DERIVING MEANING FROM NOTHINGNESS

Implicit Anaphora and Computational Semantics

Matthew Kuhn - Hot & Odd Topics in Semantics - SS 2018

What is anaphora?

- Classical Linguistics: Referring to an entity elsewhere in the discourse
- Backward (Anaphora), Forward (Cataphora), Outside (Exophora)
- Intrasentential (within sentences) vs. Intersentential (between sentences)
- Examples: Pronouns, possessive adjectives, deictic phrases

Background of Semantic Anaphora

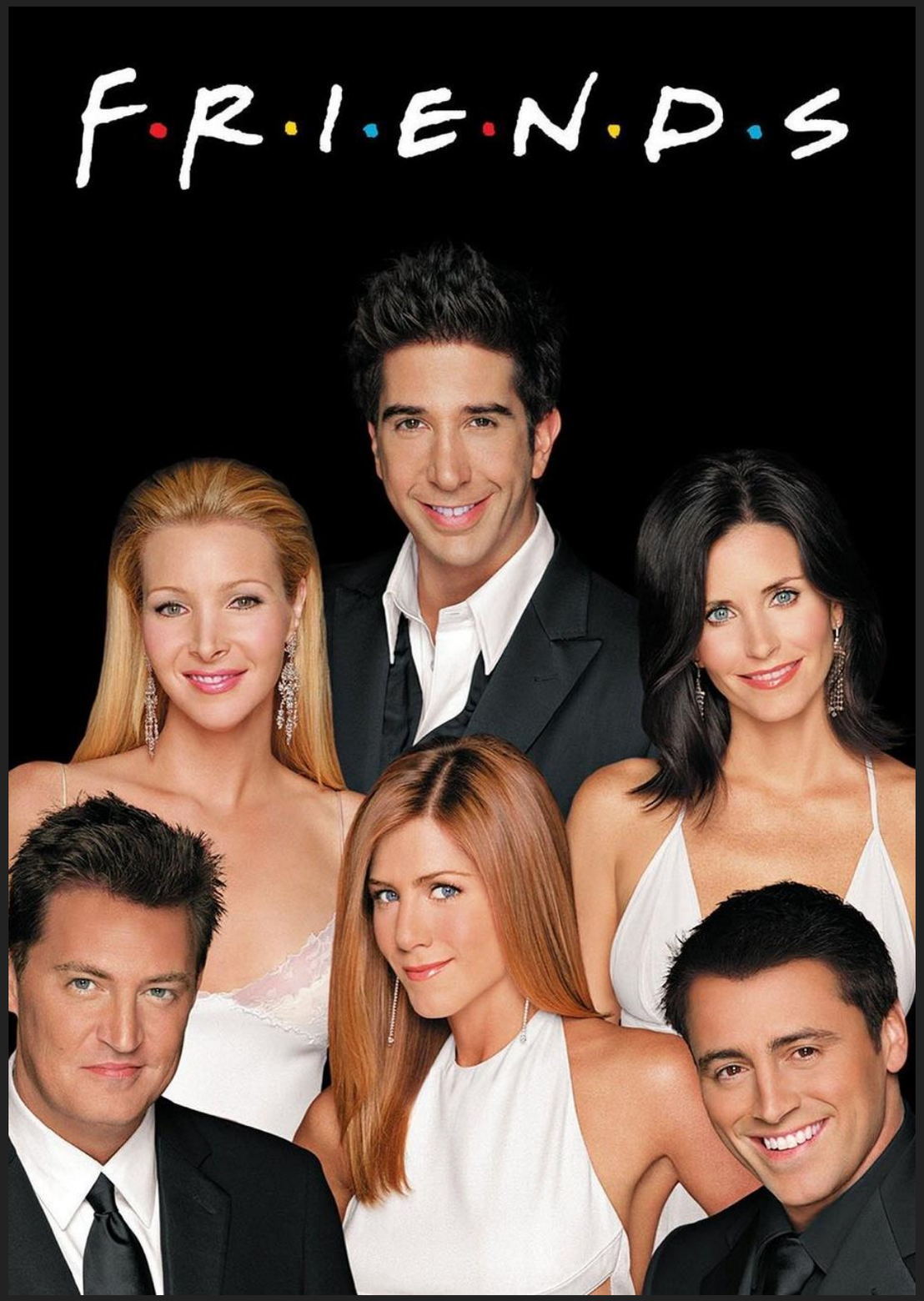
- 1950's: Noam Chomsky
  - Generative Grammar
- 1960's: Richard Montague
  - Lambda Calculus, Compositionality
- 1981: Hans Kamp
  - Discourse Representation Theory

Background of Semantic Anaphora

- 1986: Charles Fillmore
  - Frame Semantics
  - Null Complements
    - Indefinite: “Unknown or a matter of indifference”
    - Definite: must be retrieved from context
- 1986: Palmer et al.
  - PUNDIT - first attempt to automatically identify antecedents
- 1995: Grosz et al.
  - Centering Theory

After that history, Let's practice!

- Because Chandler was hot, he removed his sweater.
- After work, the friends will go to Central Perk.
- Ross is standing over there.
- Which of these was the hardest to understand?



more complicated anaphora

- Associative Anaphora
  - Information present without direct syntactic link
  - When Monica and Rachel opened the apartment door, they knew something was wrong. The walls were wet, and water was pooling on the floor.
- Implicit Anaphora
  - Purely inferential
  - “The Big Bang Theory” is a long-running sitcom. “Friends” comes close, though.

Implicit anaphora

Challenges of Anaphora for NLP

- Discourse Modeling
  - Salience is a concern
  - Ross asks, “Where is Rachel?” Phoebe says, “Rachel is at work at Central Perk in Manhattan.”
- Information Extraction
  - Identification of referents
    - How can we model this?

computational applications

Inducing Implicit Arguments via Cross-Document Alignment

- Semantic Role Labeling
  - Goal: induce predicate-argument relationships automatically from text
  - Argument ← Predicate →Argument
  - Most research is strictly local

- How should we handle non-local information?
  - 1) **Identify** that the information exists
  - 2) **Link** the implicit arguments to their antecedents
- Problem: Few annotated corpora
  - Solution: Comparable Texts

RESOURCES

- FrameNet
- Based on Frame Semantics (Fillmore 1976)
  - Frames →Core/non-Core→Participants and Properties
  - Work, V., sense, Being-Employed
    An Employee has a Position doing work in a particular Field, or on a particular Task, for which an Employer gives Compensation to the Employee.
- Advantages: correction across words, missing arguments are marked as Null-Instantiated
- Disadvantages: Poorer connection between word senses, abstraction from syntax, discourse antecedents unannotated

RESOURCES

- PropBank
  - Word Senses → Arguments
  - Syntactically derived
  - Annotation Scheme: role labels
    Ex: A0 Agent, A1 Patient, A2 beneficiary/instrument/attribute/end state
  - Advantages: More directly related to natural language, good connection between word senses
  - Disadvantages: Hard to relate predicates to one another, missing arguments are unmarked, only intrasentential

complementary paradigms

- Intuition: combine PropBank and FrameNet
- PropBank parsers offer better precision and coverage
  - Label semantic roles with PropBank
  - Merge Predicate-Argument Structures across documents
- Map induced implicit arguments to FrameNet
- Apply to existing models for linking

ROTH AND FRANK 2015

- Two Parts
- 1) Generate a Corpus of Implicit Arguments Linked to Antecedents
- 2) Apply the Corpus to Existing Methods

Generate a Corpus of Implicit Arguments Linked to Antecedents

- Step 1: Identify pairs of comparable texts
- Step 2: Detect shared information
- Step 3: Induce implicit arguments

Step 1: identifying comparable texts

- Start with Gigaword corpus

Step 1: identifying comparable texts

- Pair articles by source
- Compare headlines
- Headlines should be published in same 2-day window

$$\cos(doc_1, doc_2) = \frac{\vec{doc_1} \cdot \vec{doc_2}}{\|\vec{doc_1}\| * \|\vec{doc_2}\|}$$

cosine similarity

$$\text{TF-IDF}_{doc_1}(w) = |\{w \in hl_1\}| * \log \frac{|\{hl' \in pair corpus\}|}{|\{hl' \in pair corpus | w \in hl'\}|}$$

Term Frequency - Inverse Document Frequency

Step 1: identifying comparable texts

	Iraqi parliament approves new unity government	Iraqi parliament approves new government
	$\downarrow$	$\downarrow$
Headline vectors:	$[1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1]$	$[1 \quad 1 \quad 1 \quad 1 \quad 0 \quad 1]$
	$\downarrow$	$\downarrow$
TF-IDF weighted:	$\vec{doc_1} = [6 \quad 6 \quad 6 \quad 3 \quad 8 \quad 5]$	$\vec{doc_2} = [6 \quad 6 \quad 6 \quad 3 \quad 0 \quad 5]$

$$\Rightarrow cos(doc_1, doc_2) = 0.73$$

Example of weighted headlines

- Resulting dataset:
  - 167728 document pairs
  - >50 million tokens
  - Manual annotation of 70 pairs: 98.6% accuracy

Step 2: detecting shared information through Predicate argument structures

- Alignment Task
- Automatically identify PAS with PropBank (Bohnet 2010, Björkelund et al. 2010)
- Gold Standard: 70 manually annotated datasets

work	--A0--	The Russian military
	--A1--	to save a small submarine
	--MNR--	desparately
	--TMP--	on Friday
save	--A0--	The Russian military
	--A1--	a small submarine

Figure 6.1.: Parser output for the sentence "The Russian military worked desperately on Friday to save a small submarine." Here, predicates are displayed on the left, argument labels in the center and argument realizations on the right.

Step 2: Detecting Shared information

- Judging similarity

Predicate-specific measures		
Similarity in WordNet	X	-
Similarity in VerbNet	X	(X)
Similarity in a Semantic Space	X	-
Argument-specific measures		
Bag-of-Words Similarity	-	X
Head of Arguments Similarity	-	X
Discourse-specific measures		
Relative Discourse Position	-	X
Context Similarity	-	X

Type to enter a caption.

Step 2: Detecting shared information

$$\mathrm{sim}_{\mathrm{WN}}(p_1, p_2) = \max_{s_1 \in \mathrm{synsets}(p_1),\ s_2 \in \mathrm{synsets}(p_2)} \frac{\mathrm{IC}(\mathrm{lcs}(s_1, s_2))}{\mathrm{IC}(s_1) * \mathrm{IC}(s_2)}$$

WordNet Similarity

$$\mathrm{sim}_{\mathrm{VN}}(p_1, p_2) = \begin{cases} 1.0 & \text{if } \exists C : p_1, p_2 \in C \\ 0.8 & \text{if } \exists C, C_s : C_s \in \mathrm{sub}(C) \\ & \quad \wedge\ ((p_1 \in C, p_2 \in C_s) \vee (p_1 \in C_s, p_2 \in C)) \\ \textit{default} & \text{else} \end{cases}$$

VerbNet Similarity

$$\mathrm{pmi}(p, c) = \frac{\mathrm{freq}(p, c)}{\mathrm{freq}(p) * \mathrm{freq}(c)}$$

Pointwise Mutual Information

$$\mathrm{sim}_{Dist}(p_1, p_2) = \frac{\vec{p_1} \cdot \vec{p_2}}{\|\vec{p_1}\| * \|\vec{p_2}\|}$$

Cosine Similarity

$$\vec{p} = (\mathrm{pmi}(p, c_1), \mathrm{pmi}(p, c_2), \ldots, \mathrm{pmi}(p, c_{2,000}))$$

Similarity in Semantic Space: PMI as vectors

Step 2: Detecting shared information

$$\mathrm{sim}_{\mathrm{ABoW}}(p_1, p_2) = \frac{\sum_{w \in A_1 \cap A_2} \mathrm{idf}(w)}{\sum_{w \in A_1} \mathrm{idf}(w) + \sum_{w \in A_2} \mathrm{idf}(w)}$$

Bag-of-Words similarity

$$\mathrm{sim}_{\mathrm{AHeads}}(p_1, p_2) = \frac{\sum_{\{a_1, a_2 | \mathrm{label}(a_1) == \mathrm{label}(a_2)\}} \mathrm{simWN}(\mathrm{head}(a_1), \mathrm{head}(a_2))}{|\{a_1, a_2 | \mathrm{label}(a_1) == \mathrm{label}(a_2)\}|}$$

Head of Arguments similarity

$$\mathrm{sim}_{DPos}(p_1, p_2) = 1 - \left(\left| \frac{\mathrm{sentence\_index}(p_1)}{\mathrm{length}(d_1)} - \frac{\mathrm{sentence\_index}(p_2)}{\mathrm{length}(d_2)} \right| \right.$$

Relative Discourse Position

$$\mathrm{sim}_{DCon}(p_1, p_2) = \frac{\mathrm{context}(p_1) \cap \mathrm{context}(p_2)}{\mathrm{context}(p_1) \cup \mathrm{context}(p_2)}, \text{ with}$$

$$\mathrm{context}(p) = \{p' | index(p') \in [index(p) - n : index(p) + n]\}$$

Context similarity

THAT WAS A LOT OF MATH! HERE'S A PUPPY

Step 2: detecting shared information

- Alignment via Clustering
  - Deriving the weights

$$w_{p_1 p_2} = \Sigma_i \quad \lambda_i * \mathrm{sim}_i(p_1, p_2)$$

Type to enter a caption.

Step 2: detecting shared information

- Clustering
- Maximize sum of weights within clusters, minimize without
- Minimum Cuts
  - Find edge with lowest weight
  - Cut it such that the sum of removed weights is minimal, and the nodes formerly connected are in different graphs

Step 2: Detecting shared information

	P	R	F_1
	Baselines		
LemmaId	40.3**	60.3*	48.3**
Greedy	12.5**	27.6**	17.2**
WordAlign	19.7**	15.2**	17.2**
	Previous work		
Wolfe et al.	52.4**	**64.0****	57.6
EMNLP'12	58.7**	46.6	52.0
	This thesis		
Full	**71.8**	48.9	**58.2**

	P	R	F_1
$-\mathbf{sim_{WN}}$	78.8	44.6**	57.0
$-\mathbf{sim_{VN}}$	78.7	44.6**	57.0
$-\mathbf{sim_{Dist}}$	77.3	45.5**	57.3
$-\mathbf{sim_{ABoW}}$	67.6*	49.3	57.0
$-\mathbf{sim_{Aheads}}$	68.9	**52.9****	**59.8**
$-\mathbf{sim_{DPos}}$	64.2	43.0**	51.6**
$-\mathbf{sim_{DCon}}$	69.3*	52.2**	59.6
Full	71.8	48.9	58.2
+HighPrec	**86.2**	29.1**	43.5**

Table 6.4.: Results for discourse-level alignment in terms of precision (P), recall (R) and F_1-score (all numbers in %); left: comparison of the **Full** model to baselines and previous work; right: impact of removing individual measures and using a tuned weighting scheme; results that significantly differ from **Full** are marked with asterisks (* $p<0.05$; ** $p<0.01$).

Step 3: inducing implicit arguments

- Pre-processing
  - CoreNLP: tokenization and sentence splitting
  - MATEtools: determine local predicate structure
  - Martschat et al. 2012: resolve pronouns

Step 3: inducing implicit arguments

- Cross-Document Coreference
  - Stanford Coreference System (Lee et al. 2013)
- Automatic Identification and Linking: Restrictions
  - Mislabeled arguments
  - Incorrectly resolved pronouns
  - Missed arguments

Step 3: inducing implicit arguments

[T-Online$_i$], the leading Internet services provider in Europe and a unit of Deutsche Telekom, said Thursday its net loss more than doubled last year owing to its foreign activities and goodwill writedowns. (. . .)

The [$\emptyset_i$]$_{A0}$ [operating]$_{A3}$ **loss**, as measured by earnings before interest, tax, depreciation and amortisation, widened to 189 million euros last year from 121.6 million euros a year earlier.

[T-Online's$_i$]$_{A0}$ [operating]$_{A3}$ **loss** – earnings before financial items such as interest, taxes, depreciation and amortization – also widened, to 189 million euros (dlrs 167 million) in 2001 from 122 million (dlrs 108 million).

Example Results

- Manual analysis: 89% accurate

Applications Task 1: Linking Implicit Arguments in Discourse

- How can we use this corpus to improve existing SRL?
- SemEval 2010: Linking Events and their Participants in Discourse
  - 1) Identify implicit arguments
  - 2) Classify implicit arguments as definite or indefinite null instantiations
  - 3) Find an appropriate antecedent if possible
  - System used: Silberer and Frank 2012

Results

	Precision	Recall	F_1-score
Chen et al. (2010)[2]	**0.25**	0.01	0.02
Tonelli and Delmonte (2011)	0.13	0.06	0.08
Laparra and Rigau (2012)	0.15	**0.25**	**0.19**
Laparra and Rigau (2013b)	0.14	0.18	0.16
Gorinski et al. (2013)[3]	0.14	0.12	0.13
S&F (no additional data)	0.06	0.09	0.07
S&F (best additional data)	0.09	0.11	0.10
This thesis	0.21	0.08	0.12

Type to enter a caption.

Applications Task 2: Modeling Local Coherence

- Can the intuitions of human annotators be computationally reproduced?
  - Exploit predicate-argument structure alignments to identify explicit/implicit arguments
  - Train a model to determine if arguments should be realized or not

Applications Task 2: Modeling Local Coherence

- Corpus: document pairs that differ in only a single argument realization
- Gold Standard: human annotation
- Baselines: Entity-grid, pronoun-based, and discourse-newness models

Applications Task 2: Modeling Local Coherence

- System: LIBSVM (Chang and Lin 2011)
- Features:
  - Complexity of predicate-argument structure
  - Coherence-specific features
  - Discourse-level features

Applications Task 2: Modeling Local Coherence

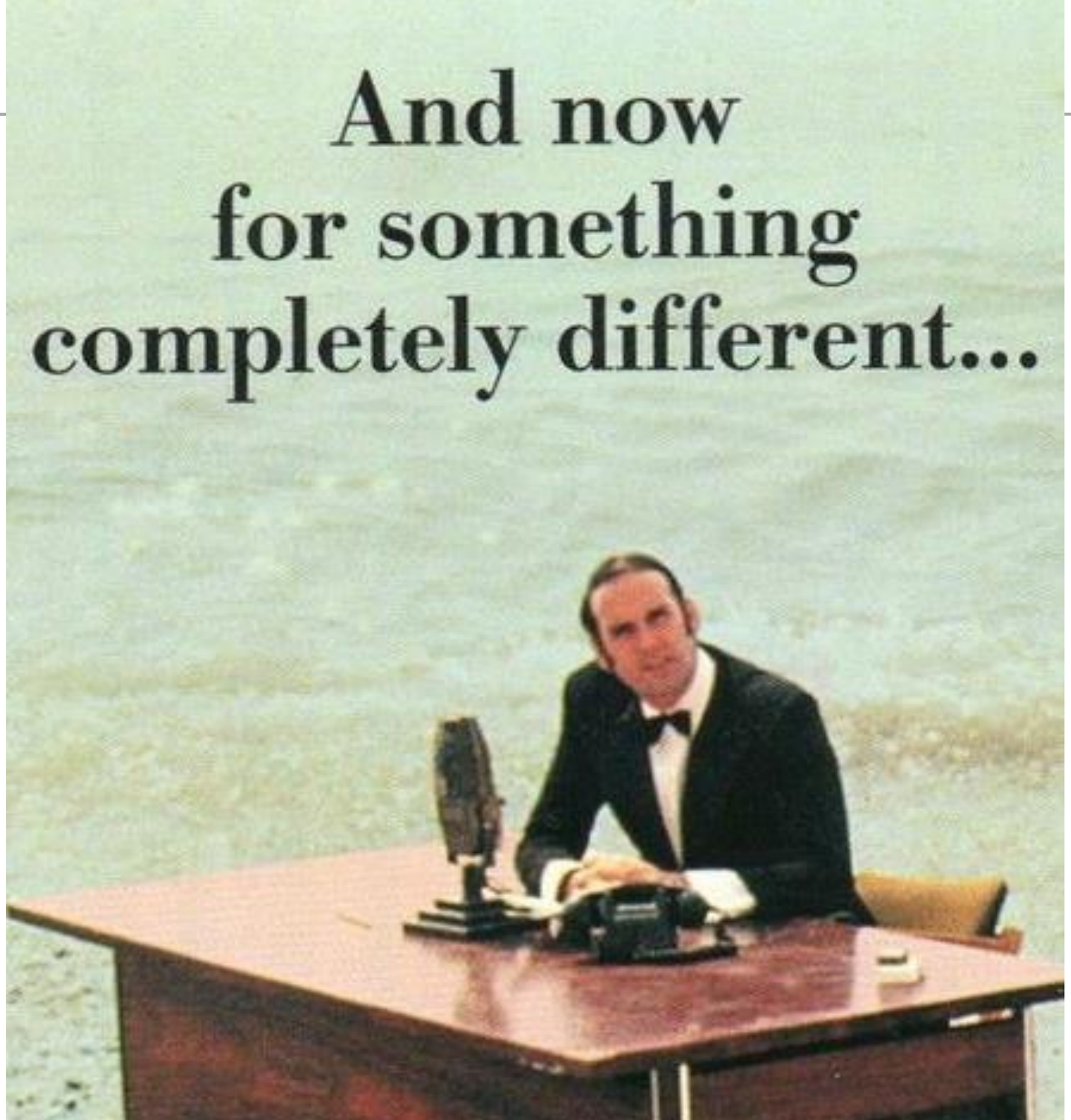
	$P_{implicit}$	$P_{explicit}$	$P_{overall}$	$R_{overall}$	F_1-score
Entity grid models	–	–	–	–	–
Baseline entity grid	0.50	0.05	0.15**	0.14**	0.15**
Extended entity grid	0.56	0.00	0.19**	0.17**	0.18**
Topical entity grid	0.86	0.20	0.34**	0.34**	0.34**
Other models	–	–	–	–	–
Pronouns	0.60	0.37	0.43**	0.34**	0.38**
Discourse-newness	**1.00**	0.25	0.48**	0.48**	0.48**
This thesis	–	–	–	–	–
Our (full) model	0.78	0.95	**0.90**	**0.90**	**0.90**
Simplified model	0.56	**1.00**	0.83	0.83	0.83
Majority class	0.00	1.00	0.69*	0.69*	0.69*

Table 8.2.: Results in Precision, Recall and F_1-score for correctly predicting argument realization; Significant differences in overall scores from our (full) model are marked with asterisks (* $p<0.1$; ** $p<0.01$)

Type to enter a caption.

Summary

- It is possible to automatically generate comparable texts
  - Work intensive, but still easier than obtaining hand-annotated texts of sufficient size
- Intersentential implicit arguments can be induced from data
- New corpora have many applications, but are probably best applied to existing models


And now
for something
completely different...

implicit argument prediction with event knowledge

- Idea: Use an easily automatically generated corpus to train a neural network to identify implicit arguments
- Ex: “More than 2600 people have been infected by Ebola in Liberia, Guinea, Sierra Leone, and Nigeria since the outbreak began in December, according to the World Health Organization. Nearly 1500 have died.”
- What killed them?

- Key Insights
  - Event Knowledge
  - Salience of missing arguments

Corpus generation

- Automated Cloze Task
  - Remove an argument that isn’t unique from each text
  - System must determine what argument fits in the hole
- Drawn from English Wikipedia
  - Pre-processed with CoreNLP
  - 2 sets: 40 million and 8 million tokens

Corpus generation

Manville Corp. said it will build a $ 24 million power plant to provide electricity to its Igaras pulp and paper mill in Brazil .

The company said the plant will ensure that it has adequate energy for the mill and will reduce the mill's energy costs .

(a) A piece of raw text from OntoNotes corpus.

x_0 = The company x_1 = mill x_2 = power plant

e_0: (*build-pred*, x_0-*subj*, x_2-*dobj*, –)
e_1: (*provide-pred*, –, *electricity-dobj*, x_1-*prep_to*)
e_2: (*ensure-pred*, x_2-*subj*, –, –)
e_3: (*has-pred*, x_0-*subj*, *energy-dobj*, x_1-*prep_for*)
e_4: (*reduce-pred*, x_2-*subj*, *cost-dobj*, –)

(b) Extracted events (e_0~e_4) and entities (x_0~x_2), using gold annotations from OntoNotes.

e_0, e_2, e_3, e_4: same as above
e_1: (*provide-pred*, –, *electricity-dobj*, ??-*prep_to*)

x_0 = The company x_1 = mill x_2 = power plant

(c) Example of an argument cloze task for *prep_to* of e_1.

Figure 1: Example of automatically extracted events and entities and an argument cloze task.

Type to enter a caption

Modeling Narrative Coherence: Event Composition

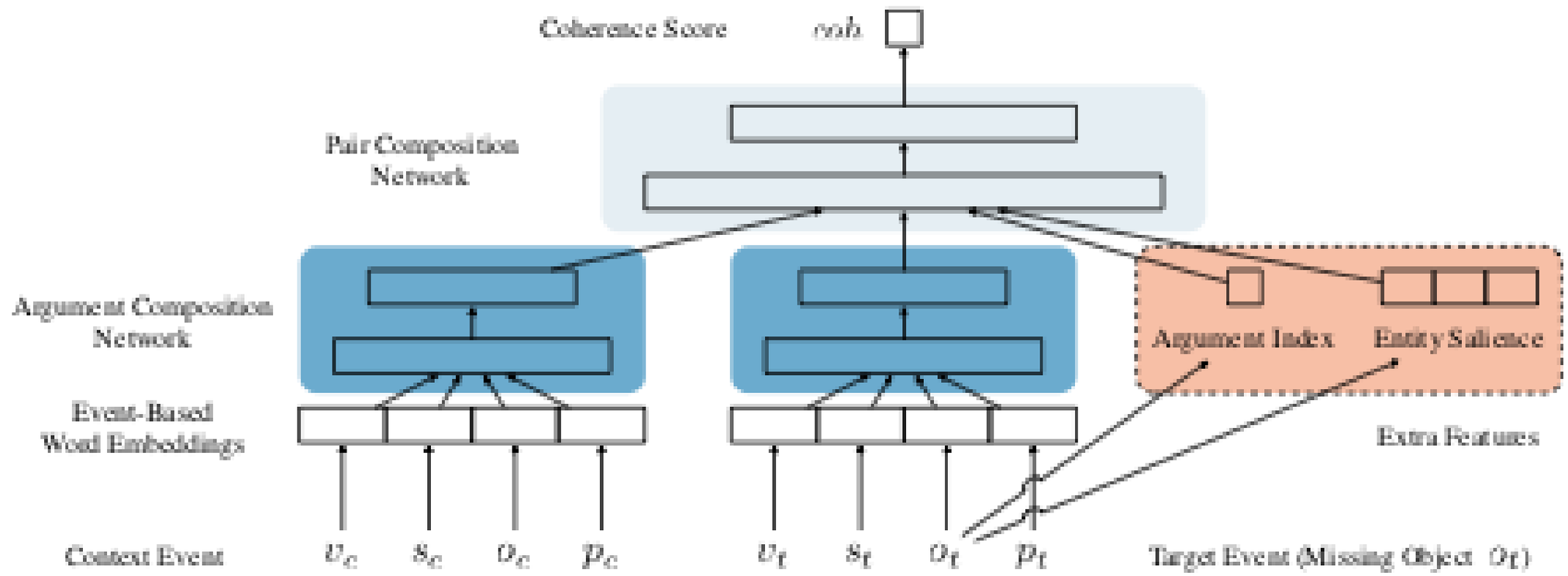


Figure 2: Diagram for event composition model. ***Input***: a context event and a target event. ***Event-Based Word Embeddings***: embeddings for components of both events that encodes event knowledge. ***Argument Composition Network***: produces an event representation from its components. ***Pair Composition Network***: computes a coherence score *coh* from two event representations. ***Extra Features***: argument index and entity salience features as additional input to the pair composition network.

Type to enter a caption.

Experimental Details

- NN Objective Function
- Entity Salience: 3 Features
  - Index of sentence where entity is first mentioned
  - Count of appearances of head word
  - Vector containing counts of named, nominal, pronominal, and total mentions
  - Evaluated on Argument Cloze and results from Gerber & Chai 2012

$$\frac{1}{m}\sum_{i=1}^{m} -\log(coh(e_{ci}, e_{pi})) - \log(1 - coh(e_{ci}, e_{ni})) \quad (3)$$

Type to enter a caption.

Evaluation on Argument Cloze

- 3 Baselines
  - Randomly select an entity
  - Always take most frequent entity
  - Event-based word embeddings (cosine similarity)

	Accuracy (%)
RANDOM	8.29
MOSTFREQ	22.76
EVENTWORD2VEC	38.40
EVENTCOMP-8M	38.26
+ entity salience	45.05
EVENTCOMP-40M	41.89
+ entity salience	**47.75**

Table 3: Evaluation on ON-SHORT.

Features	Accuracy (%)
no entity salience feature	38.26
– *mentions*	39.02
– *head_count*	**45.71**
– *1st_loc*	**45.65**
all entity salience features	45.05

Table 4: Ablation test on entity salience features. (Using EVENTCOMP-8M on ON-SHORT.)

Type to enter a caption.

Final Remarks

- Semantic parsing is non-trivial!
- Global anaphora is computationally complex
- Many open questions remain

Type to enter a caption.

THANKS!

QUESTIONS?