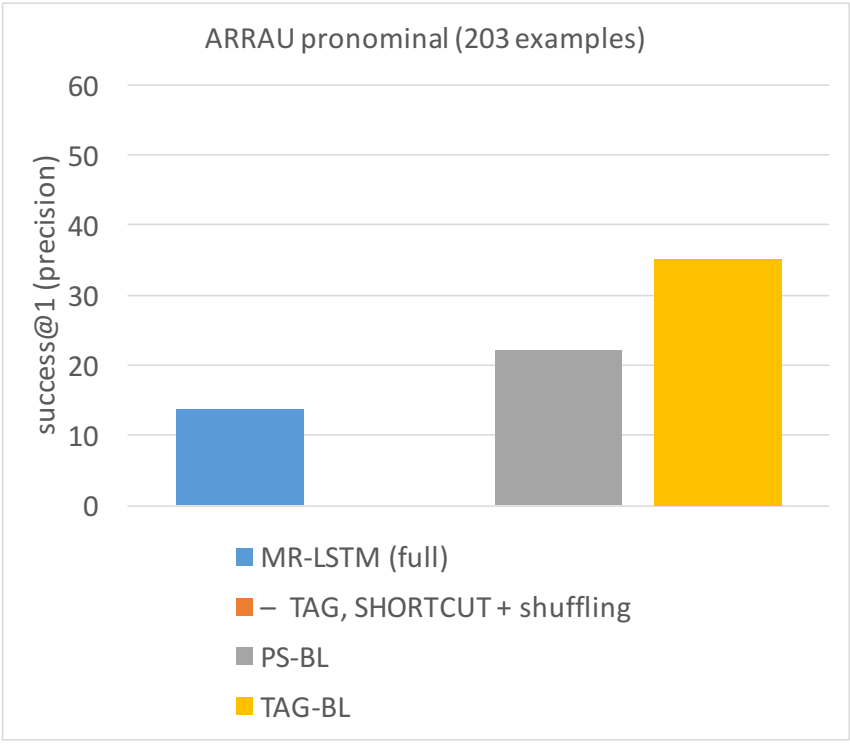




Experiment 2: Unrestricted AAR – all vs. nominal. vs. pronominal – Full MR-LSTM Model

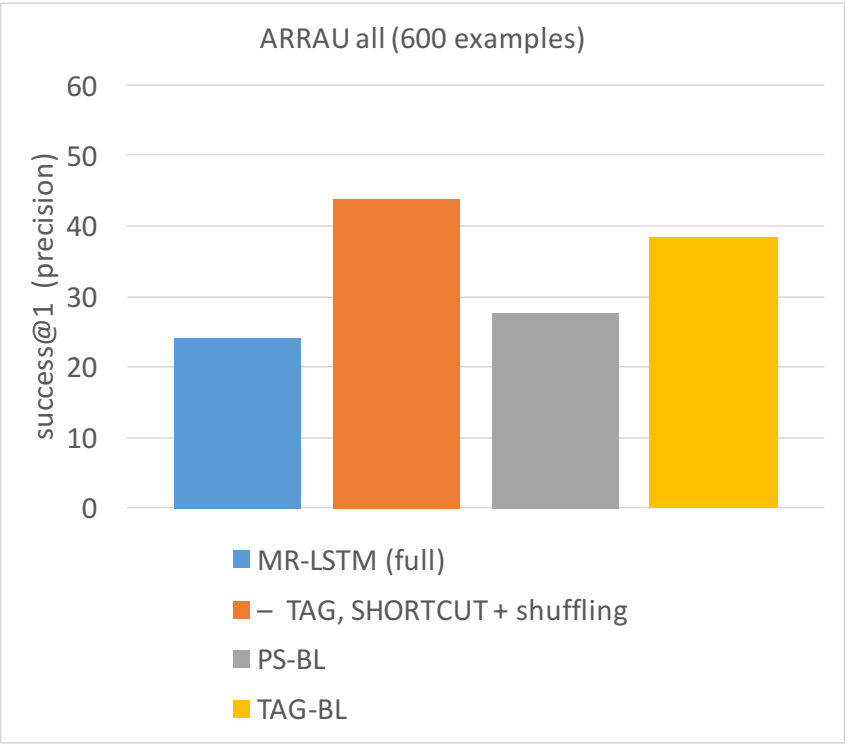


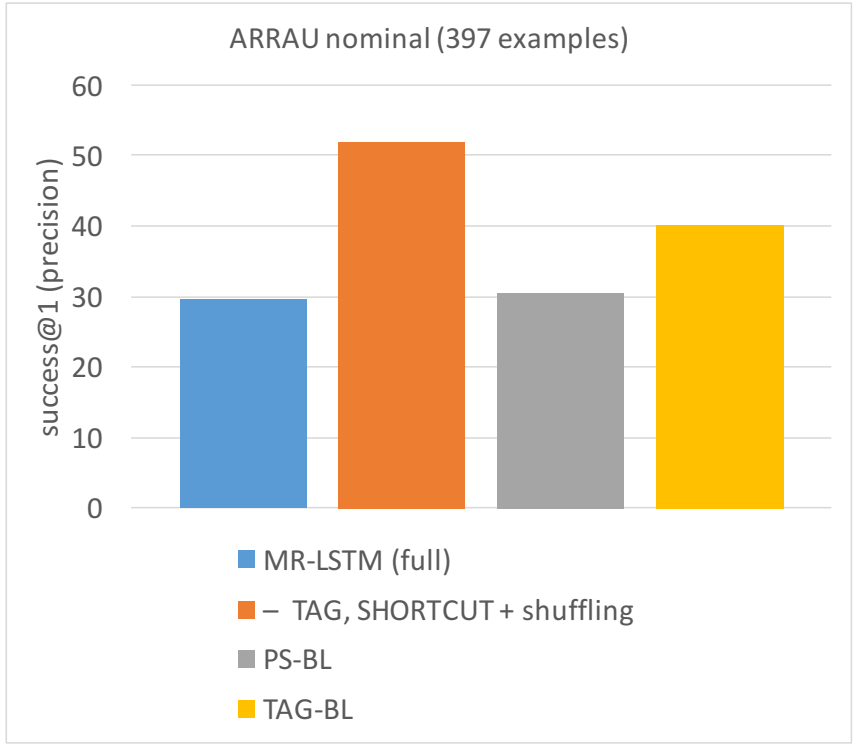


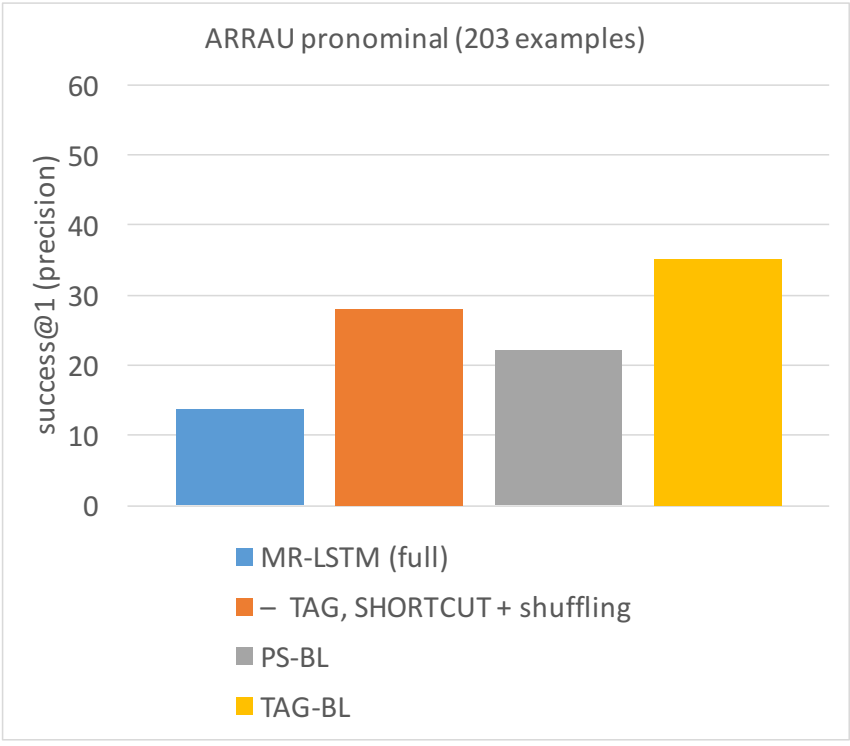


- the full MR-LSTM model is not better than the baselines

Experiment 2: Unrestricted AAR – all vs. nominal. vs. pronominal
– Full MR-LSTM Model without the Syntactic Information + Shuffling of Train Data







- omitting syntactic information & properly shuffling training data
  ⇒ MR-LSTM beats both BLs for nominals and overall, and the preceding sentence BL for pronominals

Taking stock

GOAL

> an unified approach for **unrestricted** AAR

a first step towards the goal is made

APPROACH

> learning what is the correct antecedent for a given AA ≈
learning the relation between the sentence with AA and the antecedent

Siamese Mention-Ranking Model (MR-LSTM)

OBSTACLE

> **scarce** training and evaluation data ⇒ extract training data

extracted training data

CHALLENGES

> **MR-LSTM beats PS-BL for pronouns, but not the TAG-BL**

- Should nominal and pronominal anaphors be learned independently?
- Is harvested data noisy?
- Are natural and extracted data similar enough?

- MR-LSTM outperforms prior work in shell noun resolution
- on ARRAU-AA beats all BLs for nominals

AAR with a Mention-Ranking Siamese NN and Extracted Training Data

EMNLP'17

1. Solving the resource bottleneck – by harvesting training data
2. Resolving Abstract Anaphors in a Relational Model
   – A Mention-Ranking Siamese Network Model

ONGOING

3. A closer look at training data extraction for AAR
4. Ongoing Challenges

Quality of Extracted Data – First Trial

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

- randomly sampled **10 examples per type**
- **two curators** rated the **quality** of constructed data:
  - sound
  - not usable
  - marginal acceptability due to **unnatural sounding anaphora** expression or **position in the sentence**

Quality of Extracted Data – Filtering Noise

- **Filtering embedded clauses introduced by WH-phrases:**
  1. WHNP: That selling of futures contract by elevators [is [**what** [helps keep downward pressure on crop prices during the harves]$_S$]$_{SBAR}$]$_{VP}$.
  2. WHADJP: But some analysts [wonder [**how** [strong the recovery will be]$_S$]$_{SBAR}$]$_{VP}$.
  3. WHADVP: Predictions for limited dollar losses are based largely on the pound's weak state after Mr. Lawson's resignation and the yen's inability to [strengthen substantially [**when** [there are dollar retreats]$_S$]$_{SBAR}$]$_{VP}$.
  4. WHPP: He said, while dialogue is important, enough forums already [exist [in **which** [different interests can express themselve]$_S$]$_{SBAR}$]$_{VP}$.

Quality of Extracted Data – Filtering Noise

- **Filtering embedded clauses introduced by WH-phrases:**
  1. WHNP: That selling of futures contract by elevators [is [**what** [helps keep downward pressure on crop prices during the harves]$_S$]$_{SBAR}$]$_{VP}$.
  2. WHADJP: But some analysts [wonder [**how** [strong the recovery will be]$_S$]$_{SBAR}$]$_{VP}$.
  3. WHADVP: Predictions for limited dollar losses are based largely on the pound's weak state after Mr. Lawson's resignation and the yen's inability to [strengthen substantially [**when** [there are dollar retreats]$_S$]$_{SBAR}$]$_{VP}$.
  4. WHPP: He said, while dialogue is important, enough forums already [exist [in **which** [different interests can express themselve]$_S$]$_{SBAR}$]$_{VP}$.

- **Filtering relative clauses – *that* is the head of SBAR clause and VP has more than 2 children:**
  - The Wall Street Journal [is [an excellent publication]$_{NP}$ [**that** [I enjoy reading and must read daily]$_S$]$_{SBAR}$]$_{VP}$.

Quality of Extracted Data – Filtering Noise

- **Filtering embedded clauses introduced by WH-phrases:**
  1. WHNP: That selling of futures contract by elevators [is [**what** [helps keep downward pressure on crop prices during the harves]$_S$]$_{SBAR}$]$_{VP}$.
  2. WHADJP: But some analysts [wonder [**how** [strong the recovery will be]$_S$]$_{SBAR}$]$_{VP}$.
  3. WHADVP: Predictions for limited dollar losses are based largely on the pound's weak state after Mr. Lawson's resignation and the yen's inability to [strengthen substantially [**when** [there are dollar retreats]$_S$]$_{SBAR}$]$_{VP}$.
  4. WHPP: He said, while dialogue is important, enough forums already [exist [in **which** [different interests can express themselve]$_S$]$_{SBAR}$]$_{VP}$.

- **Filtering relative clauses – *that* is the head of SBAR clause and VP has more than 2 children:**
  - The Wall Street Journal [is [an excellent publication]$_{NP}$ [**that** [I enjoy reading and must read daily]$_S$]$_{SBAR}$]$_{VP}$.

- **With caution**: *since, as, if*
  - The Indian stock markets [have been on a five-year high, with dips and corrections, [**since** [prime minister Rajiv Gandhi started liberalizing industry]$_S$]$_{SBAR}$]$_{VP}$.
  - Rraditionally, boiler rooms [operate on the cheap, [**since** [few, if any, customers ever visit their offices]$_S$]$_{SBAR}$]$_{VP}$.

use PRP, TMP & ADV parse attributes (gold parse only)

Quality of Extracted Data – Filtering Noise

- **Filtering embedded clauses introduced by WH-phrases:**
  1. WHNP: That selling of futures contract by elevators [is [**what** [helps keep downward pressure on crop prices during the harves]$_S$]$_{SBAR}$]$_{VP}$.
  2. WHADJP: But some analysts [wonder [**how** [strong the recovery will be]$_S$]$_{SBAR}$]$_{VP}$.
  3. WHADVP: Predictions for limited dollar losses are based largely on the pound's weak state after Mr. Lawson's resignation and the yen's inability to [strengthen substantially [**when** [there are dollar retreats]$_S$]$_{SBAR}$]$_{VP}$.
  4. WHPP: He said, while dialogue is important, enough forums already [exist [in **which** [different interests can express themselve]$_S$]$_{SBAR}$]$_{VP}$.

- **Filtering relative clauses – *that* is the head of SBAR clause and VP has more than 2 children:**
  - The Wall Street Journal [is [an excellent publication]$_{NP}$ [**that** [I enjoy reading and must read daily]$_S$]$_{SBAR}$]$_{VP}$.

- **With caution**: *since, as, if*
  - The Indian stock markets [have been on a five-year high, with dips and corrections, [**since** [prime minister Rajiv Gandhi started liberalizing industry]$_S$]$_{SBAR}$]$_{VP}$.
  - Rraditionally, boiler rooms [operate on the cheap, [**since** [few, if any, customers ever visit their offices]$_S$]$_{SBAR}$]$_{VP}$.

- **Remove *while*: often result with contrastive reading**
  - Kellogg's current share is believed to [be slightly under 40 % [**while** [general Mills' share is about 27 %]$_S$]$_{SBAR}$]$_{VP}$.

Quality of Extracted Data – Filtering Noise

- Filtering embedded clauses introduced by WH-phrases:
- Filtering relative clauses – *that* is the head of SBAR clause and VP has more than 2 children:
  - The Wall Street Journal [is [an excellent publication]$_{NP}$ [**that** [I enjoy reading and must read daily]$_{S}$]$_{SBAR}$]$_{VP}$.
- With caution: *since, as, if*
  - The Indian stock markets [have been on a five-year high, with dips and corrections, [**since** [prime minister Rajiv Gandhi started liberalizing industry]$_{S}$]$_{SBAR}$]$_{VP}$.
  - Rraditionally, boiler rooms [operate on the cheap, [**since** [few, if any, customers ever visit their offices]$_{S}$]$_{SBAR}$]$_{VP}$.
- Remove *while*: often result with contrastive reading
  - Kellogg's current share is believed to [be slightly under 40 % [**while** [general Mills' share is about 27 %]$_{S}$]$_{SBAR}$]$_{VP}$.
- With caution: *whether*
  - But the test may prove to be more sensitive in [determining [**whether** [a tumor has spread **or** returned following treatment]$_{S}$]$_{SBAR}$]$_{VP}$.

Quality of Extracted Data – Second Trial

- filtered noise
- **added two new replacement types:** after (replacement: after this/that/it) and until (replacement: until this/that/it)
- repeated evaluation
- remove *until*

Making Harvested Data More Similar to Natural Data – Preprocessing

additional dataset (abstract events)

more training data extracted from NYT

train	ASN	CoNLL12-Ev	WSJ (manual)	NYT (parsed)	
	1563	1720			
test	ASN	CoNLL12-Ev	ARRAU	ARRAU-nom	ARRAU-pro
	373	243	600	397	203
dev	ASN	CoNLL12-Ev			
	300	180			

- Silver
- Gold

- # candidates at least 20 (30) for instances extracted from WSJ (NYT)
- anaphoric sentence at least 15 tokens long
- remove says/saying/said/say instances from NYT and leave only 100 from WSJ

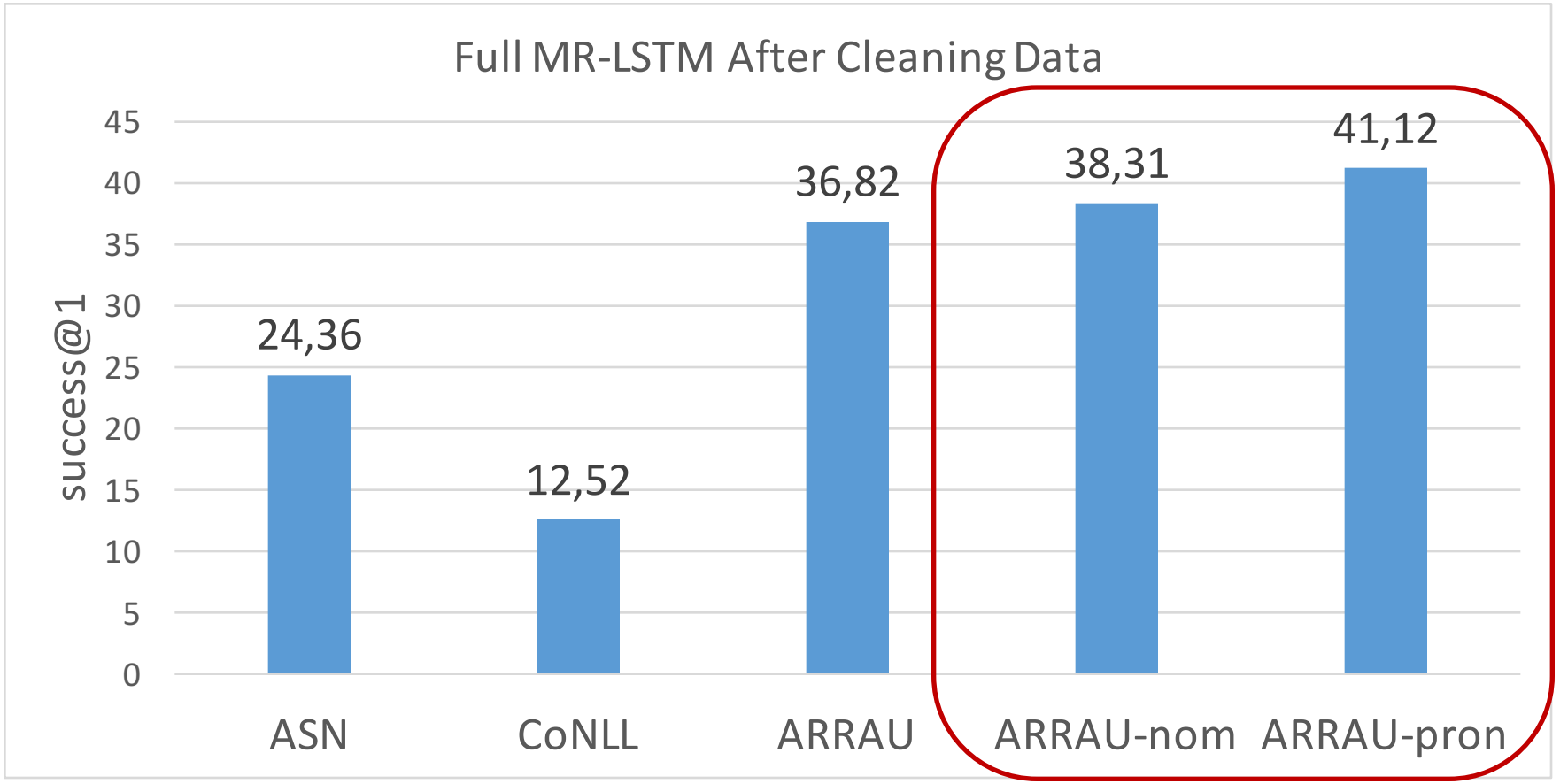
Making Harvested Data More Similar to Natural Data – Preprocessing

Assure that candidate which contains the governing verb never ends up in candidate list:

- Although British Air is waiting to see what the buy-out group comes up with, Mr. Stevens [**said** [a revised transaction with less debt leverage is likely to be more attractive to banks]$_{S}$]$_{SBAR}$]$_{VP}$.
- **antecedent sentence:** Although British Air is waiting to see what the buy-out group comes up with, Mr. Stevens **said** this.
- **candidate**: **said** a revised transaction with less debt leverage is likely to be more attractive to banks

Experiment 3: Training Full MR-LSTM with Cleaned Data

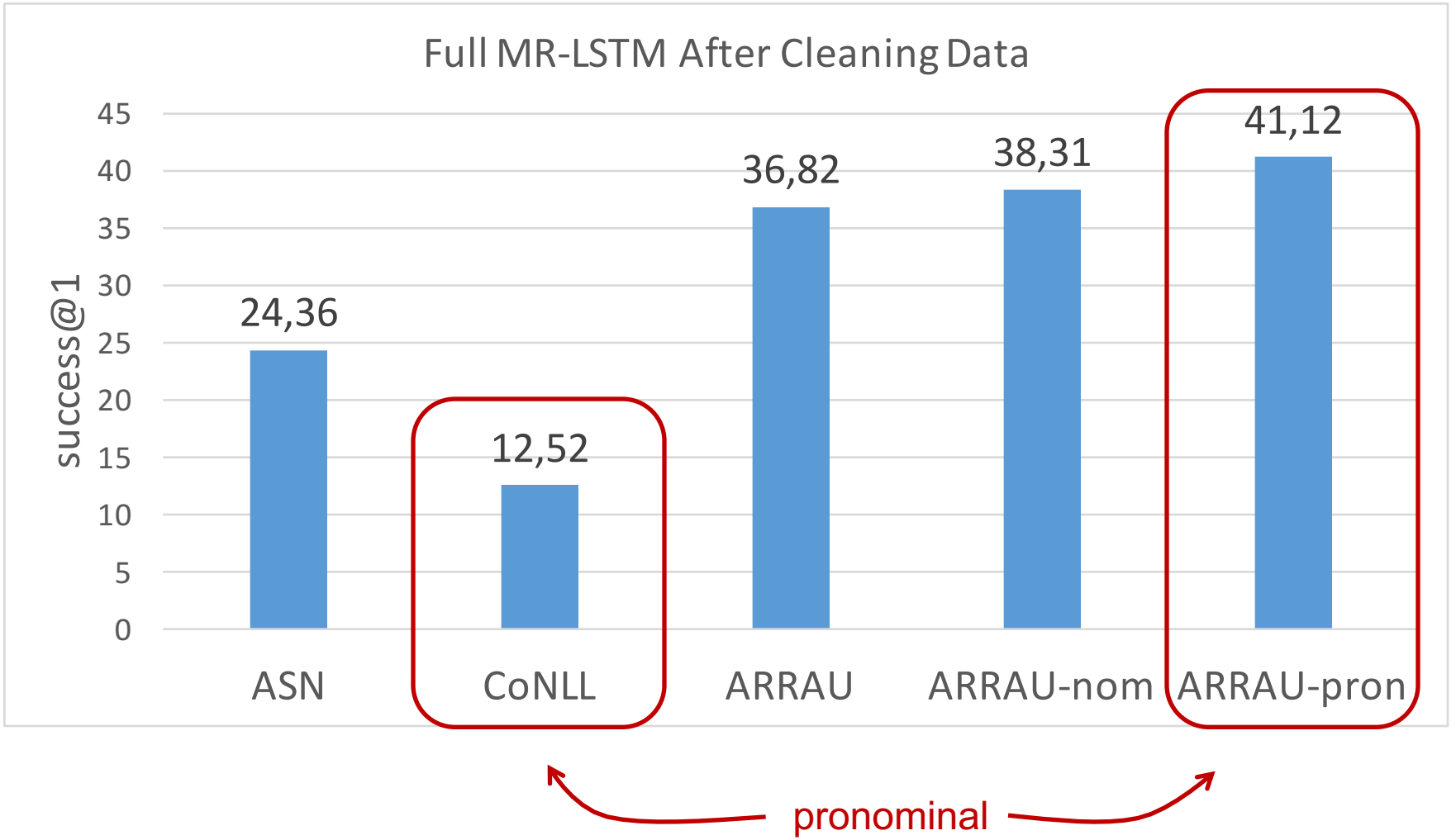
–Preliminary Results



difference between nominal and pronominal cases in only few s@1 points (before 20+)

Experiment 3: Training Full MR-LSTM with Cleaned Data

–Preliminary Results



Taking stock

GOAL

> an unified approach for **unrestricted** AAR

CHALLENGES

> MR-LSTM beats TAG-BL for pronouns, but not the preceding sentence BL:

- Should nominal and pronominal anaphors be learned independently? NO!
- Is harvested data noisy? IT WAS NOISY. IT IS CLEAN NOW.
- Are natural and extracted data similar enough? STILL HAS TO BE FIGURED OUT.

- Are **properties** of all types of abstract anaphors **homogenous enough**?
- Do we need **specifically adapted training data**?
- Are all anaphoric types **inherently relational**?
- How does MR-LSTM model perform in the **realistic task setup**?

AAR with a Mention-Ranking Siamese NN and Extracted Training Data

EMNLP'17

1. Solving the resource bottleneck – by harvesting training data
2. Resolving Abstract Anaphors in a Relational Model
   – A Mention-Ranking Siamese Network Model

ONGOING

3. A closer look at training data extraction for AAR
4. Ongoing Challenges

Question 1: Are natural and extracted data similar enough?

–MR-LSTM with Adversarial Training

Why Are We Still Worried About Data Differences?

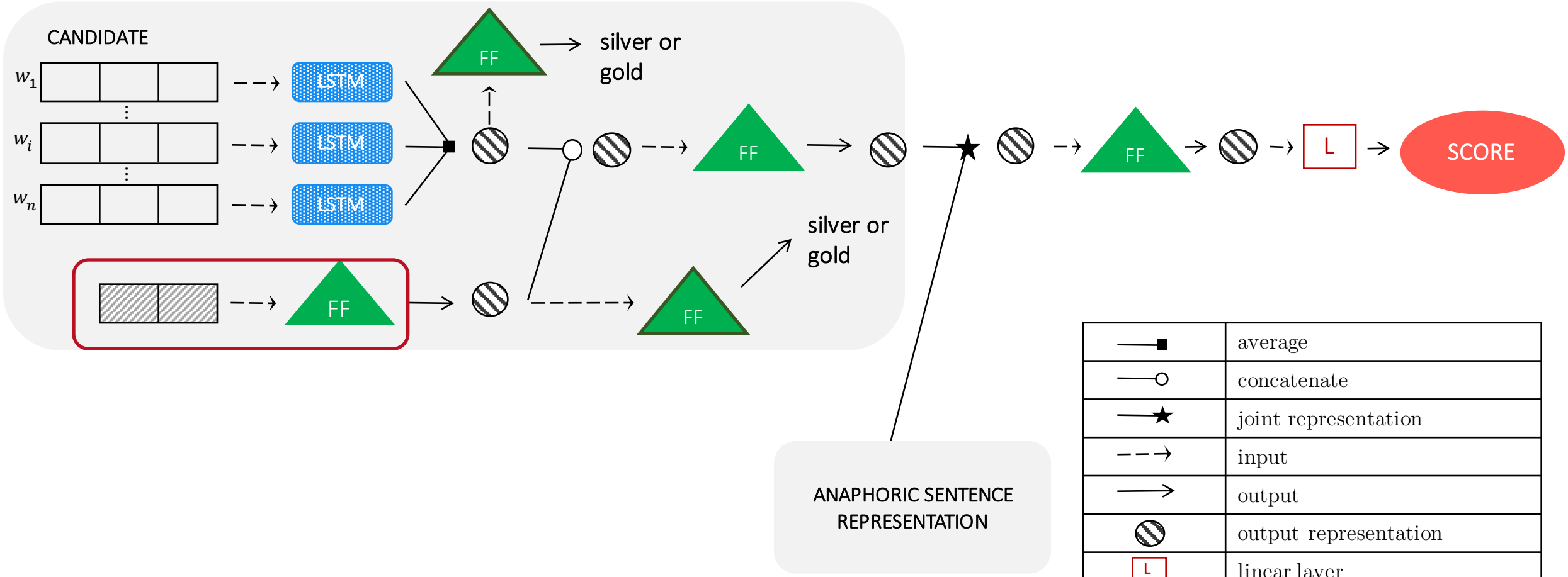
- Pre-processing done using only the ASN (shell nouns) and the CoNLL12-Ev (event) data.
- Harvested and natural data may differ in some complex features we did not cover.
- Even natural datasets differ.

re-think using ARRAU only for testing

Can we reduce remaining biases with adversarial training?

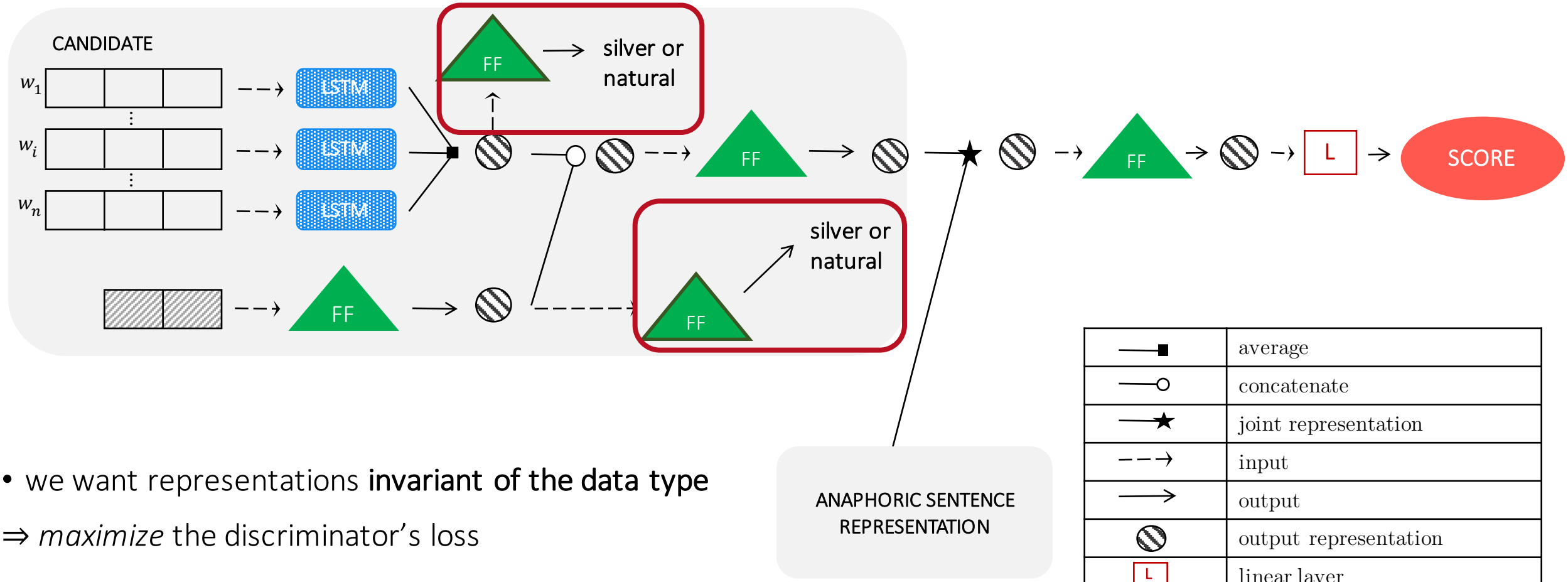
Question 1: Are natural and extracted data similar enough?

–MR-LSTM with Adversarial Training



Question 1: Are natural and extracted data similar enough?

–MR-LSTM with Adversarial Training



	average
	concatenate
	joint representation
	input
	output
	output representation
L	linear layer
	ELU feed-forward layer

- we want representations **invariant of the data type**

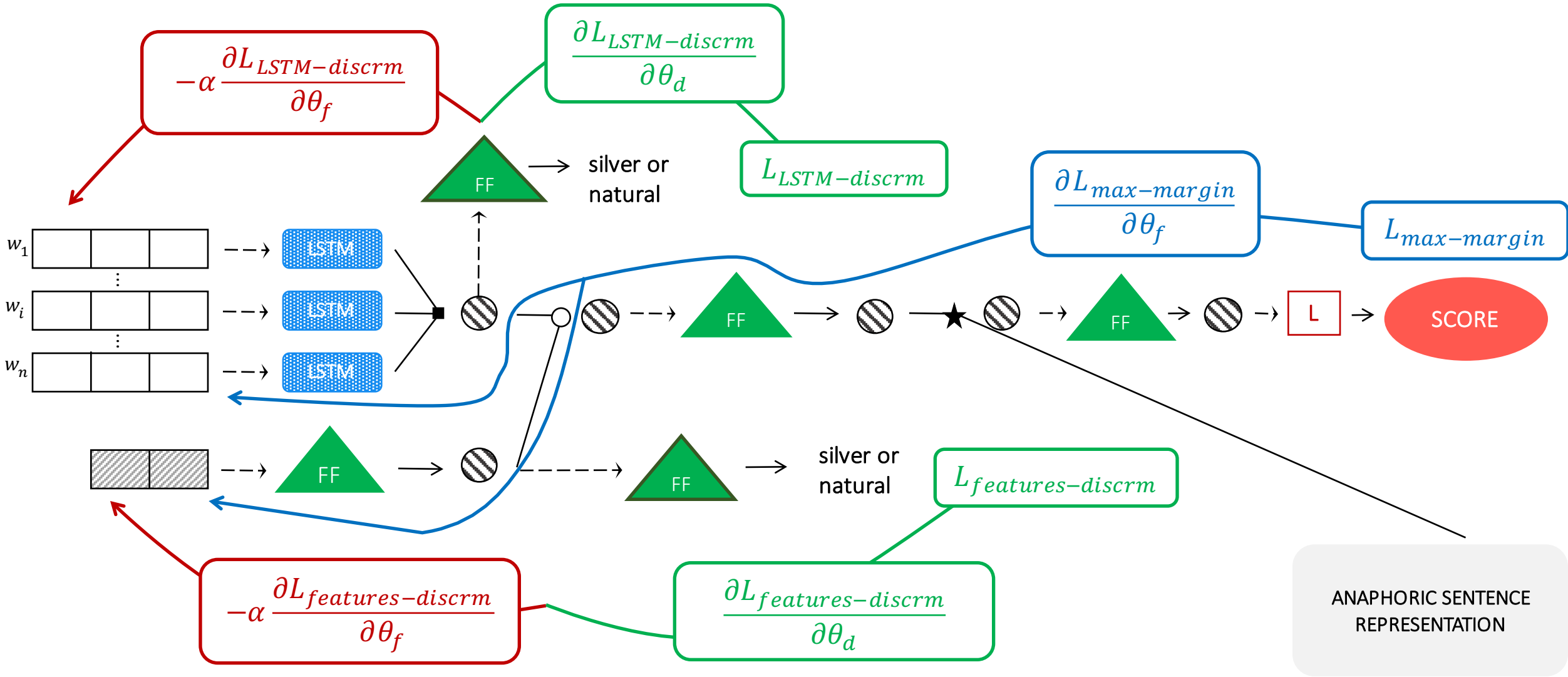
⇒ *maximize* the discriminator's loss

- we want that the **discriminator challenges the model**

⇒ *minimize* the discriminator's loss

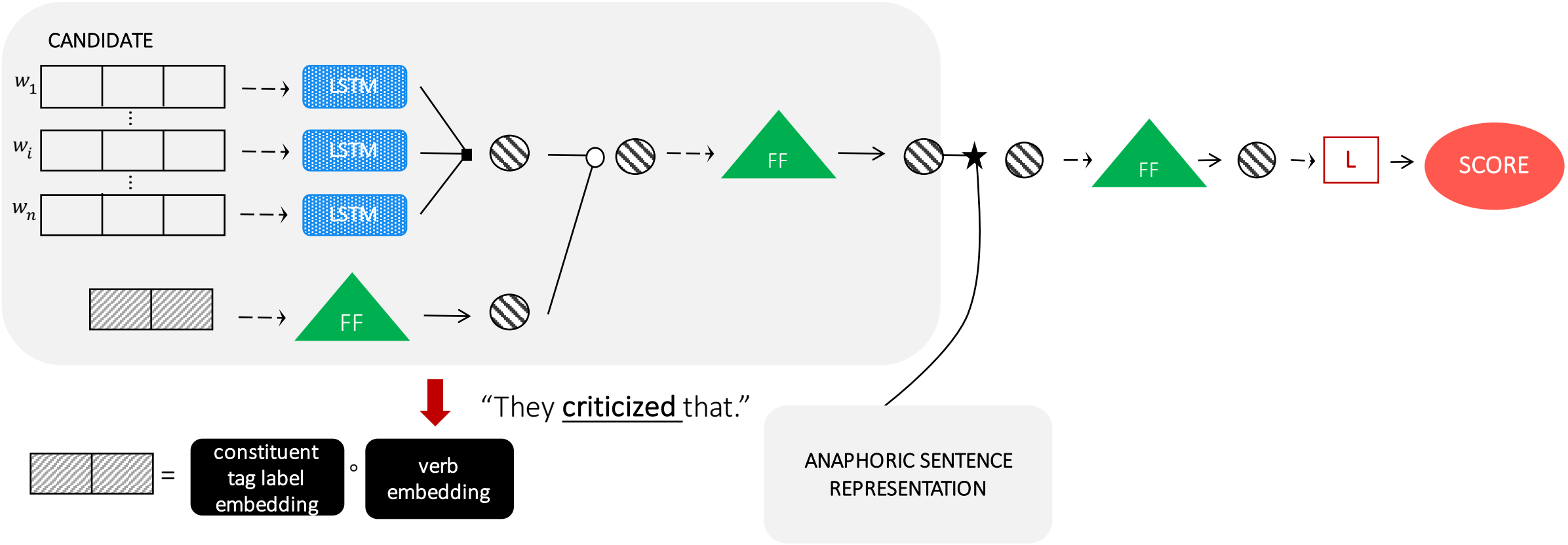
Question 1: Are natural and extracted data similar enough?

–MR-LSTM with Adversarial Training



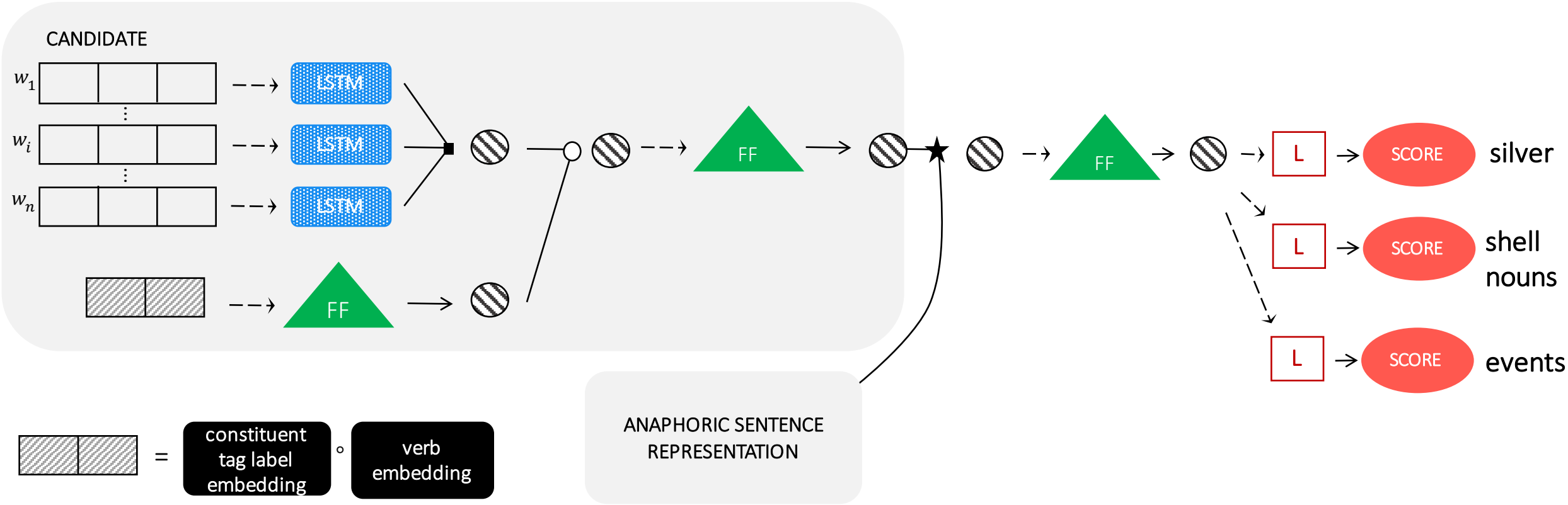
Question 2: Are properties of all types of abstract anaphors homogenous enough?

– MR-LSTM with Multi-Task Learning



Question 2: Are properties of all types of abstract anaphors homogenous enough?

– MR-LSTM with Multi-Task Learning



Question 3: How To Train MR-LSTM to Select Antecedent from Wider Context?

Current task setup

U.S. car dealers had an average of 59 days supply of cars in their lots at the end of September, according to Ward's Automotive Reports. But Mr. Tonkin said dealers should slash stocks to between 15 and 30 days to reduce the costs of financing inventory. His message is getting a chilly reception in Detroit, where the Big Three auto makers are already being forced to close plants because of soft sales and reduced dealer orders. Even before Mr. Tonkin's broadside, some large dealers said they were cutting inventories. Ford Motor Co. and Chrysler Crp. representatives criticized Mr. Tonkin's plan as unworkable.

Selecting the antecedent from wider context:
Related work* => Distance is an important feature!

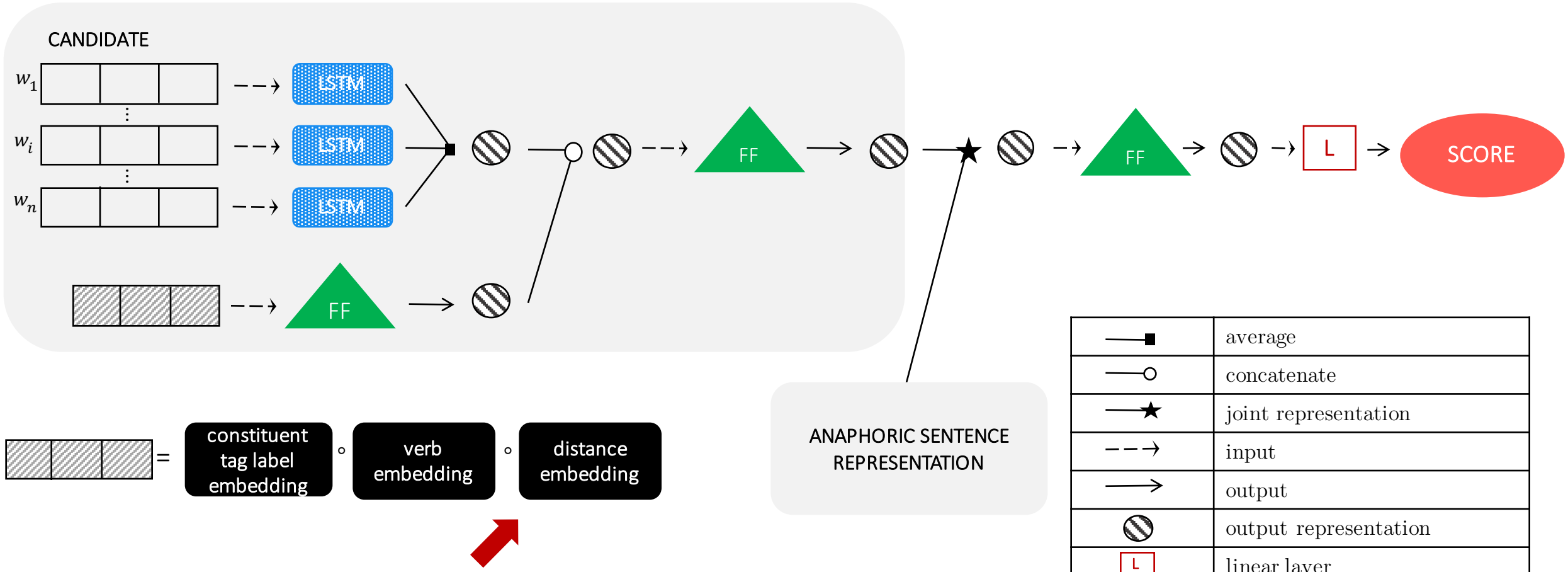
More realistic task setup

U.S. car dealers had an average of 59 days supply of cars in their lots at the end of September, according to Ward's Automotive Reports. But Mr. Tonkin said dealers should slash stocks to between 15 and 30 days to reduce the costs of financing inventory. His message is getting a chilly reception in Detroit, where the Big Three auto makers are already being forced to close plants because of soft sales and reduced dealer orders. Even before Mr. Tonkin's broadside, some large dealers said they were cutting inventories. Ford Motor Co. and Chrysler Crp. representatives criticized Mr. Tonkin's plan as unworkable.

*Jauhar et al. (2015), Anand and Hardt (2016), among others.

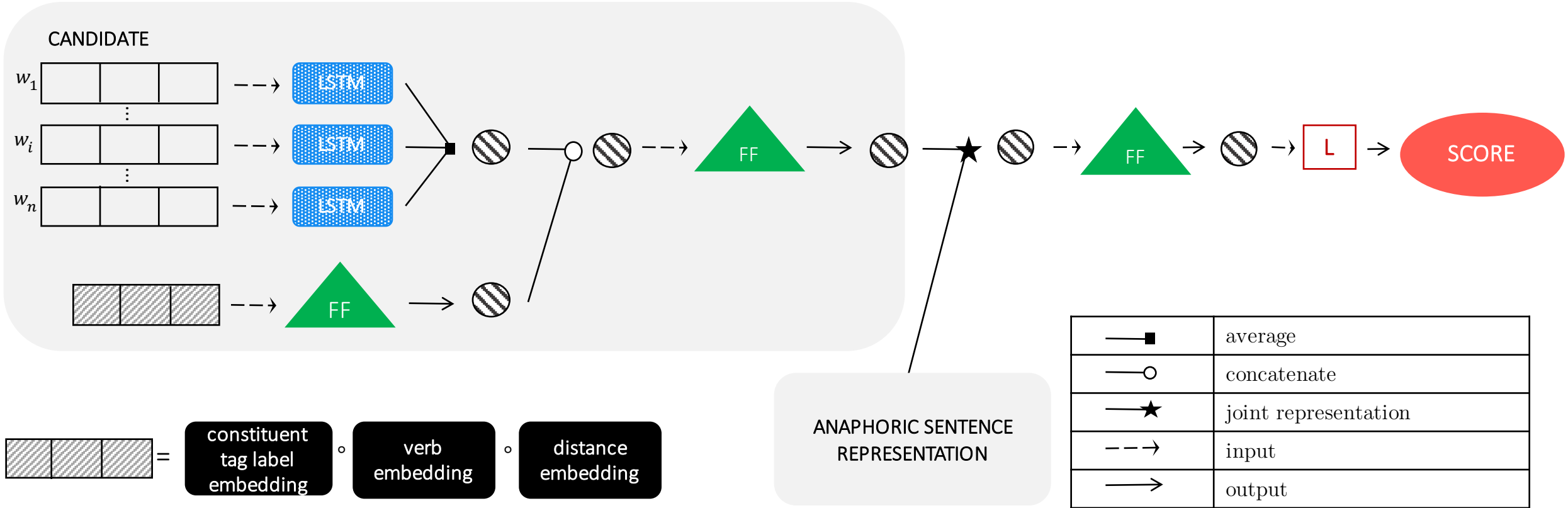
Question 3: How To Train MR-LSTM to Select Antecedent from Wider Context?

– Equipping the Relational Model



Question 3: How To Train MR-LSTM to Select Antecedent from Wider Context?

– Equipping the Relational Model



	average
	concatenate
	joint representation
	input
	output
	output representation
L	linear layer
	ELU feed-forward layer

Harvested training data does **not** contain **natural distances**

⇒ Sample **distance** from a distribution calculated from the natural data

Outlook

GOAL

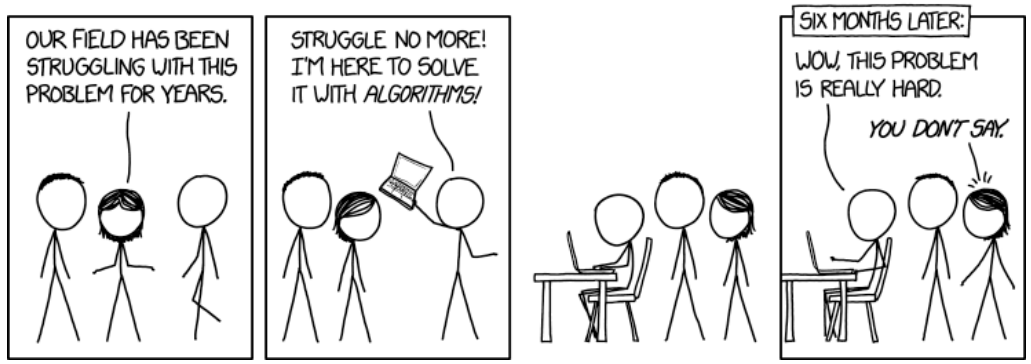
> an unified approach for **unrestricted** AAR

Made good progress until now

- stable baseline neural model for ”core phenomena”
- learned how to integrate harvested training data

The higher we get, the steeper it gets

- differences in harvested and natural data
- differences between anaphora types
- selecting antecedent from wider context



©Photo courtesy of xkcd

Thank you for your attention!

ADAPTIVE INFORMATIONSAUFBEREITUNG AUS HETEROGENEN QUELLEN

RUPRECHT-KARLS-
UNIVERSITÄT
HEIDELBERG

Heidelberger Institut für
Theoretische Studien

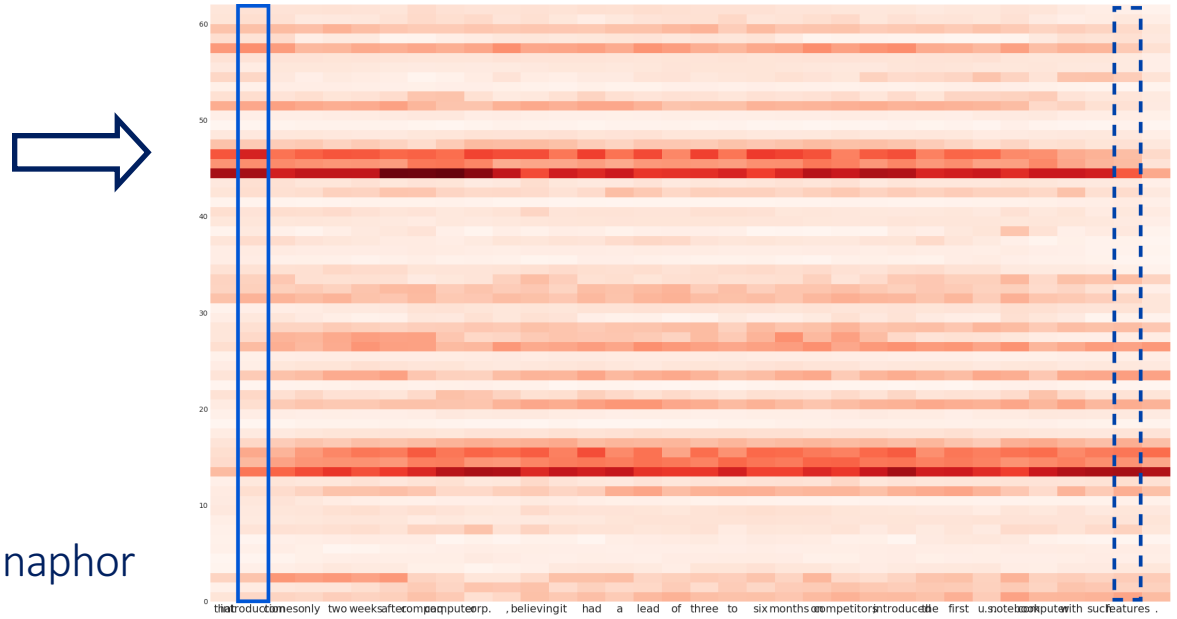
TECHNISCHE
UNIVERSITÄT
DARMSTADT

References

- Anand, P., and Hardt, D. (2016). Antecedent Selection for Sluicing: Structure and Content. In Proceedings of *EMNLP.*
- Clark, K. and Manning, C. D. (2015). Entity-centric coreference resolution with model stacking. In Proceedings of the ACL. Beijing, China.
- Dipper, S., & Zinsmeister, H. (2012). Annotating abstract anaphora. *Language Resources and Evaluation*, *46*(1), 37-52.
- Jauhar, S. K., Guerra, R., Pellicer, E. G., & Recasens, M. (2015). Resolving discourse-deictic pronouns: A two-stage approach to do it. In *Proceedings of the Fourth Joint Conference on Lexical and Computational Semantics*.
- 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*.
- Lu, J. and Ng, V. (2017). Joint Learning for Event Coreference Resolution. In Proceedings of the ACL.
- Marasović, A., Born, L., Opitz, J., & Frank, A. (2017). A Mention-Ranking Model for Abstract Anaphora Resolution. *In Proceedings of the EMNLP.*
- Poesio, M., & Artstein, R. (2008, May). Anaphoric Annotation in the ARRAU Corpus. In Proceedings of the *LREC*.
- Stede, M., and Grishina, Y. Anaphoricity in Connectives: A Case Study on German. In Proceedings of the *CORBON@HLT-NAACL.*
- 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*.
- Vieira, R., Salmon-Alt, S., Gasperin, C., Schang, E., & Othero, G. (2005). Coreference and anaphoric relations of demonstrative noun phrases in multilingual corpus. *Anaphora Processing: linguistic, cognitive and computational modeling*, 385-403.
- Webber, B. L. (1991). Structure And Sstension In The Interpretation Of Discourse Deixis. *Language and Cognitive processes*, *6*(2), 107-135.

Experiment 2: 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?



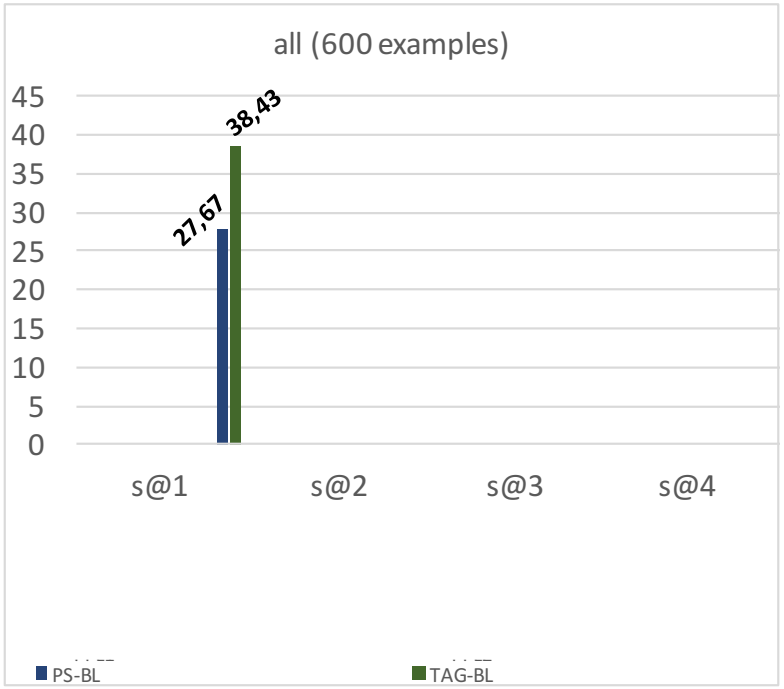
first anaphor

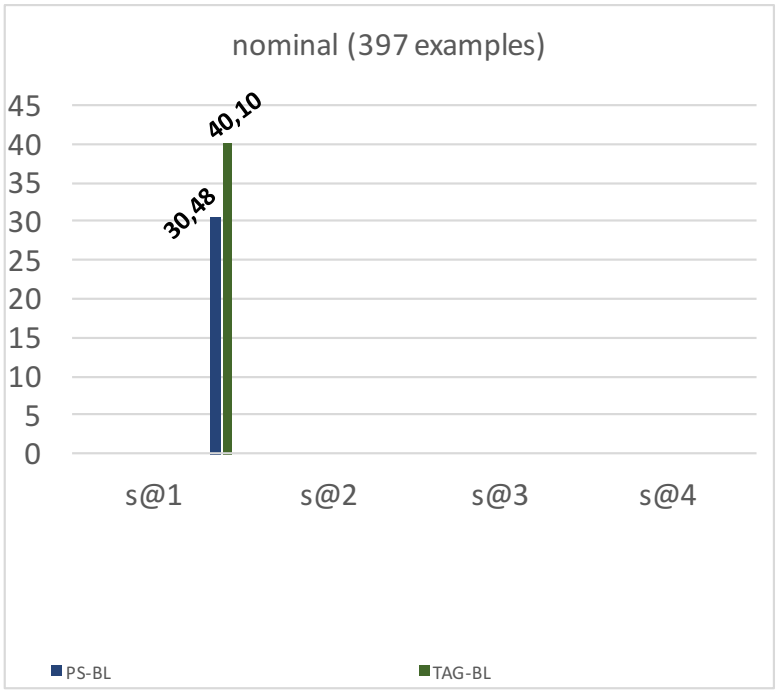
second anaphor

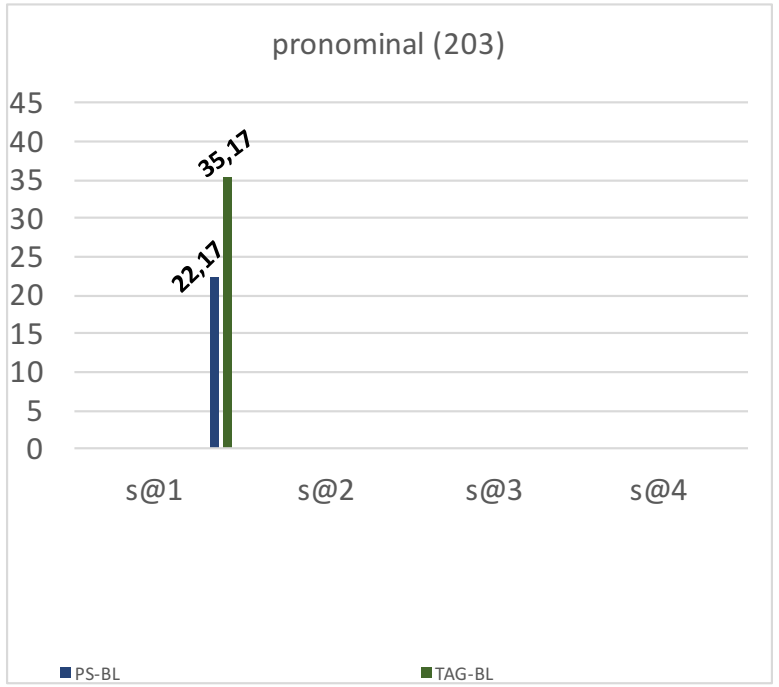
The heat-maps illustrates the difference in output of the bi-LSTM for the same **anaphoric sentence** with **two** anaphors when the **first** vs. **second** anaphor is considered.

the representations differ ⇒ the joint representations with the candidate differ as well

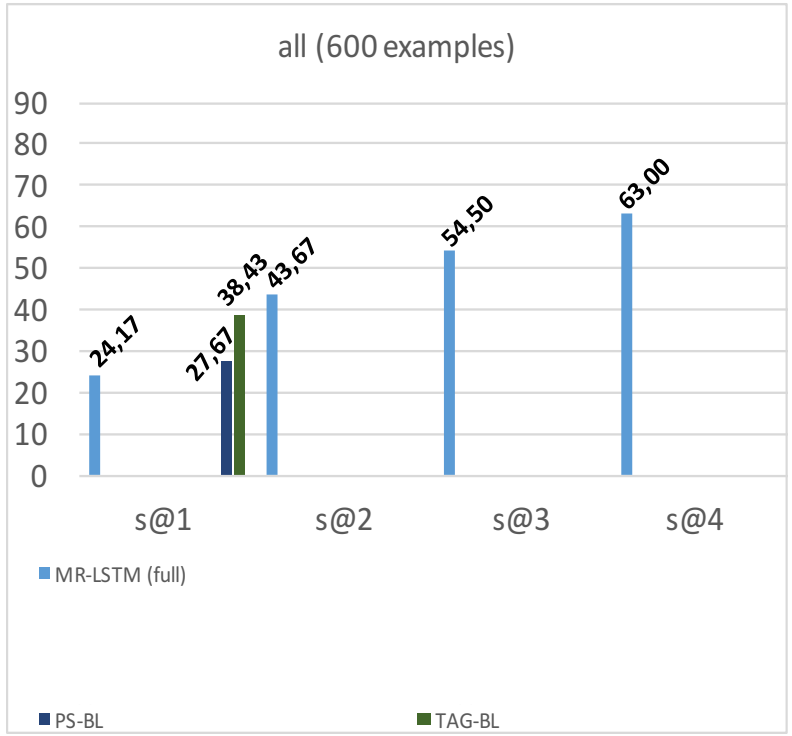
Experiment 2: Unrestricted AAR – all vs. nominal. vs. pronominal – baselines

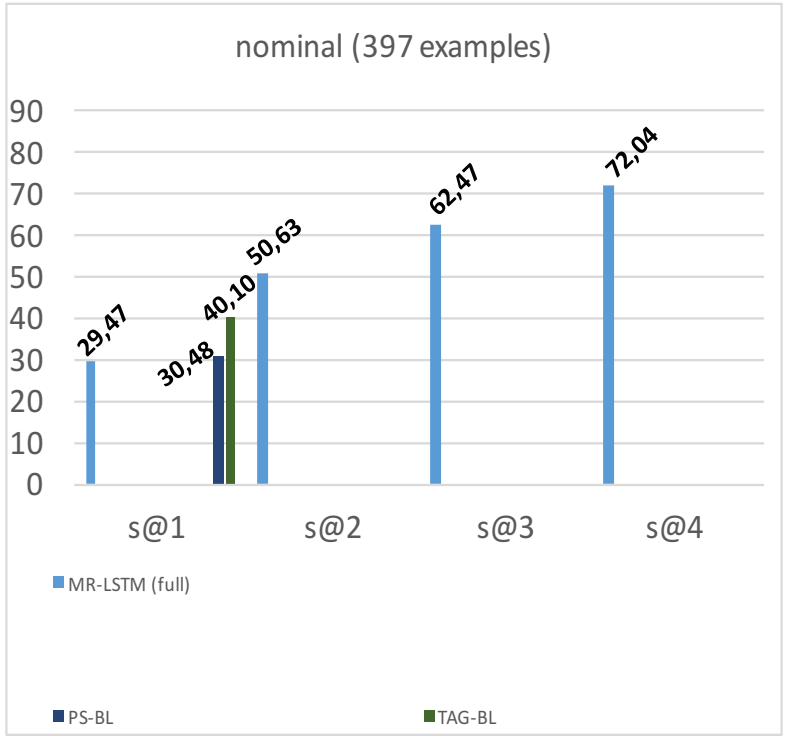


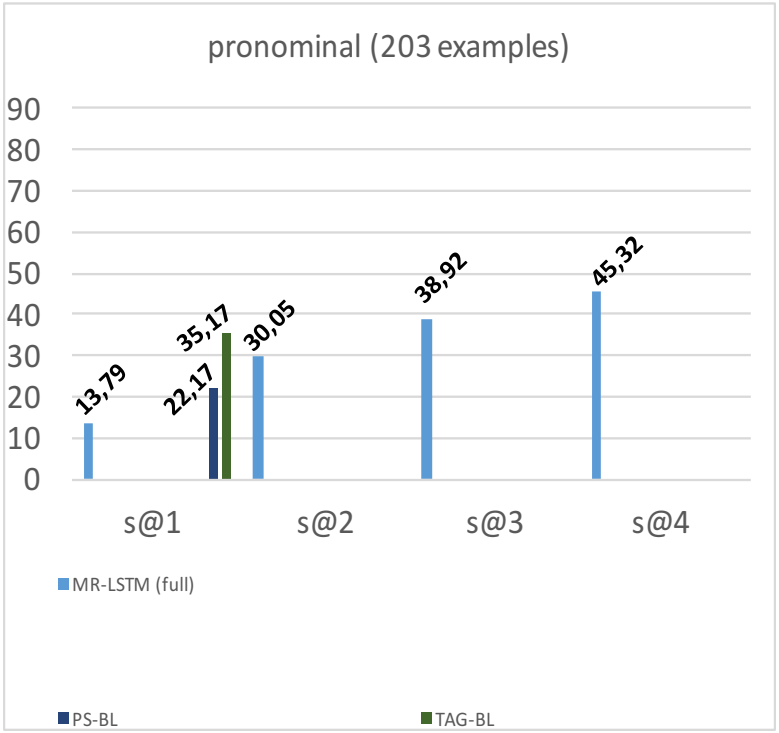




Experiment 2: Unrestricted AAR – all vs. nominal. vs. pronominal – full MR-LSTM model

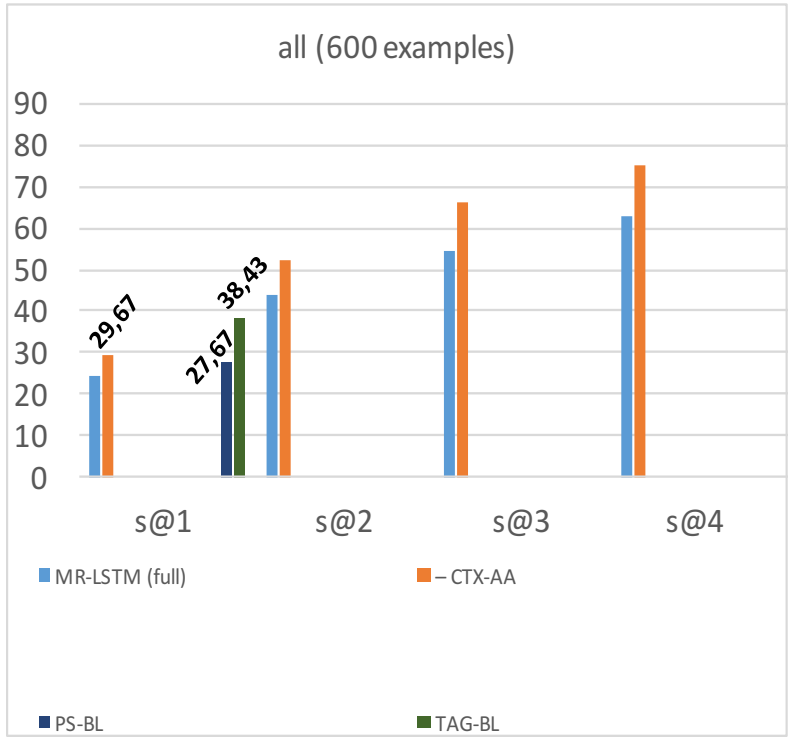


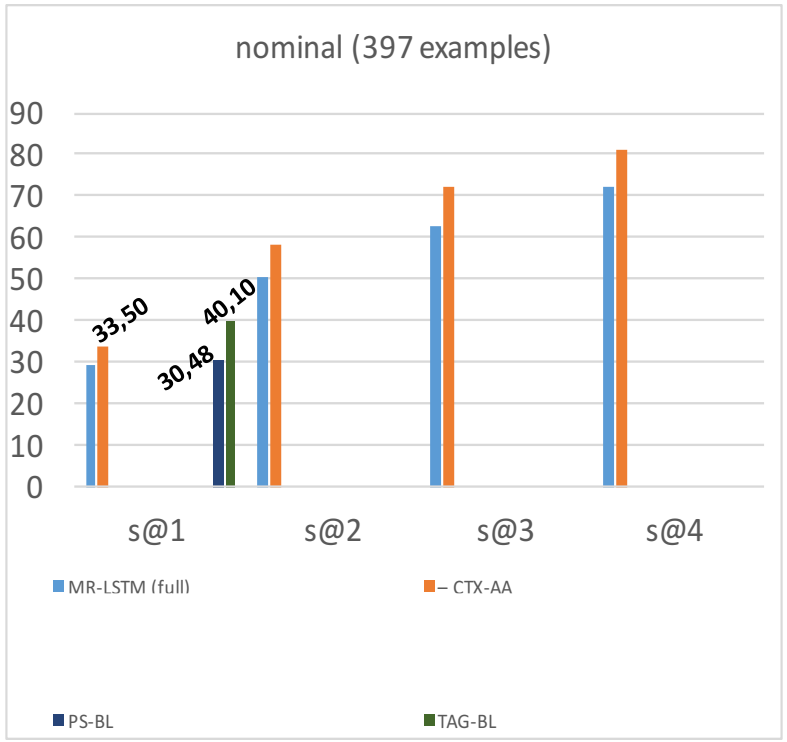


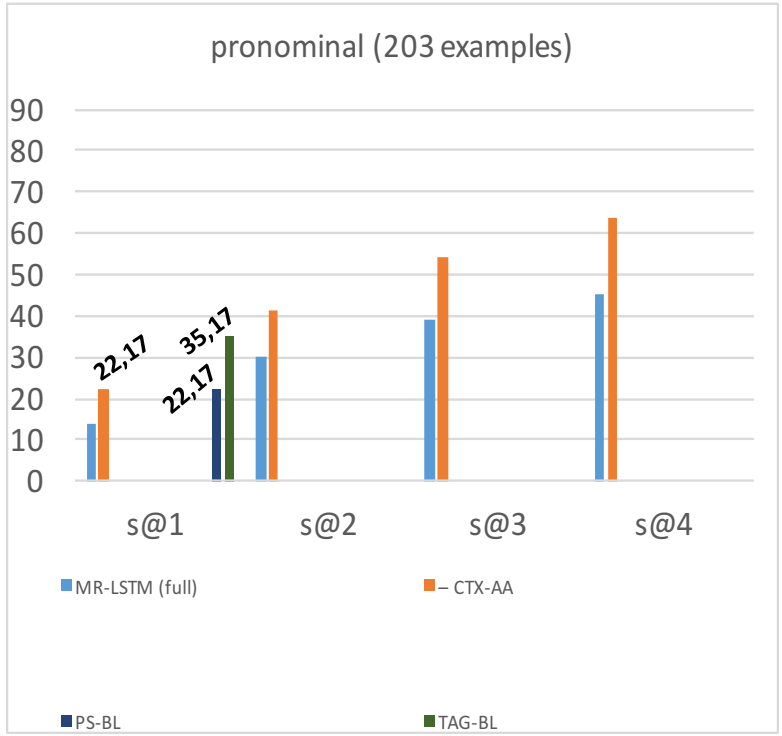


- the full MR-LSTM model is not better than the baselines

Experiment 2: Unrestricted AAR – all vs. nominal. vs. pronominal – full MR-LSTM model without the context of the anaphor

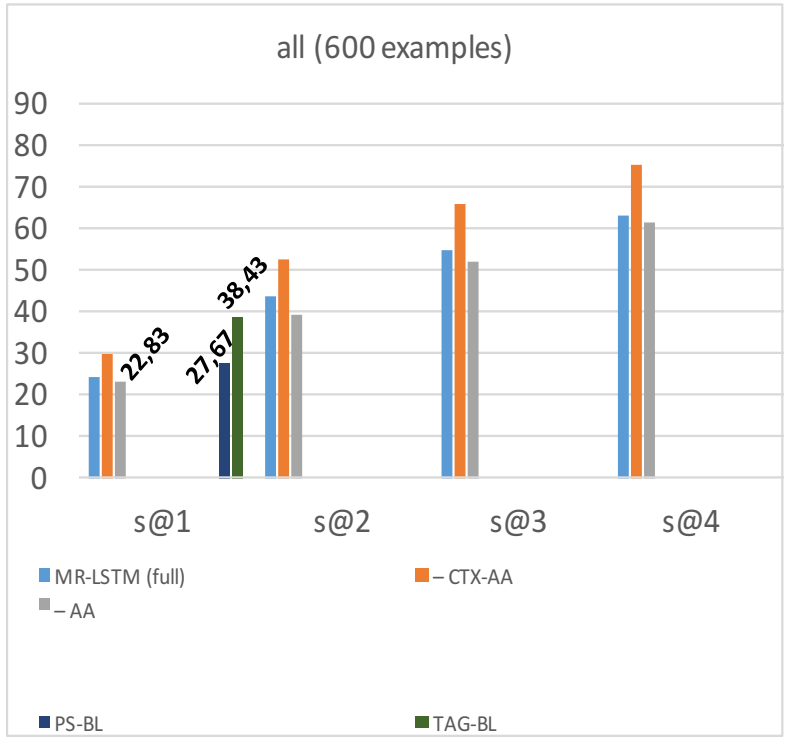


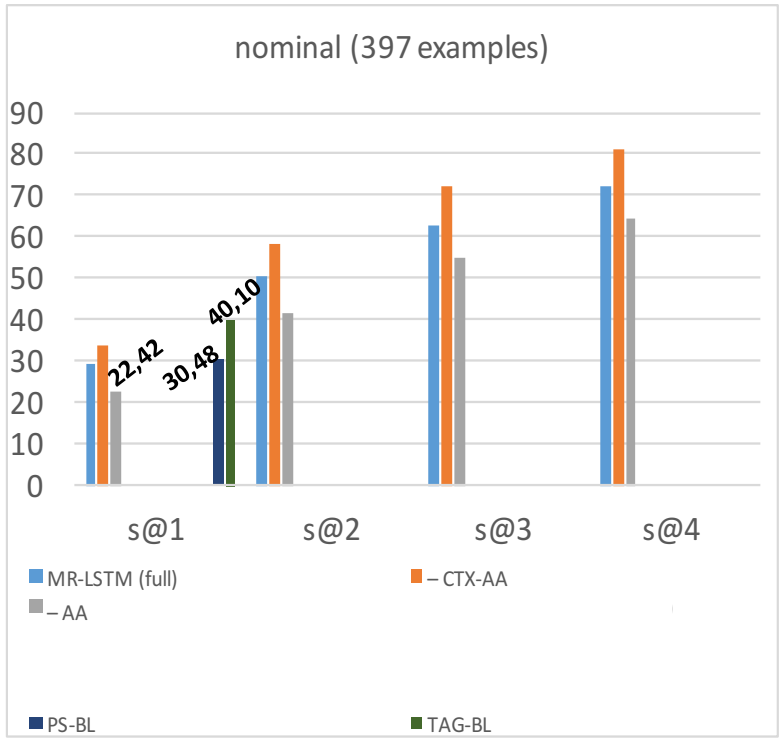


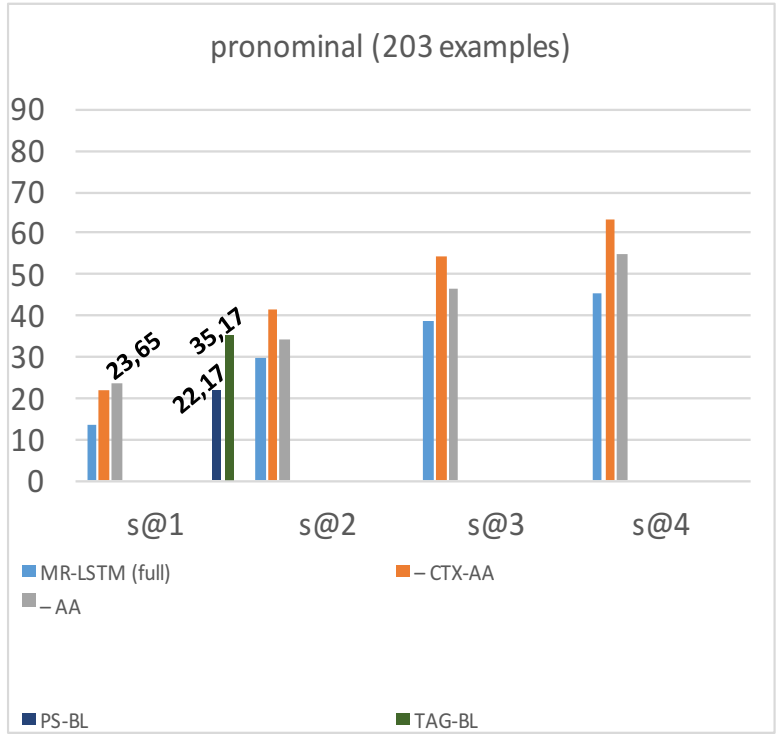


- omitting the context of the anaphor embedding $\Longrightarrow$ MR-LSTM better than TAG-BL

Experiment 2: Unrestricted AAR – all vs. nominal. vs. pronominal
– full MR-LSTM model without the head of the anaphor

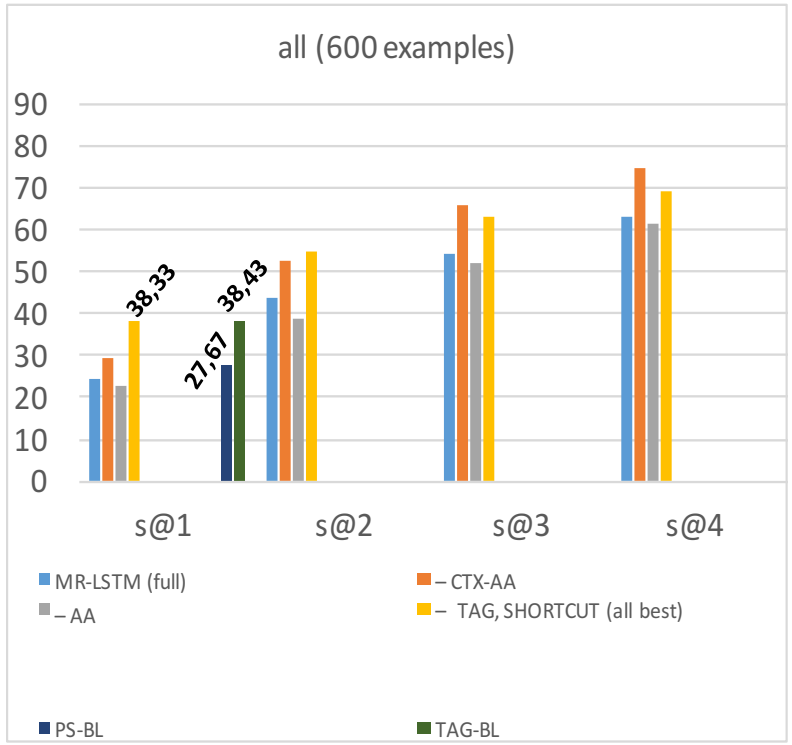


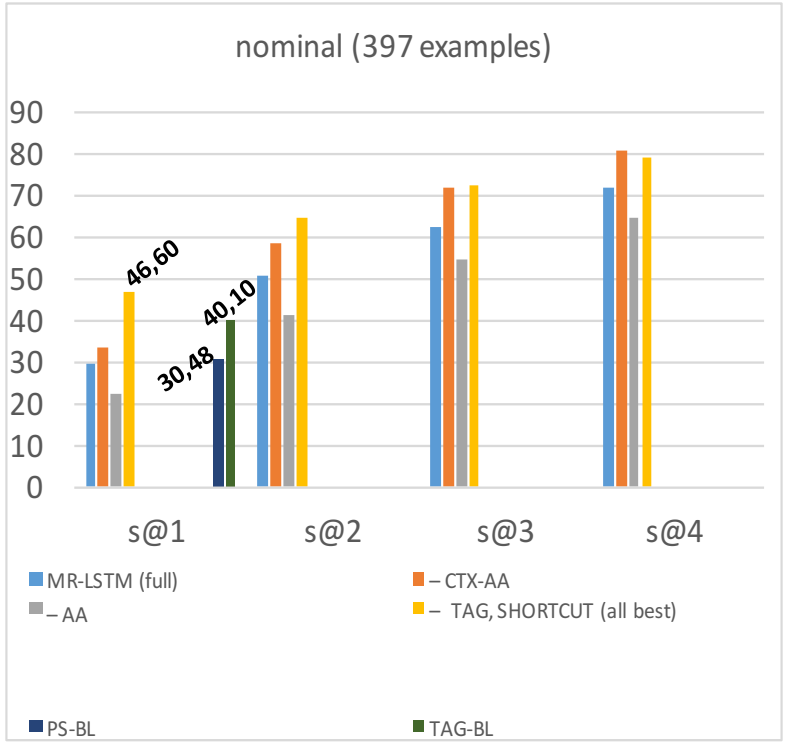


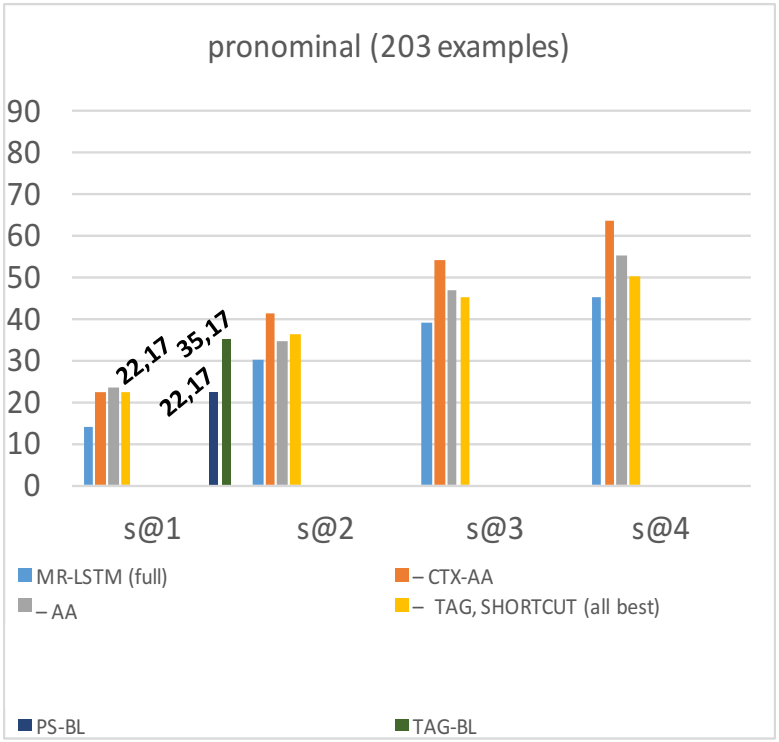


- the head of the anaphor is useful for nominal but not for pronominals

Experiment 2: Unrestricted AAR – all vs. nominal. vs. pronominal
– full MR-LSTM model without the syntactic information

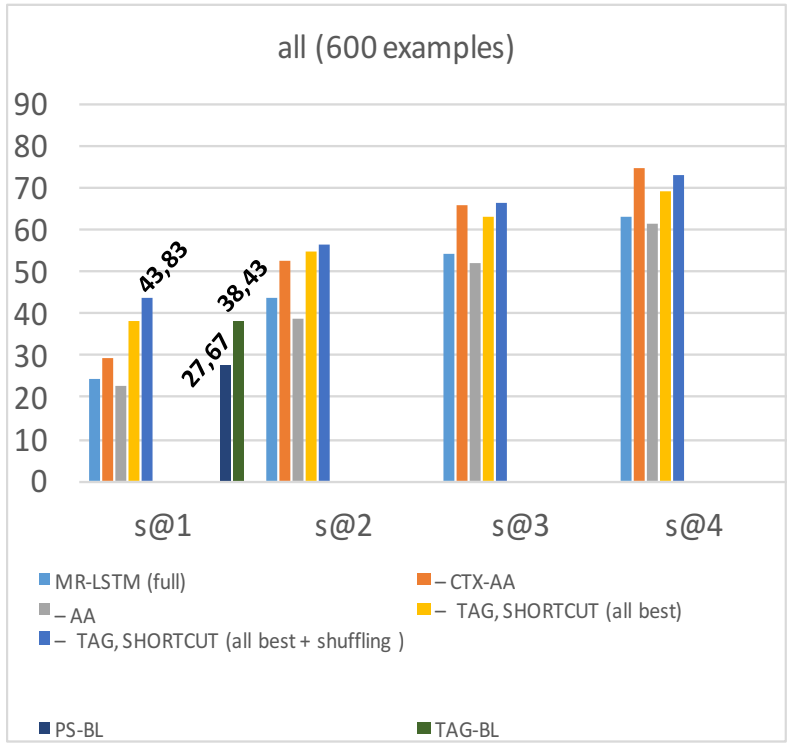


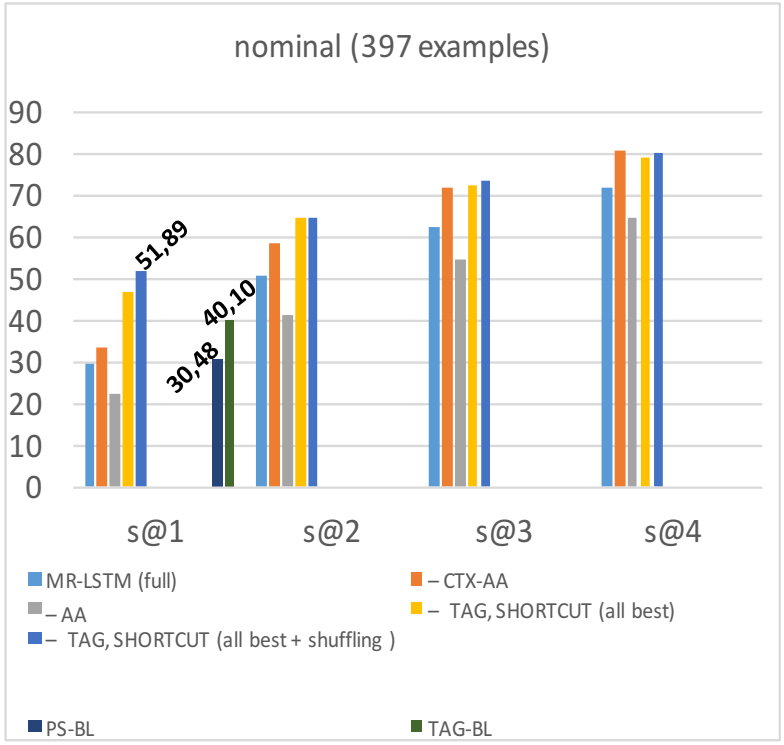


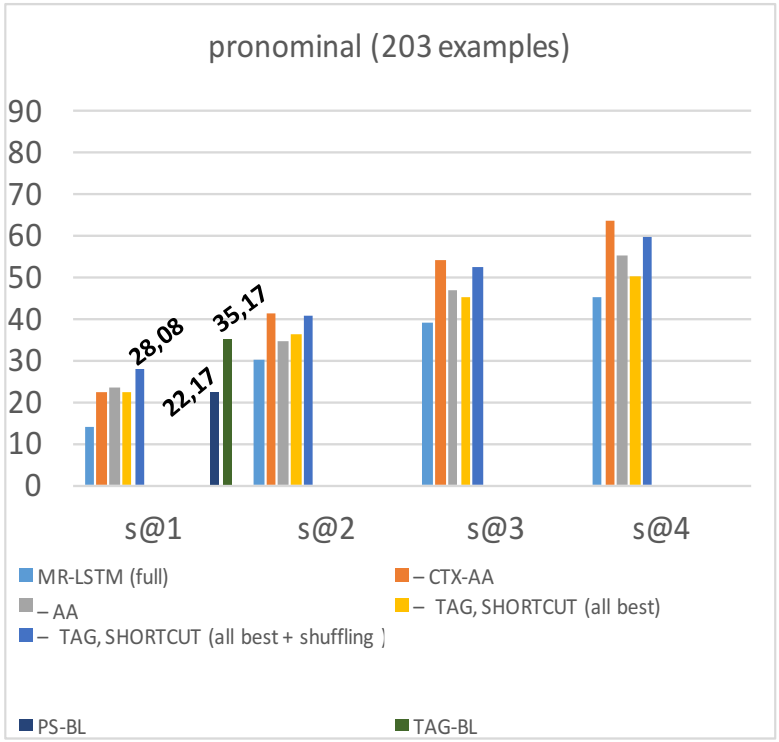


- omitting syntactic information ⟹ MR-LSTM is better than the preceding sentence BL (nominals, all) and better than the TAG-BL for nominals

Experiment 2: Unrestricted AAR – all vs. nominal. vs. pronominal
– full MR-LSTM model without the syntactic information + shuffling of train data

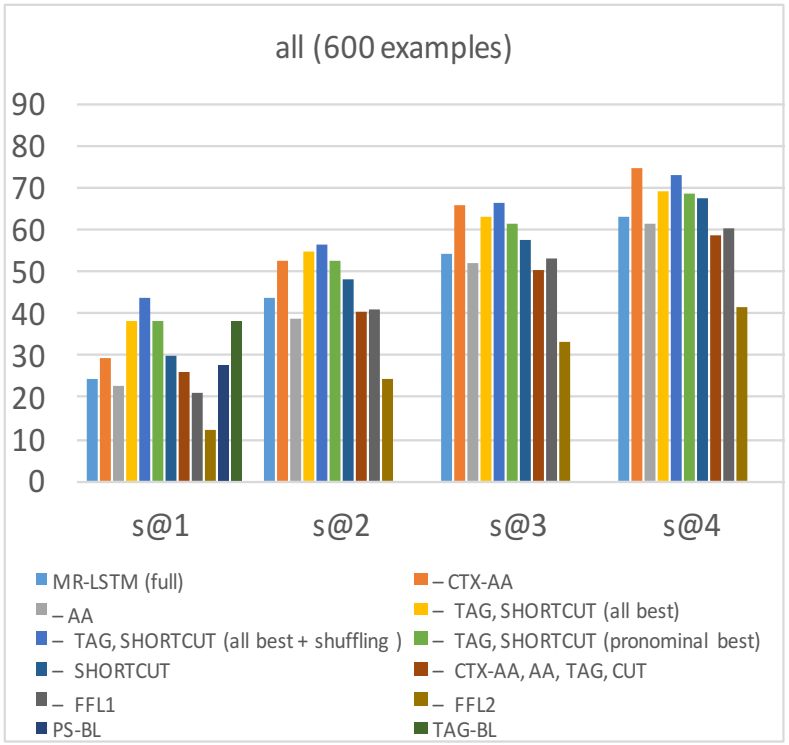


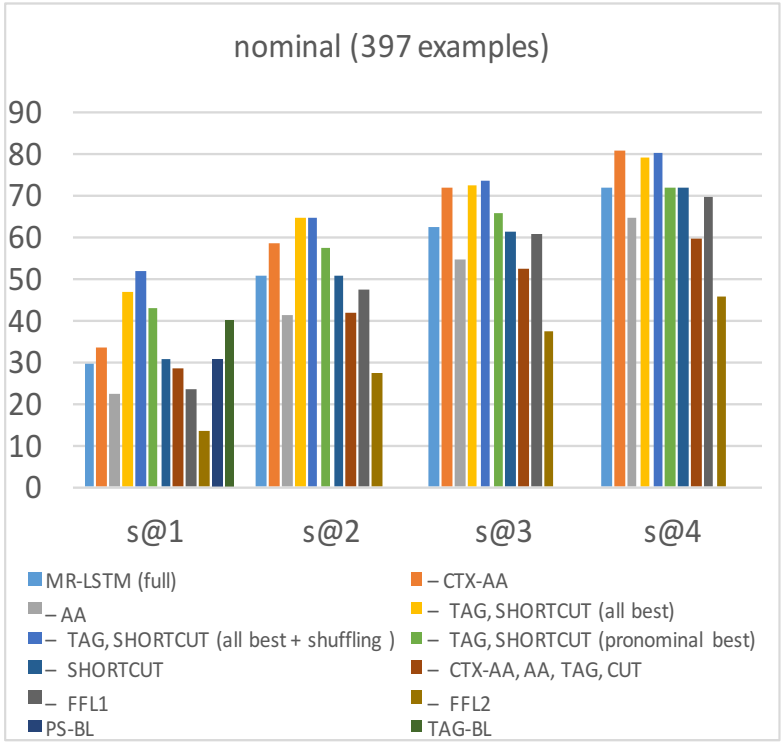


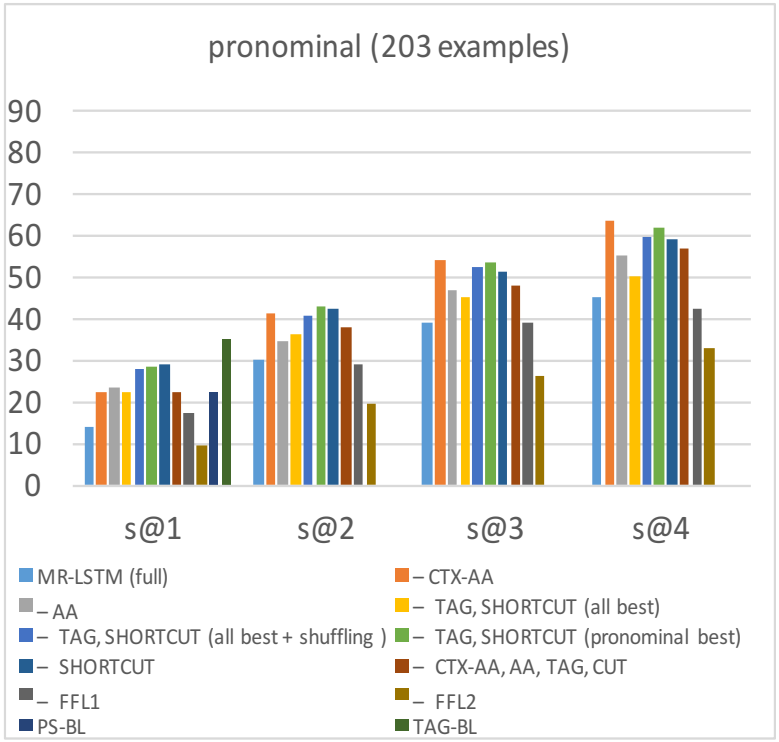


- training data is properly shuffled ⇒ MR-LSTM beats both BLs for nominals and overall, and than the preceding sentence BL for pronominals

Experiment 2: Unrestricted AAR – all vs. nominal. vs. pronominal – full MR-LSTM model and all ablations







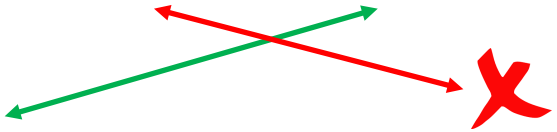
What's wrong with ***until***?

original sentence	anaphoric sentence	antecedent
They just keep [digging me in deeper [**until** [I reach the point where I give up and go away]$_S$]$_{SBAR}$]$_{VP}$.	They just keep digging me in deeper **until that** .	I reach the point where I give up and go away.

S_{t_0} until $S'_{t':fut(t_0)}$:

I am staying at home until he arrives.

stay-at-home(t_0) > he-arrives(t') > not-stay-at-home($t'' > t'$)



$S'_{t':fut(t_0)}$; S_{t_0}

He arrives. I'm staying at home until ***that*** *(happens).*

he-arrives(t_{-1}) > stay-at-home(t_0) > he-arrives($t_?$) > not-stay-at-home($t'' > t'$))

What about ***since*** **(tmp)?**

original sentence	anaphoric sentence	antecedent
Mr. Tonkin , who has been [feuding with the big three [**since** [he took office earlier this year$_S$]$_{SBAR}$]$_{VP}$, said that with half of the nation's dealers losing money or breaking even, it was time for emergency action.	Mr. Tonkin, who has been feuding with the big three **since then**, said that with half of the nation's dealers losing money or breaking even, it was time for emergency action .	He took office earlier this year.

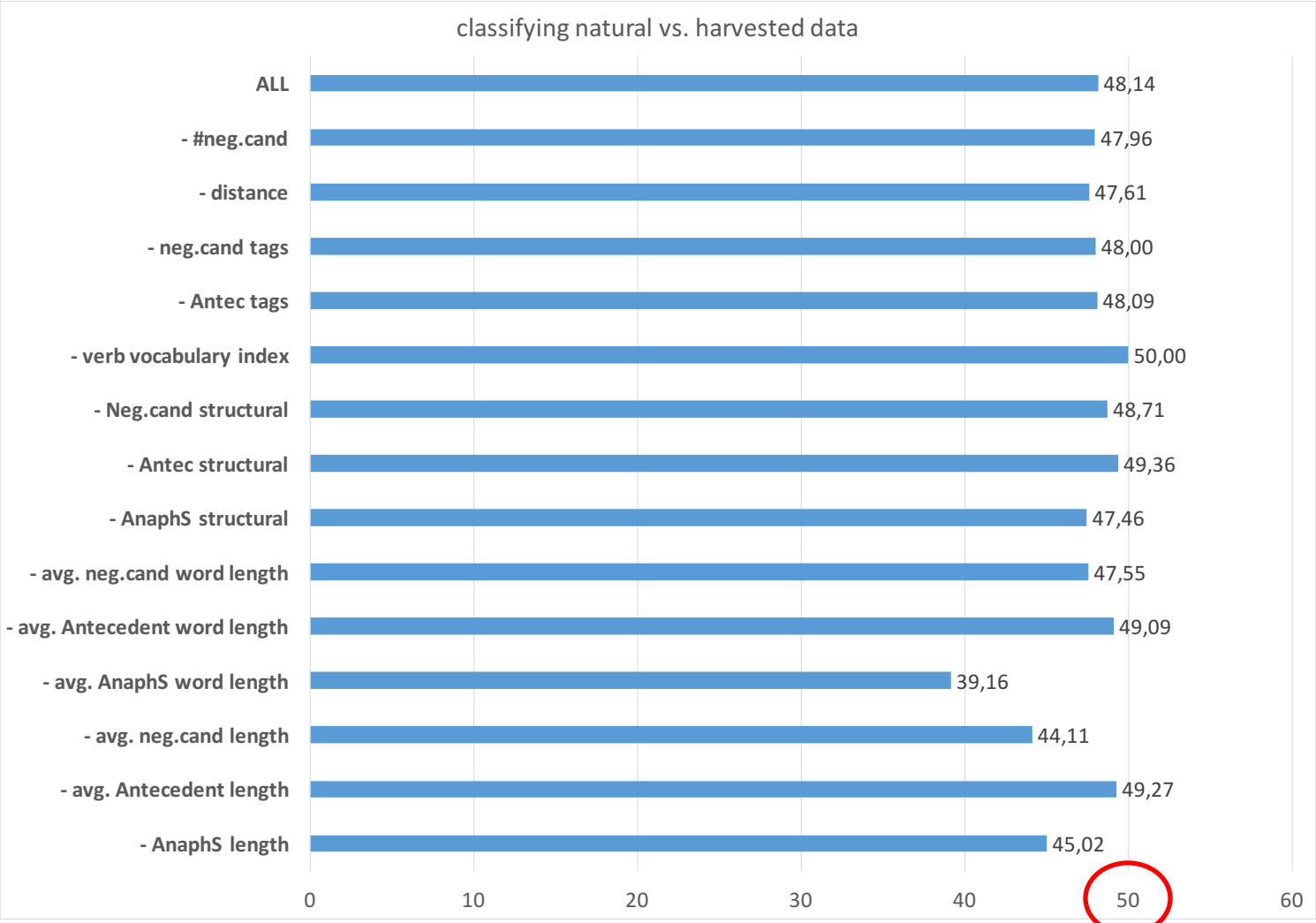
$S_{t':fut(t_{-1})}$ since $S'_{t_{-1}}$: *I played tennis since I entered school.*

enter-school (t_{-1}) > play-tennis (t'), $t': fut(t_{-1})$

$S'_{t_{-1}}$; $S_{t':fut(t_{-1})}$ *I entered school. I played tennis* ***since then / that time.***

enter-school(t_{-1}) > play-tennis (t'), $t': fut(t_{-1})$

Making Harvested Data More Similar to Natural Data
– SVM that classifies natural vs. harvested data



- **Linear SVM classifier** – compare against random BL (50.00 acc)

- **Train data for SVM**: train parts of ASN (KZH 13) and of CoNLL12-Ev and equal number of harvested data
- **Test data for SVM**: dev parts of ASN (KZH13) and CoNLL12-Ev and equal number of harvest data

- Avg. accuracy: 47.03

Experiment 3: Training MR-LSTM with Data Mixtures –Preliminary Results

Mixed silver & gold

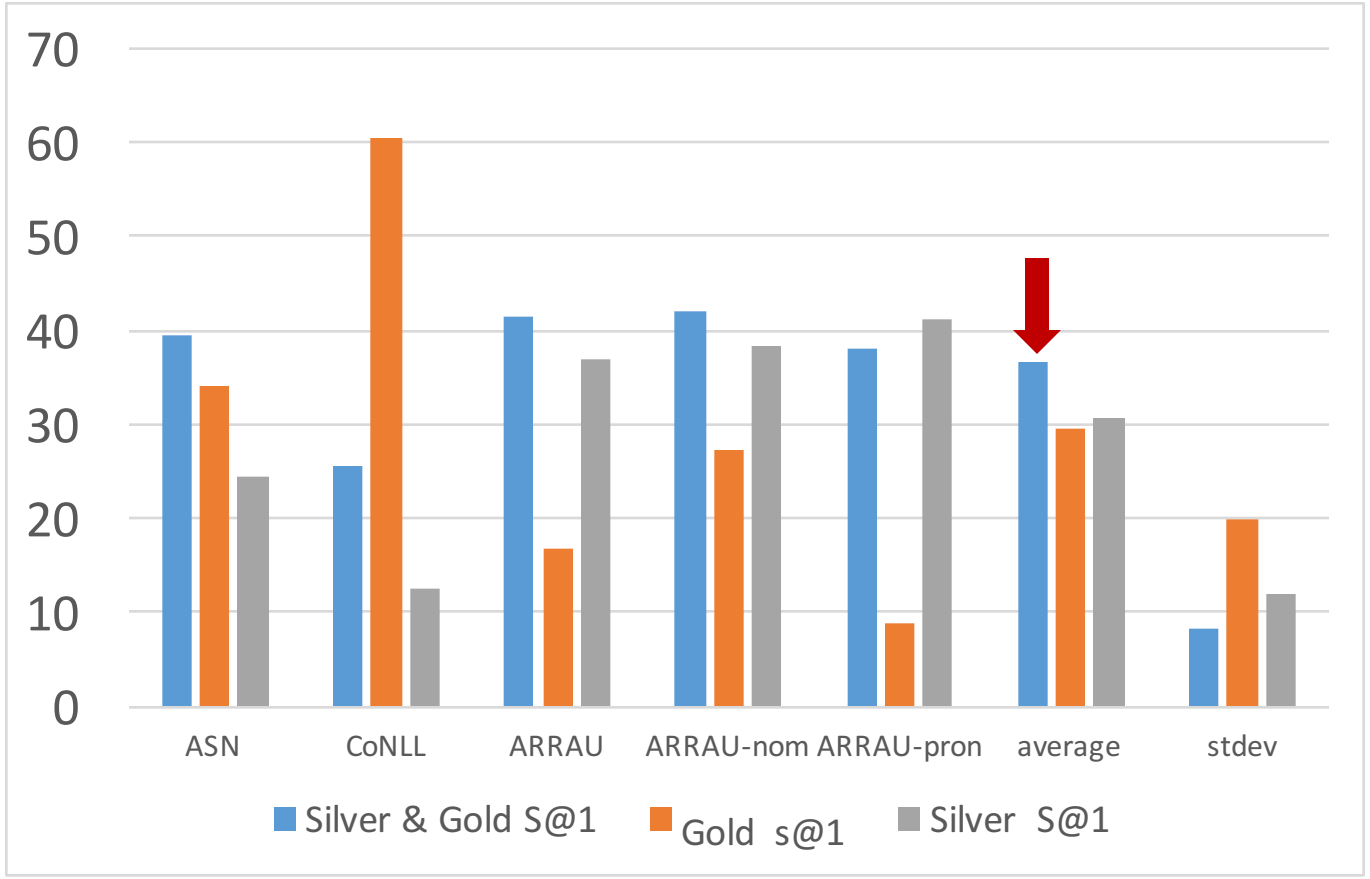
- **best overall**
- **best for** ASN, ARRAU-all & ARRAU-nom
- not perfect for ARRAU-pron (ASN!)
- not suited for CoNLL (neither is silver)

Silver-only

- best for ARRAU-pron (no ASN!)
- 2nd best for ARRAU-all and -nom

Gold-only

- best for CoNLL (silver data does not work)
- worst for ARRAU-pron (= nominal ASN data)



Experiment 3: Training MR-LSTM with Data Mixtures –Preliminary Results

Mixed silver & gold

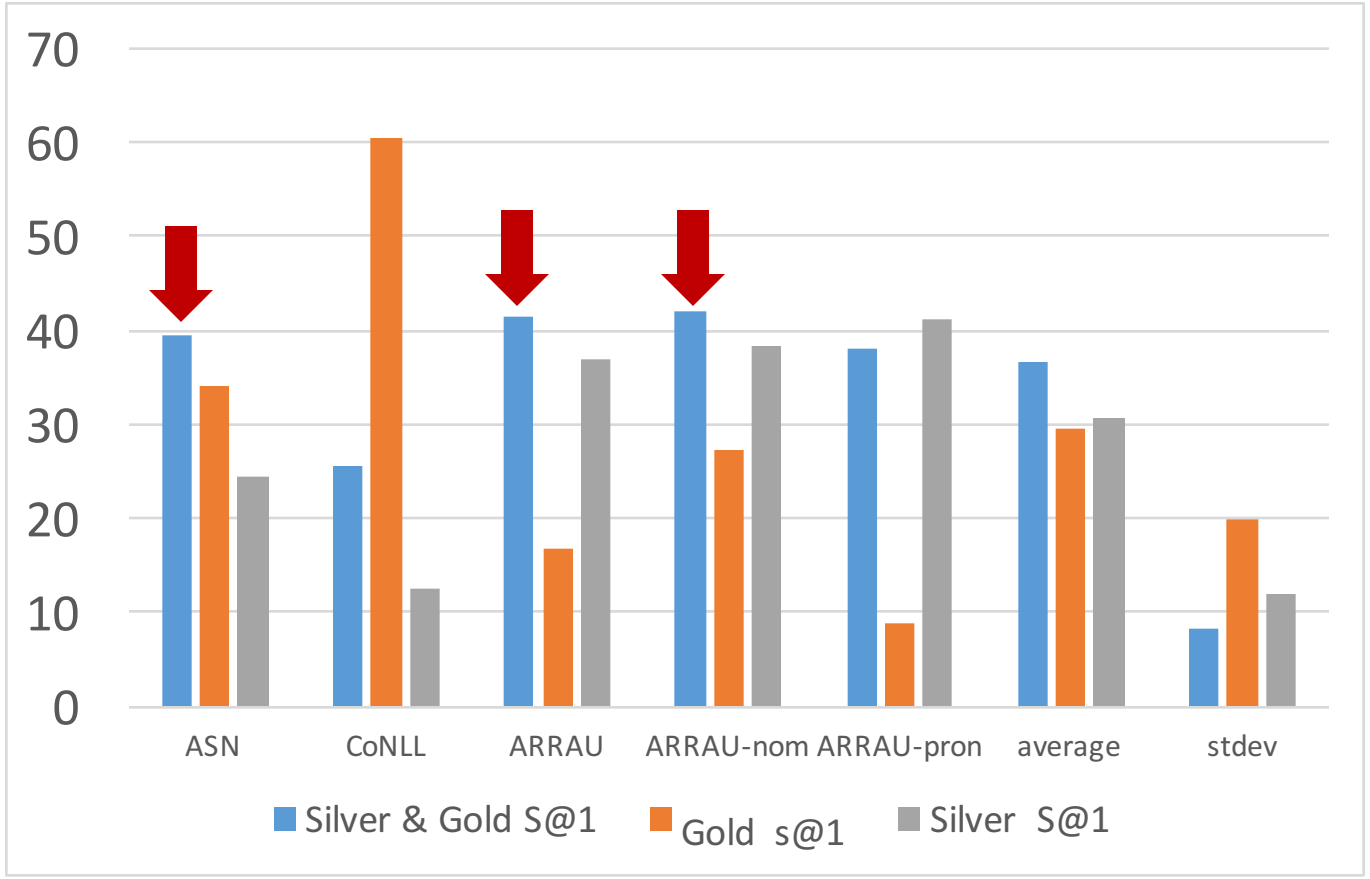
- **best overall**
- **best for** ASN, ARRAU-all & ARRAU-nom
- not perfect for ARRAU-pron (ASN!)
- not suited for CoNLL (neither is silver)

Silver-only

- best for ARRAU-pron (no ASN!)
- 2nd best for ARRAU-all and -nom

Gold-only

- best for CoNLL (silver data does not work)
- worst for ARRAU-pron (= nominal ASN data)



Experiment 3: Training MR-LSTM with Data Mixtures –Preliminary Results

Mixed silver & gold

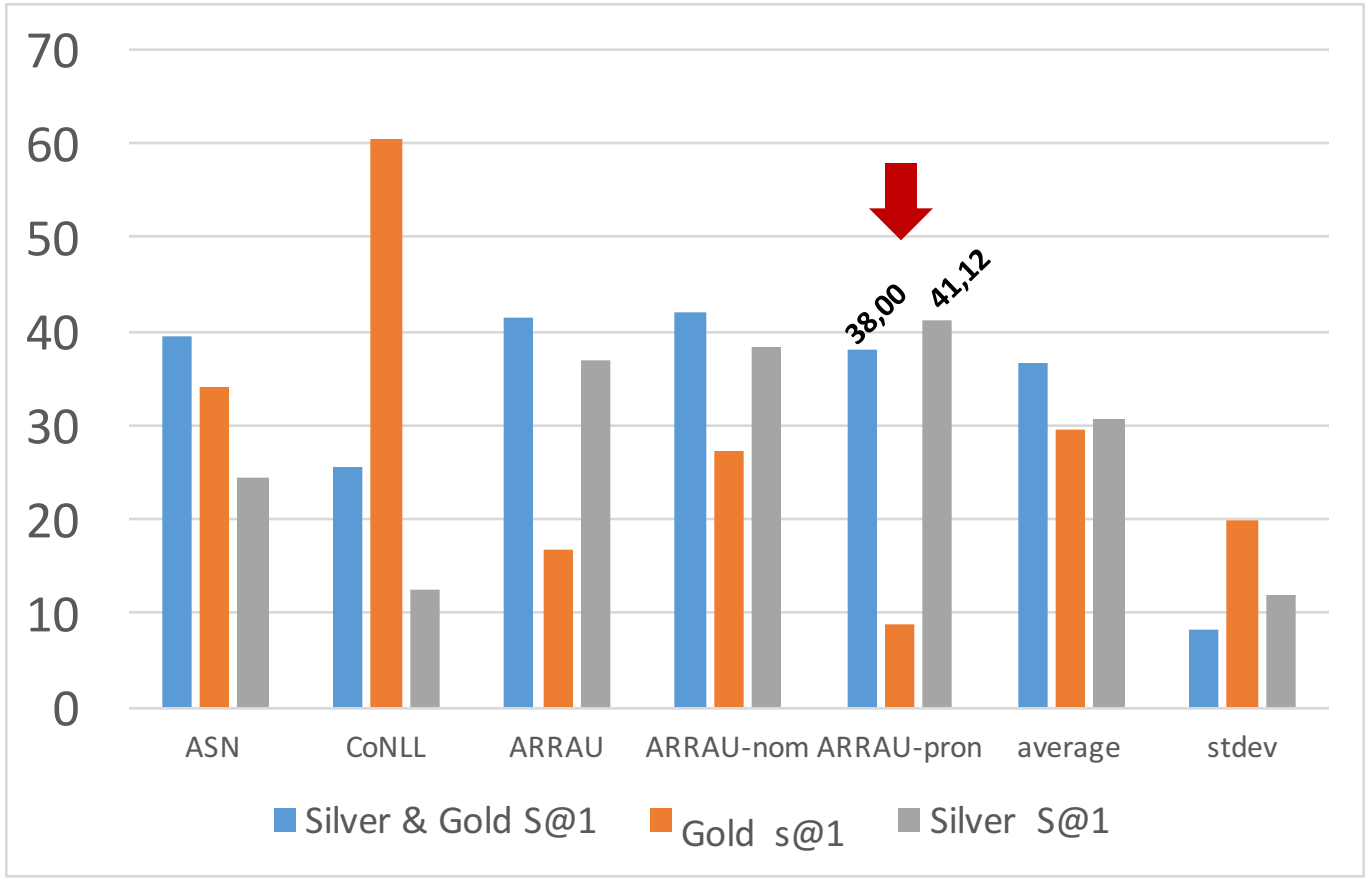
- **best overall**
- **best for** ASN, ARRAU-all & ARRAU-nom
- not perfect for ARRAU-pron (ASN!)
- not suited for CoNLL (neither is silver)

Silver-only

- best for ARRAU-pron (no ASN!)
- 2nd best for ARRAU-all and -nom

Gold-only

- best for CoNLL (silver data does not work)
- worst for ARRAU-pron (= nominal ASN data)



ADV-MR after training for a long time:
55.73 s@1 for CoNLL-Ev test (now: 25.25)

Experiment 3: Training MR-LSTM with Data Mixtures –Preliminary Results

Mixed silver & gold

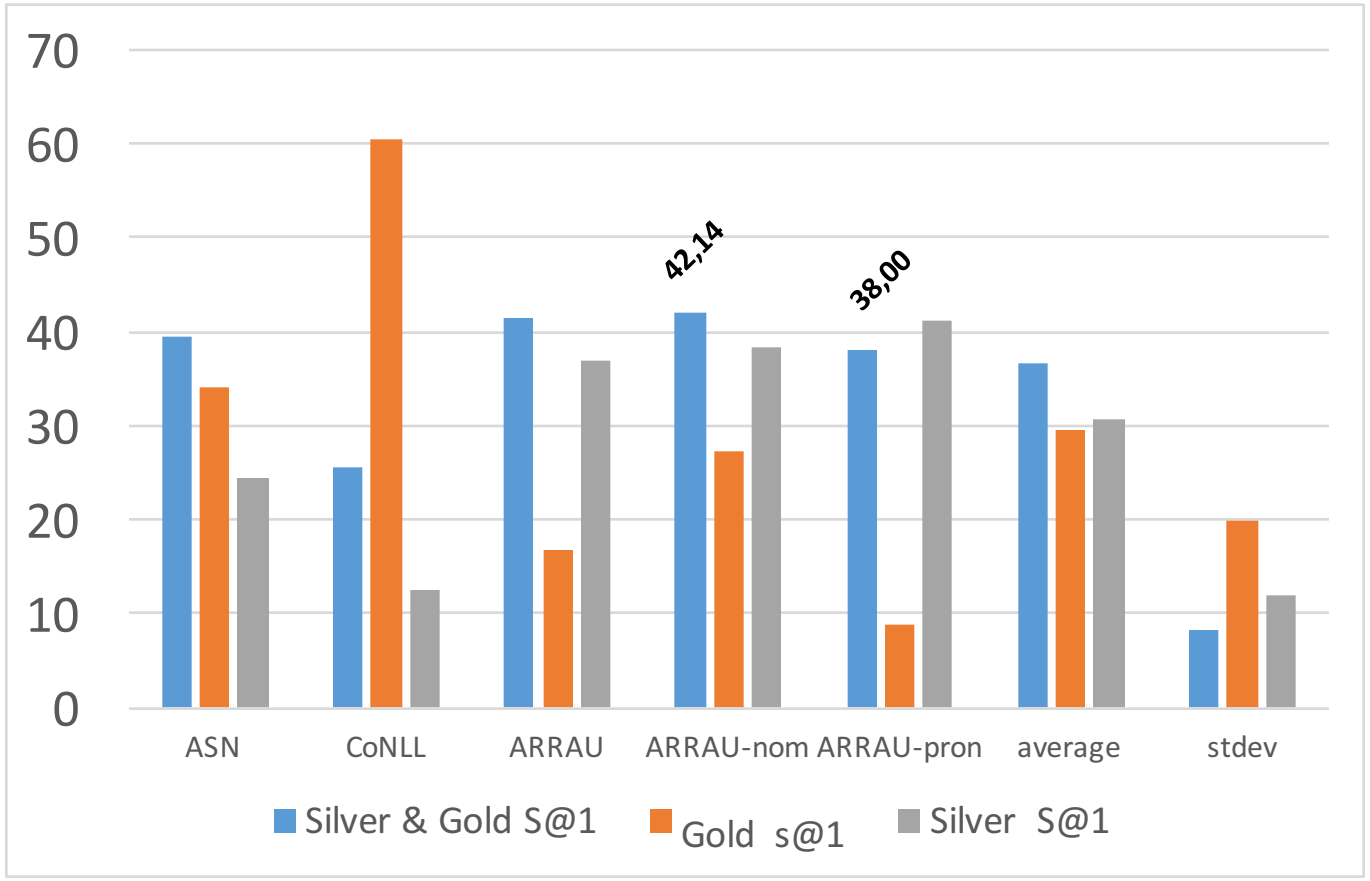
- **best overall**
- **best for** ASN, ARRAU-all & ARRAU-nom
- not perfect for ARRAU-pron (ASN!)
- not suited for CoNLL (neither is silver)

Silver-only

- best for ARRAU-pron (no ASN!)
- 2nd best for ARRAU-all and -nom

Gold-only

- best for CoNLL (silver data does not work)
- worst for ARRAU-pron (= nominal ASN data)



Experiment 3: Training MR-LSTM with Data Mixtures –Preliminary Results

Mixed silver & gold

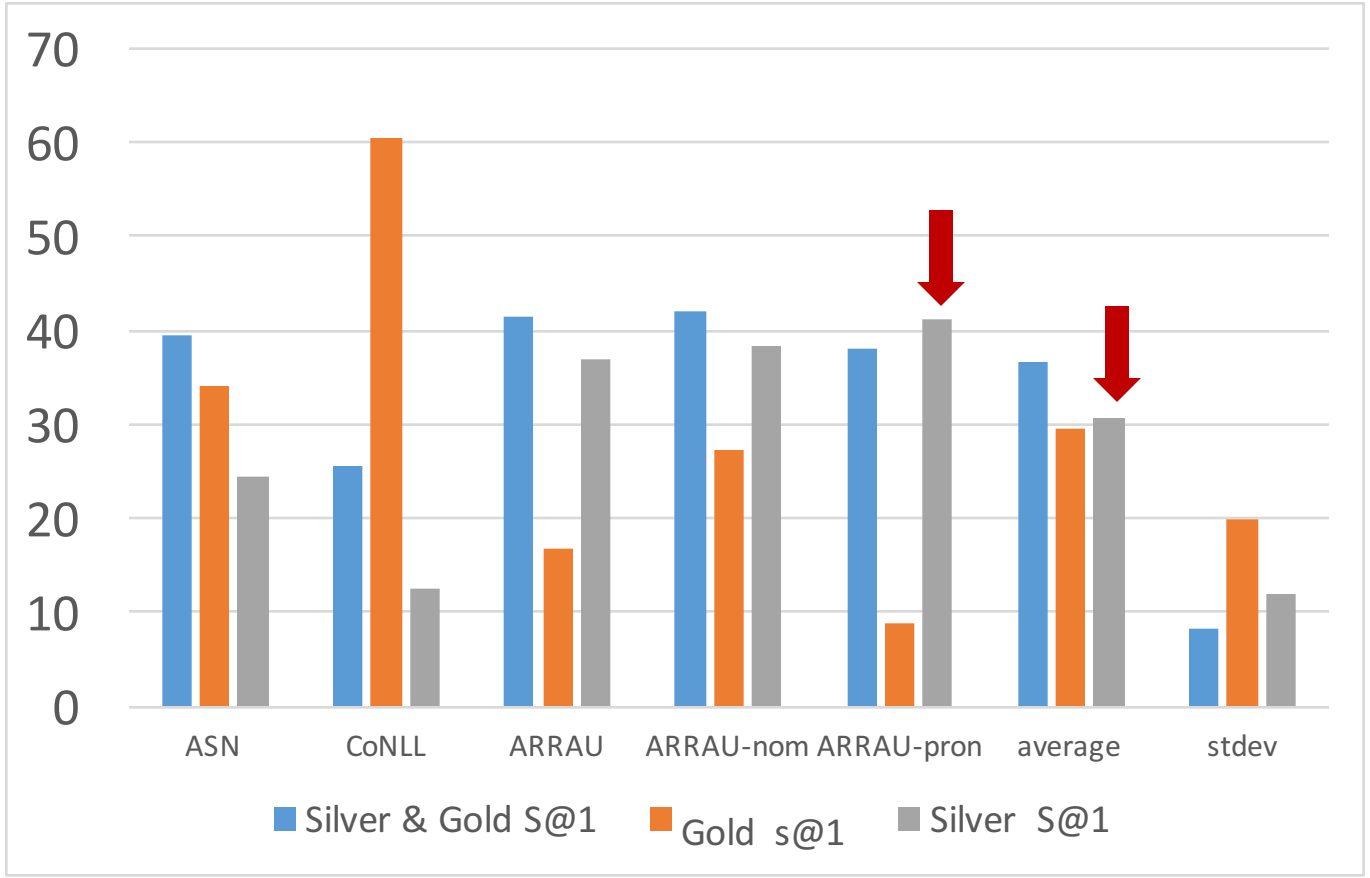
- **best overall**
- **best for** ASN, ARRAU-all & ARRAU-nom
- not perfect for ARRAU-pron (ASN!)
- not suited for CoNLL (neither is silver)

Silver-only

- best for ARRAU-pron (no ASN!)
- 2nd best for ARRAU-all and -nom

Gold-only

- best for CoNLL (silver data does not work)
- worst for ARRAU-pron (= nominal ASN data)



Experiment 3: Training MR-LSTM with Data Mixtures –Preliminary Results

Mixed silver & gold

- **best overall**
- **best for** ASN, ARRAU-all & ARRAU-nom
- not perfect for ARRAU-pron (ASN!)
- not suited for CoNLL (neither is silver)

Silver-only

- best for ARRAU-pron (no ASN!)
- 2nd best for ARRAU-all and -nom

Gold-only

- best for CoNLL (silver data does not work)
- worst for ARRAU-pron (= nominal ASN data)