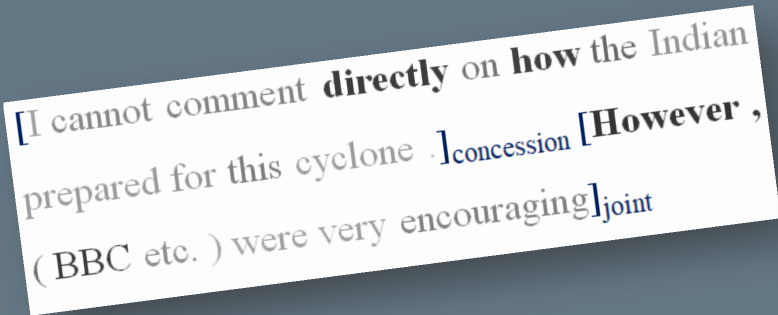


Feature Rich Models for Discourse Signaling

Amir Zeldes

Georgetown University
amir.zeldes@georgetown.edu

CLSP Seminar, JHU, 2018-02-23

Plan

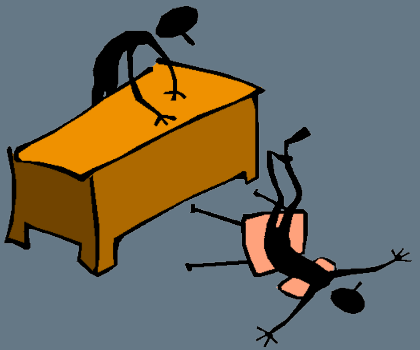
I. Discourse Relations
   - What are discourse relations?
   - Rhetorical Structure Theory (RST) in a nutshell
   - Datasets – RST-DT and GUM

II. Relation Signaling in Textual Data
   - Explicit and implicit relations
   - The RST Signalling Corpus
   - Finding signals in a text based model

III. Rich features
   - Should we add annotations to embeddings?
   - Ablation studies with feature rich models

Discourse relations

- What relations exist between utterances as a text unfolds?



1. a. [[*John pushed Mary.*]$_{\text{cause}}$ *She fell.*]
   b. [*Mary fell.* [*John pushed her.*]$_{\text{cause}}$]

(see Webber 1988, Asher & Lascarides 2003)

2. [[[*They left lights on*]$_{\text{cause}}$ *so Ellie got mad.*] [*That's totally unreasonable*]$_{\text{evaluation}}$]

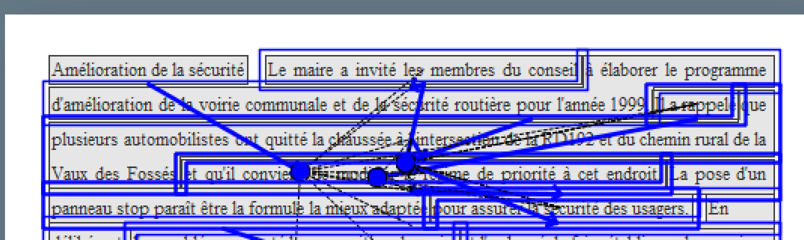
Discourse relations

- Some questions:
  - What relations exist? (Knott 1996, Knott & Sanders 1998)
    - Cross-linguistically? (van der Vliet & Radeker 2014)
    - In genres? (Taboada & Lavid 2003)
  - How are relations marked? (Taboada & Das 2013)
    - Explicit signals: *“on the other hand”* or *“although”*
    - Implicit signals: coreferent mentions, genre conventions, …

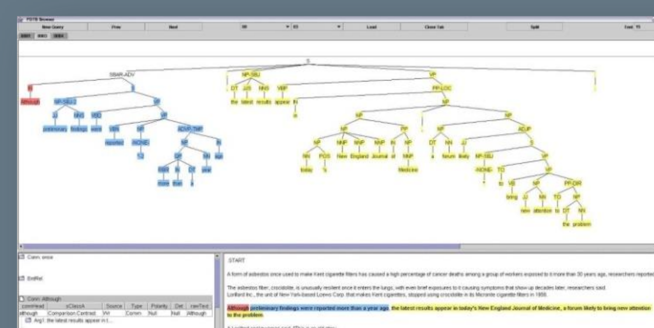
- To answer these questions we build **discourse annotated corpora**

Discourse annotation

- The task – given an arbitrary text:
  - Segment into ‘units’ (a.k.a. **E**lementary **D**iscourse **U**nits)
  - Establish the connections between these EDUs
  - Classify these connections
- Three main frameworks have implemented these tasks:
  - **Penn Discourse Treebank** (PDTB, Prasad et al. 2008) – partial parses
  - **Segmented Discourse Representation Theory** (SDRT, Asher & Lascarides 2003) – complete DAGs
  - **Rhetorical Structure Theory** (Mann & Thompson 1988) – complete trees



SDRT – Annodis corpus (Afantenos et al. 2010)



PDTB V2 (Prasad et al. 2008)

Rhetorical Structure Theory

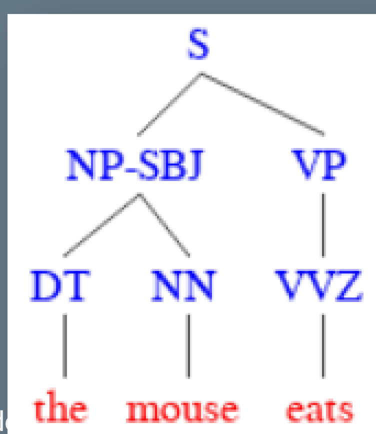
- In RST, a text is a tree of clauses

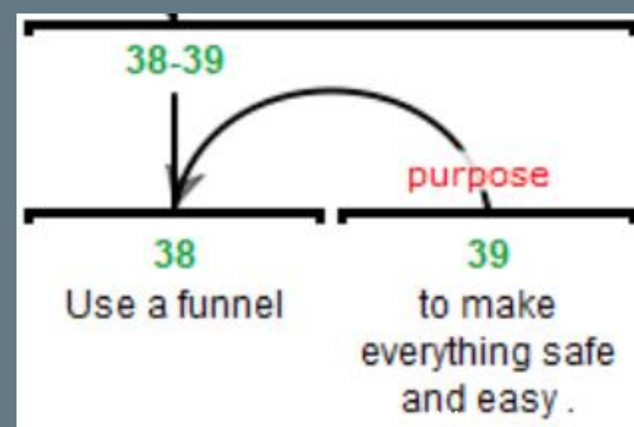
Syntax trees

- head > expansion
- Leaf = token
- Non-terminal = phrase
- Grammatical function

RST trees

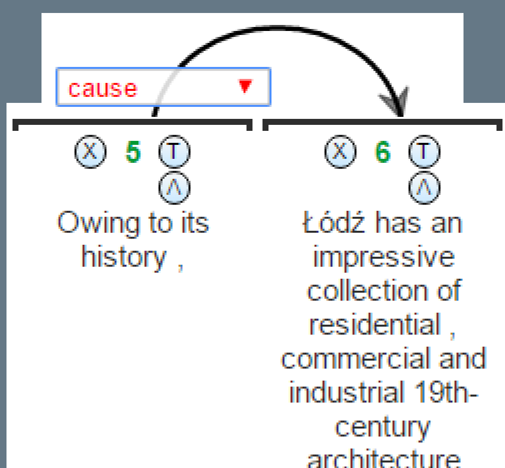
- nucleus > satellite
- Leaf = EDU
- Non-terminal = span
- Discourse function

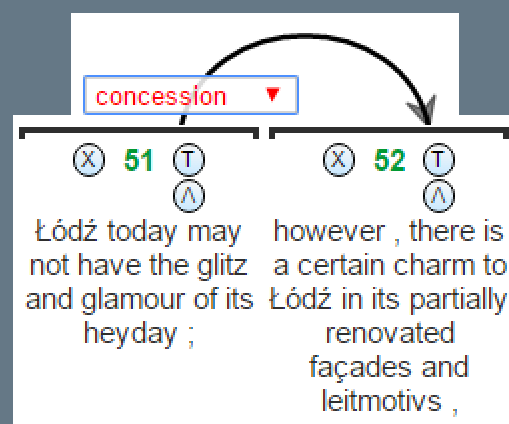


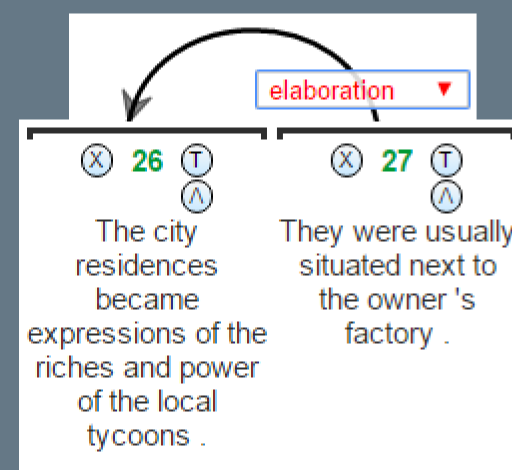


Trying it out - what are these?

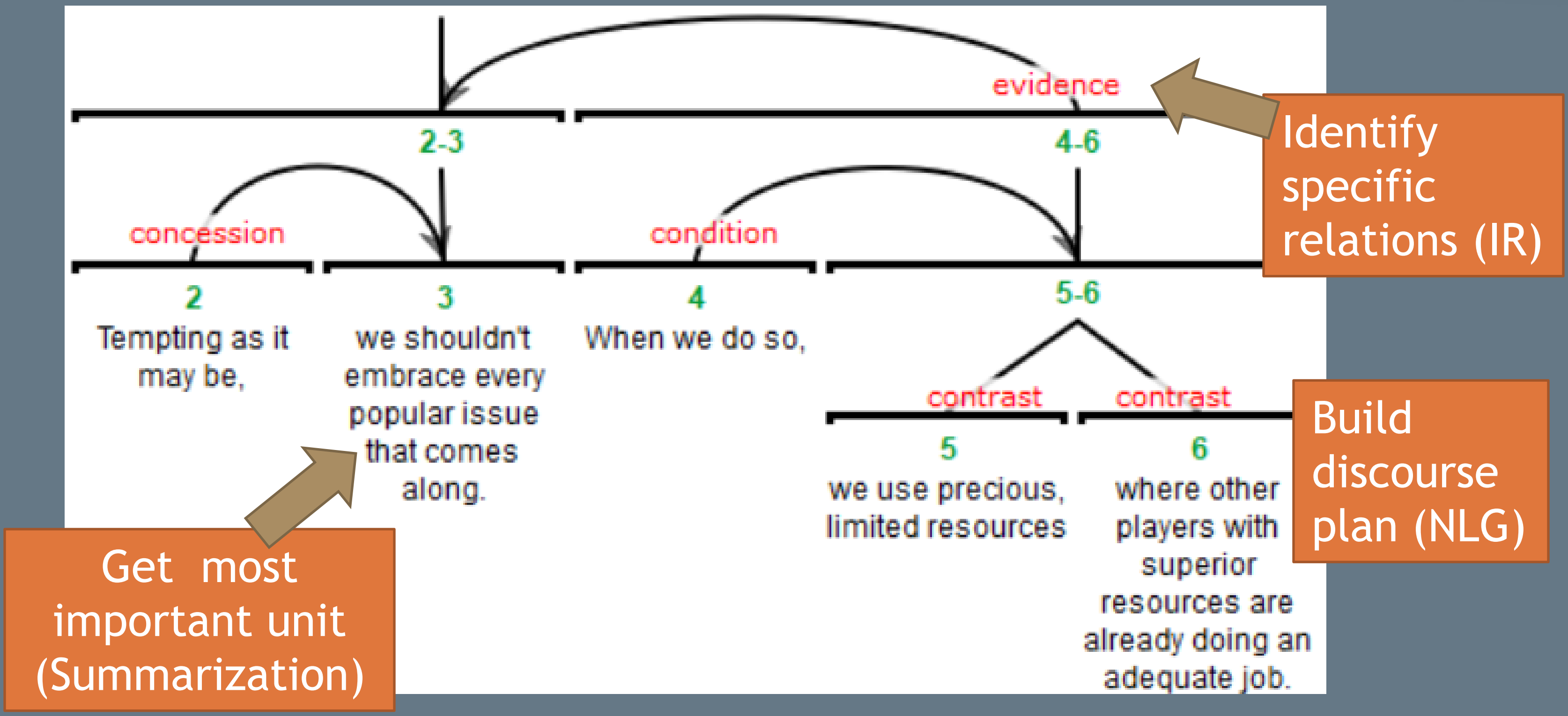
- Find the direction and label – choose from:
  - *cause*
  - *purpose*
  - *elaboration*
  - *concession*







Why is this important?



(example from RST Website: http://www.sfu.ca/rst/)

Topics for today

- What information identifies relations?
  - For humans
  - For NLP
  - How much of the information is ‘explicit’?
    (cf. Sporleder & Lascarides 2005, 2008, Taboada 2009)
- Can we identify relations directly from text?
  - Do machine learning algorithms and humans notice the same signals?
  - If/when not, why? What features do we miss?
  - Can we add them as new layers to our corpus data?
- What data can we use?

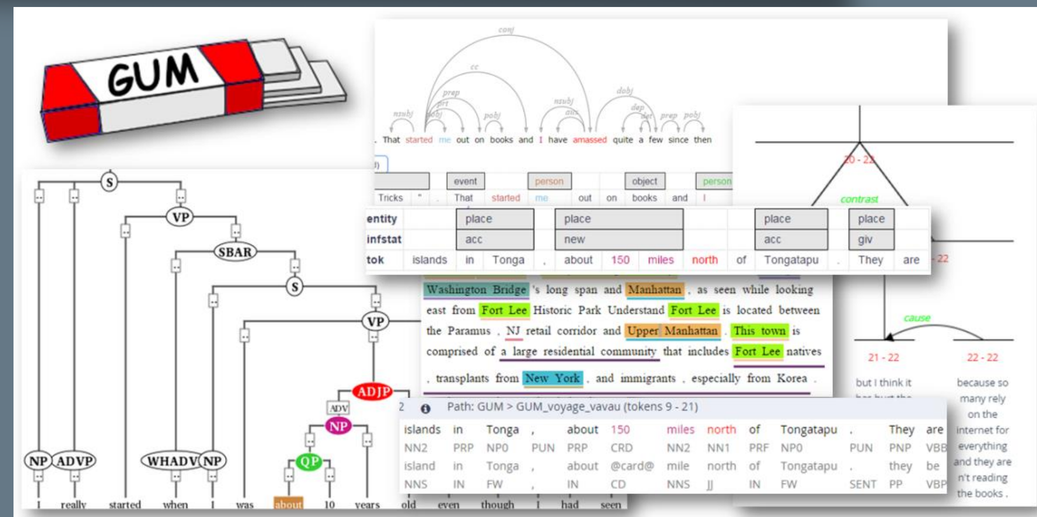
RST Discourse Treebank (Carlson et al. 2003)

- 180K tokens (WSJ)
- POS + syntax trees
- 60% overlap with OntoNotes: (Hovy et al. 2006)
  - NER (named entities only)
  - Partial coreference (no singletons, indefinites)
  - PropBank annotations

Georgetown University Multilayer corpus
(Zeldes 2017)

http://corpling.uis.georgetown.edu/gum/

- POS tagging (PTB, CLAWS, TT, UPOS)
- Sentence type (SPAAC++)
- Document structure (TEI)
- Syntax trees (PTB + Stanford + UD)
- Information status (SFB632)
- (Non-) named entity types
- Coreference + bridging
- **Rhetorical Structure Theory**
- Speaker information, ISO time…



text type	source	texts	tokens
Academic	Various	6	5,210
Biographies	Wikipedia	6	5,049
Fiction	Small Beer Press	7	5,912
Interviews	Wikinews	19	18,037
News	Wikinews	21	14,093
Travel guides	Wikivoyage	17	14,955
Forum discussions	reddit	6	5,174
How-to guides	wikiHow	19	16,920
Total		101	85,350

creative commons

II. Relation Signaling

Explicit signals

- Can we identify relations in data? (Sanders et al. 1992, Knott & Dale 1994, Taboada & Lavid 2003, Stede & Grishina 2016)
  - Discourse markers – ***however, but, if, and, as well as***
  - Adverbials – ***clearly, supposedly, in reality…***
  - Content words – ***good*** (signals evaluation?), ***last year*** (signals temporal sequence? Circumstance?)

- Annotators use cue words as diagnostics:
  - *“could I connect these with ‘because’?”*

Frequentist approaches

- Studies often cross-tabulate: ***words ~ relations***
- Problems:
  - Frequency thresholds
  - Ambiguity ("and" may not be associated with relations and appears with all relations – not a Discourse Marker?)
  - Context sensitivity – some words are cues in specific environments

Relation type	Freq	marker	translation
Elaboration	150	*kotoryj*	"which, that"
Joint	119	*i, takzhe*	and, as well
Attribution	118	*zajavil, soobschil*	report, announce etc.
Contrast	62	*Odnako, a, no*	However, but
Cause-Effect	47	*Poetomu, V+prichina*	so, accordingly, V+*cause*
Purpose	39	*Chtoby, dlya*	In order that, for
Interpretation-Evaluation	34	Nouns and verbs expressing opinion	
Background	31	No dominant marker	
Condition	27	*esli*	if

Table 1. Relations with their most frequent markers

Toldova et al. 2017

Text to labels approach

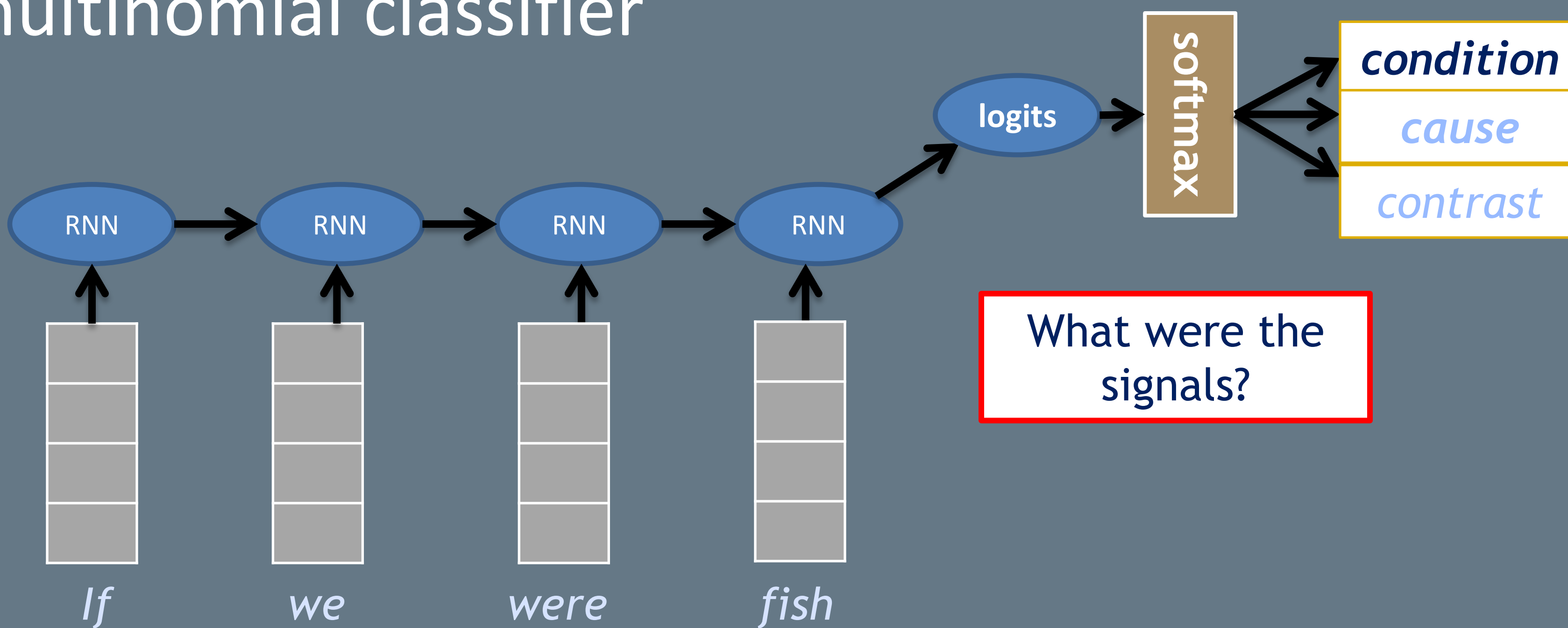
- *The core idea of our work is to learn a transformation from a bag-of-words surface representation into a latent space in which discourse relations are easily identifiable.*
  - Ji & Eisenstein (2014:13)

- Echoed in much NLP in recent years:
  - text -> labels
  - But really:
    text -> embeddings <--> labels
    (cf. Braud et al. 2016)

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

Relation classification with RNNs

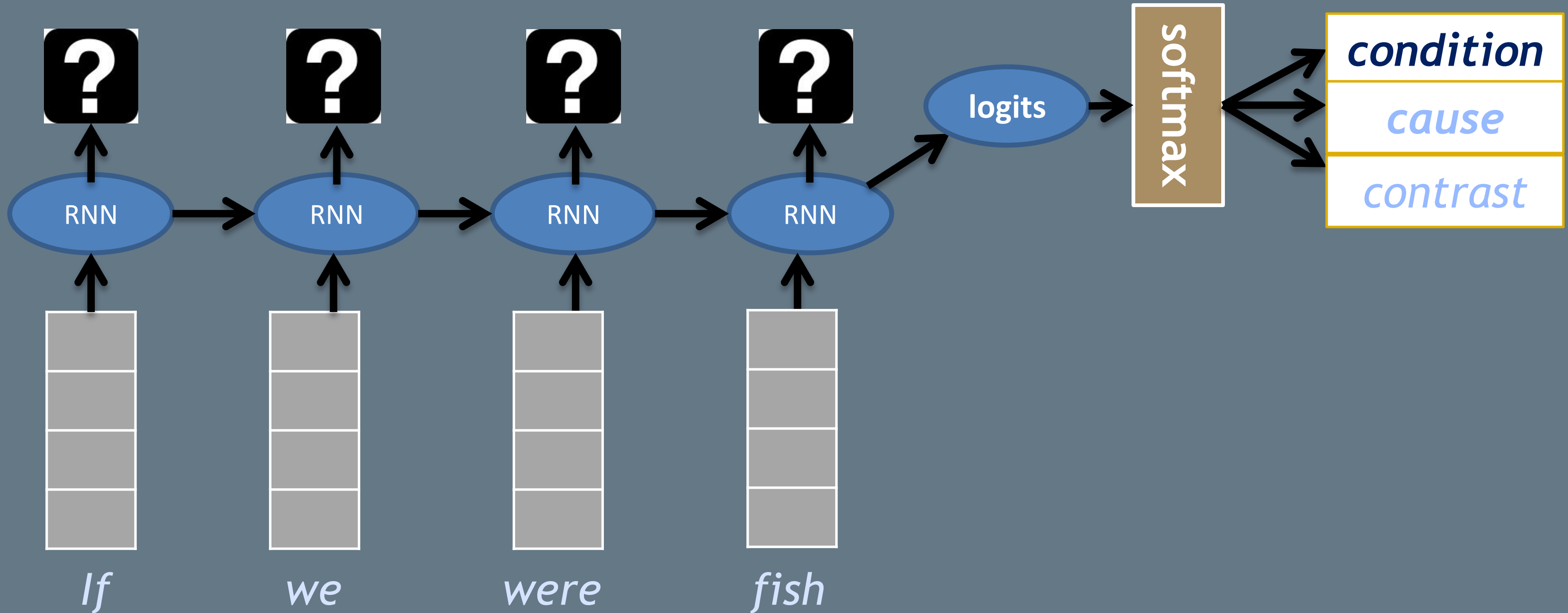
- RNNs can recognize relations from text (Braud et al. 2017; cf. entailment work, Rocktäschel et al. 2016)
- Can use encoder architecture, single output multinomial classifier



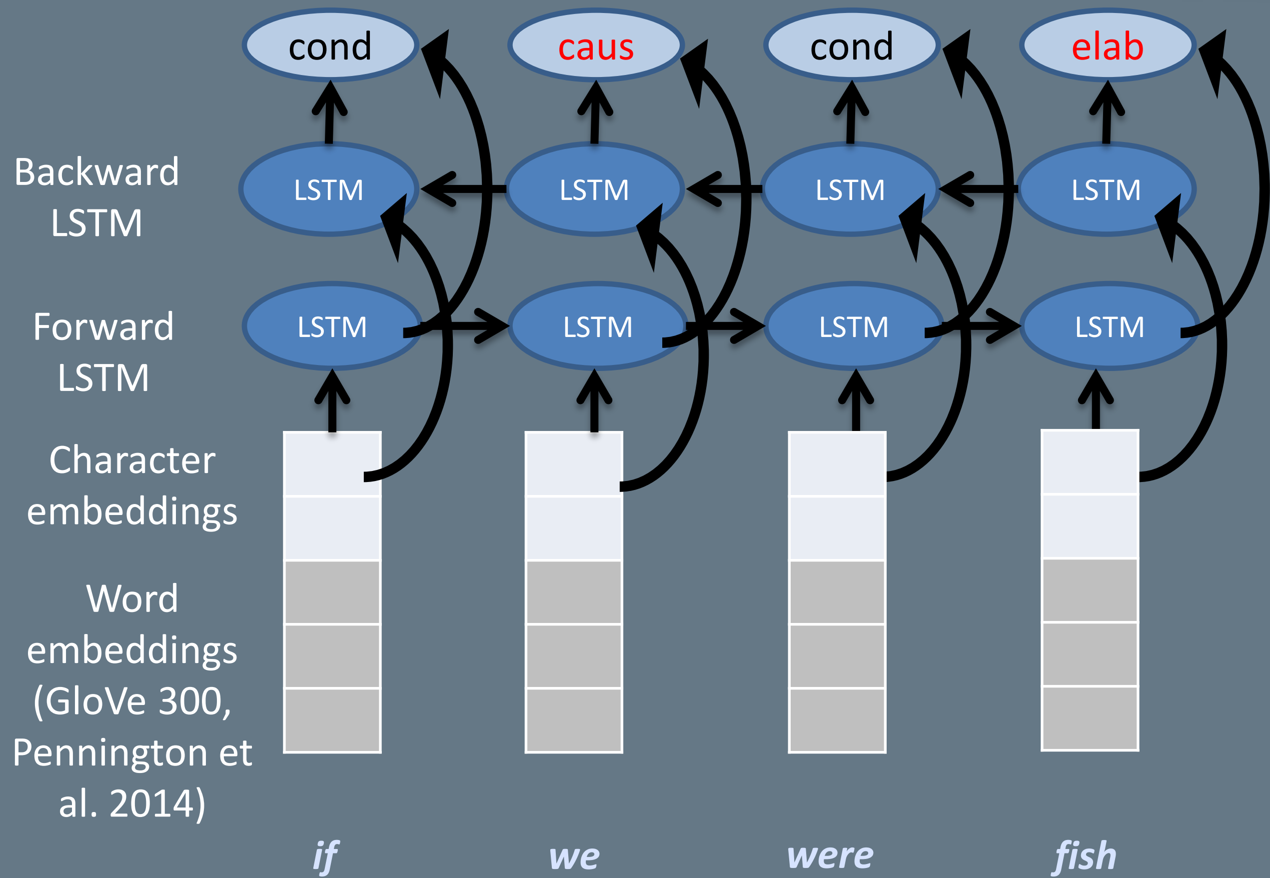
What were the signals?

A neural approach to signals with RNNs

- The RNN probably already had it at ***If***...
- To find signals, we can listen to output at every token (but loss still based on EDU relation)



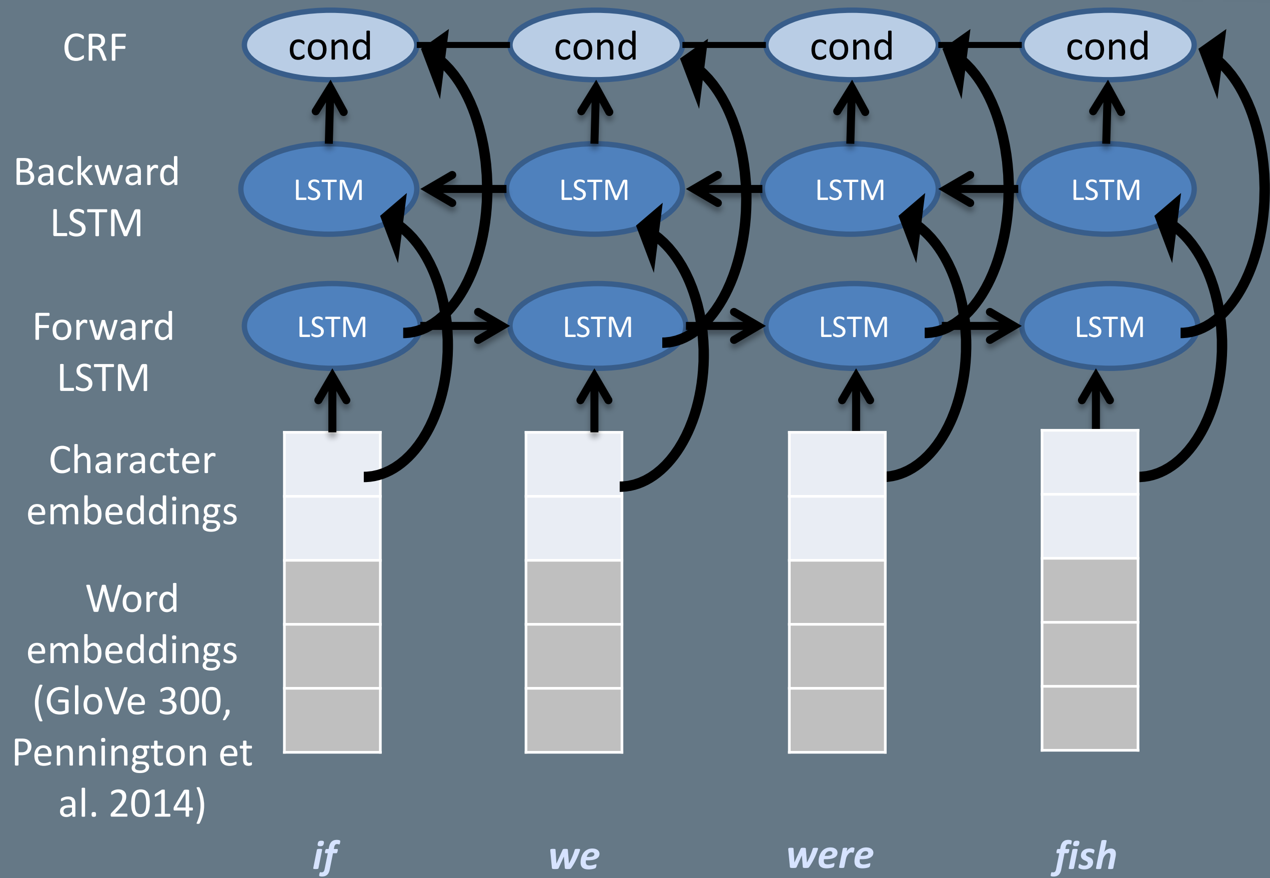
Implemented with Bi-LSTM (TensorFlow)



Hidden:	200
Optimizer:	Adam
(rec.) dropout:	0.5
Minibatch:	20
Activation:	tanh

Batch normalization
Trainable embeddings

Adding CRF (Huang et al. 2015, Ma & Hovy 2016)



Hidden:	200
Optimizer:	Adam
(rec.) dropout:	0.5
Minibatch:	20
Activation:	tanh

Batch normalization
Trainable embeddings

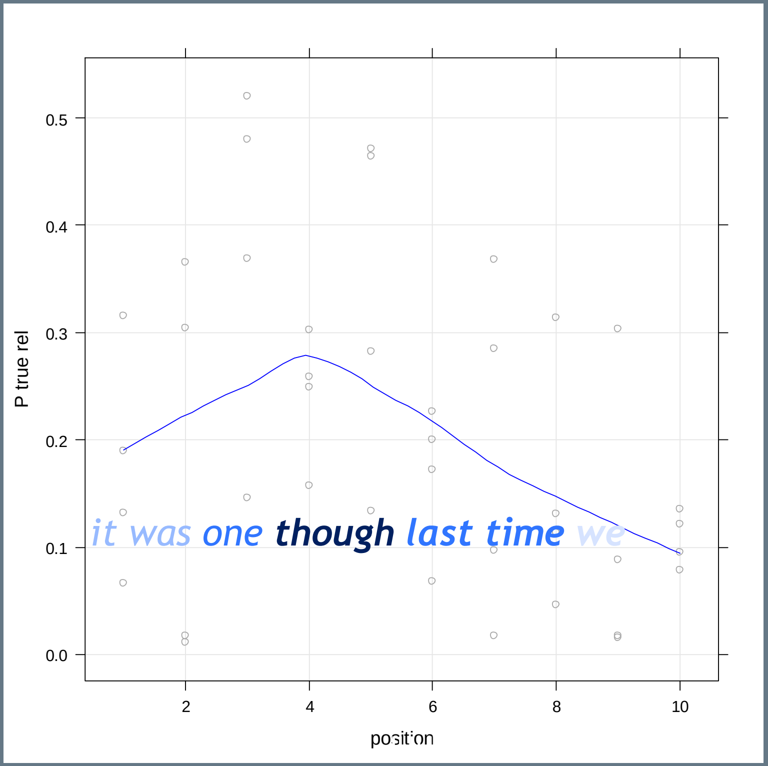
Single output performance

- Not so interesting, but:
  - RSTDT – relation accuracy by tokens:
    acc: 47.43% | f1: 41.44
    - Standard train/test split
    - 60 relations [some very rare] – note majority baseline is ~33%

- State of the art on RSTDT, hard to compare:
  - Ji & Eisenstein (2014), using engineered features, full parsing: 61.75% (by EDUs, 18 relations)
  - Braud et al. (2016), (2017) with RNNs, pretraining on PDTB, coref and more:
    60.01% (by EDUs, 18 relations)

Visualizing token-wise softmax

- Basic idea – find the most 'convincing' tokens:
  - Use tokens' softmax probability of correct relation
  - Shade by:
    - Proportion of maximum softmax probability in **sentence**
    - Proportion of maximum softmax probability in **document**



Visualizing token-wise softmax

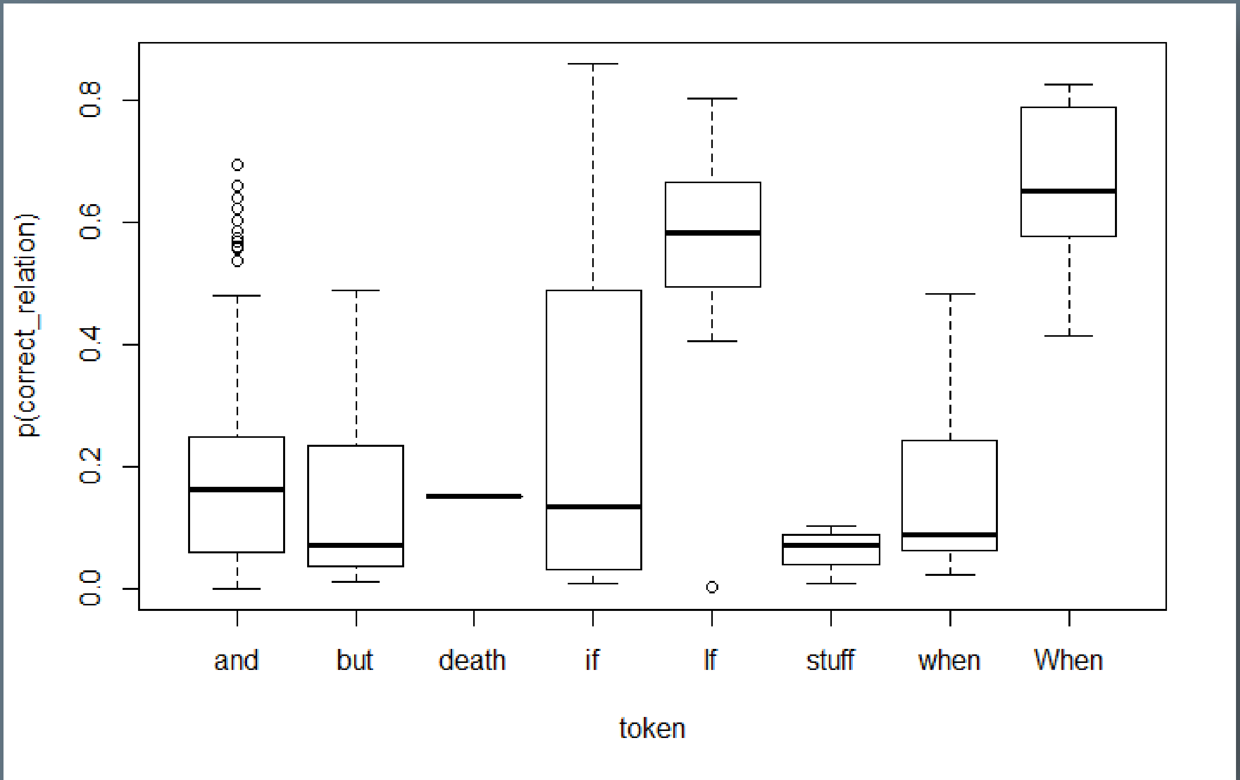
- [**This** occurs for two reasons **:**]$_{preparation}$ [**As** it moves over land **,**]$_{circumstance}$ [it is cut off from the source of energy driving the storm …]$_{cause}$
- [**Combine** 50 milliliters of hydrogen peroxide and a liter of distilled water in a **mixing** bowl .]$_{sequence}$ [A ceramic bowl **will** work **best** ,]$_{elaboration}$ [**but** plastic works **too** .]$_{concession}$

Ambiguous?

GUM data

Addressing ambiguity

- We can get ambiguity scores based on range of softmax probabilities (data: GUM)



Addressing ambiguity

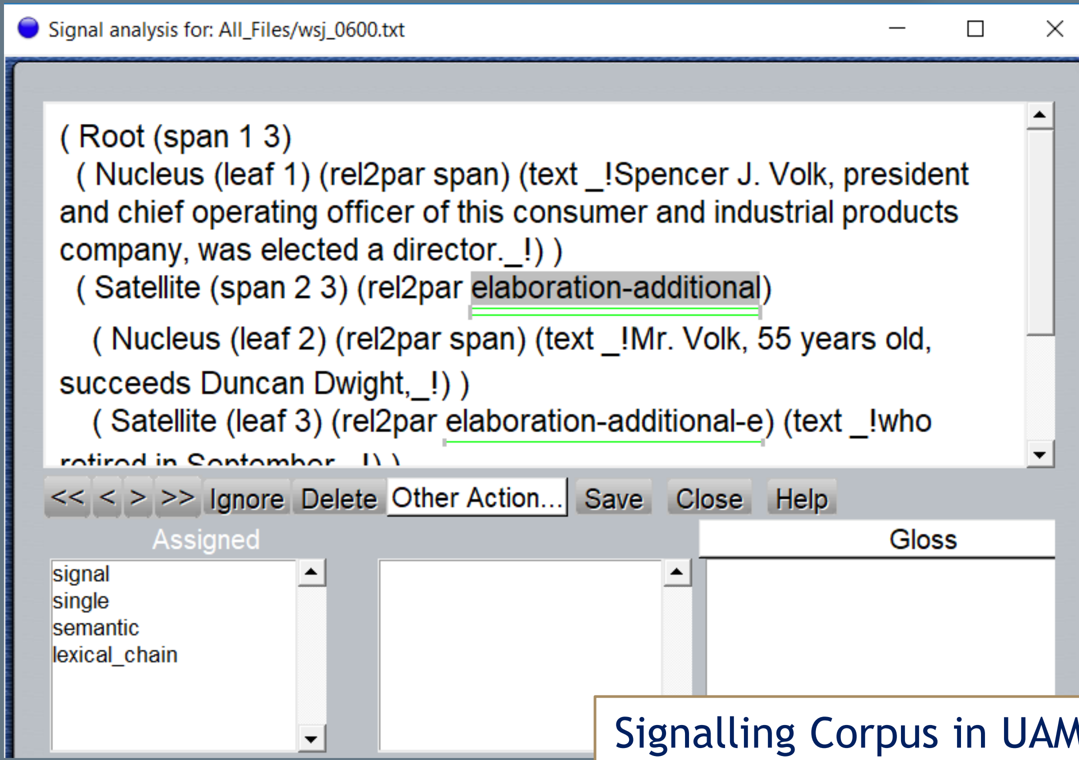
- **Irrelevant 'and's: (RST-DT)**
  - [**but** will continue as a director and chairman of the executive committee .]$_{\text{elaboration}}$
  - [and one began trading on the Nasdaq/National Market System last **week** .]$_{\text{inverted}}$
- **Important 'and's: (RST-DT)**
  - [**and** is involved in claims adjustments for insurance companies .]$_{\text{List}}$
  - [-- **and** from state and local taxes too , for in-state investors .]$_{\text{elaboration}}$

Evaluating plain text signals

- There results are qualitative, non-systematic
- Ideal scenario - compare to 'gold standard'
  - Use RST-DT Signalling Corpus (Taboada & Das 2013)
  - Open ended annotation of any kind of relation signal:
    - Discourse markers, other expressions
    - Syntactic devices, cohesion
    - Genre conventions…

Evaluating signals

- Problems:
  - Signals annotated at **node** level
  - Non trivial to associate with specific EDUs
  - Location of signal in words is not specified



Signalling Corpus in UAM (O'Donnell 2008)

Toy evaluation

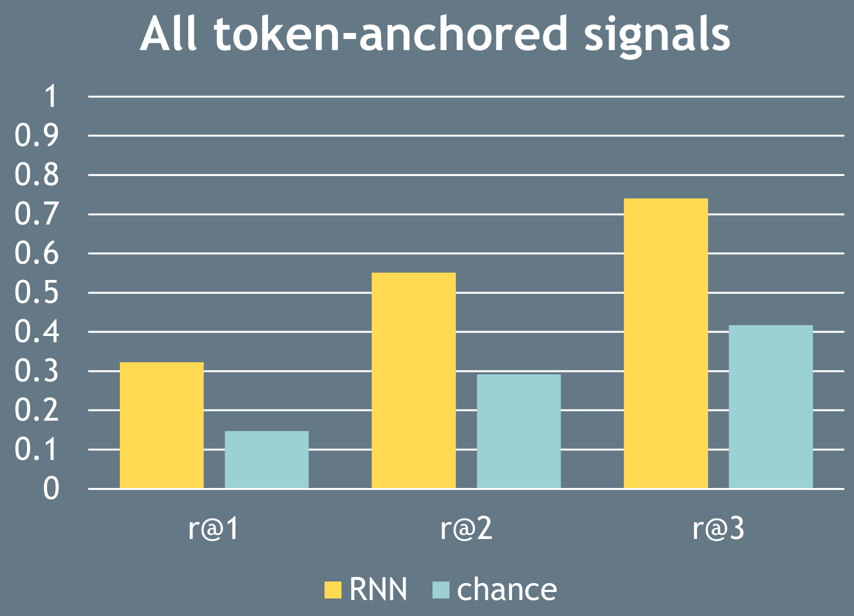
- 3 documents from Signalling Corpus (RST-DT/test)
  - 113 EDUs
  - 210 nodes
  - 153 signals manually inspected
    - Only 83 attributable to a/some tokens
      (not, e.g.: genre, zero relative, graphical layout…)

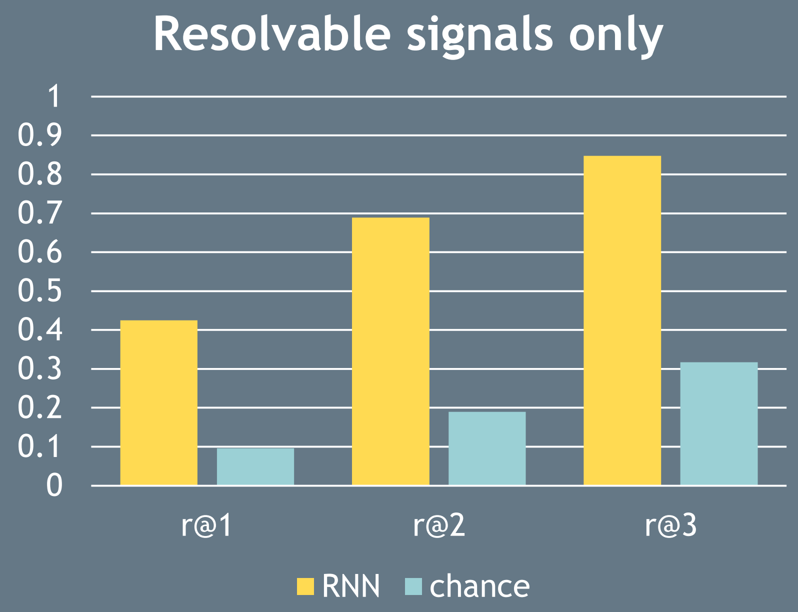
      In a remark [*someone should remember this time next year,*]
    - Only 47 reasonably detectable by net
      (not, e.g.: lexical chain, syntactic parallelism)

      Congress gave Senator Byrd's state … [***Senator Byrd** is chairman..*]

Results

- Network ranks all words (low precision if 0 signals)
- Use ***recall rate @k*** to evaluate

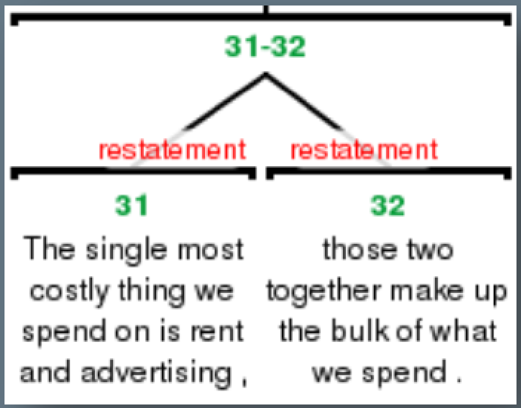




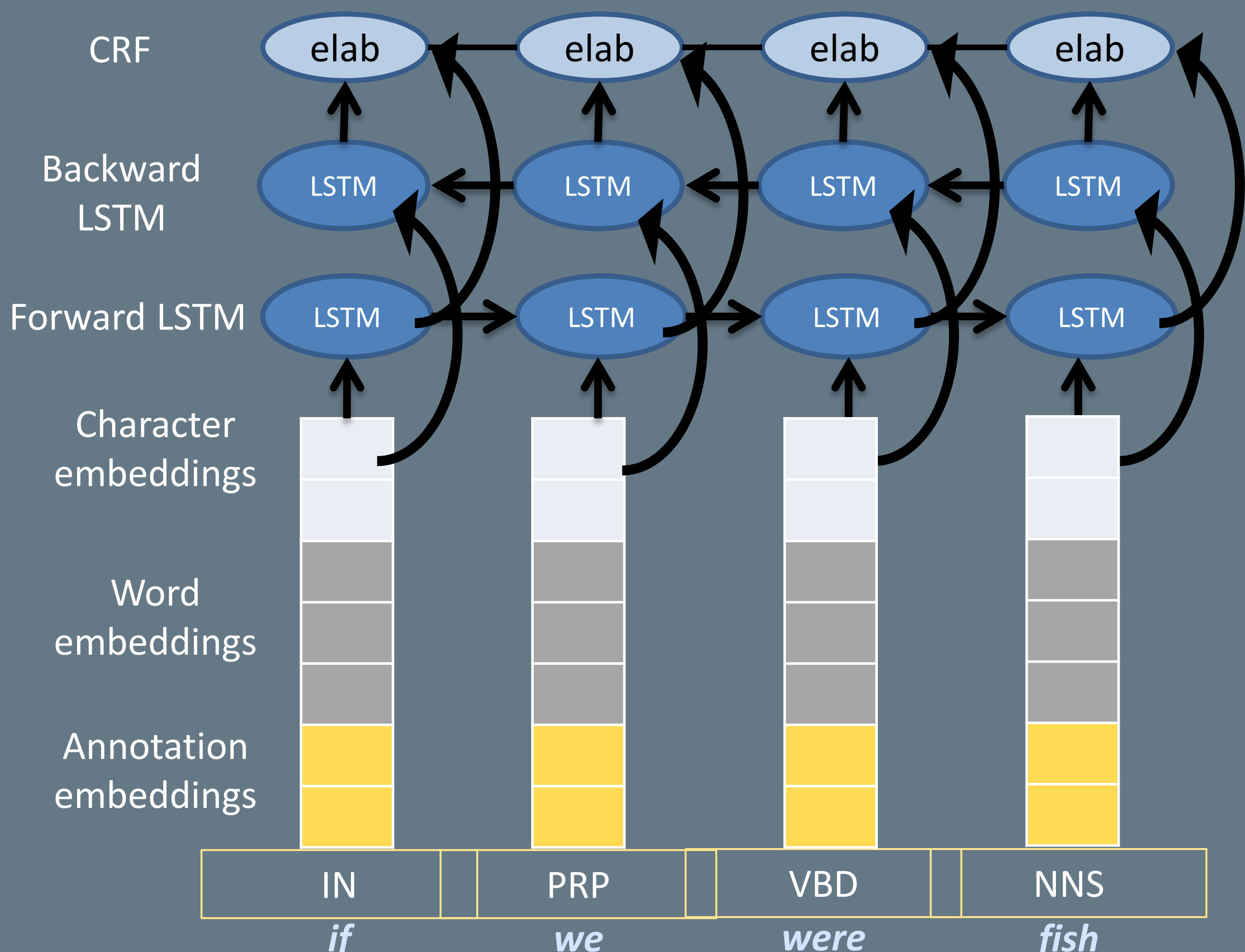
III. Feature rich models

Can we get at 'non-resolvable' cases?

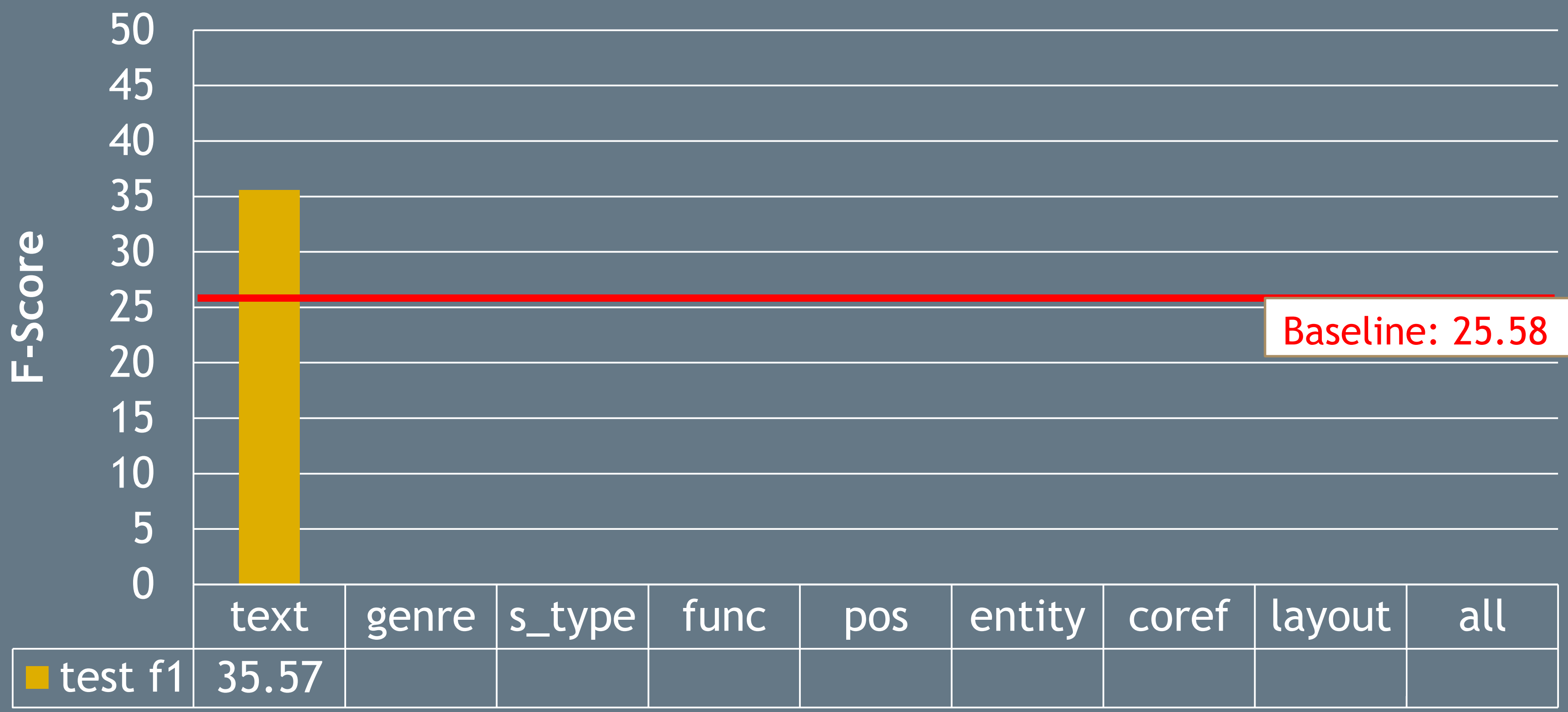
- A plain text RNN can't see many things:
  - Repetition
    - Lexical entity coreference
    - Pronoun resolution
    - Restatements…
  - Non-token signals
    - Syntax clause types and attachment
    - Zero relatives, other 'meaningful absences'
  - Genre (is that 'inside' the text already?)
  - Graphical layout (images, fonts, headings, …)
  - …



Adding annotations to vectors



Adding annotations to vectors

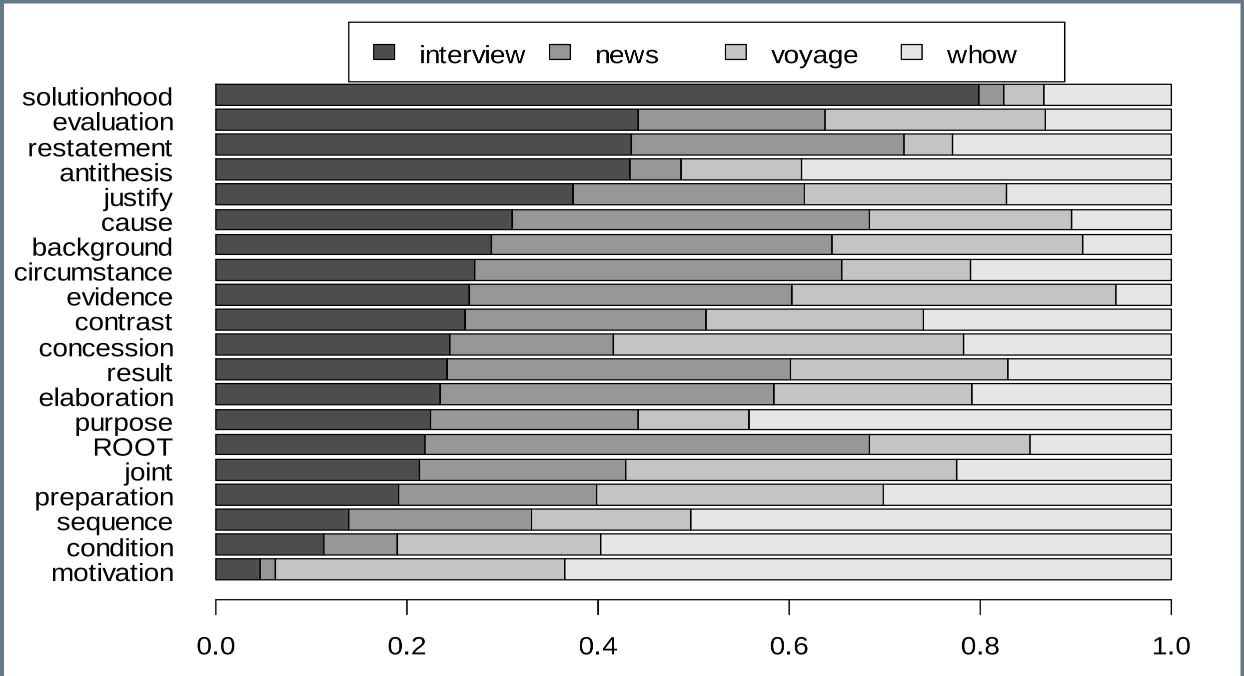


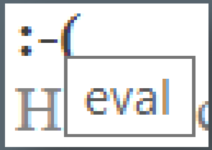
Adding annotations to vectors

	text	genre	s_type	func	pos	entity	coref	layout	all
test f1	35.57	36.65							

Genre

- Genres vary significantly in communicative means
- Prior likelihoods of relations vary:





Quiz: guess which!
- Academic
- Bio
- Fiction
- Interview
- News
- Reddit
- Voyage
- Wikihow

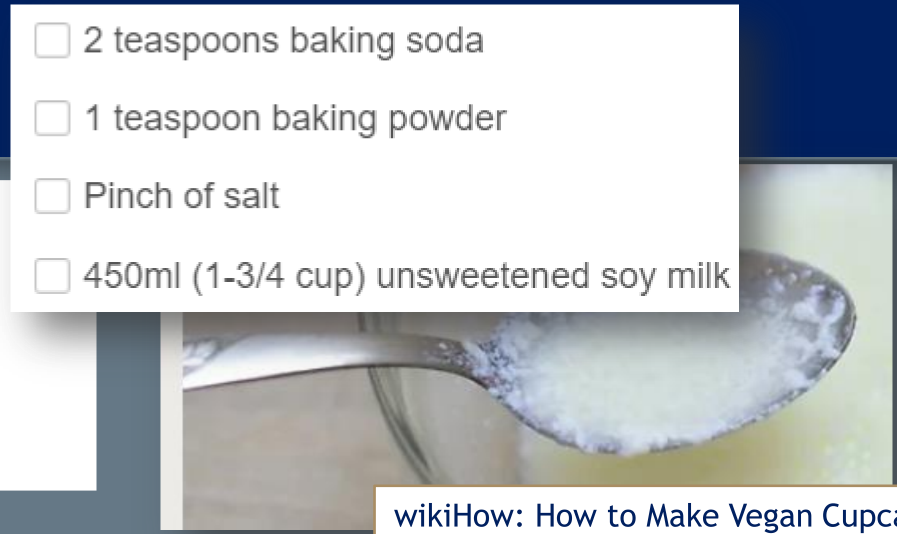
Genre

Plain:

[1 **teaspoon** baking **powder**]$_{\text{joint}}$

+Genre: (whow)

[**1 teaspoon** baking powder]$_{\text{joint}}$



wikiHow: How to Make Vegan Cupcakes

Plain:

[It has lots of local **boutiques**…]$_{\text{elab}}$

+Genre: (voyage)

[It has lots of **local** boutiques…]$_{\text{elab}}$

Wikivoyage: Oakland

Plain:

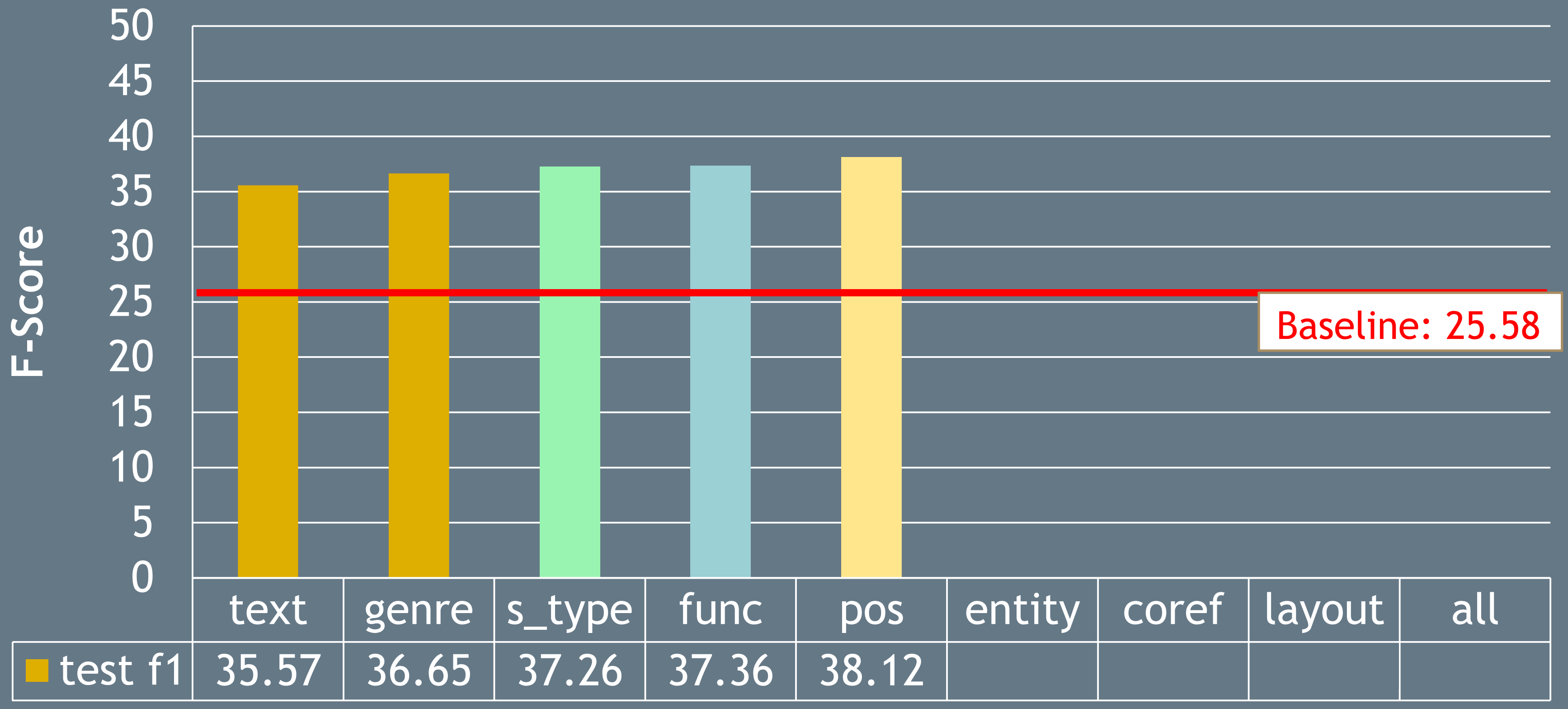
[I **do** n't like the doctor **,**]$_{\text{elab}}$

+Genre: (fiction)

[**I** do n't **like** the doctor ,]$_{\text{eval}}$

"Oversite" by
Maureen F. McHugh

Adding annotations to vectors



POS and dependency function

- The same strings can mean different things:
  - *meaning/NN is self-contained within the text*
  - *meaning/VVG as a first strike weapon*

  (cf. also 'like')

- Similarly for grammatical function:

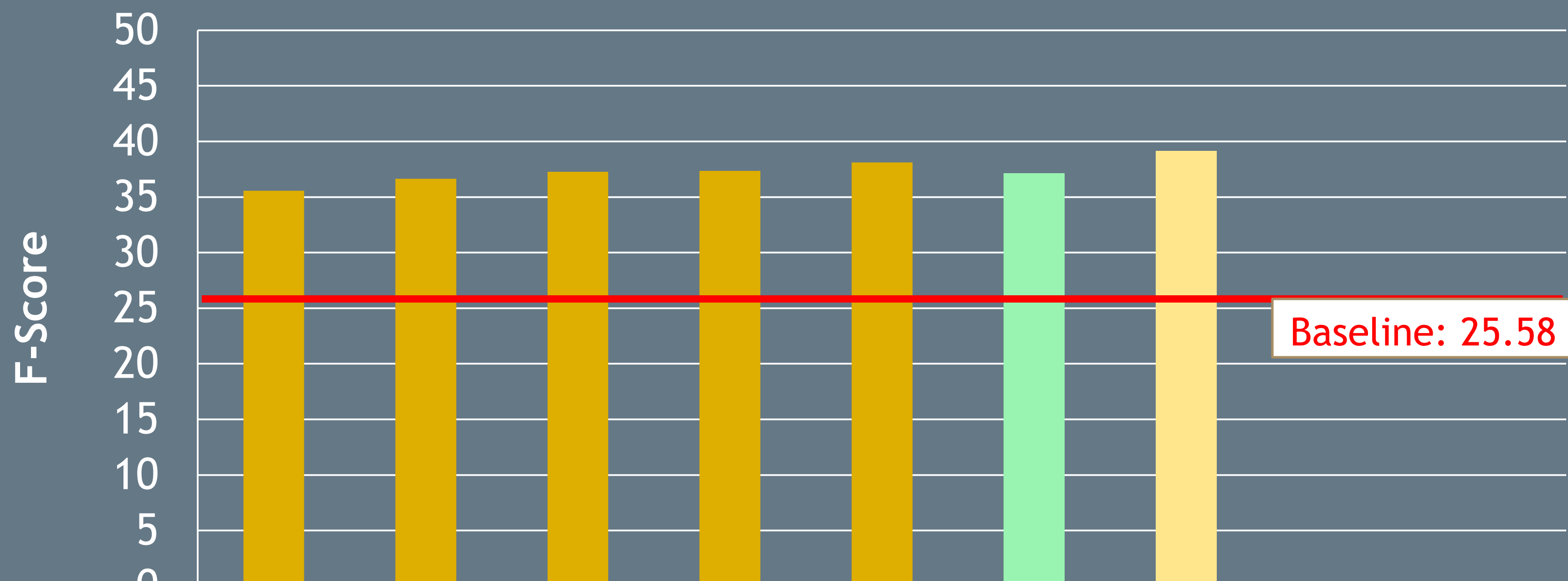
Emperor Joshua Norton; Wikipedia

He reemerged in **September 1859** …

Plain:
[laying claim **to** the position of **Emperor** of the United States . $]_{seq}$

+Deprel:
[laying claim **to** the position of Emperor of the United States . $]_{seq}$

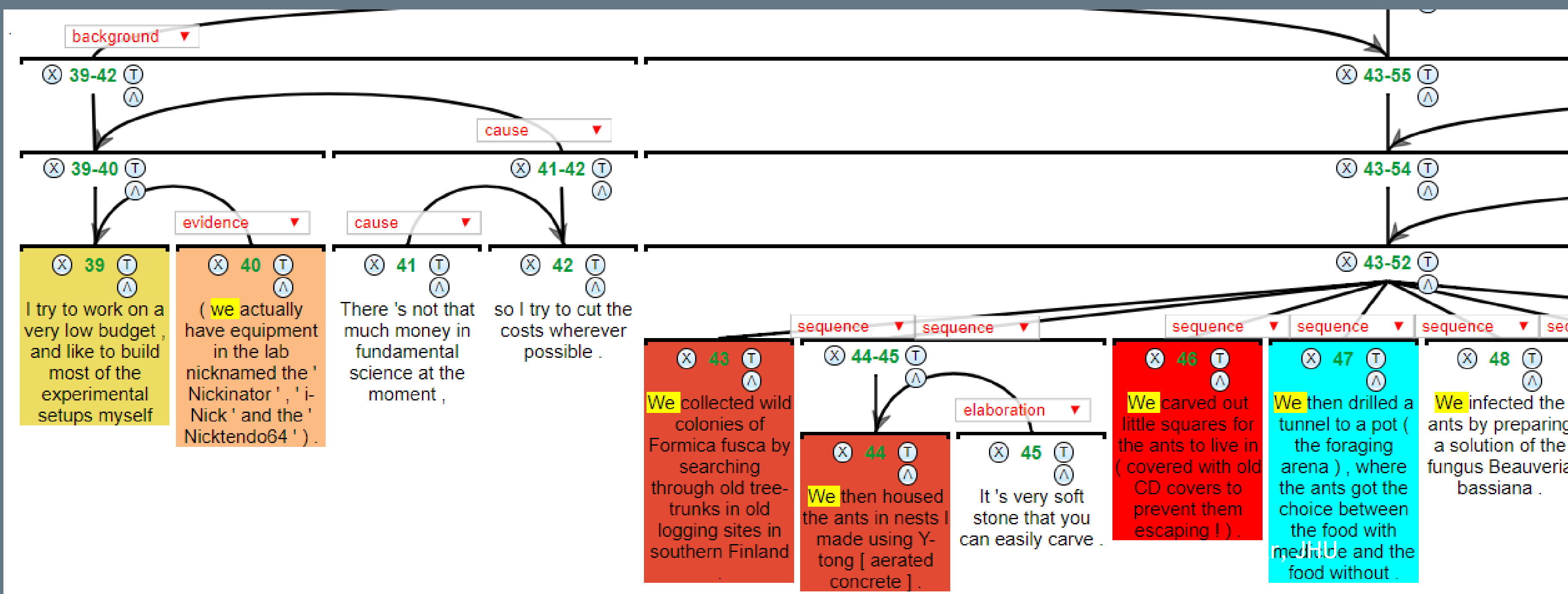
Adding annotations to vectors



	text	genre	s_type	func	pos	entity	coref	layout	all
test f1	35.57	36.65	37.26	37.36	38.12	37.14	39.14		

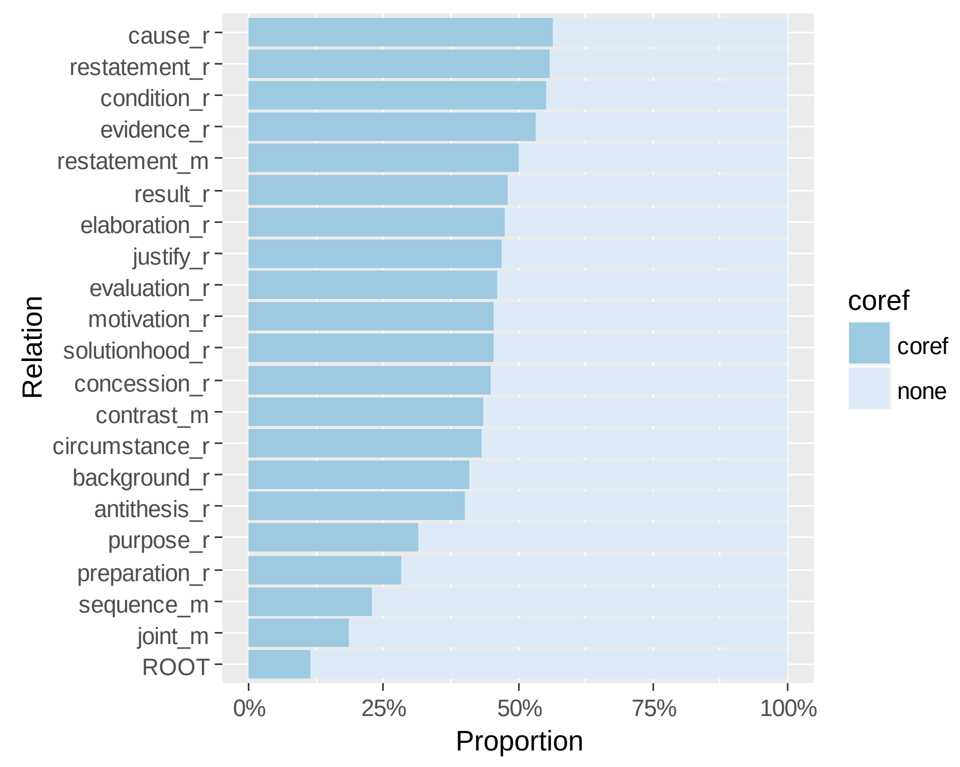
Coreference and entities

- Relationship between referential accessibility and RST graph (Veins Theory, Cristea et al. 1998)
- Coreference likelihood can be predicted by discourse parse (Zeldes 2017b)



Coreference and entities

- Again, different priors:



Coref and entity resolution:

- Know pronoun entities
- Mentioned in RST parent?

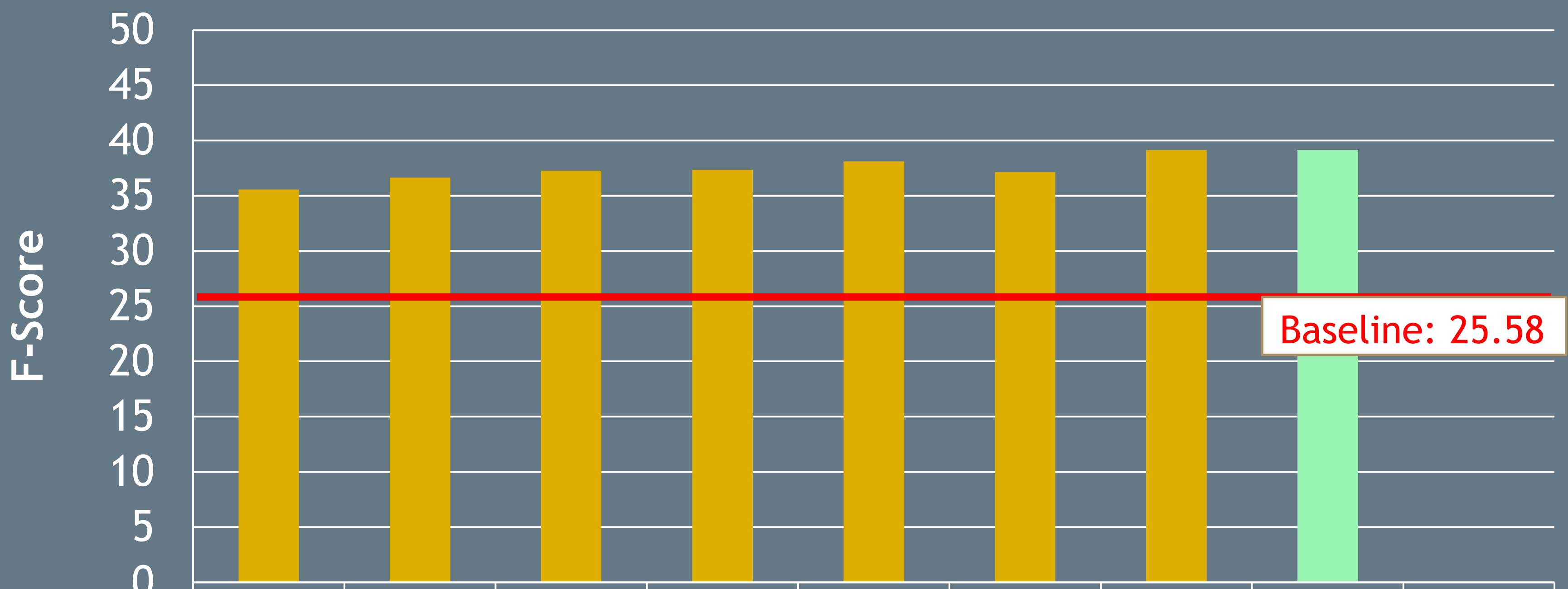
Plain:

[based **on** the knowledge and skills they feel librarians need **;**]$_{elab}$

+Coref+Entities:

[based on the knowledge and skills **they**$_{person}$ feel librarians need **;**]$_{elab}$

Adding annotations to vectors



	text	genre	s_type	func	pos	entity	coref	layout	all
test f1	35.57	36.65	37.26	37.36	38.12	37.14	39.14	39.14	

Graphical layout

- We have TEI XML tags for:
  - Paragraphs
  - Headings
  - Images and captions
  - Ordered / unordered lists
  - Beginning / end of list items
  - …

```
<list type="ordered">
<item n="1">
<head>
<s type="other">
Method  NN   method
One     CD   One
of      IN   of
Two     CD   Two
:       :    :
…
```

Graphical layout

Plain:

[For this question I **do** n't **know** the ' preparedness ' of the Indian gov't to deal with this . $]_{joint}$

+Layout annotations:

[**For** this question I do n't **know** the ' preparedness ' of the Indian gov't to deal with this . $]_{joint}$

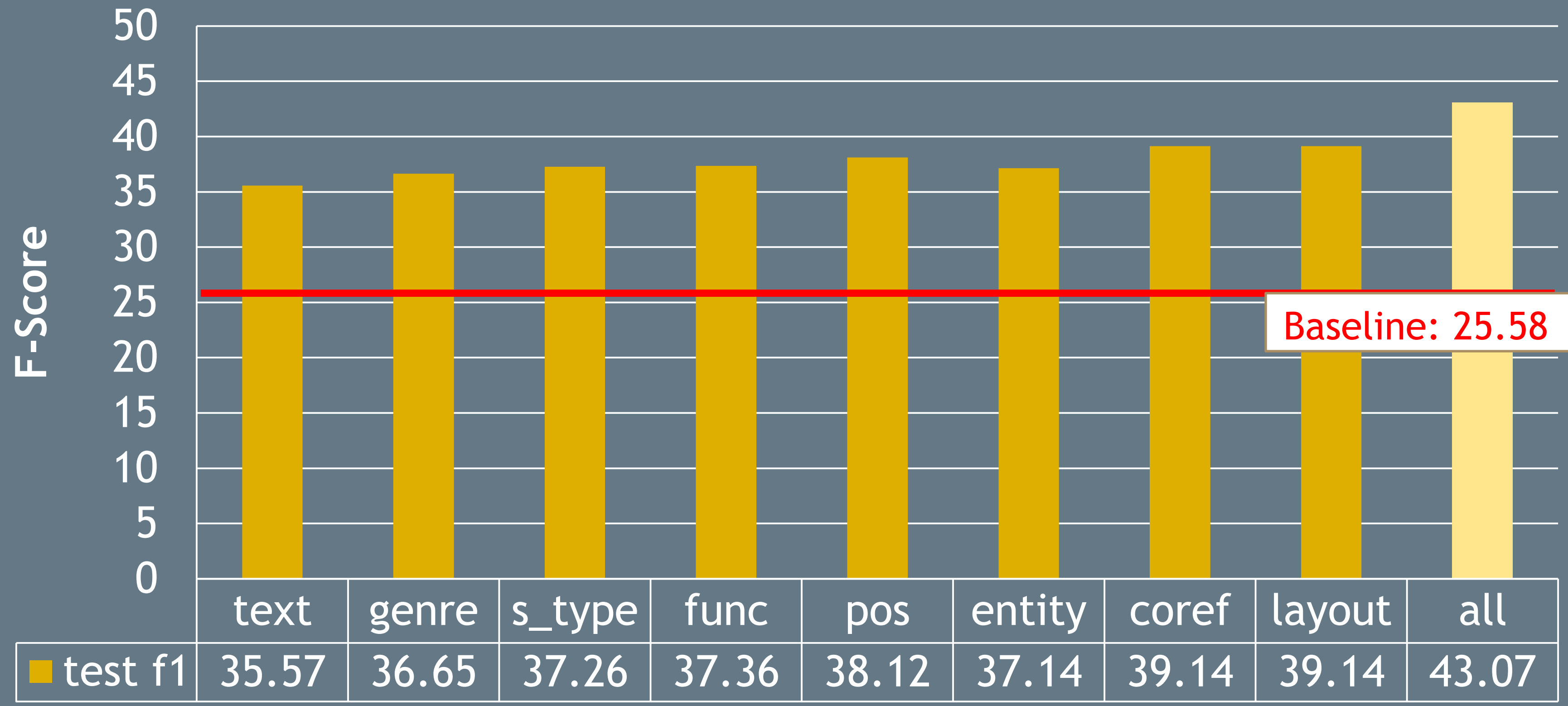
Plain:

[Listen up , **kids :**$]_{prep}$

+Layout annotations:

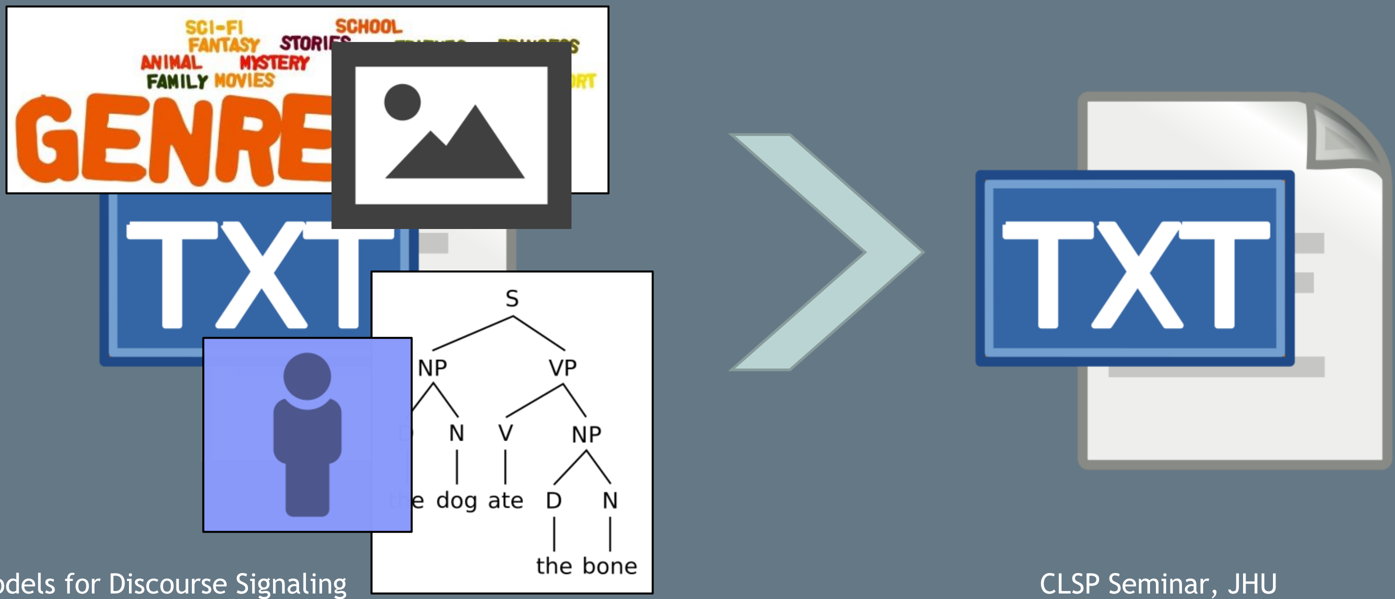
[**Listen** up , **kids** : $]_{prep}$

Adding annotations to vectors



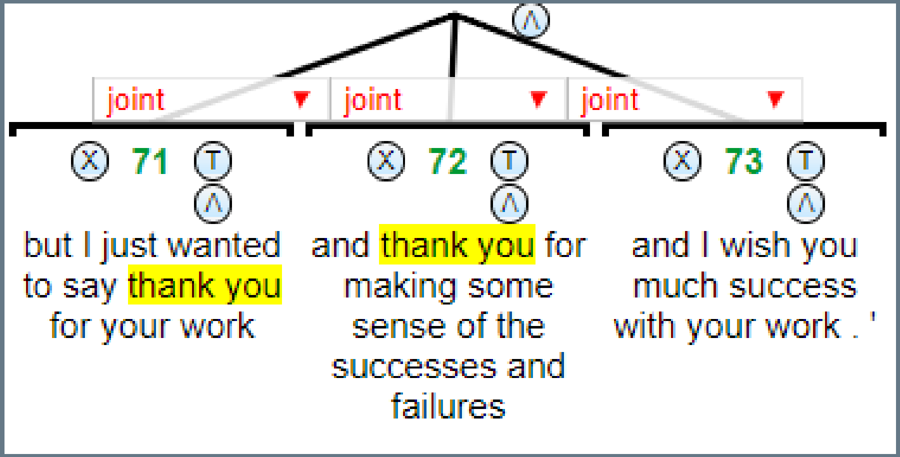
Can we get everything from text?

- Maybe not:
  - Humans use more than just text
  - Some things don't 'anchor' well to text (text!=embeddings)
  - Sometimes text is identical – but other categories matter
  - More than text may be more efficient either way



Conclusion

- Good times to be working on discourse!
- Multiple layers expose complex interdependencies
- Older ideas in computational discourse models are now more feasible:
  - From co-occurrence statistics to contextualized RNN outputs
  - Integrating cues from different levels without overfitting
- We still need new data and new learning approaches!


joint
joint
joint
71
72
73
but I just wanted to say thank you for your work
and thank you for making some sense of the successes and failures
and I wish you much success with your work . '

Thanks!

References

- Afantenos, S./Denis, P./Muller, P./Danlos, L. 2010. Learning recursive segments for discourse parsing. *LREC 2010*, 3578–3584.
- Asher, N./Lascarides, A. 2003. *Logics of Conversation*. Cambridge: Cambridge University Press.
- Braud, C./Coavoux, M./Søgaard, A. 2017. Cross-lingual RST discourse parsing. *EACL 2017*. Valencia, Spain, 292–304.
- Carlson, L./Marcu, D./Okurowski, M. E. 2003. Building a discourse-tagged corpus in the framework of rhetorical structure theory. In *Current and New Directions in Discourse and Dialogue*. Dordrecht: Kluwer, 85–112.
- Cristea, D./Ide, N./Romary, L. 1998. Veins theory: A model of global discourse cohesion and coherence. *ACL 98*, 281–285.
- Hovy, E./Marcus, M./Palmer, M./Ramshaw, L./Weischedel, R. 2006. OntoNotes: The 90% solution. In *Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers*. New York, 57–60.
- Ji, Y./Eisenstein, J. 2014. Representation learning for text-level discourse parsing. *ACL 2014*. Baltimore, MD, 13–24.
- Knott, A./Sanders, T. 1998. The classification of coherence relations and their linguistic markers. *Journal of Pragmatics* 30(2), 135–175.
- Mann, W. C./Thompson, S. A. 1988. Rhetorical Structure Theory. *Text* 8(3), 243–281.
- Prasad, R./Dinesh, N./Lee, A./Miltsakaki, E./Robaldo, L./Joshi, A./Webber, B. 2008. The Penn discourse treebank 2.0. *LREC 2008*. Marrakesh, Morocco.
- Rocktäschel, T./Grefenstette, E./Hermann, K. M./Kočiský, T./Blunsom, P. 2016. Reasoning about entailment with neural attention. *International Conference on Learning Representations (ICLR 2016)*.
- Stede, M./Grishina, Y. 2016. Anaphoricity in connectives: A case study on German. *CORBON 2016.* San Diego, 41–46.
- Taboada, M./Das, D. 2013. Annotation upon annotation: Adding signalling information to a corpus of discourse relations. *Dialogue and Discourse* 4(2), 249–281.
- Taboada, M./Lavid, J. 2003. Rhetorical and thematic patterns in scheduling dialogues. *Functions of Language* 10(2), 147–179.
- van der Vliet, N./Redeker, G. 2014. Explicit and implicit coherence relations in Dutch texts. In Gruber, H./Redeker, G. (eds.) *The Pragmatics of Discourse Coherence: Theories and applications*. Amsterdam/Philadelphia: John Benjamins, 23–52.
- Webber, B. L. 1988. *Discourse Deixis: Reference to Discourse Segments*. Technical Report, University of Pennsylvania.
- Zeldes, A. 2017a. A distributional view of discourse encapsulation: Multifactorial prediction of coreference density in RST. In *6th Workshop on Recent Advances in RST and Related Formalisms at INLG*. Santiago de Compostela, Spain, 20–28.
- Zeldes, A. 2017b. The GUM corpus: Creating multilayer resources in the classroom. *LRE Journal* 51(3), 581–612.

Thanks to GUM annotators! (so far)

- *Adrienne Isaac*
- *Akitaka Yamada*
- *Amani Aloufi*
- *Amelia Becker*
- *Andrea Price*
- *Andrew O'Brien*
- *Anna Runova*
- *Anne Butler*
- *Ayan Mandal*
- *Brandon Tullock*
- *Brent Laing*
- *Candice Penelton*
- *Chenyue Guo*
- *Colleen Diamond*
- *Connor O'Dwyer*
- *Dan Simonson*
- *Didem Ikizoglu*
- *Edwin Ko*
- *Emily Pace*
- *Emma Manning*
- *Han Bu*
- *Hang Jiang*
- *Hanwool Choe*
- *Hassan Munshi*
- *Ho Fai Cheng*
- *Jakob Prange*
- *Jehan al-Mahmoud*
- *Jemm Excelle Dela Cruz*
- *Joaquin Gris Roca*
- *John Chi*
- *Jongbong Lee*
- *Juliet May*
- *Katarina Starcevic*
- *Katherine Vadella*
- *Lara Bryfonski*
- *Lindley Winchester*
- *Logan Peng*
- *Lucia Donatelli*
- *Margaret Anne Rowe*
- *Margaret Borowczyk*
- *Maria Stoianova*
- *Mariko Uno*
- *Mary Henderson*
- *Maya Barzilai*
- *Md. Jahurul Islam*
- *Michaela Harrington*
- *Minnie Annan*
- *Mitchell Abrams*
- *Mohammad Ali Yektaie*
- *Naomee-Minh Nguyen*
- *Nicholas Workman*
- *Nicole Steinberg*
- *Rachel Thorson*
- *Rebecca Childress*
- *Ruizhong Li*
- *Ryan Murphy*
- *Sakol Suethanapornkul*
- *Sean Macavaney*
- *Sean Simpson*
- *Shannon Mooney*
- *Siddharth Singh*
- *Siyu Liang*
- *Stephanie Kramer*
- *Sylvia Sierra*
- *Timothy Ingrassia*
- *Wenxi Yang*
- *Xiaopei Wu*
- *Yang Liu*
- *Yilun Zhu*
- *Yingzhu Chen*
- *Yiran Xu*
- *Young-A Son*
- *Yushi Zhao*
- *Zhuxin Wang*
- *… and others who wish to remain anonymous!*