

Word Senses

Interactive Exercise

Consider the word `mouse`.

1. What meanings of `mouse` do you know?

Word Senses

Interactive Exercise

Consider the word `mouse`.

1. What meanings of `mouse` do you know?

- ☐ `mouse`¹: animal
- ☐ `mouse`²: computer input device

2. What is a **word sense**?

Word Senses

Interactive Exercise

Consider the word `mouse`.

1. What meanings of `mouse` do you know?

- ❑ `mouse`¹: animal
- ❑ `mouse`²: computer input device

2. What is a **word sense**?

- ❑ A word sense is a distinct meaning of a word.

3. How can different senses be distinguished in notation?

Word Senses

Interactive Exercise

Consider the word `mouse`.

1. What meanings of `mouse` do you know?

- ❑ `mouse`¹: animal
- ❑ `mouse`²: computer input device

2. What is a **word sense**?

- ❑ A word sense is a distinct meaning of a word.

3. How can different senses be distinguished in notation?

- ❑ Different senses have superscript numbers, e.g., `mouse`¹, `mouse`².

4. What is a **gloss**?

Word Senses

Interactive Exercise

Consider the word `mouse`.

1. What meanings of `mouse` do you know?

- ❑ `mouse`¹: animal
- ❑ `mouse`²: computer input device

2. What is a **word sense**?

- ❑ A word sense is a distinct meaning of a word.

3. How can different senses be distinguished in notation?

- ❑ Different senses have superscript numbers, e.g., `mouse`¹, `mouse`².

4. What is a **gloss**?

- ❑ A gloss is an informal textual description of a sense.

Sense	Gloss
<code>mouse</code> ¹	Small rodent with a pointed snout and long tail.
<code>mouse</code> ²	Handheld computer input device used to control the cursor.

Synsets

Interactive Exercise

1. What is a **synset**?

Synsets

Interactive Exercise

1. What is a **synset**?

- ❑ set of near-synonymous *senses* for the same lemma
- ❑ each synset has topic (“supersense”)

2. Which entries below belong to the same synset?

Sense	Gloss
bass ¹	The lowest part of the musical range.
bass ¹ , basso	An adult male singer with the lowest voice.
bass ² , sea bass	The lean flesh of a saltwater fish.

Synsets

Interactive Exercise

1. What is a **synset**?

- ❑ set of near-synonymous *senses* for the same lemma
- ❑ each synset has topic (“supersense”)

2. Which entries below belong to the same synset?

Sense	Gloss
bass ¹	The lowest part of the musical range.
bass ¹ , basso	An adult male singer with the lowest voice.
bass ² , sea bass	The lean flesh of a saltwater fish.

- ❑ bass¹ and basso¹ form one synset with supersense “music”.
- ❑ sea bass² forms another with supersense “food”.

3. What is a **supersense**?

Synsets

Interactive Exercise

1. What is a **synset**?

- ❑ set of near-synonymous *senses* for the same lemma
- ❑ each synset has topic (“supersense”)

2. Which entries below belong to the same synset?

- ❑ `bass1` and `basso1` form one synset with supersense “music”.
- ❑ `sea bass2` forms another with supersense “food”.

3. What is a **supersense**?

- ❑ semantic category (topic) assigned to a synset
- ❑ from a controlled vocabulary

Synset	Topic	Gloss
<code>bass¹</code>	music	The lowest part of the musical range.
<code>basso¹</code>	music	An adult male singer with the lowest voice.
<code>bass², sea bass</code>	food	The lean flesh of a saltwater fish.

Interactive Exercise: Fill in Senses and Glosses

Interactive Exercise

Fill in the missing missing headers.

bank ¹	Financial institution that accepts, deposits, and lends money.
bank ²	Sloping land beside a river or other body of water.

Word Sense Disambiguation

Interactive Exercise

Consider the sentence

We guarantee that your bank deposits will cover future costs.

1. What is Word Sense Disambiguation (WSD)?

Word Sense Disambiguation

Interactive Exercise

Consider the sentence

We guarantee that your bank deposits will cover future costs.

1. What is Word Sense Disambiguation (WSD)?
 - ❑ Given a text, assign the correct sense to each word from a sense lexicon.
2. Which sense of bank is intended?
 - ❑ bank¹: financial institution
 - ❑ bank²: land beside a body of water

Word Sense Disambiguation

Interactive Exercise

Consider the sentence

We guarantee that your bank deposits will cover future costs.

1. What is Word Sense Disambiguation (WSD)?
 - ❑ Given a text, assign the correct sense to each word from a sense lexicon.
2. Which sense of bank is intended?
 - ❑ **bank¹: financial institution**
 - ❑ bank²: land beside a body of water
3. How could a computer choose the correct sense?

Word Sense Disambiguation

Interactive Exercise

Consider the sentence

We guarantee that your bank deposits will cover future costs.

1. What is Word Sense Disambiguation (WSD)?
 - ❑ Given a text, assign the correct sense to each word from a sense lexicon.
2. Which sense of bank is intended?
 - ❑ **bank¹: financial institution**
 - ❑ bank²: land beside a body of water
3. How could a computer choose the correct sense?
 - ❑ Compare the context with the glosses of each candidate sense.
 - ❑ Choose the gloss with the largest semantic similarity.

Sense	Gloss
bank ¹	Financial institution that accepts deposits and lends money.
bank ²	Sloping land beside a river or other body of water.

Lesk Algorithm

Word Sense Disambiguation

Algorithm: Simplified Lesk Algorithm. [[Kilgariff and Rosenzweig 2000](#)], [[Wikipedia](#)]

Input: t_i . Target token t_i .

$d = (t_{i-k}, \dots, t_i, \dots, t_{i+k})$. Context window of t_i .

G . Dictionary of word senses.

SimplifiedLesk(t_i, d, G)

1. $G_{t_i} \subset G$ (Subset of t_i 's glosses in G)
2. Remove stop words from and lemmatize the tokens of context window d .
3. Compute semantic vector embeddings of context window d and all glosses in G_{t_i} .
4. Return the sense of G_{t_i} which has the highest semantic similarity to d .
 - The original Lesk algorithm gloss overlap as the similarity measure.

Word Sense Disambiguation

Interactive Exercise

State of the art for WSD uses similarity of embeddings instead of text overlap.

Would you choose static or contextualized embeddings for WSD with the Lesk algorithm?

Word Sense Disambiguation

Interactive Exercise

State of the art for WSD uses similarity of embeddings instead of text overlap.

Would you choose static or contextualized embeddings for WSD with the Lesk algorithm?

- ❑ Static embeddings
 - e.g., Word2Vec
 - one vector per word
 - different senses share the same vector representation
- ❑ Contextualized embeddings
 - e.g., BERT
 - different senses have different vector representations
- ❑ Contextualized embeddings are the better choice for WSD