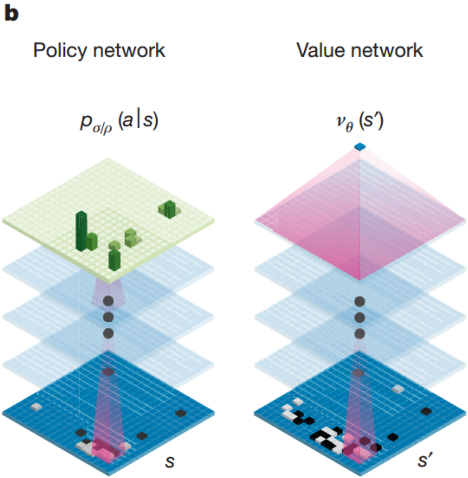
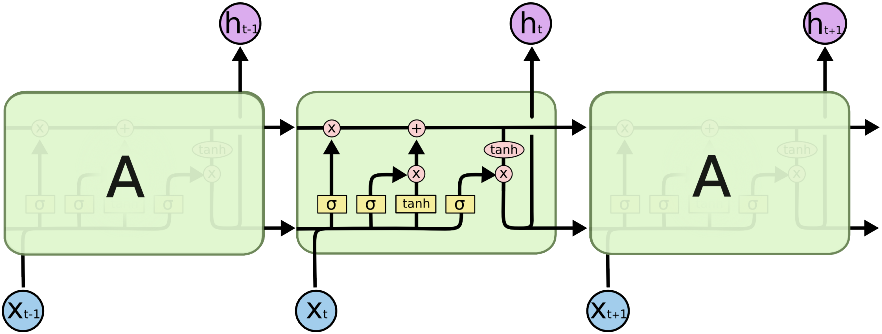
A Tour of Natural Language Applications

CS 287

AlphaGo



Review: LSTM (Hochreiter and Schmidhuber, 1997)



Review: Highway Network (Srivastava et al., 2015)

- Now add a combination at each dimension.
- $\odot$ is point-wise multiplication.

$$
\begin{aligned}
NN_{highway}(\mathbf{x}) &= (1 - \mathbf{t}) \odot \tilde{\mathbf{h}} + \mathbf{t} \odot \mathbf{x} \\
\tilde{\mathbf{h}} &= \mathrm{ReLU}(\mathbf{x}\mathbf{W}^1 + \mathbf{b}^1) \\
\mathbf{t} &= \sigma(\mathbf{x}\mathbf{W}^t + \mathbf{b}^t) \\
\mathbf{W}^t, \mathbf{W}^1 &\in \mathbb{R}^{d_{\mathrm{hid}} \times d_{\mathrm{hid}}} \\
\mathbf{b}^t, \mathbf{b}^1 &\in \mathbb{R}^{1 \times d_{\mathrm{hid}}}
\end{aligned}
$$

- $\tilde{\mathbf{h}}$; *transform* (e.g. standard MLP layer)
- $\mathbf{t}$; *carry* (dimension-specific dynamic skipping)

Review: Gated Recurrent Unit (GRU) (Cho et al 2014)

$$
\begin{aligned}
R(\mathbf{s}_{i-1}, \mathbf{x}_i) &= (1 - \mathbf{t}) \odot \tilde{\mathbf{h}} + \mathbf{t} \odot \mathbf{s}_{i-1} \\
\tilde{\mathbf{h}} &= \tanh(\mathbf{x}\mathbf{W}^{x} + (\mathbf{r} \odot \mathbf{s}_{i-1})\mathbf{W}^{s} + \mathbf{b}) \\
\mathbf{r} &= \sigma(\mathbf{x}\mathbf{W}^{xr} + \mathbf{s}_{i-1}\mathbf{W}^{sr} + \mathbf{b}^{r}) \\
\mathbf{t} &= \sigma(\mathbf{x}\mathbf{W}^{xt} + \mathbf{s}_{i-1}\mathbf{W}^{st} + \mathbf{b}^{t}) \\
\mathbf{W}^{xt}, \mathbf{W}^{xr}, \mathbf{W}^{x} &\in \mathbb{R}^{d_{\mathrm{in}} \times d_{\mathrm{hid}}} \\
\mathbf{W}^{st}, \mathbf{W}^{sr}, \mathbf{W}^{s} &\in \mathbb{R}^{d_{\mathrm{hid}} \times d_{\mathrm{hid}}} \\
\mathbf{b}^{t}, \mathbf{b} &\in \mathbb{R}^{1 \times d_{\mathrm{hid}}}
\end{aligned}
$$

- $\mathbf{t}$; dynamic skip-connections
- $\mathbf{r}$; reset gating
- $\mathbf{s}$; hidden state

Review: Long Short-Term Memory

Finally, another output gate is applied to $\mathbf{h}$

$$
\begin{aligned}
R(\mathbf{s}_{i-1}, \mathbf{x}_i) &= [\mathbf{c}_i, \mathbf{h}_i] \\
\mathbf{c}_i &= \mathbf{j} \odot \mathbf{i} + \mathbf{f} \odot \mathbf{c}_{i-1} \\
\mathbf{h}_i &= \tanh(\mathbf{c}_i) \odot \mathbf{o} \\
\mathbf{i} &= \tanh(\mathbf{x}\mathbf{W}^{xi} + \mathbf{h}_{i-1}\mathbf{W}^{hi} + \mathbf{b}^i) \\
\mathbf{j} &= \sigma(\mathbf{x}\mathbf{W}^{xj} + \mathbf{h}_{i-1}\mathbf{W}^{hj} + \mathbf{b}^j) \\
\mathbf{f} &= \sigma(\mathbf{x}\mathbf{W}^{xf} + \mathbf{h}_{i-1}\mathbf{W}^{hf} + \mathbf{b}^f) \\
\mathbf{o} &= \sigma(\mathbf{x}\mathbf{W}^{xo} + \mathbf{h}_{i-1}\mathbf{W}^{ho} + \mathbf{b}^o)
\end{aligned}
$$

Review: The Promise of RNNs

- Learn the long-range interactions of language from data.
- Example: classify true and false statements:
- **Example:**

  `Eliot house is the coolest`

  `Mather does not look like a prison.`

- Works well if you control for *exploding* or *vanishing* gradients.

Review: The Promise of RNNs

- Learn the long-range interactions of language from data.
- Example: classify true and false statements:
- **Example:**

  `Eliot house is the coolest`

  `Mather does not look like a prison.`
- Works well if you control for *exploding* or *vanishing* gradients.

Quiz

Last class we discussed the issue of the exploding gradient in RNNs. There are two practical heuristics for this problem:

- gradient clipping, i.e. bounding any gradient by a maximum value
- gradient normalization, i.e. renormalizing the the RNN gradients if they are above a fixed norm value.

Describe the positive and negatives of these approaches. How would you implement these in a system like Torch?

Today's Lecture

- High-level picture of select natural language challenges.
- Caveat: Not a representative sample of NLP.
- Meant as a final project shopping list.

Recommendations

- Sometimes datasets are private, speak to us about getting them.
- The state-of-the-art on these problems is changing quite quickly.
- Getting Started:
  - Make sure you understand the problem and the metric.
  - Read papers on the topic.
  - Experiment first with count-based or linear models
  - Reimplement other systems

Topics

High-level areas:

- Information Extraction
  - Named-Entity Recognition
  - Semantic-Role Labeling
  - Entity Linking
- Question Answering
  - Knowledge Graph (factoid)
  - Knowledge Processing (non-factoid)
  - Comprehension (non-factoid)
- Document Understanding
  - Discourse
  - Summarization
  - Coreference

Other NLP Areas

There are many other areas of NLP:

- Speech
- Syntax
- Machine Translation

Require more details [future lectures]

Other possible topics (bring-your-own domain knowledge)

- Music Processing
- Vision
- Game Playing

Contents

Abraham Lincoln

16th U.S. President

Abraham Lincoln was the 16th President of the United States, serving from March 1861 until his assassination in April 1865. Wikipedia

Born: February 12, 1809, Hodgenville, KY

Height: 6′ 4″

Spouse: Mary Todd Lincoln (m. 1842–1865)

Party: National Union Party

Children: William Wallace Lincoln, Robert Todd Lincoln, Tad Lincoln, Edward Baker Lincoln

Quotes

View 7+ more

Nearly all men can stand adversity, but if you want to test a man's character, give him power.

Whatever you are, be a good one.

Always bear in mind that your own resolution to succeed is more important than any other.

People also search for

View 15+ more


George Washington


William Wallace Lincoln
Son


John Wilkes Booth


John F. Kennedy


Mary Todd Lincoln
Spouse

Information Extraction

Goal: Map text to structured information.

Applications:

- Knowledge-base construction
- Quantitative research from freetext
- Identifying relationships

Terminology (ACE 2008 task)

Entity the underlying semantic actor.

- e.g. persons, countries, organizations, teams

Relation semantic relations between entities

- e.g. part of, located in, member of, works for

Event a semantic occurrence involving entities

- e.g. marriage, attack, takeover, visit

Mention a reference to an entity, relation, or event in text

- e.g. China, the country, it, the People's Republic of China

Problem 1: Named-Entity Recognition

Goal	Identify explicitly named entities in text.
Input	Sentence to be tagged.
Output	Mentions of identified entities and their type.

Named-Entity Recognition

Input: `U.N. official Ekeus heads for Baghdad.`

Gold: `[ORG U.N. ] official [PER Ekeus ] heads for [LOC Baghdad ] .`

Standard Dataformat (CoNLL 2003)

```
U.N.        NNP  I-NP  I-ORG
official    NN   I-NP  O
Ekeus       NNP  I-NP  I-PER
heads       VBZ  I-VP  O
for         IN   I-PP  O
Baghdad     NNP  I-NP  I-LOC
.           .    O     O
```

BIO Tagging

B-TYPE	Stop current mention and begin new mention
I-TYPE	Continue adding to current mention
O	Not part of a mention.

Example:

`[PER George Bush ] [LOC U.S. ] president is traveling to [LOC Baghdad ] .`

BIO Tagging

B-TYPE Stop current mention and begin new mention

I-TYPE Continue adding to current mention

O Not part of a mention.

Example:

`[PER George Bush ] [LOC U.S. ] president is traveling to [LOC Baghdad ] .`

Example Tag Set

- Loc
- Org
- Person
- Misc

- How big is $\mathcal{C}$ for this problem?

Problem 2: Entity Linking

Goal Identify explicit entities and link to a standard central database.

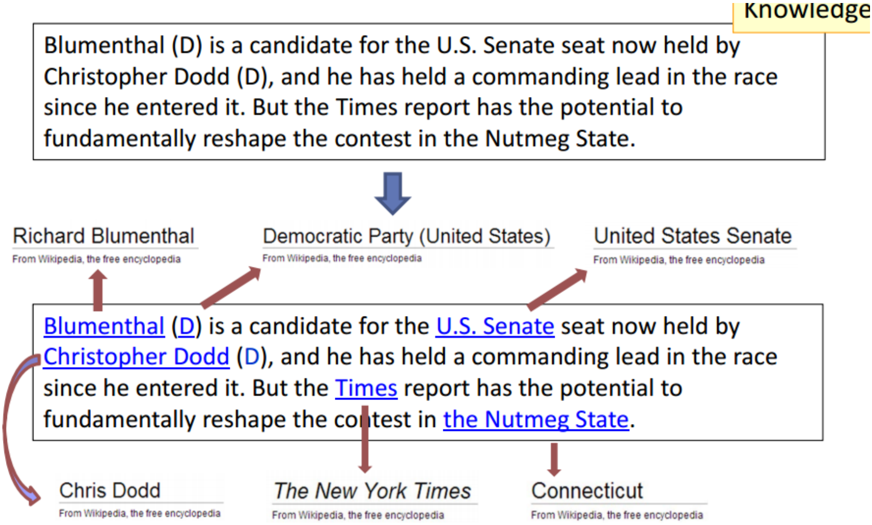
Input Sentence or Document

Output Mentions and pointer to a central source.

Entity Linking: Wikification (TAC 2014)

- Uses Wikipedia as canonical source.
- Goal is to link mentions to the correct wikipedia page.
- Classification is over a much larger space.

Wikification (Roth et al, 2014)



Medical Term Linkage (Semeval, 2015)

- Uses canonical medical thesaurus (SNOMED/UMLS)
- Goal is to link medical notes to thesaurus
- Subset of UMLS has order of 100,000 terms.

Medical Term Extraction

1. The rhythm appears to be [atrial fibrillation].
2. The [left atrium] is moderately [dilated] .
3. 53 year old man s/p [fall from ladder].

Linked terms:

- atrial fibrillation - C0004238; UMLS preferred term “atrial fibrillation”
- left atrium...dilated - C0344720; UMLS preferred term “left atrial dilatation”
- fall from ladder - C0337212; UMLS preferred term is “accidental fall from ladder”

Problem 3: Semantic Role Labeling

Goal	Mark the semantic roles of sentence elements
Input	Sentence
Output	Identify verb, its type, its arguments, and their types

Language Applications: Semantic Role Labeling

He would n't accept anything of value from those he was writing about

[A0 He] [AM-MOD would] [AM-NEG n't] [V accept] [A1 anything of value] from [A2 those he was writing about]

- V: verb (accept)
- A0: acceptor
- A1: thing accepted
- A2: accepted-from
- A3:attribute
- AM-MOD: modal
- AM-NEG: negation

SRL Requires Long-Range Interactions

Collobert approach:

- First given a verb w_i e.g. `accept`.
- Then consider a word w_j e.g. `n't`
- For a word w_k features are

$$v(w_k), v_2(cap(w_k)), v_3(i-k), v_4(j-k)$$

- Convolution over sentence is used to predict role.
- $O(n \times |verbs|)$ convolutions per sentence

Contents

Question Answering

- Big area, lots of different problems.
- Generally specific to the type of question and style of input.

  `what high school did president bill clinton attend?`

  versus

  `how many rivers in texas are longer than the red?`

- Various methods for solving:
  - Learn to map text to explicit logical query
  - Treat logical query as latent term
  - Attempt to directly map to answer

Factoid Question Answering

Goal	Map question to an answer from a knowledge base
Input	Question and knowledge-base source
Output	Answer

WebQuestions (Berant)

Questions:

- what high school did president bill clinton attend?
- what form of government does russia have today?
- what movies does taylor lautner play in?

Answers:

- Hot Springs High School
  `http://www.freebase.com/view/en/bill_clinton`
- Constitutional republic
  `http://www.freebase.com/view/en/russia`
- Eclipse, Valentine's Day, The Twilight Saga: Breaking Dawn - Part 1, New Moon
  `http://www.freebase.com/view/en/taylor_lautner`

Freetext Knowledge Sources

Goal	Map question to an answer described in text.
Input	Question and Text Source (textbooks)
Output	Answer

Biological Processes (Berant et al, 2013)

"…**Water is split**, providing a source of electrons and protons (hydrogen ions, H^+) and giving off O_2 as a by-product. **Light absorbed** by chlorophyll drives a **transfer of the electrons and hydrogen ions** from water to an acceptor called $NADP^+$ …"

Q What can the splitting of water lead to?

a Light absorption

b Transfer of ions

Non-Factoid

Goal	Map question to an answer described in casual text.
Input	Question, multiple choices, and text source (narratives)
Output	Answer

MCTest (Richardson et al, 2013)

James the Turtle was always getting in trouble. Sometimes he'd reach into the freezer and empty out all the food. Other times he'd sled on the deck and get a splinter. His aunt Jane tried as hard as she could to keep him out of trouble, but he was sneaky and got into lots of trouble behind her back.

One day, James thought he would go into town and see what kind of trouble he could get into. He went to the grocery store and pulled all the pudding off the shelves and ate two jars. Then he walked to the fast food restaurant and ordered 15 bags of fries. He didn't pay, and instead headed home.

His aunt was waiting for him in his room. She told James that she loved him, but he would have to start acting like a well-behaved turtle.After about a month, and after getting into lots of trouble, James finally made up his mind to be a better turtle.

MCTest: Questions

- What is the name of the trouble making turtle?
  1. Fries
  2. Pudding
  3. James
  4. Jane
- What did James pull off of the shelves in the grocery store?
  1. pudding
  2. fries
  3. food
  4. splinters
- Where did James go after he went to the grocery store?
  1. his deck
  2. his freezer
  3. a fast food restaurant
  4. his room

bAbI Tasks (Weston et al, preprint)

```
1 Mary moved to the bathroom.
2 John went to the hallway.
3 Where is Mary?        bathroom        1
4 Daniel went back to the hallway.
5 Sandra moved to the garden.
6 Where is Daniel?      hallway 4
7 John moved to the office.
8 Sandra journeyed to the bathroom.
9 Where is Daniel?      hallway 4
10 Mary moved to the hallway.
11 Daniel travelled to the office.
12 Where is Daniel?     office  11
```

bAbI

- Synthetic tasks to test MemNN type architectures
- Running reading comprehension from a synthetic domain.
- 20 different tasks ranging in complexity

Simple Model (Non-Questions)

- 1 Mary moved to the bathroom.

I Read in words

G Construct and append CBoW sentence representation

$$\mathbf{s}_j = G(\mathbf{x}^0) = \sum_{i=1}^{n} \mathbf{W}\mathbf{x}_i^0$$

O,R Nothing

Simple Model (Non-Questions)

- 3 Where is Mary?

I Read in words

G Construct CBoW sentence representation

$$\mathbf{x} = G(\mathbf{x}^0) = \sum_{i=1}^{n} \mathbf{W}\mathbf{x}_i^0$$

O Find best sentence match and apply hard attention

$$j^* = \arg\max_j s(\mathbf{x}, \mathbf{s}_j)$$

R Respond by classification over possible outputs

$$\hat{\mathbf{y}} = \mathrm{softmax}(NN_{MLP}([\mathbf{x}; \mathbf{s}_{j^*}]))$$

bAbI

1. 1 Mary moved to the bathroom.
2. John went to the hallway.
3. Where is Mary?

- Use CBoW to match and to help predict answer (bathroom)

Multiple Hops

Task 18: Size Reasoning

1. The football fits in the suitcase.
2. The suitcase fits in the cupboard.
3. The box is smaller than the football.
4. Will the box fit in the suitcase? A:yes
5. Will the cupboard fit in the box? A:no

Multi-Hop Model (Non-Questions)

- 3 Will the box fit in the suitcase?

I Read in words

G Construct CBoW sentence representation

$$\mathbf{x} = G(\mathbf{x}^0) = \sum_{i=1}^{n} \mathbf{W}\mathbf{x}_i^0$$

O Find best sentence match and apply hard attention

$$j^* = \arg\max_j s(\mathbf{x}, \mathbf{s}_j)$$

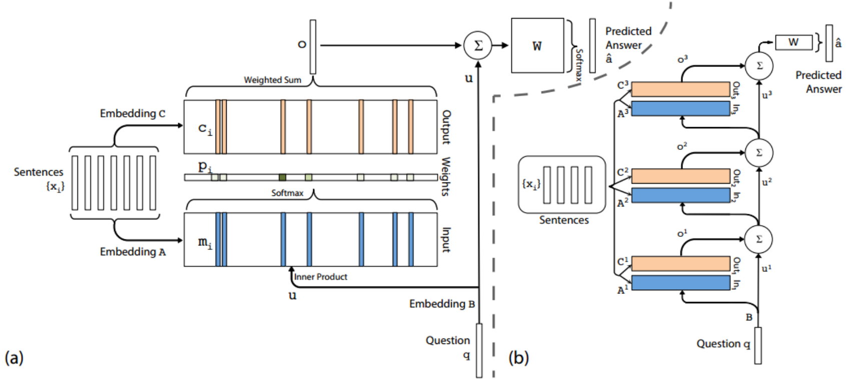
$$k^* = \arg\max_k s(\mathbf{x}, \mathbf{s}_{j^*}, \mathbf{s}_k)$$

R Respond by classification over possible outputs

$$\hat{\mathbf{y}} = \mathrm{softmax}(NN_{MLP}([\mathbf{x}; \mathbf{s}_{j^*}; \mathbf{s}_{k^*}]))$$

How do we learn the hard-attention?

- Strong supervision
  - Hard-attention is trained in O step
  - No explicit backprop through decisions
- Weak supervision (End-to-End)
  - Soft-attention with no intermediary answers given
  - Use softmax to combine possible memories.



(a)
Sentences
{x_i}
Embedding C
Embedding A
c_i
p_i
m_i
Weighted Sum
Softmax
Output
Weights
Input
Inner Product
u
o
Σ
W
Softmax
Predicted
Answer
â
Embedding B
Question
q
(b)
{x_i}
Sentences
C^3
A^3
C^2
A^2
C^1
A^1
Out_3
In_3
Out_2
In_2
Out_1
In_1
o^3
o^2
o^1
u^3
u^2
u^1
B
Question q
W
â
Predicted
Answer

Contents

Discourse Parsing

Goal Determine the discourse relation between adjacent sentences

Input Two sentences

Output One of a predefined set of relationship, e.g. compare, contrast expand.

Implicit Discourse Connectives (PDTB)

Questions:

- Financial planners often urge investors to diversify and to hold a smattering of international securities. And many emerging markets have outpaced more mature markets, such as the U.S. and Japan.
- Country funds offer an easy way to get a taste of foreign stocks without the hard research of seeking out individual companies. But it doesnt take much to get burned.
- But it doesnt take much to get burned. FOR EXAMPLE Political and currency gyrations can whipsaw the funds.

Answers:

- Expansion.Conjunction
- Comparison.Contrast
- Expansion.Restatement.Specification

Summarization

Goal	Produce a shorter version of the input.
Input	Document
Output	Extracted sentences representing the document.

Document Summarization (DUC 2003)

Tension between Turkey and Syria has risen to the point where the top Turkish military commander says the two hostile neighbors have reached "a state of undeclared war."

"We are trying to be patient," said the commander, Gen. Huseyin Kivrikoglu, "but that has a limit."

Syria has reacted angrily to Turkey's blossoming friendship with Israel.

It also makes territorial claims against Turkey and accuses Turkey of unfairly diverting water from rivers that flow through both countries.

For its part, Turkey is complaining ever more loudly about Syria's support for Kurdish insurgents in Turkey. The insurgents are said to have bases in Syria, and their leader reportedly lives in Damascus.

Turkey and Syria are moving troops and equipment to border positions, according to news reports, but no outbreak of fighting is considered imminent.

Sentence Summarization

Goal Shorter version of the original sentence.

Input Sentence

Output Shortened sentence (possibly with different words).

Sentence Summarization and Compression

Source

Russian Defense Minister Ivanov called Sunday for the creation of a joint front for combating global terrorism.

Target

Russia calls for joint front against terrorism.

Summarization Phenomena:

- Generalization
- Deletion
- Paraphrase

Sentence Summarization and Compression

Source

Russian Defense Minister Ivanov called Sunday for the creation of a joint front for combating global terrorism.

Target

Russia calls for joint front against terrorism.

Summarization Phenomena:

- Generalization
- Deletion
- Paraphrase

Sentence Summarization and Compression

Source

Russian Defense Minister Ivanov called Sunday for the creation of a joint front for combating global terrorism.

Target

Russia calls for joint front against terrorism.

Summarization Phenomena:

- Generalization
- Deletion
- Paraphrase

Sentence Summarization and Compression

Source

Russian Defense Minister Ivanov called Sunday for the creation of a joint front for combating global terrorism.

Target

Russia calls for joint front against terrorism.

Summarization Phenomena:

- Generalization
- Deletion
- Paraphrase

Sep 13, 3:17 PM EDT

GERMANY IMPLEMENTS TEMPORARY BORDER CHECKS TO LIMIT MIGRANTS

BY GEIR MOULSON AND SHAWN POGATCHNIK
ASSOCIATED PRESS

BERLIN (AP) -- Germany introduced temporary border controls Sunday to stem the tide of thousands of refugees streaming across its frontier, sending a clear message to its European partners that it needs more help with an influx that is straining its ability to cope.

AP Photo/Kay Nietfeld

Germany is a preferred destination for many people fleeing Syria's civil war and other troubled nations in the migration crisis that has bitterly divided Europe. They have braved dangerous sea crossings in flimsy

Coreference

Goal Cluster mentions based on their underlying entity.

Input Document

Output Clusters of coreferent mentions.

A Preliminary Example (CoNLL Dev Set, wsj/2404)

Cadillac posted a 3.2% increase despite new competition from Lexus, the fledgling luxury-car division of Toyota Motor Corp. Lexus sales weren't available; the cars are imported and Toyota reports their sales only at month-end.

With Coreferent Mentions Annotated

Cadillac posted a 3.2% increase despite new competition from [Lexus, the fledgling luxury-car division of [Toyota Motor Corp]] . [Lexus] sales weren't available; the cars are imported and [Toyota] reports [their] sales only at month-end.

Summary of (Informal) Terminology

Cadillac posted a 3.2% increase despite new competition from [Lexus, the fledgling luxury-car division of [Toyota Motor Corp]] . [Lexus] sales weren't available; the cars are imported and [Toyota] reports [their] sales only at month-end.

- **mention**: a span of text that can refer or be referred to
- **anaphoric**: a mention is anaphoric if it is coreferent with a previous mention
- **antecedent**: a mention to which an anaphoric mention refers