Instance-based disambiguation of English -*ment* derivatives

Marios Andreou[1], Lea Kawaletz[1], Max Kisselew[2], Gabriella Lapesa[2], Sebastian Pado[2], Ingo Plag[1]

[1]Heinrich Heine Universität Düsseldorf, [2]Universität Stuttgart

15.09.2016

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

SFB 991

Universität Stuttgart

Institut für
Maschinelle
Sprachverarbeitung

Introduction

- One of the central problems in the semantics of derived words is polysemy (Lieber 2004, Rainer 2014)

- Affixes are frequently semantically underspecified, and subject to polysemy and meaning extensions of various sorts (Bauer, Lieber & Plag 2013: 641)

- Context does not always fully determine the reading of a given derived word (Kawaletz and Plag, 2015; Plag, Andreou & Kawaletz, to appear).

- To which extent does context determine the reading of derived words?

Introduction

- One of the central problems in the semantics of derived words is polysemy (Lieber 2004, Rainer 2014)

- Affixes are frequently semantically underspecified, and subject to polysemy and meaning extensions of various sorts (Bauer, Lieber & Plag 2013: 641)

- Context does not always fully determine the reading of a given derived word (Kawaletz and Plag, 2015; Plag, Andreou & Kawaletz, to appear).

- To which extent does context determine the reading of derived words?

Introduction

- One of the central problems in the semantics of derived words is polysemy (Lieber 2004, Rainer 2014)
- Affixes are frequently semantically underspecified, and subject to polysemy and meaning extensions of various sorts (Bauer, Lieber & Plag 2013: 641)
- Context does not always fully determine the reading of a given derived word (Kawaletz and Plag, 2015; Plag, Andreou & Kawaletz, to appear).
- To which extent does context determine the reading of derived words?

- One of the central problems in the semantics of derived words is polysemy (Lieber 2004, Rainer 2014)
- Affixes are frequently semantically underspecified, and subject to polysemy and meaning extensions of various sorts (Bauer, Lieber & Plag 2013: 641)
- Context does not always fully determine the reading of a given derived word (Kawaletz and Plag, 2015; Plag, Andreou & Kawaletz, to appear).
- To which extent does context determine the reading of derived words?

Aim

- We explore the problem of disambiguating newly derived words in context, using the Distributional Semantics methodology (Firth 1957).

- We use corpus-extracted data to interpret deverbal *-ment* nominalizations:

    - event-denoting (e.g. *assessment*)

    - object-denoting (e.g. *pavement*).

Aim

- We explore the problem of disambiguating newly derived words in context, using the Distributional Semantics methodology (Firth 1957).
- We use corpus-extracted data to interpret deverbal *-ment* nominalizations:
  - event-denoting (e.g. *assessment*)
  - object-denoting (e.g. *pavement*).

Aim

- We explore the problem of disambiguating newly derived words in context, using the Distributional Semantics methodology (Firth 1957).
- We use corpus-extracted data to interpret deverbal *-ment* nominalizations:
  - event-denoting (e.g. *assessment*)
  - object-denoting (e.g. *pavement*).

Aim

- We explore the problem of disambiguating newly derived words in context, using the Distributional Semantics methodology (Firth 1957).
- We use corpus-extracted data to interpret deverbal *-ment* nominalizations:
    - event-denoting (e.g. *assessment*)
    - object-denoting (e.g. *pavement*).

The suffix *-ment*

- Bauer et. al (2013), Kawaletz and Plag (2015): *-ment* derives nominals of various readings, among which
  - events (e.g. *assessment*)
  - results (e.g. *containment*)
  - states (e.g. *contentment*)
  - products (e.g. *pavement*)
  - instruments (e.g. *entertainment*)
  - locations (e.g. *embankment*)

The suffix *-ment*

- Bauer et. al (2013), Kawaletz and Plag (2015): *-ment* derives nominals of various readings, among which
  - events (e.g. *assessment*)
  - results (e.g. *containment*)
  - states (e.g. *contentment*)
  - products (e.g. *pavement*)
  - instruments (e.g. *entertainment*)
  - locations (e.g. *embankment*)

- Bauer et. al (2013), Kawaletz and Plag (2015): *-ment* derives nominals of various readings, among which
  - events (e.g. *assessment*)
  - results (e.g. *containment*)
  - states (e.g. *contentment*)
  - products (e.g. *pavement*)
  - instruments (e.g. *entertainment*)
  - locations (e.g. *embankment*)

The suffix *-ment*

- Bauer et. al (2013), Kawaletz and Plag (2015): *-ment* derives nominals of various readings, among which
  - events (e.g. *assessment*)
  - results (e.g. *containment*)
  - states (e.g. *contentment*)
  - products (e.g. *pavement*)
  - instruments (e.g. *entertainment*)
  - locations (e.g. *embankment*)

The suffix *-ment*

- Bauer et. al (2013), Kawaletz and Plag (2015): *-ment* derives nominals of various readings, among which
  - events (e.g. *assessment*)
  - results (e.g. *containment*)
  - states (e.g. *contentment*)
  - products (e.g. *pavement*)
  - instruments (e.g. *entertainment*)
  - locations (e.g. *embankment*)

The suffix *-ment*

- Bauer et. al (2013), Kawaletz and Plag (2015): *-ment* derives nominals of various readings, among which
  - events (e.g. *assessment*)
  - results (e.g. *containment*)
  - states (e.g. *contentment*)
  - products (e.g. *pavement*)
  - instruments (e.g. *entertainment*)
  - locations (e.g. *embankment*)

- Bauer et. al (2013), Kawaletz and Plag (2015): *-ment* derives nominals of various readings, among which
  - events (e.g. *assessment*)
  - results (e.g. *containment*)
  - states (e.g. *contentment*)
  - products (e.g. *pavement*)
  - instruments (e.g. *entertainment*)
  - locations (e.g. *embankment*)

- Polysemy in word-formation can be disambiguated in context (see among others, Lieber in press)

- Object-denoting nominalization:
  "I set down the scrap of doll's dress, a **bedragglement** of loose lace hem" (COCA FIC 1999)

- Event-denoting nominalization:
  "In many places, **emplacement** of granite plutons is synchronous to volcanic eruptions" (Google WEB 1995)

Polysemy and context

- Polysemy in word-formation can be disambiguated in context (see among others, Lieber in press)
- Object-denoting nominalization:
  “I set down the scrap of doll’s dress, a **bedragglement** of loose lace hem” (COCA FIC 1999)
- Event-denoting nominalization:
  “In many places, **emplacement** of granite plutons is synchronous to volcanic eruptions” (Google WEB 1995)

Polysemy and context

- Polysemy in word-formation can be disambiguated in context (see among others, Lieber in press)

- Object-denoting nominalization:
  “I set down the scrap of doll’s dress, a **bedragglement** of loose lace hem” (COCA FIC 1999)

- Event-denoting nominalization:
  “In many places, **emplacement** of granite plutons is synchronous to volcanic eruptions” (Google WEB 1995)

- Ambiguous readings:

  - “After 8 weeks of hydrolytic degradation, the nonwoven fabric was broken. There is an obvious **embrittlement** and cracking on the nonwoven fabric (Figure 6.5b).” (Google ACAD 2014)

  - “There is a persuasive legitimacy in this hatred of a war when it is evoked by a man who has suffered its most horrible **debauchments**.” (Google FIC 1965)

Polysemy and context

- Ambiguous readings:
  - “After 8 weeks of hydrolytic degradation, the nonwoven fabric was broken. There is an obvious **embrittlement** and cracking on the nonwoven fabric (Figure 6.5b).” (Google ACAD 2014)
  - “There is a persuasive legitimacy in this hatred of a war when it is evoked by a man who has suffered its most horrible **debauchments**.” (Google FIC 1965)

- Ambiguous readings:
  - “After 8 weeks of hydrolytic degradation, the nonwoven fabric was broken. There is an obvious **embrittlement** and cracking on the nonwoven fabric (Figure 6.5b).” (Google ACAD 2014)
  - “There is a persuasive legitimacy in this hatred of a war when it is evoked by a man who has suffered its most horrible **debauchments**.” (Google FIC 1965)

Dataset

- Low frequency *-ment* derivatives extracted from corpora such as the *Corpus of Contemporary American English*, the *Corpus of GlobalWeb-Based English*, and *WebCorp*.

- 56 types, 401 tokens

- 4 verb classes

  - 13 change-of-state verbs (e.g. *congeal*)
  - 10 force verbs (e.g. *coerce*)
  - 19 psych verbs (e.g. *annoy*)
  - 14 putting verbs (e.g. *embed*)

Dataset

- Low frequency *-ment* derivatives extracted from corpora such as the *Corpus of Contemporary American English*, the *Corpus of GlobalWeb-Based English*, and *WebCorp*.
- 56 types, 401 tokens
- 4 verb classes
  - 13 change-of-state verbs (e.g. *congeal*)
  - 10 force verbs (e.g. *coerce*)
  - 19 psych verbs (e.g. *annoy*)
  - 14 putting verbs (e.g. *embed*)

Dataset

- Low frequency *-ment* derivatives extracted from corpora such as the *Corpus of Contemporary American English*, the *Corpus of GlobalWeb-Based English*, and *WebCorp*.

- 56 types, 401 tokens

- 4 verb classes

  - 13 change-of-state verbs (e.g. *congeal*)

  - 10 force verbs (e.g. *coerce*)

  - 19 psych verbs (e.g. *annoy*)

  - 14 putting verbs (e.g. *embed*)

Dataset

- Low frequency *-ment* derivatives extracted from corpora such as the *Corpus of Contemporary American English*, the *Corpus of GlobalWeb-Based English*, and *WebCorp*.
- 56 types, 401 tokens
- 4 verb classes
  - 13 change-of-state verbs (e.g. *congeal*)
  - 10 force verbs (e.g. *coerce*)
  - 19 psych verbs (e.g. *annoy*)
  - 14 putting verbs (e.g. *embed*)

Dataset

- Low frequency *-ment* derivatives extracted from corpora such as the *Corpus of Contemporary American English*, the *Corpus of GlobalWeb-Based English*, and *WebCorp*.
- 56 types, 401 tokens
- 4 verb classes
  - 13 change-of-state verbs (e.g. *congeal*)
  - 10 force verbs (e.g. *coerce*)
  - 19 psych verbs (e.g. *annoy*)
  - 14 putting verbs (e.g. *embed*)

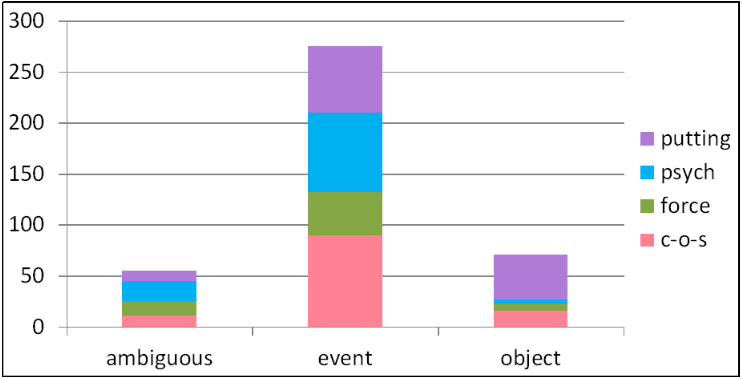
Dataset

- Low frequency *-ment* derivatives extracted from corpora such as the *Corpus of Contemporary American English*, the *Corpus of GlobalWeb-Based English*, and *WebCorp*.
- 56 types, 401 tokens
- 4 verb classes
  - 13 change-of-state verbs (e.g. *congeal*)
  - 10 force verbs (e.g. *coerce*)
  - 19 psych verbs (e.g. *annoy*)
  - 14 putting verbs (e.g. *embed*)

Dataset

- Low frequency *-ment* derivatives extracted from corpora such as the *Corpus of Contemporary American English*, the *Corpus of GlobalWeb-Based English*, and *WebCorp*.
- 56 types, 401 tokens
- 4 verb classes
  - 13 change-of-state verbs (e.g. *congeal*)
  - 10 force verbs (e.g. *coerce*)
  - 19 psych verbs (e.g. *annoy*)
  - 14 putting verbs (e.g. *embed*)

Distribution of readings



- Can we use Distributional Semantics tools to successfully disambiguate derived words in context?
- Tasks:
  - Can we distinguish between event and object readings?
  - How do we classify the ambiguous cases?

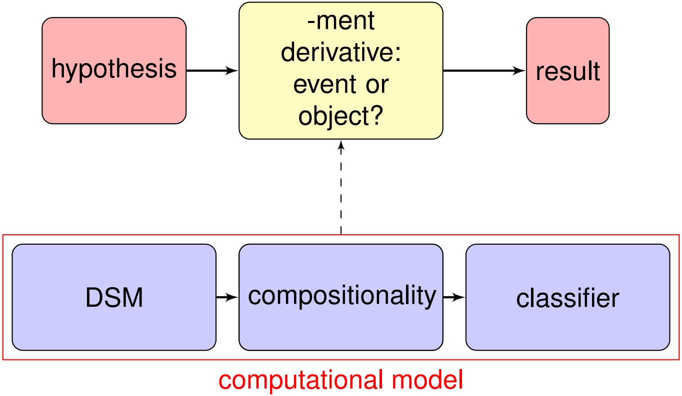
- Can we use Distributional Semantics tools to successfully disambiguate derived words in context?
- Tasks:
  - Can we distinguish between event and object readings?
  - How do we classify the ambiguous cases?

Research questions

- Can we use Distributional Semantics tools to successfully disambiguate derived words in context?
- Tasks:
  - Can we distinguish between event and object readings?
  - How do we classify the ambiguous cases?

Research questions

- Can we use Distributional Semantics tools to successfully disambiguate derived words in context?
- Tasks:
  - Can we distinguish between event and object readings?
  - How do we classify the ambiguous cases?


hypothesis
-ment derivative: event or object?
result
DSM
compositionality
classifier
computational model

- **Distributional Hypothesis (Firth, 1957; Harris, 1954)**

  difference in meaning $\iff$ difference in distribution

- **Distributional Semantic Models:**

  DSM meaning of w = list of words which co-occur with w

	law	wear
judge	8	2
t-shirt	1	8
tie	3	6

Distance between word vectors $\iff$ semantic similarity

empirical correlate of the amount of shared meaning

- **Distributional Hypothesis (Firth, 1957; Harris, 1954)**
  difference in meaning $\Longleftrightarrow$ difference in distribution
- **Distributional Semantic Models**:
  DSM meaning of w = list of words which co-occur with w

	law	wear
judge	8	2
t-shirt	1	8
tie	3	6

Distance between word vectors $\Longleftrightarrow$ semantic similarity
empirical correlate of the amount of shared meaning

- **Distributional Hypothesis (Firth, 1957; Harris, 1954)**
  difference in meaning $\Longleftrightarrow$ difference in distribution
- **Distributional Semantic Models**:
  DSM meaning of w = list of words which co-occur with w

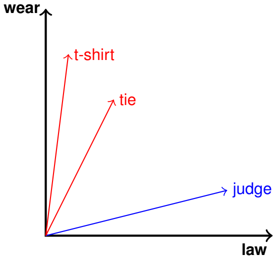
	law	**wear**
judge	8	2
t-shirt	1	8
tie	3	6

Distance between word vectors $\Longleftrightarrow$ semantic similarity
empirical correlate of the amount of shared meaning

Distributional Semantic Models

- **Distributional Hypothesis (Firth, 1957; Harris, 1954)**
  difference in meaning $\Longleftrightarrow$ difference in distribution
- **Distributional Semantic Models**:
  DSM meaning of w = list of words which co-occur with w

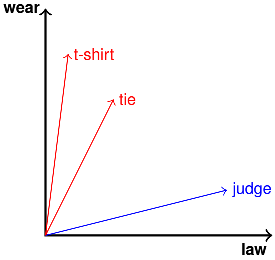
	law	**wear**
judge	8	2
t-shirt	1	8
tie	3	6



Distance between word vectors $\Longleftrightarrow$ semantic similarity
empirical correlate of the amount of shared meaning

- **Distributional Hypothesis (Firth, 1957; Harris, 1954)**
  difference in meaning $\Longleftrightarrow$ difference in distribution
- **Distributional Semantic Models**:
  DSM meaning of w = list of words which co-occur with w

	law	**wear**
judge	8	2
t-shirt	1	8
tie	3	6



Distance between word vectors $\Longleftrightarrow$ semantic similarity
empirical correlate of the amount of shared meaning

Distributional Semantic Models

- DSM vectors as usage-based **lexical entries**:
  - Contain **contextual features** for target words:
    - Typical actions/actors (*birds – fly*), coordination (*cats – dogs*), script knowledge (*cook – eat*)
  - Co-occurrence quantifies the **salience** of distributional features for a specific target

- Well established **properties**:
  - Successful in modeling semantic similarity:
    - NLP, Lexical Semantics, Psycholinguistic modeling
    - Recent developments: compositionality and multimodality
  - Known weaknesses:
    - Low frequency words: vectors are unreliable
    - Polysemous words: senses are conflated in a unique vector

- DSM vectors as usage-based **lexical entries**:
  - Contain **contextual features** for target words:
    - Typical actions/actors (*birds – fly*), coordination (*cats – dogs*), script knowledge (*cook – eat*)
  - Co-occurrence quantifies the **salience** of distributional features for a specific target
- Well established **properties**:
  - Successful in modeling semantic similarity:
    - NLP, Lexical Semantics, Psycholinguistic modeling
    - Recent developments: compositionality and multimodality
  - Known weaknesses:
    - Low frequency words: vectors are unreliable
    - Polysemous words: senses are conflated in a unique vector

- DSM vectors as usage-based **lexical entries**:
  - Contain **contextual features** for target words:
    - Typical actions/actors (*birds – fly*), coordination (*cats – dogs*), script knowledge (*cook – eat*)
  - Co-occurrence quantifies the **salience** of distributional features for a specific target
- Well established **properties**:
  - Successful in modeling semantic similarity:
    - NLP, Lexical Semantics, Psycholinguistic modeling
    - Recent developments: compositionality and multimodality
  - Known weaknesses:
    - Low frequency words: vectors are unreliable
    - Polysemous words: senses are conflated in a unique vector

Distributional Semantic Models

- DSM vectors as usage-based **lexical entries**:
  - Contain **contextual features** for target words:
    - Typical actions/actors (*birds – fly*), coordination (*cats – dogs*), script knowledge (*cook – eat*)
  - Co-occurrence quantifies the **salience** of distributional features for a specific target

- Well established **properties**:
  - Successful in modeling semantic similarity:
    - NLP, Lexical Semantics, Psycholinguistic modeling
    - Recent developments: compositionality and multimodality
  - Known weaknesses:
    - Low frequency words: vectors are unreliable
    - Polysemous words: senses are conflated in a unique vector

- DSM vectors as usage-based **lexical entries**:
  - Contain **contextual features** for target words:
    - Typical actions/actors (*birds – fly*), coordination (*cats – dogs*), script knowledge (*cook – eat*)
  - Co-occurrence quantifies the **salience** of distributional features for a specific target

- Well established **properties**:
  - Successful in modeling semantic similarity:
    - NLP, Lexical Semantics, Psycholinguistic modeling
    - Recent developments: compositionality and multimodality
  - Known weaknesses:
    - Low frequency words: vectors are unreliable
    - Polysemous words: senses are conflated in a unique vector

- DSM vectors as usage-based **lexical entries**:
  - Contain **contextual features** for target words:
    - Typical actions/actors (*birds – fly*), coordination (*cats – dogs*), script knowledge (*cook – eat*)
  - Co-occurrence quantifies the **salience** of distributional features for a specific target

- Well established **properties**:
  - Successful in modeling semantic similarity:
    - NLP, Lexical Semantics, Psycholinguistic modeling
    - Recent developments: compositionality and multimodality
  - Known weaknesses:
    - Low frequency words: vectors are unreliable
    - Polysemous words: senses are conflated in a unique vector

- DSM vectors as usage-based **lexical entries**:
  - Contain **contextual features** for target words:
    - Typical actions/actors (*birds – fly*), coordination (*cats – dogs*), script knowledge (*cook – eat*)
  - Co-occurrence quantifies the **salience** of distributional features for a specific target

- Well established **properties**:
  - Successful in modeling semantic similarity:
    - NLP, Lexical Semantics, Psycholinguistic modeling
    - Recent developments: compositionality and multimodality
  - Known weaknesses:
    - Low frequency words: vectors are unreliable
    - Polysemous words: senses are conflated in a unique vector

- DSM vectors as usage-based **lexical entries**:
  - Contain **contextual features** for target words:
    - Typical actions/actors (*birds – fly*), coordination (*cats – dogs*), script knowledge (*cook – eat*)
  - Co-occurrence quantifies the **salience** of distributional features for a specific target

- Well established **properties**:
  - Successful in modeling semantic similarity:
    - NLP, Lexical Semantics, Psycholinguistic modeling
    - Recent developments: compositionality and multimodality
  - Known weaknesses:
    - Low frequency words: vectors are unreliable
    - Polysemous words: senses are conflated in a unique vector

- DSM vectors as usage-based **lexical entries**:
  - Contain **contextual features** for target words:
    - Typical actions/actors (*birds – fly*), coordination (*cats – dogs*), script knowledge (*cook – eat*)
  - Co-occurrence quantifies the **salience** of distributional features for a specific target

- Well established **properties**:
  - Successful in modeling semantic similarity:
    - NLP, Lexical Semantics, Psycholinguistic modeling
    - Recent developments: compositionality and multimodality
  - Known weaknesses:
    - Low frequency words: vectors are unreliable
    - Polysemous words: senses are conflated in a unique vector

- DSM vectors as usage-based **lexical entries**:
  - Contain **contextual features** for target words:
    - Typical actions/actors (*birds – fly*), coordination (*cats – dogs*), script knowledge (*cook – eat*)
  - Co-occurrence quantifies the **salience** of distributional features for a specific target

- Well established **properties**:
  - Successful in modeling semantic similarity:
    - NLP, Lexical Semantics, Psycholinguistic modeling
    - Recent developments: compositionality and multimodality
  - Known weaknesses:
    - Low frequency words: vectors are unreliable
    - Polysemous words: senses are conflated in a unique vector

Distributional Semantic Models

- DSM vectors as usage-based **lexical entries**:
  - Contain **contextual features** for target words:
    - Typical actions/actors (*birds – fly*), coordination (*cats – dogs*), script knowledge (*cook – eat*)
  - Co-occurrence quantifies the **salience** of distributional features for a specific target

- Well established **properties**:
  - Successful in modeling semantic similarity:
    - NLP, Lexical Semantics, Psycholinguistic modeling
    - Recent developments: compositionality and multimodality
  - Known weaknesses:
    - Low frequency words: vectors are unreliable
    - Polysemous words: senses are conflated in a unique vector

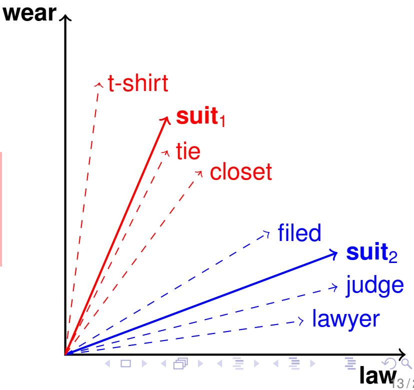
- **Problem**: the *-ment* derivatives in our dataset have low frequency and are potentially ambiguous!

- **Problem**: the *-ment* derivatives in our dataset have low frequency and are potentially ambiguous!
- **Solution**: words in the context as an approximation of the meaning of the *-ment* derivatives

- **Problem**: the *-ment* derivatives in our dataset have low frequency and are potentially ambiguous!
- **Solution**: words in the context as an approximation of the meaning of the *-ment* derivatives
  $\rightarrow$ **sentence vectors**: average of the vectors of the context words (Schütze, 1997)

- **Problem**: the *-ment* derivatives in our dataset have low frequency and are potentially ambiguous!
- **Solution**: words in the context as an approximation of the meaning of the *-ment* derivatives
  → **sentence vectors**: average of the vectors of the context words (Schütze, 1997)

Target: *suit*
suit$_1$: *The suit was in the closet, with the tie and the t-shirt*
suit$_2$: *The lawyer filed a suit to the judge*



Supervised classification:

- Given an **observation** and a set of **categories**, assign the observation to one of the categories:
  - Observations: *-ment* sentences – Categories: object vs. event
- How do we classify?
  - We identify a set of **training examples** (pairs of observations and categories), and build a **generalization** which we can use to classify new observations (test data)
  - The generalization is our **classifier**
  - The classifier is applied to unseen data
    → is the classification correct?

Supervised classification:

- Given an **observation** and a set of **categories**, assign the observation to one of the categories:
  - Observations: *-ment* sentences – Categories: object vs. event
- How do we classify?
  - We identify a set of **training examples** (pairs of observations and categories), and build a **generalization** which we can use to classify new observations (test data)
  - The generalization is our **classifier**
  - The classifier is applied to unseen data
    → is the classification correct?

Supervised classification:

- Given an **observation** and a set of **categories**, assign the observation to one of the categories:
    - Observations: *-ment* sentences – Categories: object vs. event

- How do we classify?
    - We identify a set of **training examples** (pairs of observations and categories), and build a **generalization** which we can use to classify new observations (test data)
    - The generalization is our **classifier**
    - The classifier is applied to unseen data
      → is the classification correct?

Supervised classification:

- Given an **observation** and a set of **categories**, assign the observation to one of the categories:
  - Observations: *-ment* sentences – Categories: object vs. event

- How do we classify?
  - We identify a set of **training examples** (pairs of observations and categories), and build a **generalization** which we can use to classify new observations (test data)
  - The generalization is our **classifier**
  - The classifier is applied to unseen data
    → is the classification correct?

Classifier

Supervised classification:

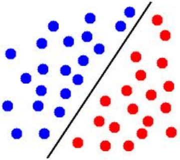
- Given an **observation** and a set of **categories**, assign the observation to one of the categories:
  - Observations: *-ment* sentences – Categories: object vs. event

- How do we classify?
  - We identify a set of **training examples** (pairs of observations and categories), and build a **generalization** which we can use to classify new observations (test data)
  - The generalization is our **classifier**
  - The classifier is applied to unseen data
    $\rightarrow$ is the classification correct?

Supervised classification:

- Given an **observation** and a set of **categories**, assign the observation to one of the categories:
  - Observations: *-ment* sentences – Categories: object vs. event

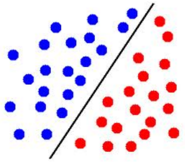
- How do we classify?
  - We identify a set of **training examples** (pairs of observations and categories), and build a **generalization** which we can use to classify new observations (test data)
  - The generalization is our **classifier**
  - The classifier is applied to unseen data → is the classification correct?



Supervised classification:

- Given an **observation** and a set of **categories**, assign the observation to one of the categories:
  - Observations: *-ment* sentences – Categories: object vs. event

- How do we classify?
  - We identify a set of **training examples** (pairs of observations and categories), and build a **generalization** which we can use to classify new observations (test data)
  - The generalization is our **classifier**
  - The classifier is applied to unseen data
    → is the classification correct?



- **Training data**:
  - CoreLex (Buitelaar, 1998): 196 EVENT and ARTIFACT seed nouns
    - No polysemy: *lunch* (EVENT+OBJECT) is not an EVENT seed
    - Frequency > 100
    - No *-ment* derivatives
  - 100 sentences per seed (randomly sampled)
  - We built vector representations for the seed sentences, and used them as (hopefully) unambiguous cases to train the classifier

- **Output:** probabilities of category assignment: e.g. 80% event vs. 20% object

We experimented with different classifier settings. Today, we discuss results achieved with a SVM degree 3, medium regularization (R package e1071)

Classifier

- **Training data**:
  - CoreLex (Buitelaar, 1998): 196 EVENT and ARTIFACT seed nouns
    - No polysemy: *lunch* (EVENT+OBJECT) is not an EVENT seed
    - Frequency > 100
    - No *-ment* derivatives
  - 100 sentences per seed (randomly sampled)
  - We built vector representations for the seed sentences, and used them as (hopefully) unambiguous cases to train the classifier
- **Output**: probabilities of category assignment: e.g. 80% event vs. 20% object

We experimented with different classifier settings. Today, we discuss results achieved with a SVM degree 3, medium regularization (R package e1071).

- **Training data**:
  - CoreLex (Buitelaar, 1998): 196 EVENT and ARTIFACT seed nouns
    - No polysemy: *lunch* (EVENT+OBJECT) is not an EVENT seed
    - Frequency > 100
    - No *-ment* derivatives
  - 100 sentences per seed (randomly sampled)
  - We built vector representations for the seed sentences, and used them as (hopefully) unambiguous cases to train the classifier

- **Output**: probabilities of category assignment: e.g. 80% event vs. 20% object

We experimented with different classifier settings. Today, we discuss results achieved with a SVM degree 3, medium regularization (R package e1071).

- **Training data**:
  - CoreLex (Buitelaar, 1998): 196 EVENT and ARTIFACT seed nouns
    - No polysemy: *lunch* (EVENT+OBJECT) is not an EVENT seed
    - Frequency > 100
    - No *-ment* derivatives
  - 100 sentences per seed (randomly sampled)
  - We built vector representations for the seed sentences, and used them as (hopefully) unambiguous cases to train the classifier

- **Output**: probabilities of category assignment: e.g. 80% event vs. 20% object

We experimented with different classifier settings. Today, we discuss results achieved with a SVM degree 3, medium regularization (R package e1071)

Classifier

- **Training data**:
  - CoreLex (Buitelaar, 1998): 196 EVENT and ARTIFACT seed nouns
    - No polysemy: *lunch* (EVENT+OBJECT) is not an EVENT seed
    - Frequency > 100
    - No *-ment* derivatives
  - 100 sentences per seed (randomly sampled)
  - We built vector representations for the seed sentences, and used them as (hopefully) unambiguous cases to train the classifier
- **Output**: probabilities of category assignment: e.g. 80% event vs. 20% object

We experimented with different classifier settings. Today, we discuss results achieved with a SVM degree 3, medium regularization (R package e1071)

Classifier

- **Training data**:
  - CoreLex (Buitelaar, 1998): 196 EVENT and ARTIFACT seed nouns
    - No polysemy: *lunch* (EVENT+OBJECT) is not an EVENT seed
    - Frequency > 100
    - No *-ment* derivatives
  - 100 sentences per seed (randomly sampled)
  - We built vector representations for the seed sentences, and used them as (hopefully) unambiguous cases to train the classifier
- **Output:** probabilities of category assignment: e.g. 80% event vs. 20% object

We experimented with different classifier settings. Today, we discuss results achieved with a SVM degree 3, medium regularization (R package e1071)

Classifier

- **Training data**:
  - CoreLex (Buitelaar, 1998): 196 EVENT and ARTIFACT seed nouns
    - No polysemy: *lunch* (EVENT+OBJECT) is not an EVENT seed
    - Frequency > 100
    - No *-ment* derivatives
  - 100 sentences per seed (randomly sampled)
  - We built vector representations for the seed sentences, and used them as (hopefully) unambiguous cases to train the classifier

- **Output**: probabilities of category assignment: e.g. 80% event vs. 20% object

We experimented with different classifier settings. Today, we discuss results achieved with a SVM degree 3, medium regularization (R package e1071)

Classifier

- **Training data**:
  - CoreLex (Buitelaar, 1998): 196 EVENT and ARTIFACT seed nouns
    - No polysemy: *lunch* (EVENT+OBJECT) is not an EVENT seed
    - Frequency > 100
    - No *-ment* derivatives
  - 100 sentences per seed (randomly sampled)
  - We built vector representations for the seed sentences, and used them as (hopefully) unambiguous cases to train the classifier

- **Output**: probabilities of category assignment: e.g, 80% event vs. 20% object

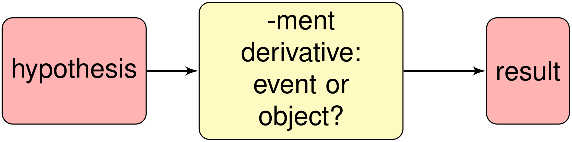
We experimented with different classifier settings. Today, we discuss results achieved with a SVM degree 3, medium regularization (R package e1071)

Classifier

- **Training data**:
  - CoreLex (Buitelaar, 1998): 196 EVENT and ARTIFACT seed nouns
    - No polysemy: *lunch* (EVENT+OBJECT) is not an EVENT seed
    - Frequency > 100
    - No *-ment* derivatives
  - 100 sentences per seed (randomly sampled)
  - We built vector representations for the seed sentences, and used them as (hopefully) unambiguous cases to train the classifier

- **Output**: probabilities of category assignment: e.g, 80% event vs. 20% object

We experimented with different classifier settings. Today, we discuss results achieved with a `svm` degree 3, medium regularization (R package `e1071`)



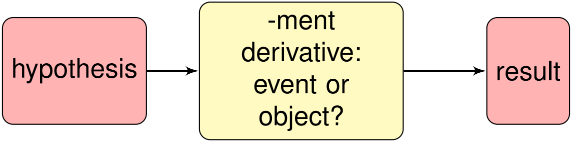
Hypothesis: the algorithm can tell objects from events, which means:

- It assigns high probability of an event reading to event nominalizations
- It assigns low probability of an event reading to object nominalizations

Question: do the semantic classes the base verbs differ in term of the predicted semantic type of the output?

?

Theoretical predictions and expected results



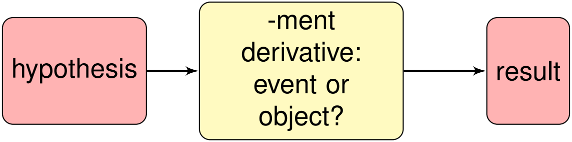
Hypothesis: the algorithm can tell objects from events, which means:

- It assigns high probability of an event reading to event nominalizations
- It assigns low probability of an event reading to object nominalizations

Question: do the semantic classes the base verbs differ in term of the predicted semantic type of the output?

?

Theoretical predictions and expected results



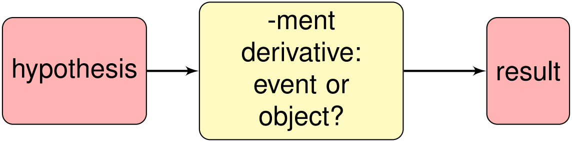
Hypothesis: the algorithm can tell objects from events, which means:

- It assigns high probability of an event reading to event nominalizations
- It assigns low probability of an event reading to object nominalizations

Question: do the semantic classes the base verbs differ in term of the predicted semantic type of the output?

?

Theoretical predictions and expected results



Hypothesis: the algorithm can tell objects from events, which means:

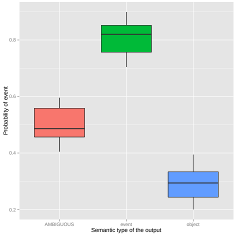
- It assigns high probability of an event reading to event nominalizations
- It assigns low probability of an event reading to object nominalizations

Question: do the semantic classes the base verbs differ in term of the predicted semantic type of the output?

?

Theoretical predictions and expected results

hypothesis → -ment derivative: event or object? → result



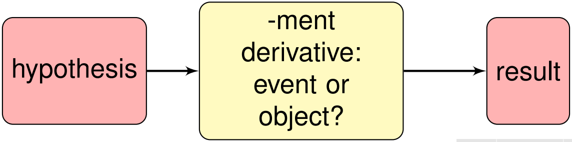
Hypothesis: the algorithm can tell objects from events, which means:

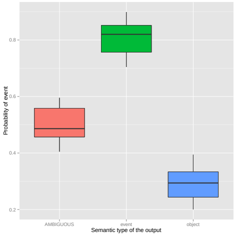
- It assigns high probability of an event reading to event nominalizations
- It assigns low probability of an event reading to object nominalizations

Question: do the semantic classes the base verbs differ in term of the predicted semantic type of the output?

?

Theoretical predictions and expected results





Hypothesis: the algorithm can tell objects from events, which means:

- It assigns high probability of an event reading to event nominalizations
- It assigns low probability of an event reading to object nominalizations

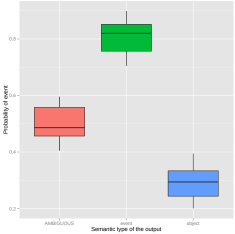
Question: do the semantic classes the base verbs differ in term of the predicted semantic type of the output?

Theoretical predictions and expected results

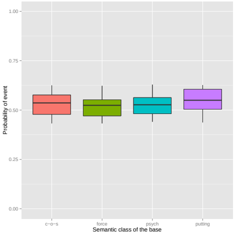
hypothesis → -ment derivative: event or object? → result

Hypothesis: the algorithm can tell objects from events, which means:

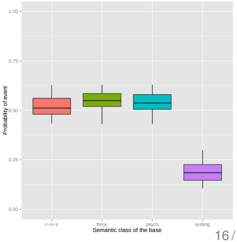
- It assigns high probability of an event reading to event nominalizations
- It assigns low probability of an event reading to object nominalizations



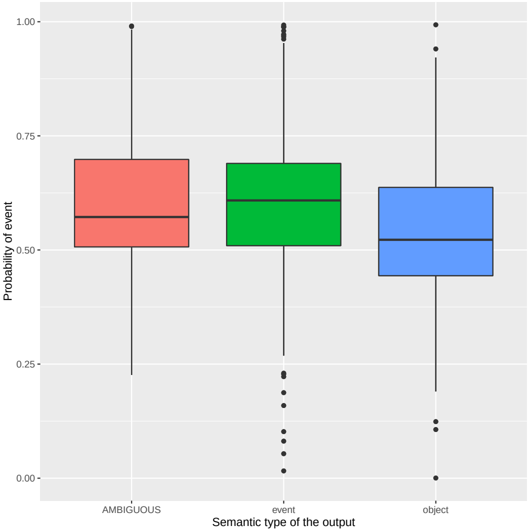
Question: do the semantic classes the base verbs differ in term of the predicted semantic type of the output?



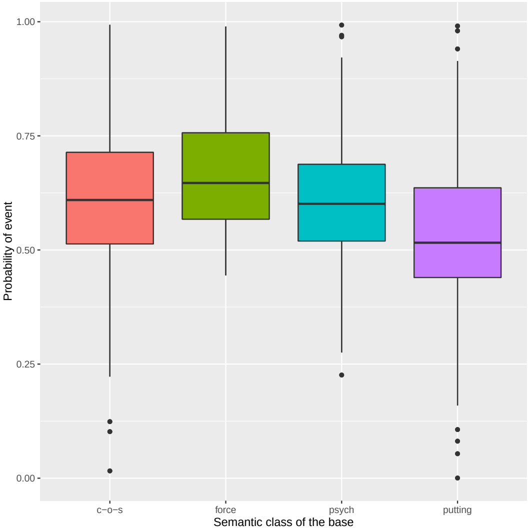
?



Results: semantic type of the output



Results: semantic class of the base



Results: overview

- A weak positive result $\rightarrow$ events are distinguishable from objects
- Ambiguous cases pattern with events
  - Explanation: within the ambiguous class, there seems to be a preponderance of possible eventive readings
- A difference between *force* (event) and *putting* (object)
- Next steps:
  1. What are the factors which influence the prediction? $\rightarrow$ regression analysis
  2. How can we improve the classifier?

Results: overview

- A weak positive result $\rightarrow$ events are distinguishable from objects
- Ambiguous cases pattern with events
  - Explanation: within the ambiguous class, there seems to be a preponderance of possible eventive readings
- A difference between *force* (event) and *putting* (object)
- Next steps:
  1. What are the factors which influence the prediction? $\rightarrow$ regression analysis
  2. How can we improve the classifier?

Results: overview

- A weak positive result → events are distinguishable from objects
- Ambiguous cases pattern with events
  - Explanation: within the ambiguous class, there seems to be a preponderance of possible eventive readings
- A difference between *force* (event) and *putting* (object)
- Next steps:
  1. What are the factors which influence the prediction? → regression analysis
  2. How can we improve the classifier?

Results: overview

- A weak positive result $\rightarrow$ events are distinguishable from objects
- Ambiguous cases pattern with events
  - Explanation: within the ambiguous class, there seems to be a preponderance of possible eventive readings
- A difference between *force* (event) and *putting* (object)
- Next steps:
  1. What are the factors which influence the prediction?
     $\rightarrow$ regression analysis
  2. How can we improve the classifier?

Results: overview

- A weak positive result $\rightarrow$ events are distinguishable from objects
- Ambiguous cases pattern with events
  - Explanation: within the ambiguous class, there seems to be a preponderance of possible eventive readings
- A difference between *force* (event) and *putting* (object)
- Next steps:
  1. What are the factors which influence the prediction? $\rightarrow$ regression analysis
  2. How can we improve the classifier?

Results: overview

- A weak positive result $\rightarrow$ events are distinguishable from objects
- Ambiguous cases pattern with events
  - Explanation: within the ambiguous class, there seems to be a preponderance of possible eventive readings
- A difference between *force* (event) and *putting* (object)
- Next steps:
  1. What are the factors which influence the prediction? $\rightarrow$ regression analysis
  2. How can we improve the classifier?

Results: overview

- A weak positive result $\rightarrow$ events are distinguishable from objects
- Ambiguous cases pattern with events
  - Explanation: within the ambiguous class, there seems to be a preponderance of possible eventive readings
- A difference between *force* (event) and *putting* (object)
- Next steps:
  1. What are the factors which influence the prediction? $\rightarrow$ regression analysis
  2. How can we improve the classifier?

Regression analysis is employed to test the effect of a number of factors (predictors) on a variable of interest (dependent variable)

- **Dependent variable**: probability of event ← classifier output
- **Variables of interest**:
  - Semantic type of the *-ment* derivative ← dataset annotation
    - Object (reference), event, ambiguous
  - Semantic class of the base ← dataset annotation
    - Putting (reference), force, psych, change of state
- **Covariates**:
  - Frequency of the *-ment* derivative in COCA (*freq-coca*)
  - Average frequency of the context words in the sentence (*context-frequency*)
  - How many words did we use to calculate the sentence vector? (*coverage*)

Regression analysis

Regression analysis is employed to test the effect of a number of factors (predictors) on a variable of interest (dependent variable)

- **Dependent variable**: probability of event ← classifier output
- **Variables of interest**:
  - Semantic type of the *-ment* derivative ← dataset annotation
    - Object (reference), event, ambiguous
  - Semantic class of the base ← dataset annotation
    - Putting (reference), force, psych, change of state
- **Covariates**:
  - Frequency of the *-ment* derivative in COCA (*freq-coca*)
  - Average frequency of the context words in the sentence (*context-frequency*)
  - How many words did we use to calculate the sentence vector? (*coverage*)

Regression analysis is employed to test the effect of a number of factors (predictors) on a variable of interest (dependent variable)

- **Dependent variable**: probability of event ← classifier output
- **Variables of interest**:
  - Semantic type of the *-ment* derivative ← dataset annotation
    - Object (reference), event, ambiguous
  - Semantic class of the base ← dataset annotation
    - Putting (reference), force, psych, change of state
- **Covariates**:
  - Frequency of the *-ment* derivative in COCA (*freq-coca*)
  - Average frequency of the context words in the sentence (*context-frequency*)
  - How many words did we use to calculate the sentence vector? (*coverage*)

Regression analysis is employed to test the effect of a number of factors (predictors) on a variable of interest (dependent variable)

- **Dependent variable**: probability of event ← classifier output
- **Variables of interest**:
  - Semantic type of the *-ment* derivative ← dataset annotation
    - Object (reference), event, ambiguous
  - Semantic class of the base ← dataset annotation
    - Putting (reference), force, psych, change of state
- **Covariates**:
  - Frequency of the *-ment* derivative in COCA (*freq-coca*)
  - Average frequency of the context words in the sentence (*context-frequency*)
  - How many words did we use to calculate the sentence vector? (*coverage*)

Regression analysis is employed to test the effect of a number of factors (predictors) on a variable of interest (dependent variable)

- **Dependent variable**: probability of event ← classifier output
- **Variables of interest**:
  - Semantic type of the *-ment* derivative ← dataset annotation
    - Object (reference), event, ambiguous
  - Semantic class of the base ← dataset annotation
    - Putting (reference), force, psych, change of state
- **Covariates**:
  - Frequency of the *-ment* derivative in COCA (*freq-coca*)
  - Average frequency of the context words in the sentence (*context-frequency*)
  - How many words did we use to calculate the sentence vector? (*coverage*)

Regression analysis

Regression analysis is employed to test the effect of a number of factors (predictors) on a variable of interest (dependent variable)

- **Dependent variable**: probability of event ← classifier output
- **Variables of interest**:
  - Semantic type of the *-ment* derivative ← dataset annotation
    - Object (reference), event, ambiguous
  - Semantic class of the base ← dataset annotation
    - Putting (reference), force, psych, change of state
- **Covariates**:
  - Frequency of the *-ment* derivative in COCA (*freq-coca*)
  - Average frequency of the context words in the sentence (*context-frequency*)
  - How many words did we use to calculate the sentence vector? (*coverage*)

Regression analysis

Regression analysis is employed to test the effect of a number of factors (predictors) on a variable of interest (dependent variable)

- **Dependent variable**: probability of event $\leftarrow$ classifier output
- **Variables of interest**:
  - Semantic type of the *-ment* derivative $\leftarrow$ dataset annotation
    - Object (reference), event, ambiguous
  - Semantic class of the base $\leftarrow$ dataset annotation
    - Putting (reference), force, psych, change of state
- **Covariates**:
  - Frequency of the *-ment* derivative in COCA (*freq-coca*)
  - Average frequency of the context words in the sentence (*context-frequency*)
  - How many words did we use to calculate the sentence vector? (*coverage*)

Regression analysis is employed to test the effect of a number of factors (predictors) on a variable of interest (dependent variable)

- **Dependent variable**: probability of event $\leftarrow$ classifier output
- **Variables of interest**:
  - Semantic type of the *-ment* derivative $\leftarrow$ dataset annotation
    - Object (reference), event, ambiguous
  - Semantic class of the base $\leftarrow$ dataset annotation
    - Putting (reference), force, psych, change of state
- **Covariates**:
  - Frequency of the *-ment* derivative in COCA (*freq-coca*)
  - Average frequency of the context words in the sentence (*context-frequency*)
  - How many words did we use to calculate the sentence vector? (*coverage*)

Regression analysis is employed to test the effect of a number of factors (predictors) on a variable of interest (dependent variable)

- **Dependent variable**: probability of event ← classifier output
- **Variables of interest**:
  - Semantic type of the *-ment* derivative ← dataset annotation
    - Object (reference), event, ambiguous
  - Semantic class of the base ← dataset annotation
    - Putting (reference), force, psych, change of state
- **Covariates**:
  - Frequency of the *-ment* derivative in COCA (*freq-coca*)
  - Average frequency of the context words in the sentence (*context-frequency*)
  - How many words did we use to calculate the sentence vector? (*coverage*)

Regression analysis

Regression analysis is employed to test the effect of a number of factors (predictors) on a variable of interest (dependent variable)

- **Dependent variable**: probability of event ← classifier output
- **Variables of interest**:
  - Semantic type of the *-ment* derivative ← dataset annotation
    - Object (reference), event, ambiguous
  - Semantic class of the base ← dataset annotation
    - Putting (reference), force, psych, change of state
- **Covariates**:
  - Frequency of the *-ment* derivative in COCA (*freq-coca*)
  - Average frequency of the context words in the sentence (*context-frequency*)
  - How many words did we use to calculate the sentence vector? (*coverage*)

Regression analysis

Regression analysis is employed to test the effect of a number of factors (predictors) on a variable of interest (dependent variable)

- **Dependent variable**: probability of event ← classifier output
- **Variables of interest**:
  - Semantic type of the *-ment* derivative ← dataset annotation
    - Object (reference), event, ambiguous
  - Semantic class of the base ← dataset annotation
    - Putting (reference), force, psych, change of state
- **Covariates**:
  - Frequency of the *-ment* derivative in COCA (*freq-coca*)
  - Average frequency of the context words in the sentence (*context-frequency*)
  - How many words did we use to calculate the sentence vector? (*coverage*)

Predictor	Effect	Significance
H: SEMANTIC TYPE (REF: OBJECT)		
event	+	*
ambiguous		
Q: BASE SEMANTIC CLASS (REF: PUTTING)		
change of state		
force	+	**
psych		
COVARIATES		
freq-coca	+	.
context-frequency	+	***
coverage		

$+$ = prediction pulled toward the event reading
R^2=13.7%, no collinearities

Need for more discriminative contexts

- More frequent words are notoriously less discriminative
- Our predictions show that lack of discriminative contexts leads to an eventive reading
- To improve the performance of our classifier on objects:
  - Use only lower frequency words in training and testing
  - From whole sentence to smaller window

Need for more discriminative contexts

- More frequent words are notoriously less discriminative
- Our predictions show that lack of discriminative contexts leads to an eventive reading
- To improve the performance of our classifier on objects:
  - Use only lower frequency words in training and testing
  - From whole sentence to smaller window

Need for more discriminative contexts

- More frequent words are notoriously less discriminative
- Our predictions show that lack of discriminative contexts leads to an eventive reading
- To improve the performance of our classifier on objects:
  - Use only lower frequency words in training and testing
  - From whole sentence to smaller window

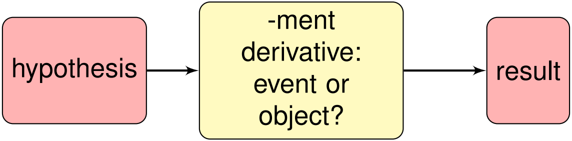
Need for more discriminative contexts

- More frequent words are notoriously less discriminative
- Our predictions show that lack of discriminative contexts leads to an eventive reading
- To improve the performance of our classifier on objects:
  - Use only lower frequency words in training and testing
  - From whole sentence to smaller window

Need for more discriminative contexts

- More frequent words are notoriously less discriminative
- Our predictions show that lack of discriminative contexts leads to an eventive reading
- To improve the performance of our classifier on objects:
  - Use only lower frequency words in training and testing
  - From whole sentence to smaller window

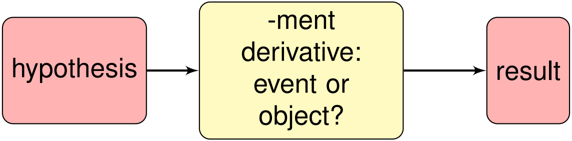
Summary



- Can we use Distributional Semantics tools to successfully disambiguate derived words in context?

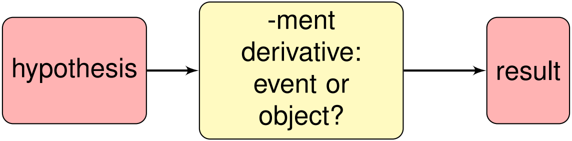
  - Can we distinguish between event and object readings?
    Yes, but we could do better
  - How do we classify the ambiguous cases?
    As events, but it is not so clear why

Summary



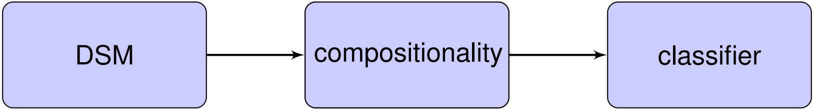
- Can we use Distributional Semantics tools to successfully disambiguate derived words in context?
  - Can we distinguish between event and object readings?
    Yes, but we could do better
  - How do we classify the ambiguous cases?
    As events, but it is not so clear why

Summary



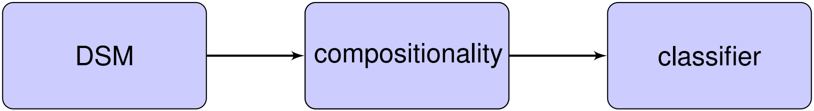
- Can we use Distributional Semantics tools to successfully disambiguate derived words in context?
    - Can we distinguish between event and object readings?
      Yes, but we could do better
    - How do we classify the ambiguous cases?
      As events, but it is not so clear why

Where do we go from here?



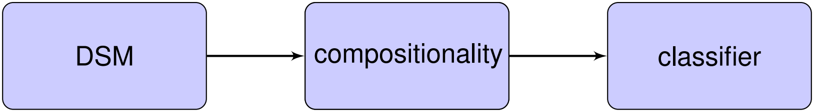
- How can we improve the classifier?

- DSM: reliable
- Compositionality (how we compute sentence vectors): further work
  - Use only lower frequency words in training and testing
  - From whole sentence to smaller window
- Classifier: computationally reliable, but we may improve the properties of the training data



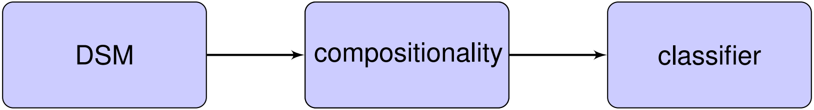
- How can we improve the classifier?
  - DSM: reliable
  - Compositionality (how we compute sentence vectors): further work
    - Use only lower frequency words in training and testing
    - From whole sentence to smaller window
  - Classifier: computationally reliable, but we may improve the properties of the training data

Where do we go from here?



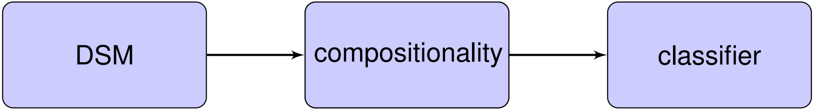
- How can we improve the classifier?
    - DSM: reliable
    - Compositionality (how we compute sentence vectors): further work
        - Use only lower frequency words in training and testing
        - From whole sentence to smaller window
    - Classifier: computationally reliable, but we may improve the properties of the training data

Where do we go from here?



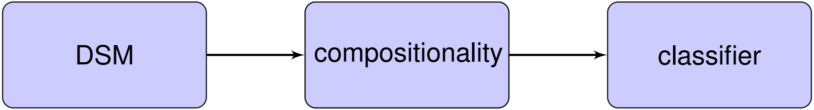
- How can we improve the classifier?
  - DSM: reliable
  - Compositionality (how we compute sentence vectors): further work
    - Use only lower frequency words in training and testing
    - From whole sentence to smaller window
  - Classifier: computationally reliable, but we may improve the properties of the training data

Where do we go from here?



- How can we improve the classifier?
  - DSM: reliable
  - Compositionality (how we compute sentence vectors): further work
    - Use only lower frequency words in training and testing
    - From whole sentence to smaller window
  - Classifier: computationally reliable, but we may improve the properties of the training data

Where do we go from here?



- How can we improve the classifier?
  - DSM: reliable
  - Compositionality (how we compute sentence vectors): further work
    - Use only lower frequency words in training and testing
    - From whole sentence to smaller window
  - Classifier: computationally reliable, but we may improve the properties of the training data

Thank you!

References

Buitelaar, P. (1998). *CoreLex: Systematic Polysemy and Underspecification*. PhD Thesis, Computer Science Brandeis University.

Firth, J. (1957). A synopsis of linguistic theory 1930-1955. *Studies in Linguistic Analysis (special volume of the Philological Society), 1952-59*, 1-32.

Harris, Z. (1954). Distributional structure. *Word*, 10(23), 146Ð162.

Kawaletz, L., & Plag, I. (2015). Predicting the semantics of English nominalizations: A frame-based analysis of -ment suffixation. In L. Bauer, L. Körtvélyessy, & P. Štekauer(Eds.), *Semantics of complex words* (Vol. 3, p. 289-319). Dordrecht: Springer.

Lieber, R. (in press). *English nouns: The ecology of nominalization*. Cambridge: Cambridge

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient estimation of word representations in vector space*. CoRR.

Plag, I., Andreou, M., & Kawaletz, L., (2016). A frame-semantic approach to polysemy in affixation. In *The lexeme in descriptive and theoretical morphology*.

Schütze, H. (1998). Automatic word sense discrimination. *Computational Linguistics* 27(1), 97-123.

Technical details: DSM

- BNC + UkWac: 3.6 bln tokens
- Inflected words, frequency > 14 (800k target words)
- Window size: symmetric window, 5 words; state-of-the-art extraction algorithm (Mikolov et al., 2013)