Polysemy in Spoken Conversations and Written Texts

Aina Garí Soler, Matthieu Labeau, Chloé Clavel

Textual Polysemy

Textual Polysemy

We use polysemous words (i.e., words with multiple senses) all the time

Textual Polysemy

We use polysemous words (i.e., words with multiple senses) all the time

Textual Polysemy

We use polysemous words (i.e., words with multiple senses) all the time

Sometimes even in different senses within one discourse

Textual Polysemy

We use polysemous words (i.e., words with multiple senses) all the time

Sometimes even in different senses within one discourse

- This does not necessarily hinder communication

Textual Polysemy

We use polysemous words (i.e., words with multiple senses) all the time

Sometimes even in different senses within one discourse

- This does not necessarily hinder communication

- For example, speakers may exploit polysemy for metaphors, or to showcase their sense of humor with irony and jokes... (Nerlich and Clarke, 2001)

Textual Polysemy

We use polysemous words (i.e., words with multiple senses) all the time

Sometimes even in different senses within one discourse

- This does not necessarily hinder communication

- For example, speakers may exploit polysemy for metaphors, or to showcase their sense of humor with irony and jokes... (Nerlich and Clarke, 2001)

- ... or also to illustrate a point in this presentation

Textual Polysemy

We use polysemous words (i.e., words with multiple senses) all the time

Sometimes even in different senses within one discourse

- This does not necessarily hinder communication

- For example, speakers may exploit polysemy for metaphors, or to showcase their sense of humor with irony and jokes... (Nerlich and Clarke, 2001)

- ... or also to illustrate a point in this presentation

- ... and to do it twice, to the point where the audience has had enough of it.

Textual Polysemy

We use polysemous words (i.e., words with multiple senses) all the time

Sometimes even in different senses within one discourse

- This does not necessarily hinder communication

- For example, speakers may exploit polysemy for metaphors, or to showcase their sense of humor with irony and jokes... (Nerlich and Clarke, 2001)

- ... or also to illustrate a point in this presentation

- ... and to do it twice, to the point where the audience has had enough of it.

Textual Polysemy: Background

Textual Polysemy: Background

- "**One Sense per Discourse**" (OSD) hypothesis (Gale et al., 1992)

Textual Polysemy: Background

- "**One Sense per Discourse**" (OSD) hypothesis (Gale et al., 1992)

  ➔ Used as a heuristic for Word Sense Disambiguation (WSD), Named Entity Recognition and Machine Translation (Cucerzan,2007; Ture et al., 2012; Pilehvar and Navigli, 2015; Chaplot and Salakhutdinov, 2018...)

Textual Polysemy: Background

- "**One Sense per Discourse**" (OSD) hypothesis (Gale et al., 1992)

  - ➔ Used as a heuristic for Word Sense Disambiguation (WSD), Named Entity Recognition and Machine Translation (Cucerzan,2007; Ture et al., 2012; Pilehvar and Navigli, 2015; Chaplot and Salakhutdinov, 2018...)
  - ➔ But also put into question (Krovetz 1998, Leacock et al., 1998)

Textual Polysemy: Background

- "**One Sense per Discourse**" (OSD) hypothesis (Gale et al., 1992)

  - ➜ Used as a heuristic for Word Sense Disambiguation (WSD), Named Entity Recognition and Machine Translation (Cucerzan,2007; Ture et al., 2012; Pilehvar and Navigli, 2015; Chaplot and Salakhutdinov, 2018...)
  - ➜ But also put into question (Krovetz 1998, Leacock et al., 1998)

- Potential polysemy of a corpus can be used as an indication of its **WSD-difficulty**

Textual Polysemy: Background

- "**One Sense per Discourse**" (OSD) hypothesis (Gale et al., 1992)

  - ➔ Used as a heuristic for Word Sense Disambiguation (WSD), Named Entity Recognition and Machine Translation (Cucerzan,2007; Ture et al., 2012; Pilehvar and Navigli, 2015; Chaplot and Salakhutdinov, 2018...)
  - ➔ But also put into question (Krovetz 1998, Leacock et al., 1998)

- Potential polysemy of a corpus can be used as an indication of its **WSD-difficulty**

- Work on quantifying textual polysemy has focused on **written**, **monolog-like discourse**
  (McCarthy et al., 2007; Pasini and Camacho-Collados, 2020; Barba et al., 2021)

Textual Polysemy: Background

- "**One Sense per Discourse**" (OSD) hypothesis (Gale et al., 1992)

  - ➔ Used as a heuristic for Word Sense Disambiguation (WSD), Named Entity Recognition and Machine Translation (Cucerzan,2007; Ture et al., 2012; Pilehvar and Navigli, 2015; Chaplot and Salakhutdinov, 2018...)
  - ➔ But also put into question (Krovetz 1998, Leacock et al., 1998)

- Potential polysemy of a corpus can be used as an indication of its **WSD-difficulty**

- Work on quantifying textual polysemy has focused on **written**, **monolog-like discourse** (McCarthy et al., 2007; Pasini and Camacho-Collados, 2020; Barba et al., 2021)
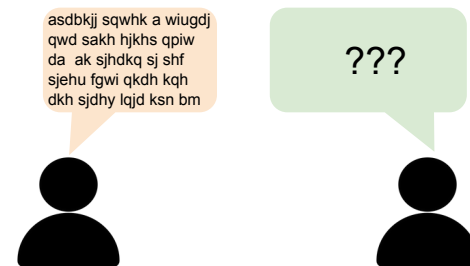
What about spoken language and dialog?

Dialog vs Monolog

Dialog vs Monolog

- Differences between speakers (background, world knowledge, idiolect, language level, opinion...)

- In dialog: misunderstandings, disagreements

asdbkjj sqwhk a wiugdj qwd sakh hjkhs qpiw da ak sjhdkq sj shf sjehu fgwi qkdh kqh dkh sjdhy lqjd ksn bm
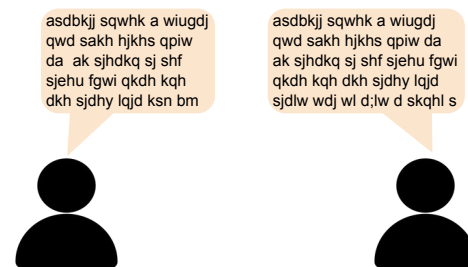
???

Dialog vs Monolog

- Differences between speakers (background, world knowledge, idiolect, language level, opinion…)

- In dialog: misunderstandings, disagreements

- The monolog speaker may be unaware of potential ambiguities in their discourse

- In dialog, speakers often align in the way they talk (Pickering and Garrod, 2004)

What we do

What we do

➜ We compare the observed polysemy level of texts of different nature

What we do

➔ We compare the observed polysemy level of texts of different nature

- We perform **automatic WSD** on multiple different datasets,

- we calculate different **polysemy measures** (and propose our own),

- we **compare** the results for different kinds of texts.

Data

(Spoken) Data

(Spoken) Data

- 2020 US presidential **debate** (Joe Biden vs Donald Trump)

(Spoken) Data

- 2020 US presidential **debate** (Joe Biden vs Donald Trump)

15 Monologs
(longer topic-initial turns)

6 Dialogs
(subsequent interaction about a topic)

(Spoken) Data

- 2020 US presidential **debate** (Joe Biden vs Donald Trump)

15 Monologs
(longer topic-initial turns)

6 Dialogs
(subsequent interaction about a topic)

- **Iemocap** (Busso et al., 2008): hypothetical emotional conversations between actors

(Spoken) Data

- 2020 US presidential **debate** (Joe Biden vs Donald Trump)

15 Monologs
(longer topic-initial turns)

6 Dialogs
(subsequent interaction about a topic)

- **Iemocap** (Busso et al., 2008): hypothetical emotional conversations between actors

71 Scripted

80 Spontaneous

(Spoken, Spontaneous, Dialog) Data

(Spoken, Spontaneous, Dialog) Data

- **JUSThink** (Norman et al., 2021) 10 task-oriented conversations between children

- **Switchboard** (Stolcke et al., 2000) 1,126 conversations on a provided topic

- **HCRC MapTask Corpus** (Thompson et al., 1993) 128 task-oriented conversations

- **BT Oasis Corpus** (Leech and Weisser, 2003) 378 calls to British Telecom and Trainline operator services

(Written) Data

(Written) Data

Three WSD evaluation campaings (Raganato et al., 2017)

- **Senseval-2** (Edmonds and Cotton, 2001) (3 texts)
- **Senseval-3** task 1 (Snyder and Palmer, 2004) (3 texts)
- **SemEval-15** task 13 (Moro and Navigli, 2015) (4 texts)

Methodology

Automatic WSD Annotation

Automatic WSD Annotation

1. Text preprocessing (tokenization, pos-tagging, lemmatization)

Automatic WSD Annotation

1.   Text preprocessing (tokenization, pos-tagging, lemmatization)

2.   Automatic WSD annotation with ESCHER (Barba et al., 2021) of nouns, verbs and adjectives

Automatic WSD Annotation

1. Text preprocessing (tokenization, pos-tagging, lemmatization)

2. Automatic WSD annotation with ESCHER (Barba et al., 2021) of nouns, verbs and adjectives

Automatic Annotation Quality

Written: automatic evaluation

Automatic WSD Annotation

1. Text preprocessing (tokenization, pos-tagging, lemmatization)

2. Automatic WSD annotation with ESCHER (Barba et al., 2021) of nouns, verbs and adjectives

Automatic Annotation Quality   **Accuracy**

Written: automatic evaluation   **73.4%** - **79.9%**

Automatic WSD Annotation

1. Text preprocessing (tokenization, pos-tagging, lemmatization)

2. Automatic WSD annotation with ESCHER (Barba et al., 2021) of nouns, verbs and adjectives

Automatic Annotation Quality

Accuracy

| Written: automatic evaluation | **73.4% - 79.9%** |

| Spoken: manual verification of 5 texts |

Automatic WSD Annotation

1. Text preprocessing (tokenization, pos-tagging, lemmatization)

2. Automatic WSD annotation with ESCHER (Barba et al., 2021) of nouns, verbs and adjectives

Automatic Annotation Quality   **Accuracy**

Written: automatic evaluation	**73.4%** - **79.9%**
Spoken: manual verification of 5 texts	**75.0% - 88.5%**

Potential Polysemy Measures

(independent of sense annotations)

Potential Polysemy Measures

(independent of sense annotations)

Potential Ambiguity (PA)

Potential Polysemy Measures

(independent of sense annotations)

Potential Ambiguity (PA)

average number of Wordnet senses of instances in a text

Potential Polysemy Measures

(independent of sense annotations)

Potential Ambiguity (PA)

average$_6$ number$_{12}$ of Wordnet senses$_5$ of instances$_2$ in a text$_4$

Potential Polysemy Measures

(independent of sense annotations)

Potential Ambiguity (PA)

average$_6$ number$_{12}$ of Wordnet senses$_5$ of instances$_2$ in a text$_4$

Potential Polysemy Measures

(independent of sense annotations)

Potential Ambiguity (PA)

average$_6$ number$_{12}$ of Wordnet senses$_5$ of instances$_2$ in a text$_4$

$$\frac{6+12+5+2+4}{5} = 5.8$$

Potential Polysemy Measures

(independent of sense annotations)

Potential Ambiguity (PA)

average$_6$ number$_{12}$ of Wordnet senses$_5$ of instances$_2$ in a text$_4$

$$\frac{6+12+5+2+4}{5} = 5.8$$

Percentage of polysemous words (pct-poly)

Potential Polysemy Measures

(independent of sense annotations)

Potential Ambiguity (PA)

average$_6$ number$_{12}$ of Wordnet senses$_5$ of instances$_2$ in a text$_4$

$$\frac{6+12+5+2+4}{5} = 5.8$$

Percentage of polysemous words (pct-poly)

proportion of polysemous word instances out of all word instances in a text

Potential Polysemy Measures

(independent of sense annotations)

Potential Ambiguity (PA)

average$_6$ number$_{12}$ of Wordnet senses$_5$ of instances$_2$ in a text$_4$

$$\frac{6+12+5+2+4}{5} = 5.8$$

Percentage of polysemous words (pct-poly)

proportion$_5$ of polysemous$_1$ word$_{10}$ instances$_2$ out of all word$_{10}$ instances$_2$ in a text$_4$

Potential Polysemy Measures

(independent of sense annotations)

Potential Ambiguity (PA)

average$_6$ number$_{12}$ of Wordnet senses$_5$ of instances$_2$ in a text$_4$

$$\frac{6+12+5+2+4}{5} = 5.8$$

Percentage of polysemous words (pct-poly)

proportion$_5$ of polysemous$_1$ word$_{10}$ instances$_2$ out of all word$_{10}$ instances$_2$ in a text$_4$

Potential Polysemy Measures

(independent of sense annotations)

Potential Ambiguity (PA)

average$_6$ number$_{12}$ of Wordnet senses$_5$ of instances$_2$ in a text$_4$

$$\frac{6+12+5+2+4}{5} = 5.8$$

Percentage of polysemous words (pct-poly)

proportion$_5$ of polysemous$_1$ word$_{10}$ instances$_2$ out of all word$_{10}$ instances$_2$ in a text$_4$

$$\frac{6}{7} = 0.86$$

Observed Polysemy Measures

Observed Polysemy Measures

More than One Sense per Discourse (MOSD)

Observed Polysemy Measures

More than One Sense per Discourse (MOSD)

Percentage of repeated polysemous words used in more than one sense in a text
(Krovetz, 1998)

Observed Polysemy Measures

More than One Sense per Discourse (MOSD)

Percentage of repeated polysemous words used in more than one sense in a text
(Krovetz, 1998)

AVGSENSES

Observed Polysemy Measures

More than One Sense per Discourse (MOSD)

Percentage of repeated polysemous words used in more than one sense in a text
(Krovetz, 1998)

AVGSENSES

Average number of senses in which words in a discourse are used. It is similar to PA, but the number of senses of a word is the observed one and not the potential one taken from WordNet.

Observed Polysemy Measures

More than One Sense per Discourse (MOSD)

Percentage of repeated polysemous words used in more than one sense in a text
(Krovetz, 1998)

AVGSENSES

Average number of senses in which words in a discourse are used. It is similar to PA, but the number of senses of a word is the observed one and not the potential one taken from WordNet.

Observed Polysemy Measures

More than One Sense per Discourse (MOSD)

Percentage of repeated polysemous words used in more than one sense in a text
(Krovetz, 1998)

AVGSENSES

Average number of senses in which words in a discourse are used. It is similar to PA, but the number of senses of a word is the observed one and not the potential one taken from WordNet.

Sense Dispersion Measure

Sense Dispersion Measure

ENTROPY

Sense Dispersion Measure

ENTROPY

Average of entropies of the observed sense distributions of repeated words in a text.
Based on Pasini and Navigli (2018)

Sense Dispersion Measure

ENTROPY

Average of entropies of the observed sense distributions of repeated words in a text.
Based on Pasini and Navigli (2018)

	s_1	s_2
w_1	8	1

Sense Dispersion Measure

ENTROPY

Average of entropies of the observed sense distributions of repeated words in a text.
Based on Pasini and Navigli (2018)

	s_1	s_2
w_1	8	1

	s_1	s_2
w_2	3	2

Sense Dispersion Measure

ENTROPY

Average of entropies of the observed sense distributions of repeated words in a text.
Based on Pasini and Navigli (2018)

	s_1	s_2
w_1	8	1

$\longrightarrow \mathcal{H} = 0.35$

	s_1	s_2
w_2	3	2

$\longrightarrow \mathcal{H} = 0.67$

Sense Dispersion Measure

ENTROPY

Average of entropies of the observed sense distributions of repeated words in a text.
Based on Pasini and Navigli (2018)

	s_1	s_2
w_1	8	1

$\longrightarrow \mathcal{H} = 0.35$

	s_1	s_2
w_2	3	2

$\longrightarrow \mathcal{H} = 0.67$

$$ENTROPY = 0.51$$

Sense Granularity

Sense Granularity

WordNet is very fine-grained: we can use WordNet supersenses

Sense Granularity

WordNet is very fine-grained: we can use WordNet supersenses

NOUN	Tops, act, animal, artifact, attribute, body, cognition, communication, event, feeling, food, group, location, motive, object, person, phenomenon, plant, possession, process, quantity, relation, shape, state, substance, time
VERB	body, change, cognition, communication, competition, consumption, contact, creation, emotion, motion, perception, possession, social, stative, weather
ADJ	all, pert, ppl

Sense Granularity

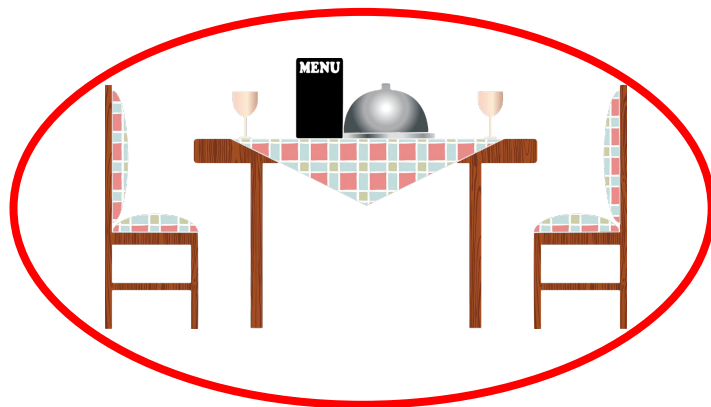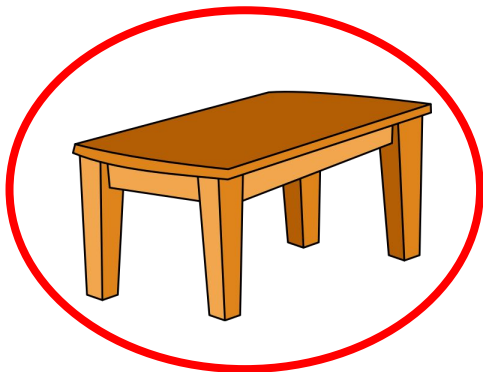WordNet is very fine-grained: we can use WordNet supersenses

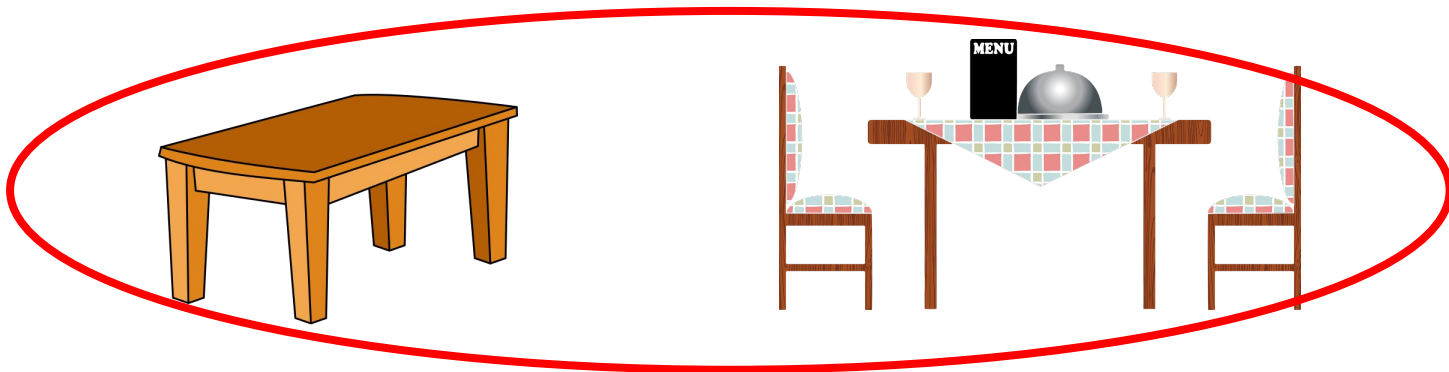| NOUN	Tops, act, animal, artifact, attribute, body, cognition, communication, event, feeling, food, group, location, motive, object, person, phenomenon, plant, possession, process, quantity, relation, shape, state, substance, time
VERB	body, change, cognition, communication, competition, consumption, contact, creation, emotion, motion, perception, possession, social, stative, weather
ADJ	all, pert, ppl

Sense Granularity

WordNet is very fine-grained: we can use WordNet supersenses

NOUN	Tops, act, animal, **artifact**, attribute, body, cognition, communication, event, feeling, food, group, location, motive, object, person, phenomenon, plant, possession, process, quantity, relation, shape, state, substance, time
VERB	body, change, cognition, communication, competition, consumption, contact, creation, emotion, motion, perception, possession, social, stative, weather
ADJ	all, pert, ppl

Sense Granularity

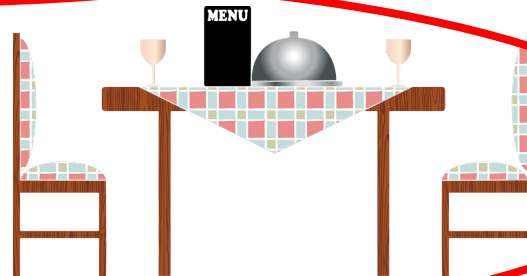WordNet is very fine-grained: we can use WordNet supersenses

NOUN	Tops, act, animal, **artifact**, attribute, body, cognition, communication, event, feeling, food, group, location, motive, object, person, phenomenon, plant, possession, process, quantity, relation, shape, state, substance, time
VERB	body, change, cognition, communication, competition, consumption, contact, creation, emotion, motion, perception, possession, social, stative, weather
ADJ	all, pert, ppl

Sense Granularity

WordNet is very fine-grained: we can use WordNet supersenses
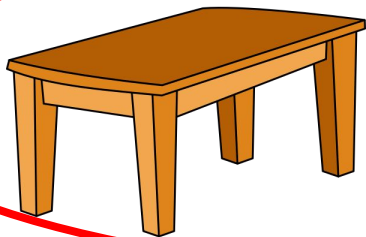
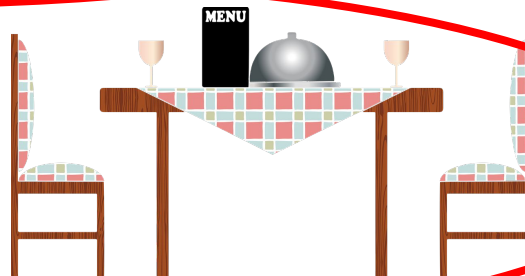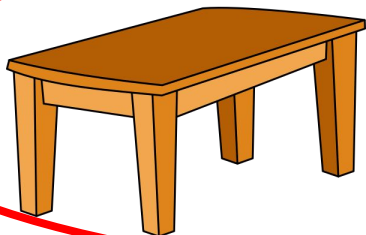|	
NOUN	Tops, act, animal, **artifact**, attribute, body, cognition, communication, event, feeling, food, **group**, location, motive, object, person, phenomenon, plant, possession, process, quantity, relation, shape, state, substance, time
VERB	body, change, cognition, communication, competition, consumption, contact, creation, emotion, motion, perception, possession, social, stative, weather
ADJ	all, pert, ppl

Results

Results

Discourse type	POTENTIAL		OBSERVED		
	PA	PCT-POLY	MOSD	AVGSENSES	ENTROPY
Spontaneous Spoken **Dialog**					
Scripted Spoken **Dialog**					
Spontaneous Spoken **Monolog**					
Scripted Written **Monolog**					

Results

Discourse type	POTENTIAL		OBSERVED		
	PA	PCT-POLY	MOSD	AVGSENSES	ENTROPY
Spontaneous Spoken **Dialog**					
Scripted Spoken **Dialog**					
Spontaneous Spoken **Monolog**					
Scripted Written **Monolog**					

Results

Discourse type	POTENTIAL		OBSERVED		
	PA	PCT-POLY	MOSD	AVGSENSES	ENTROPY
Spontaneous Spoken **Dialog**					
Scripted Spoken **Dialog**					
Spontaneous Spoken **Monolog**					
Scripted Written **Monolog**					

Results

	POTENTIAL		OBSERVED		
Discourse type	PA	PCT-POLY	MOSD	AVGSENSES	ENTROPY
Spontaneous Spoken **Dialog**					
Scripted Spoken **Dialog**					
Spontaneous Spoken **Monolog**					
Scripted Written **Monolog**					

Results

Discourse type	POTENTIAL		OBSERVED		
	PA	PCT-POLY	MOSD	AVGSENSES	ENTROPY
Spontaneous Spoken **Dialog**					
Scripted Spoken **Dialog**					
Spontaneous Spoken **Monolog**					
Scripted Written **Monolog**					

Results

Discourse type	POTENTIAL		OBSERVED		
	PA	PCT-POLY	MOSD	AVGSENSES	ENTROPY
Spontaneous Spoken **Dialog**					
Scripted Spoken **Dialog**					
Spontaneous Spoken **Monolog**					
Scripted Written **Monolog**					

Results

Discourse type	POTENTIAL		OBSERVED		
	PA	PCT-POLY	MOSD	AVGSENSES	ENTROPY
Spontaneous Spoken **Dialog**	10.6	91.9	37.1	**1.75**	0.24
Scripted Spoken **Dialog**	9.9	93.1	**44.1**	1.51	**0.32**
Spontaneous Spoken **Monolog**	**10.9**	**93.6**	33.1	1.28	0.23
Scripted Written **Monolog**	6.9	85.1	28.2	1.19	0.17

Results

Discourse type	POTENTIAL		OBSERVED		
	PA	PCT-POLY	MOSD	AVGSENSES	ENTROPY
Spontaneous Spoken **Dialog**	10.6	91.9	37.1	**1.75**	0.24
Scripted Spoken **Dialog**	9.9	93.1	**44.1**	1.51	**0.32**
Spontaneous Spoken **Monolog**	**10.9**	**93.6**	33.1	1.28	0.23
Scripted Written **Monolog**	6.9	85.1	28.2	1.19	0.17

Written text has a lower polysemy level than spoken text

Results

Discourse type	POTENTIAL		OBSERVED		
	PA	PCT-POLY	MOSD	AVGSENSES	ENTROPY
Spontaneous Spoken **Dialog**	10.6	91.9	37.1	**1.75**	0.24
Scripted Spoken **Dialog**	9.9	93.1	**44.1**	1.51	**0.32**
Spontaneous Spoken **Monolog**	**10.9**	**93.6**	33.1	1.28	0.23
Scripted Written **Monolog**	6.9	85.1	28.2	1.19	0.17

Results

Discourse type	POTENTIAL		OBSERVED		
	PA	PCT-POLY	MOSD	AVGSENSES	ENTROPY
Spontaneous Spoken **Dialog**	10.6	91.9	37.1	**1.75**	0.24
Scripted Spoken **Dialog**	9.9	93.1	**44.1**	1.51	**0.32**
Spontaneous Spoken **Monolog**	**10.9**	**93.6**	33.1	1.28	0.23
Scripted Written **Monolog**	6.9	85.1	28.2	1.19	0.17

Conclusion and Future Work

Conclusion and Future Work

❖ **Spoken** texts tend to present a **higher level of observed polysemy** than **written** discourse

❖ **Useful for WSD** of different kinds of text: the OSD heuristic may work less well on this kind of data

Conclusion and Future Work

❖ **Spoken** texts tend to present a **higher level of observed polysemy** than **written** discourse

❖ **Useful for WSD** of different kinds of text: the OSD heuristic may work less well on this kind of data

❖ What kind of words tend to be used in multiple senses in a discourse?

❖ Is observed polysemy higher when speakers disagree on a topic?

Differences in Word Usage between Stances

Differences in Word Usage between Stances

Do we use words differently depending on our opinions?

Differences in Word Usage between Stances

Do we use words differently depending on our opinions?

IN FAVOR OF Zoos

AGAINST Zoos

Differences in Word Usage between Stances

Do we use words differently depending on our opinions?

IN FAVOR OF Zoos

AGAINST Zoos

➡ BERT (Devlin et al., 2019) representations of words that are relevant to the topic (animal, zoo, live, habitat) differ the most between these two groups

Differences in Word Usage between Stances

Do we use words differently depending on our opinions?

IN FAVOR OF Zoos   **AGAINST Zoos**

➡ BERT (Devlin et al., 2019) representations of words that are relevant to the topic (animal, zoo, live, habitat) differ the most between these two groups

➡ Word representations have higher similarity **within** a stance than **between** opposing stances ➡ if we share an opinion, we use words in a similar way

Thank you!

Discourse Length and Lexical Diversity

Discourse Length and Lexical Diversity

- **Discourse length** correlates positively with **observed polysemy**
  (longer text -> more occasions to reuse a word in a different sense)

Discourse Length and Lexical Diversity

- **Discourse length** correlates positively with **observed polysemy**
  (longer text -> more occasions to reuse a word in a different sense)

- **Lexical diversity** correlates negatively with **observed polysemy**
  (lower diversity -> words are reused more often -> higher changes of encountering different senses for a word)