Final Class

Announcements

- Course evaluation: please fill out
- Final reviews:
- Final exam: 12/9, final is cumulative
  - Closed book and electronics But you can bring calculator

Today

- Research topics in the Columbia NLP group
- Guidelines for studying
- Review
  - Attention
  - Bert
  - Word sense disambiguation/ POS tagging
  - Summarization/pyramid

Columbia NLP

- Michael Collins
  - Spring: NLP
- Julia Hirschberg
  - Spring: Advanced Spoken Language Processing
- Kathy McKeown
  - Spring: independent projects
- Smara Muresan
  - Spring: Multilingual Language Technologies and Language Diversity”

Cross-Lingual Summarization for Low Resource Languages

- Cross-lingual summarization: summarize in one language a document written in another
    - Summarize and then translate the summary
    - Translate and then summarize the translation
- Low resource languages
    - Little to no data to train summarization systems
- Query-focused
    - Given a query, generate a summary that indicates whether a document is relevant to the query
    - Automatically generating training data for different kinds of queries

Funded by IARPA MATERIAL

New Directions

- What about summaries that are really abstractive?
  - Summaries of online personal narrative?
  - Summaries of debates?
- Can we develop better representations?
  - Currently embeddings don't capture salience

Analyzing social media

- Stress
- Stance
- Sentiment
- Persuasion/influence
- Argumentation
- Disinformation/fact-checking
- Hate speech/abusive language

Stress

- Dreaddit: a corpus of long-form social media text for stress analysis
  - 5 subreddits (abuse, anxiety, financial, PTSD…)

I have this feeling of dread about school right before I go to bed and I wake up with an upset stomach which lasts all day and nakes me feel like I'll throw up. This causes me to lose appetite and not wanting to drink water for fear of throwing up. I'm not sure where else to go with this, but I need help. If any of you have this, can you tell me how you deal with it? I'm tired of having this every day and feeling like I'll throw up.

Stance and Sentiment

If interested send email to
eallaway@cs.columbia.edu

- Sentiment
  - BiLSTM + trained attention[1]
  - Multi-lingual embeddings[2]
  - *Can we use visual information as well?*
- Stance
  - Can help us to identify implicit information
  - Using loaded language to help identify

*In a **dramatic** press conference, Ukraines new security chiefs say Yanukovych ordered the **mass slayings** and the snipers were under his direct leadership.*

arg1 *arg2*

Stance: anti-Russia

Sponsorship Relation

[1] Zhong et al., 2019 (arXiv); [2] Conneau et al. 2018 (ICLR)

Background: Firearm-related deaths in the US

- Violence impacts low-income cities
  - Chicago had **>3,000 shooting victims in 2015**.

- Violence exacerbated by taunting between gang members on social media: the "digital street"

- Identification of those who post about aggression or loss can help community outreach workers

Collaboration with Social Work Faculty: Desmond Patton

Case Study:
Gakirah Barnes @TyquanAssassin

- Recently deceased gang member in Chicago
- 9 killings to her name until she was killed at the age of 17
- 27,000 tweets from December 2011 to April 11, 2014
- ~ 4,200 followers on Twitter

Qualitative Analysis -> Prediction of aggression/loss

Tweet	Label
If We see a opp F[redacted] We Gne smoke em 😈	Aggression (Threat)
My bro [redacted] thirsty he jus wana 👏 sum 💩 🔫 😈 💯	Aggression (Insult)
Damn juss peeped shorty on tha news out here @USER ..smh.. crazyy.. #RIPShorty	Loss

Research Directions

- Have developed a CNN classifier for aggression/loss using context
- Extend to use information about triggering events
- Look at what happens after a loss
    - How do online interactions differ between people who adapt to loss and those who don't?

Michael Collins

- Question answering
  - Release of natural questions corpus through Google
- Computational models of the brain
  - The use of assemblies in language processing
- Variational encoders for information extraction
  - train latent variable models based on neural networks using differentiable approximations to maximum likelihood

COLUMBIA UNIVERSITY
Data Science Institute

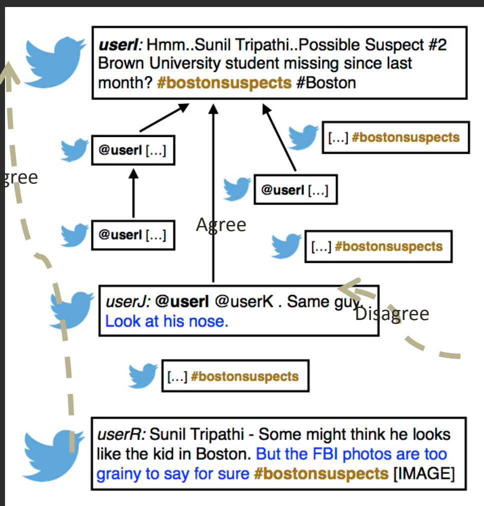
Computational models for understanding language in context

Smaranda Muresan

SMARA@COLUMBIA.EDU

COMPUTATINAL MODELS FOR UNDERSTANDING LANGUAGE IN CONTEXT

Social Context



Models of Argument

Persuasion

Fact-checking

Students' scientific writing

Understanding collective opinions

Rumor Detection

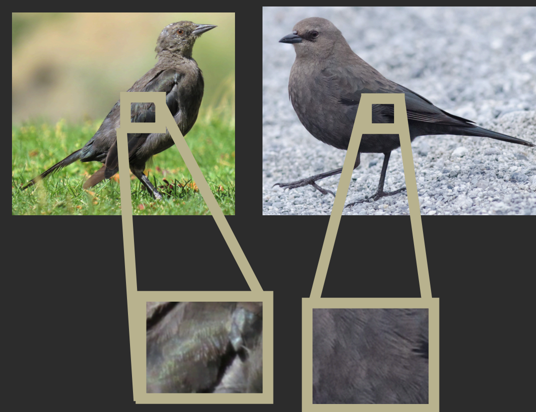
Irony & Sarcasm (sentiment/beliefs)

Abusive Language

Public Health (Suicide Risk Assessment,

Digital Epidemiology)

Visual Context



The juvenile is ***lighter brown*** *than the female*

Learning word semantics by grounding them in images

Multilingual Context

Learning word semantics in a multilingual context – focus on low resource languages

Unsupervised morphological analysis for low resource languages

Smaranda Muresan
(smara@columbia.edu)

Julia Hirschberg: Speech and Language

- Deception and trust
  - Multimodal cues, crowdsourcing data
- Mental illness in social media
- Charisma in speech
- Multimodal detection of humor
- Prosody
  - Learning from linguistic cues
  - Using prosody to improve text to speech

Industry and internships

- Amazon
- Google
- Facebook
- Start-ups
- Ivy Elkins: ivy@cs.columbia.edu, CS Career Placement

Final Exam Guidelines

- Know the circumstances under which different architectures might be used
  - No need to memorize architectures used for different tasks/ in different papers
  - No derivations on the exam
  - Given a task, be able to design an architecture
- Focus on topics since the midterm
- For prior to the midterm topics, go back to the midterm itself

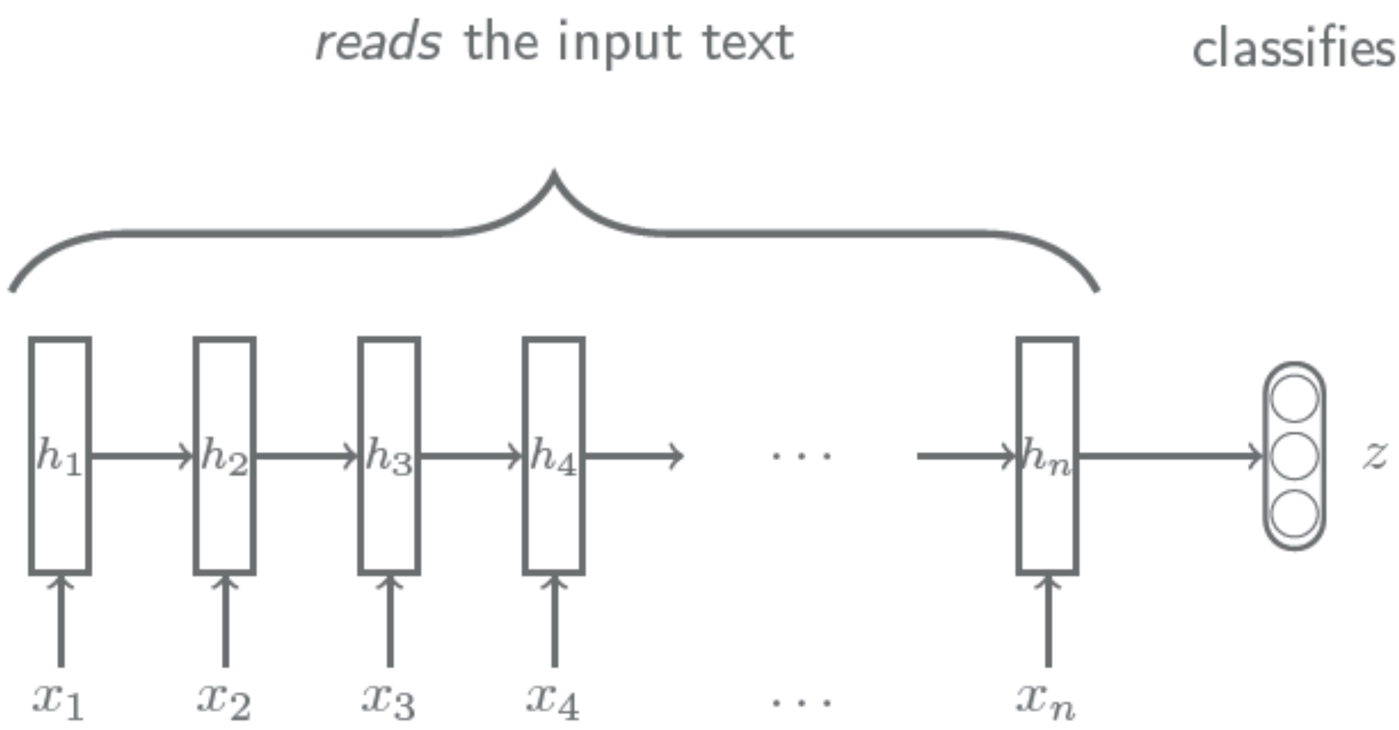
Attention

- What is the problem attention is trying to address?
  - When decoding and generating a word of output, may want to focus on specific words of the input
  - When generating word 1 of the output look at word 1 of the input

RNN classifier

Input words $x_1, \ldots, x_n$

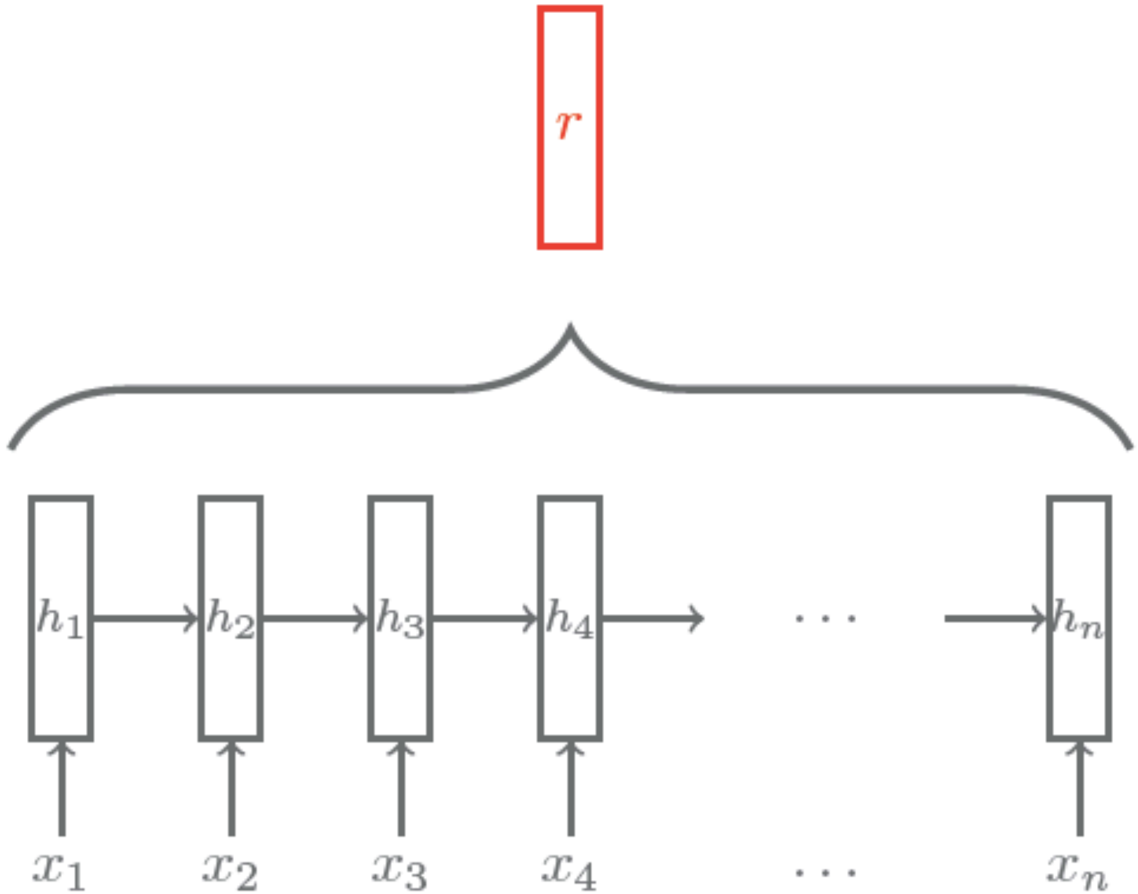
Output category label z



RNN encoder

Input words $x_1, \ldots, x_n$

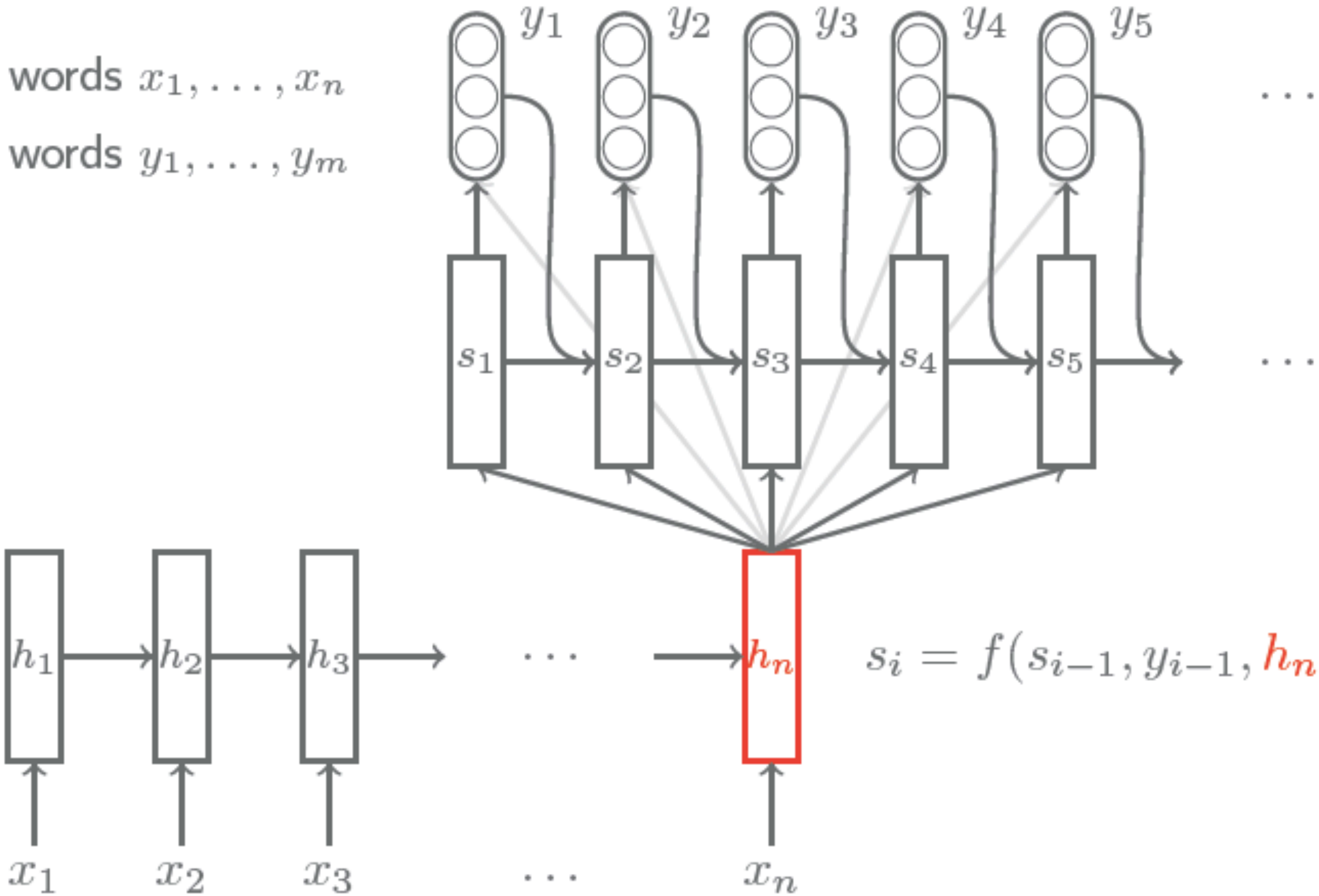
Output representation r



Sequence-to-sequence learning

Input words $x_1, \ldots, x_n$

Output words $y_1, \ldots, y_m$

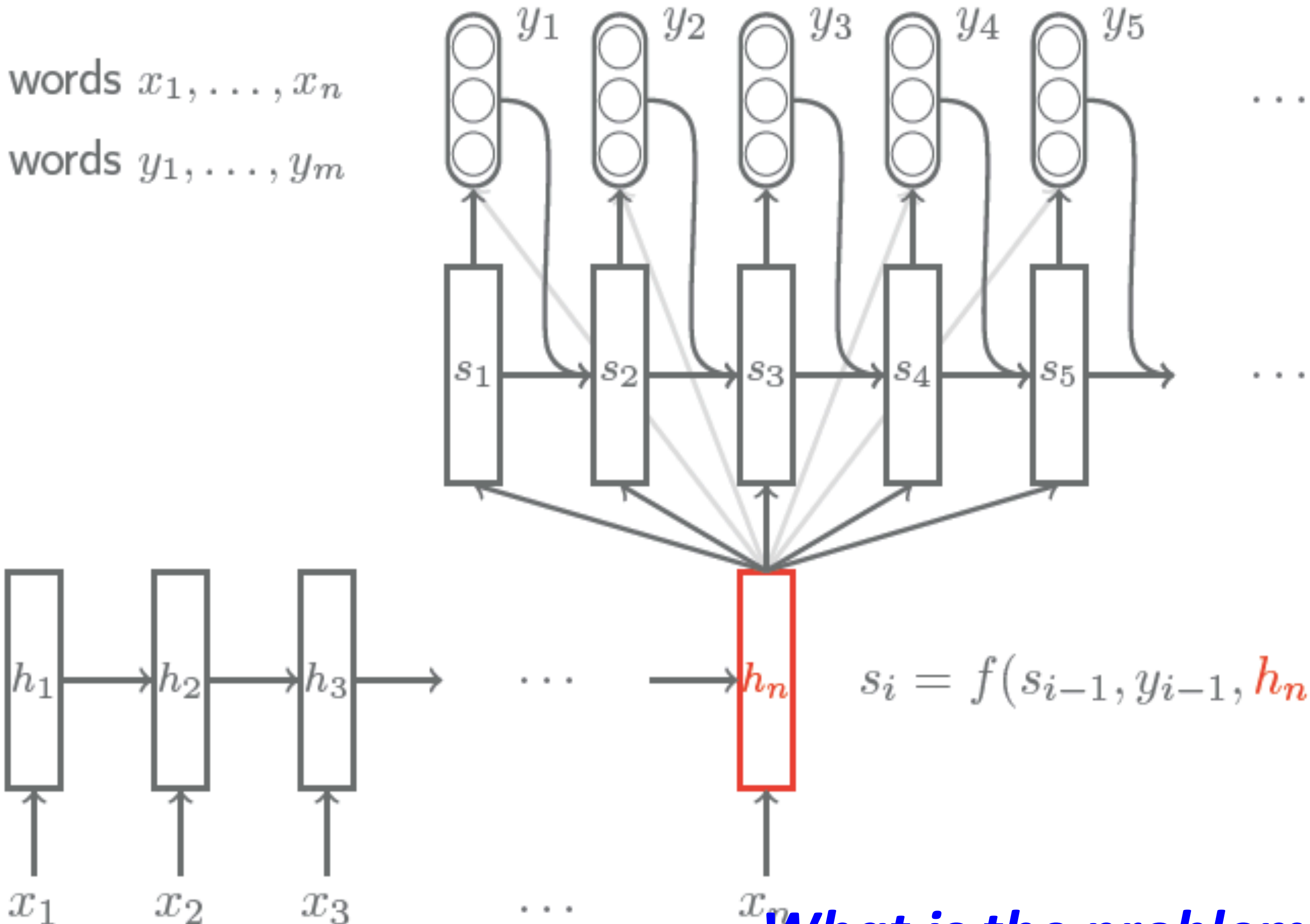


$s_i = f(s_{i-1}, y_{i-1}, h_n)$

Sequence-to-sequence learning

Input words $x_1, \ldots, x_n$

Output words $y_1, \ldots, y_m$



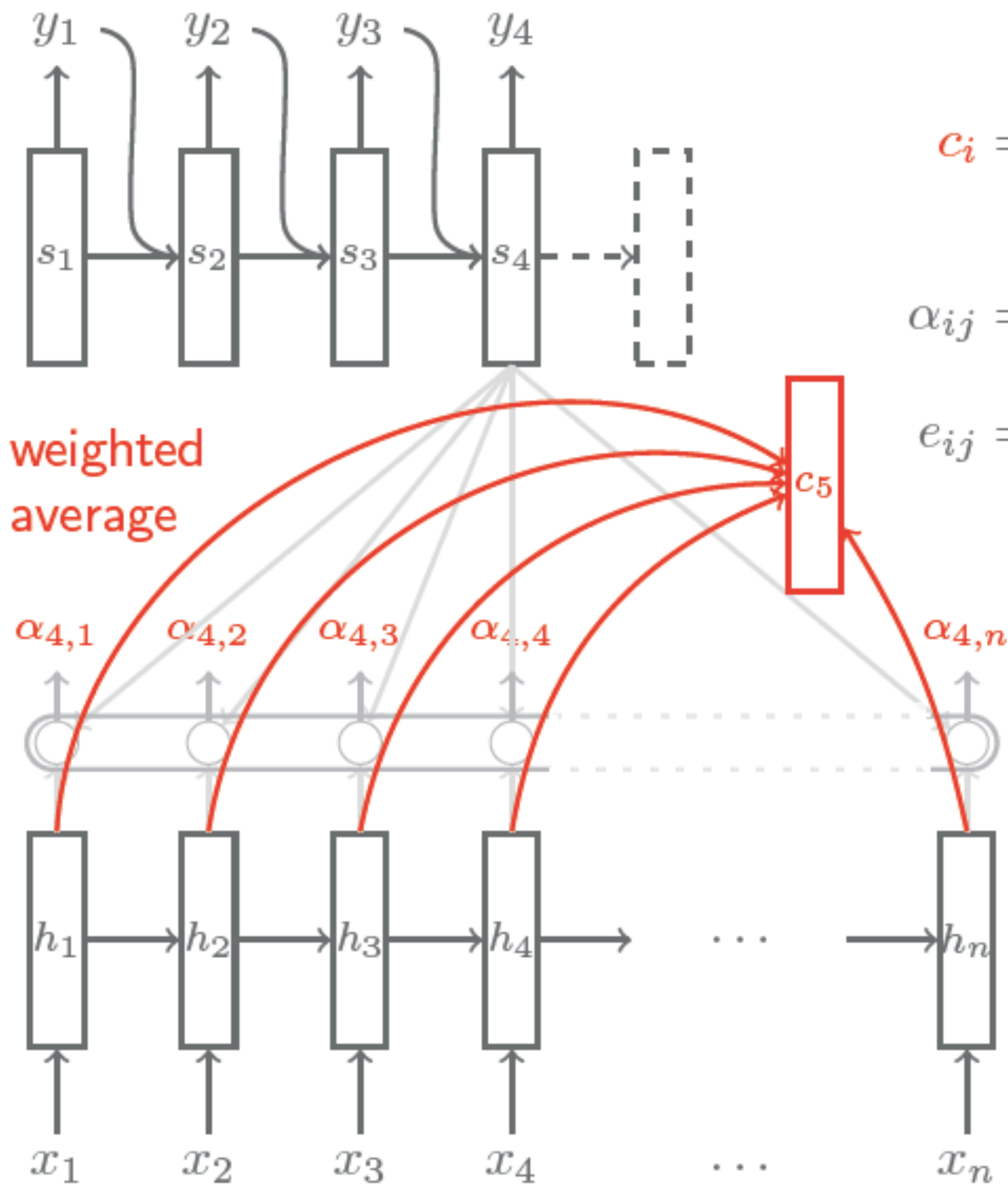
$s_i = f(s_{i-1}, y_{i-1}, h_n)$

What is the problem?

Attention mechanism

- Dynamic context vector that changes with each decoding step
- Weighted average over all encoder hidden states
- Weights (“atttention”) conditioned on current decoder hidden state
- Allows gradients to flow from errors in current decoding state directly to relevant encoder states

Attention-based translation

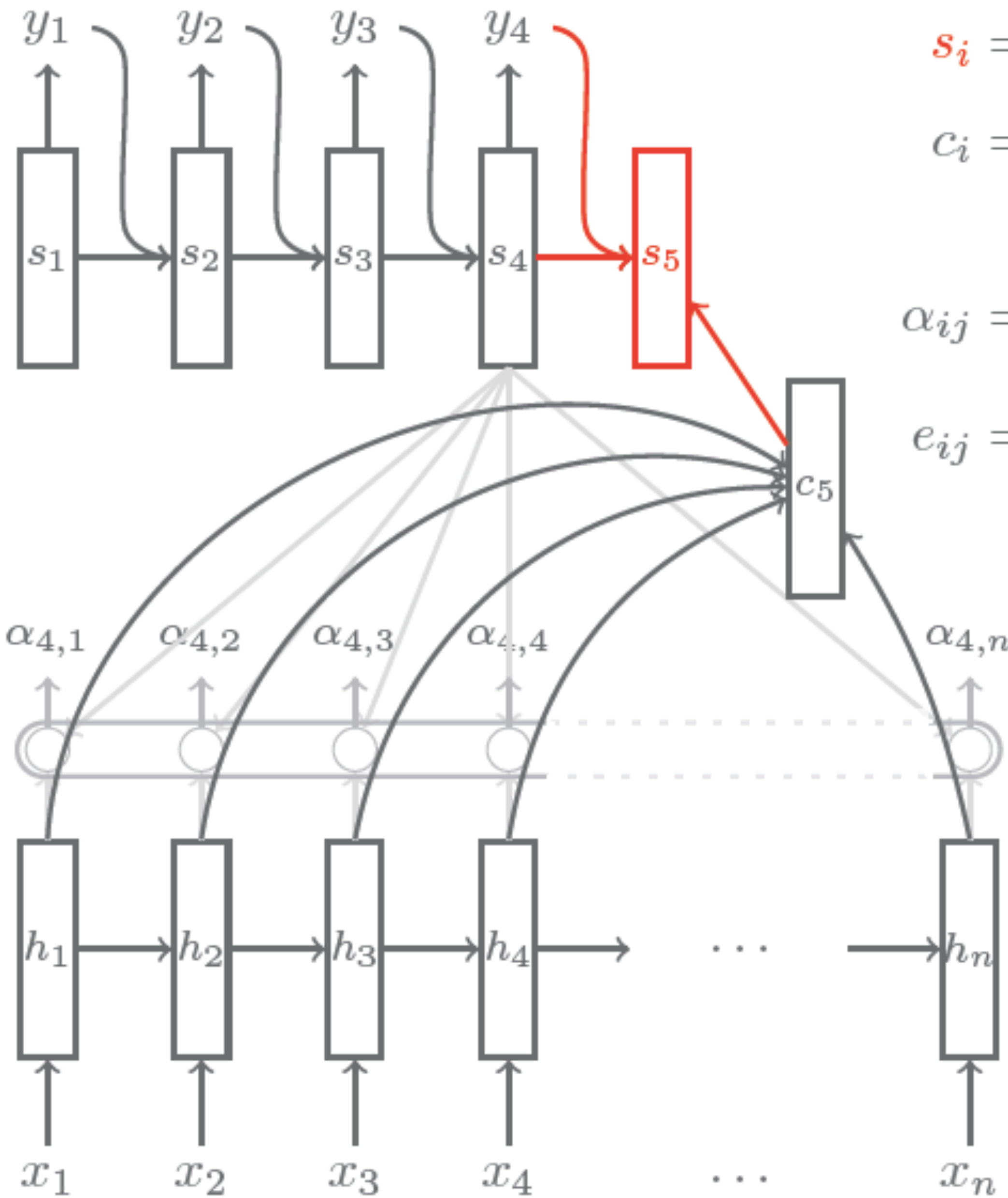


$$c_i = \sum_j \alpha_{ij} h_j$$

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_k \exp(e_{ik})}$$

$$e_{ij} = a(s_{i-1}, h_j)$$

Attention-based translation



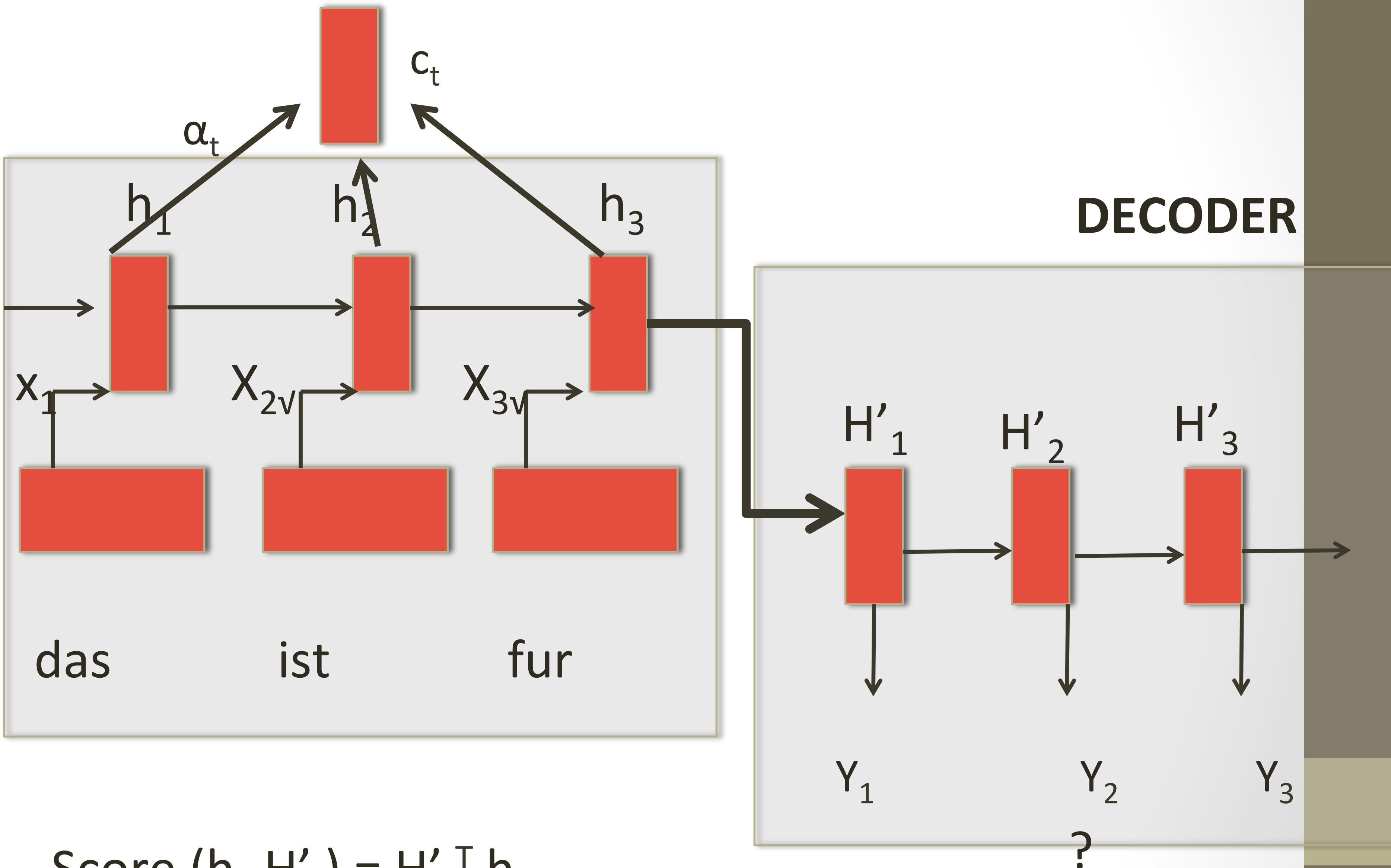
$$s_i = f(s_{i-1}, y_{i-1}, c_i)$$

$$c_i = \sum_j \alpha_{ij} h_j$$

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_k \exp(e_{ik})}$$

$$e_{ij} = a(s_{i-1}, h_j)$$

How do you score it?

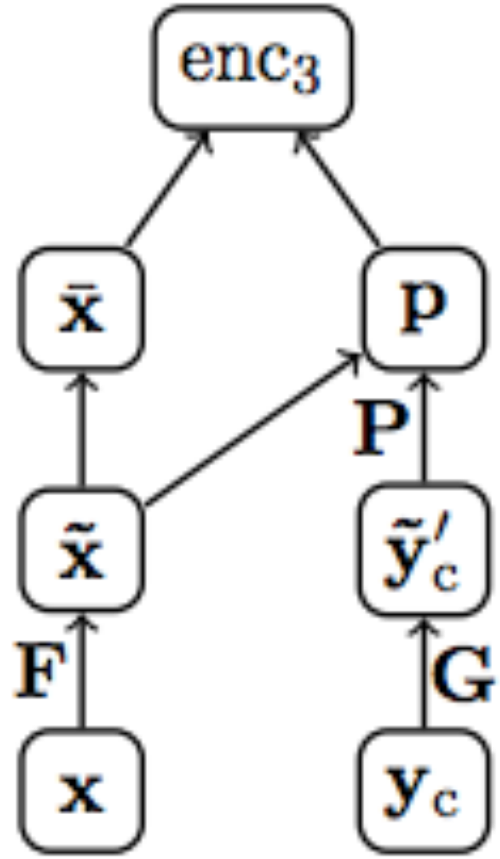


Score $(h_s, H'_t) = H'^{T}_{t}\, h_s$

or $= H'^{T}_{t}\, W_\alpha\, h_s$ (Luong et al 2015)

Attention based encoder

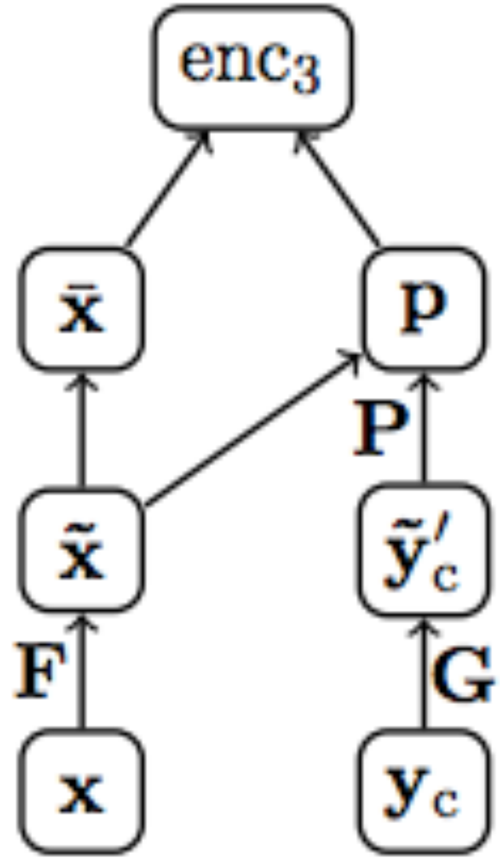
$$
\begin{aligned}
\mathrm{enc}_3(\mathbf{x}, \mathbf{y}_c) &= \mathbf{p}^\top \bar{\mathbf{x}}, \\
\mathbf{p} &\propto \exp(\tilde{\mathbf{x}} \mathbf{P} \tilde{\mathbf{y}}_c'), \\
\tilde{\mathbf{x}} &= [\mathbf{F}\mathbf{x}_1, \ldots, \mathbf{F}\mathbf{x}_M], \\
\tilde{\mathbf{y}}_c' &= [\mathbf{G}\mathbf{y}_{i-C+1}, \ldots, \mathbf{G}\mathbf{y}_i], \\
\forall i \quad \bar{\mathbf{x}}_i &= \sum_{q=i-Q}^{i+Q} \tilde{\mathbf{x}}_i / Q.
\end{aligned}
$$



Attention based encoder

X = input
Y_c = what we have generated so far. C is limited in this case to the previous c words.
P = weight matrix parameter.
F is the embedding matrix of the input x
G is an embedding matrix for output context

Attention is the dot product between x and y_c mediated by P. A learned soft alignment between input and the summary



What are three problems for which we would use a sequence to sequence decoder?

Summarization

- Extractive summarization
  - Select sentences from document to appear in summary
  - Classifier: for each sentence -> 1,0

- Abstractive summarization
  - Rewrite the input sentences
  - Compression
  - Paraphrasing
  - Fusion

Neural Summarization

- Dataset is necessary
  - Headline generation: what is the dataset?
  - Single document news summarization: what is the dataset?

Headline generation

- Seq2seq model
- Encode the input sentence
- Generate the next word y, looking at the context of the previous c generated words
  - What was the vocabulary from which y could be drawn?
  - Was the model abstractive or extractive?

The architecture

- Encoder – experimented with three models
  - Bag of words model
  - Convolutional encoder
  - Attention based model

- Decoder
  - Neural language model – any neural language model. Don't worry about the specific formulas used and the reference to Banko.

The architecture

- Encoder – experimented with three models
  - Bag of words model
  - Convolutional encoder
  - *Attention based model*

- Decoder
  - Neural language model – any neural language model. Don't worry about the specific formulas used and the reference to Banko.

Additions

- Features which encourage the decoder to choose vocabulary from the input sentence

- Beam search decoder

Without beam search, what gets generated at each step?

Why does beam search help?

At any point in generating the sentence, with a beam of k, how many sequences of words is the system examining?

- K*k
- K
- k+k
- K to the k

Word sense disambiguation

- POS tagging
  - I went to the race
  - I like to race down the block.
- Word sense disambiguation
  - I sat on the bank and enjoyed the sound of the water flowing by.
  - I went to the bank to open a checking account.

Word sense disambiguation

- POS tagging
  - I went to the race
    noun
  - I like to race down the block.
    Verb
- Word sense disambiguation
  - I sat on the bank and enjoyed the sound

    of the water flowing by.
  - I went to the bank to open a checking account.

Wordnet Synsets

- http://wordnetweb.princeton.edu/

Word sense disambiguation

- POS tagging
  - I went to the race
    noun
  - I like to race down the block.
    Verb
- Word sense disambiguation
  - I sat on the bank and enjoyed the sound
    bank#1
    of the water flowing by.
  - I went to the bank to open a checking account.
    Bank#2

- "In our house, everybody has a career and none of them includes washing *dishes*," he says.
- In her tiny kitchen at home, Ms. Chen works efficiently, stir-frying several simple *dishes,* including braised pig's ears and chcken livers with green peppers.
- Post quick and convenient *dishes* to fix when your in a hurry.
- Japanese cuisine offers a great variety of *dishes* and regional specialties

- We need more good teachers – right now, there are only a half a dozen who can play the free *bass* with ease.
- Though still a far cry from the lake's record 52-pound *bass* of a decade ago, "you could fillet these fish again, and that made people very, very happy." Mr. Paulson says.
- An electric guitar and *bass* player stand off to one side, not really part of the scene, just as a sort of nod to gringo expectations again.
- Lowe caught his *bass* while fishing with pro Bill Lee of Killeen, Texas, who is currently in 144th place with two bass weighing 2-09.

Collocational

- Position-specific information about the words in the window
- guitar and bass player stand
  - [guitar, NN, and, CC, player, NN, stand, VB]
  - Word_{n-2}, POS_{n-2}, word_{n-1}, POS_{n-1}, Word_{n+1} POS_{n+1}…
  - In other words, a vector consisting of
  - [position n word, position n part-of-speech…]

Bag-of-words

- Information about the words that occur within the window.
- First derive a set of terms to place in the vector.
- Then note how often each of those terms occurs in a given window.

Co-Occurrence Example

- Assume we've settled on a possible vocabulary of 12 words that includes guitar and player but not and and stand

- guitar and bass player stand
  - [0,0,0,1,0,0,0,0,0,1,0,0]
  - Which are the counts of words predefined as e.g.,
  - [fish,fishing,viol, guitar, double,cello…

Decision Lists: another popular method

- A case statement….

Rule		Sense
fish within window	$\Rightarrow$	**bass**^{1}
striped bass	$\Rightarrow$	**bass**^{1}
guitar within window	$\Rightarrow$	**bass**^{2}
bass player	$\Rightarrow$	**bass**^{2}
piano within window	$\Rightarrow$	**bass**^{2}
tenor within window	$\Rightarrow$	**bass**^{2}
sea bass	$\Rightarrow$	**bass**^{1}
play/V bass	$\Rightarrow$	**bass**^{2}
river within window	$\Rightarrow$	**bass**^{1}
violin within window	$\Rightarrow$	**bass**^{2}
salmon within window	$\Rightarrow$	**bass**^{1}
on bass	$\Rightarrow$	**bass**^{2}
bass are	$\Rightarrow$	**bass**^{1}

Learning Decision Lists

- Restrict the lists to rules that test a single feature (1-decisionlist rules)
- Evaluate each possible test and rank them based on how well they work.
- Glue the top-N tests together and call that your decision list.

Yarowsky

- On a binary (homonymy) distinction used the following metric to rank the tests

$$\frac{P(\mathrm{Sense}_1 \mid Feature)}{P(\mathrm{Sense}_2 \mid Feature)}$$

- This gives about 95% on this test…

How would we compute P(sense1|feature)

BERT and sense

Figure 4: Embeddings for the word "die" in different contexts, visualized with UMAP. Sample points are annotated with corresponding sentences. Overall annotations (blue text) are added as a guide.

Figure: From Coenen et al. [2019]

Good luck on the exam!

It was great having you in class!