Modeling covert event retrieval in logical metonymy: probabilistic and distributional accounts

Alessandra Zarcone[1], Jason Utt[1], Sebastian Padó[2]
[1]Institut für Maschinelle Sprachverarbeitung, Stuttgart
[2]Institut für Computerlinguistik, Heidelberg
zarconaa@ims.uni-stuttgart.de,
uttjn@ims.uni-stuttgart.de, pado@cl.uni-heidelberg.de

CMCL 2012
Montréal, June 7th 2012

Outline

Logical metonymy and covert events

Logical metonymy:

begin the newspaper	⟶	begin **reading** the newspaper
enjoy the beer	⟶	enjoy **drinking** the beer

↓

Covert Events (CE)

- not realized on the surface, but understood
- influence reading times, available for inference
- a challenge to compositionality

Logical metonymy and covert events

Logical metonymy:

begin the newspaper	⟶	begin **reading** the newspaper
enjoy the beer	⟶	enjoy **drinking** the beer

↓

Covert Events (CE)

- not realized on the surface, but understood
- influence reading times, available for inference
- a challenge to compositionality

Logical metonymy and covert events

Logical metonymy:

begin the newspaper	⟶	begin **reading** the newspaper
enjoy the beer	⟶	enjoy **drinking** the beer

↓

Covert Events (CE)

- not realized on the surface, but understood
- influence reading times, available for inference
- a challenge to compositionality

Logical metonymy and covert events

Logical metonymy:

begin the newspaper	⟶	begin **reading** the newspaper
enjoy the beer	⟶	enjoy **drinking** the beer

↓

Covert Events (CE)

- not realized on the surface, but understood
- influence reading times, available for inference
- a challenge to compositionality

Accounts of logical metonymy

begin the newspaper	⟶	begin **reading** the newspaper
enjoy the beer	⟶	enjoy **drinking** the beer

Lexical account [Pustejovsky, 1995]:

- ontological trigger: CEs triggered by a type-mismatch (event-subcat. verb + entity-denoting obj.)
- qualia structures: CEs from qualia structure in the lexicon

Pragmatic account [Fodor and Lepore, 1998, De Almeida and Dwivedi, 2008]:

- dynamic inferences (world knowledge and communication principles)
- post-lexical information

Accounts of logical metonymy

begin the newspaper	⟶	begin **reading** the newspaper
enjoy the beer	⟶	enjoy **drinking** the beer

Lexical account [Pustejovsky, 1995]:

- ontological trigger: CEs triggered by a type-mismatch (event-subcat. verb + entity-denoting obj.)
- qualia structures: CEs from qualia structure in the lexicon

Pragmatic account [Fodor and Lepore, 1998, De Almeida and Dwivedi, 2008]:

- dynamic inferences (world knowledge and communication principles)
- post-lexical information

Accounts of logical metonymy

begin the newspaper	⟶	begin **reading** the newspaper
enjoy the beer	⟶	enjoy **drinking** the beer

Lexical account [Pustejovsky, 1995]:

- ontological trigger: CEs triggered by a type-mismatch (event-subcat. verb + entity-denoting obj.)
- qualia structures: CEs from qualia structure in the lexicon

Pragmatic account [Fodor and Lepore, 1998, De Almeida and Dwivedi, 2008]:

- dynamic inferences (world knowledge and communication principles)
- post-lexical information

Accounts of logical metonymy

begin the newspaper	⟶	begin **reading** the newspaper
enjoy the beer	⟶	enjoy **drinking** the beer

Lexical account [Pustejovsky, 1995]:

- ontological trigger: CEs triggered by a type-mismatch (event-subcat. verb + entity-denoting obj.)
- qualia structures: CEs from qualia structure in the lexicon

Pragmatic account [Fodor and Lepore, 1998, De Almeida and Dwivedi, 2008]:

- dynamic inferences (world knowledge and communication principles)
- post-lexical information

Accounts of logical metonymy

begin the newspaper	⟶	begin **reading** the newspaper
enjoy the beer	⟶	enjoy **drinking** the beer

Lexical account [Pustejovsky, 1995]:

- ontological trigger: CEs triggered by a type-mismatch (event-subcat. verb + entity-denoting obj.)
- qualia structures: CEs from qualia structure in the lexicon

Pragmatic account [Fodor and Lepore, 1998, De Almeida and Dwivedi, 2008]:

- dynamic inferences (world knowledge and communication principles)
- post-lexical information

Effects of typicality / thematic fit

Selectional preferences
[Ferretti et al., 2001,
Bicknell et al., 2010]:

- $arrest \xrightarrow{agent} cop$
- $\langle journalist, check \rangle \xrightarrow{patient} spelling$
- $\langle mechanic, check \rangle \xrightarrow{patient} car$

Logical metonymy
[Zarcone and Padó, 2011,
Zarcone et al., 2012]:

- $\langle confectioner, finish, icing \rangle \xrightarrow{CE} spread$
- $\langle child, finish, icing \rangle \xrightarrow{CE} eat$

A test bed for cognitively plausible models of language:

- sensitive to context and typicality effects
- interpretation of implicit content (CEs)
- between lexical semantics and world knowledge

Effects of typicality / thematic fit

Selectional preferences

[Ferretti et al., 2001, Bicknell et al., 2010]:

- $\mathit{arrest} \xrightarrow{\mathit{agent}} \mathit{cop}$
- $\langle \mathit{journalist}, \mathit{check} \rangle \xrightarrow{\mathit{patient}} \mathit{spelling}$
- $\langle \mathit{mechanic}, \mathit{check} \rangle \xrightarrow{\mathit{patient}} \mathit{car}$

Logical metonymy

[Zarcone and Padó, 2011, Zarcone et al., 2012]:

- $\langle \mathit{confectioner}, \mathit{finish}, \mathit{icing} \rangle \xrightarrow{\mathit{CE}} \mathit{spread}$
- $\langle \mathit{child}, \mathit{finish}, \mathit{icing} \rangle \xrightarrow{\mathit{CE}} \mathit{eat}$

A test bed for cognitively plausible models of language:

- sensitive to context and typicality effects
- interpretation of implicit content (CEs)
- between lexical semantics and world knowledge

Effects of typicality / thematic fit

Selectional preferences

[Ferretti et al., 2001, Bicknell et al., 2010]:

- *arrest* $\xrightarrow{agent}$ *cop*
- $\langle$*journalist*, *check*$\rangle$ $\xrightarrow{patient}$ *spelling*
- $\langle$*mechanic*, *check*$\rangle$ $\xrightarrow{patient}$ *car*

Logical metonymy

[Zarcone and Padó, 2011, Zarcone et al., 2012]:

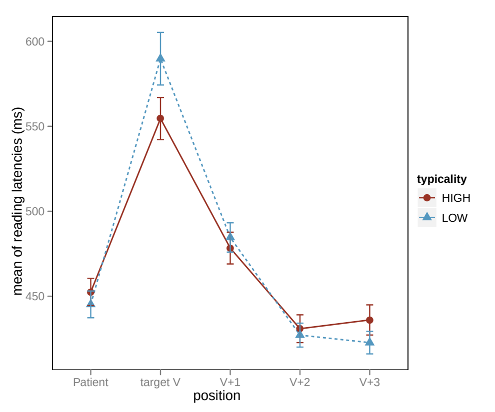
- $\langle$*confectioner*, *finish*, *icing*$\rangle$ $\xrightarrow{CE}$ *spread*
- $\langle$*child*, *finish*, *icing*$\rangle$ $\xrightarrow{CE}$ *eat*

A test bed for cognitively plausible models of language:

- sensitive to context and typicality effects
- interpretation of implicit content (CEs)
- between lexical semantics and world knowledge

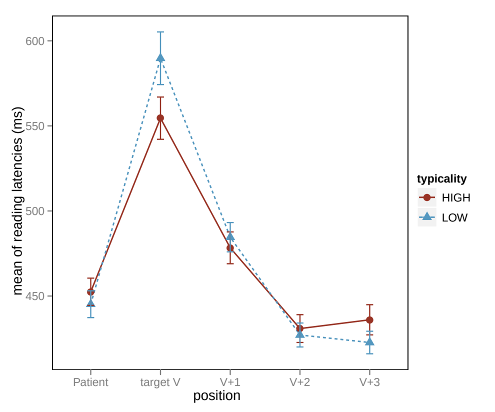
Effects of typicality / thematic fit

Der Konditor / das Kind	hörte auf,	die Glasur	aufzutragen	und fing mit..
The baker / the child	finished	the icing	to spread	and started with...



Effects of typicality / thematic fit

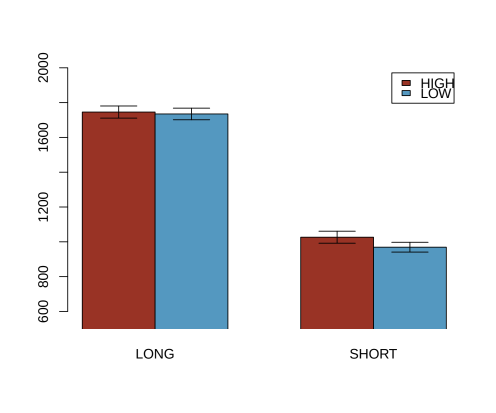
Der Konditor / **das Kind**	hörte auf,	die Glasur	aufzutragen	und fing mit..
The baker / **the child**	finished	the icing	to spread	and started with...



Self-paced reading:
↓
facilitation effect on **high typicality** CEs

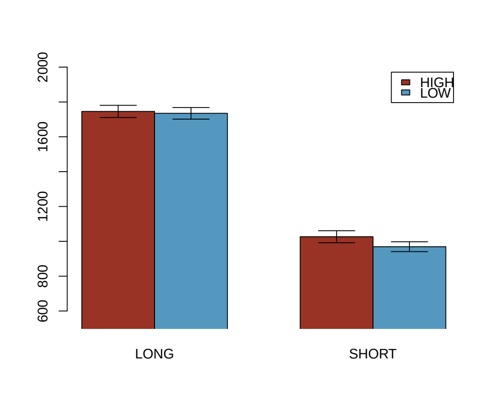
Effects of typicality / thematic fit

Der Konditor / das Kind hörte mit der Glasur auf → **AUFTRAGEN**
The baker / the child finished with the icing → **SPREAD**



Effects of typicality / thematic fit

Der Konditor / das Kind	hörte mit	der Glasur auf	→	**AUFTRAGEN**
The baker / the child	finished with	the icing	→	**SPREAD**



Probe recognition:
("was the probe in the sentence?")
↓
facilitation effect on **low typicality** CEs

Task

- **48 test sentence pairs from the psycholinguistic experiments**:

Der Braumeister vermied das Bier	→	**brauen**	/	**trinken**
The brewer avoided the beer	→	**brew**	/	**drink**

- **48 tuple pairs for the model evaluation**:

			CE	
S	V	O	high-typicality	low-typicality
Braumeister	vermeiden	Bier	brauen	trinken
Student	vermeiden	Bier	trinken	brauen

- **Evaluation task**: given S, V and O, choose the high-typicality CE over the low-typicality CE

Task

- **48 test sentence pairs from the psycholinguistic experiments**:

  Der Braumeister vermied das Bier → **brauen** / **trinken**
  The brewer avoided the beer → **brew** / **drink**

- **48 tuple pairs for the model evaluation**:

S	V	O	CE: high-typicality	CE: low-typicality
Braumeister	vermeiden	Bier	brauen	trinken
Student	vermeiden	Bier	trinken	brauen

- **Evaluation task**: given S, V and O, choose the high-typicality CE over the low-typicality CE

Task

- **48 test sentence pairs from the psycholinguistic experiments**:

  Der Braumeister vermied das Bier → **brauen** / **trinken**
  The brewer avoided the beer → **brew** / **drink**

- **48 tuple pairs for the model evaluation**:

			CE	
S	V	O	high-typicality	low-typicality
Braumeister	vermeiden	Bier	brauen	trinken
Student	vermeiden	Bier	trinken	brauen

- **Evaluation task**: given S, V and O, choose the high-typicality CE over the low-typicality CE

Two compositional models

Probabilistic models

- based on [Lapata et al., 2003] and [Lapata and Lascarides, 2003]
- first-order co-occurrence information
- most probable event

Similarity-based models

- based on [Lenci, 2011]
- higher-order co-occurrence information
- most similar event to prototypical event

Novelty

- German data
- large web corpus
- first similarity-based account of logical metonymy

Two compositional models

Probabilistic models

- based on [Lapata et al., 2003] and [Lapata and Lascarides, 2003]
- first-order co-occurrence information
- most probable event

Similarity-based models

- based on [Lenci, 2011]
- higher-order co-occurrence information
- most similar event to prototypical event

Novelty

- German data
- large web corpus
- first similarity-based account of logical metonymy

Probabilistic models

Probabilistic models

Established model for the task: probabilistic models of logical metonymy [Lapata et al., 2003, Lapata and Lascarides, 2003]

- logical metonymy interpretation as joint distribution $P(,,,)$
- two models

SOV_p: CE in a given context maximizes $P(s, v, o, e)$:

$$\hat{e} = \arg\max_e P(e)\ P(o|e)\ P(v|e)\ P(s|e)$$

SO_p: CE in a given context maximizes $P(s, o, e)$:

$$\hat{e} = \arg\max_e P(e)\ P(o|e)\ P(s|e)$$

- probabilities estimated from corpus extracted co-occurrences (simple co-occurrences)

Probabilistic models

Established model for the task: probabilistic models of logical metonymy [Lapata et al., 2003, Lapata and Lascarides, 2003]

- logical metonymy interpretation as joint distribution $P(s, v, o, e)$

the student avoided the beer $\longrightarrow$ drinking / brewing

- two models

SOV_p: CE in a given context maximizes $P(s, v, o, e)$:

$$\hat{e} = \arg\max_e P(e)\ P(o|e)\ P(v|e)\ P(s|e)$$

SO_p: CE in a given context maximizes $P(s, o, e)$:

$$\hat{e} = \arg\max_e P(e)\ P(o|e)\ P(s|e)$$

- probabilities estimated from corpus extracted co-occurrences (simple co-occurrences)

Probabilistic models

Established model for the task: probabilistic models of logical metonymy [Lapata et al., 2003, Lapata and Lascarides, 2003]

- logical metonymy interpretation as joint distribution $P(\mathbf{s}, v, o, e)$

the student avoided the beer ⟶ drinking / brewing

- two models

SOV_p: CE in a given context maximizes $P(s, v, o, e)$:

$$\hat{e} = \arg\max_e P(e)\ P(o|e)\ P(v|e)\ P(s|e)$$

SO_p: CE in a given context maximizes $P(s, o, e)$:

$$\hat{e} = \arg\max_e P(e)\ P(o|e)\ P(s|e)$$

- probabilities estimated from corpus extracted co-occurrences (simple co-occurrences)

Probabilistic models

Established model for the task: probabilistic models of logical metonymy [Lapata et al., 2003, Lapata and Lascarides, 2003]

- logical metonymy interpretation as joint distribution $P(s, \mathbf{v}, o, e)$

  the student **avoided** the beer $\longrightarrow$ drinking / brewing

- two models

  SOV_p: CE in a given context maximizes $P(s, v, o, e)$:

  $$\hat{e} = \arg\max_e P(e)\ P(o|e)\ P(v|e)\ P(s|e)$$

  SO_p: CE in a given context maximizes $P(s, o, e)$:

  $$\hat{e} = \arg\max_e P(e)\ P(o|e)\ P(s|e)$$

- probabilities estimated from corpus extracted co-occurrences (simple co-occurrences)

Probabilistic models

Established model for the task: probabilistic models of logical metonymy [Lapata et al., 2003, Lapata and Lascarides, 2003]

- logical metonymy interpretation as joint distribution $P(s, v, \mathbf{o}, e)$

the student avoided **the beer** ⟶ drinking / brewing

- two models

SOV_p: CE in a given context maximizes $P(s, v, o, e)$:

$$\hat{e} = \arg\max_e P(e)\ P(o|e)\ P(v|e)\ P(s|e)$$

SO_p: CE in a given context maximizes $P(s, o, e)$:

$$\hat{e} = \arg\max_e P(e)\ P(o|e)\ P(s|e)$$

- probabilities estimated from corpus extracted co-occurrences (simple co-occurrences)

Probabilistic models

Established model for the task: probabilistic models of logical metonymy [Lapata et al., 2003, Lapata and Lascarides, 2003]

- logical metonymy interpretation as joint distribution $P(s, v, o, \mathbf{e})$

  the student avoided the beer $\longrightarrow$ **drinking** / **brewing**

- two models

  SOV_p: CE in a given context maximizes $P(s, v, o, e)$:

  $$\hat{e} = \arg\max_e P(e)\ P(o|e)\ P(v|e)\ P(s|e)$$

  SO_p: CE in a given context maximizes $P(s, o, e)$:

  $$\hat{e} = \arg\max_e P(e)\ P(o|e)\ P(s|e)$$

- probabilities estimated from corpus extracted co-occurrences (simple co-occurrences)

Probabilistic models

Established model for the task: probabilistic models of logical metonymy [Lapata et al., 2003, Lapata and Lascarides, 2003]

- logical metonymy interpretation as joint distribution $P(s, v, o, e)$

  the student avoided the beer $\longrightarrow$ drinking / brewing

- two models

  SOV_p: CE in a given context maximizes $P(s, v, o, e)$:

  $$\hat{e} = \arg\max_e P(e)\ P(o|e)\ P(v|e)\ P(s|e)$$

  SO_p: CE in a given context maximizes $P(s, o, e)$:

  $$\hat{e} = \arg\max_e P(e)\ P(o|e)\ P(s|e)$$

- probabilities estimated from corpus extracted co-occurrences (simple co-occurrences)

Probabilistic models

Established model for the task: probabilistic models of logical metonymy [Lapata et al., 2003, Lapata and Lascarides, 2003]

- logical metonymy interpretation as joint distribution $P(s, v, o, e)$

the student avoided the beer $\longrightarrow$ drinking / brewing

- two models

SOV$_p$: CE in a given context maximizes $P(s, v, o, e)$:

$$\hat{e} = \arg\max_e P(e)\ P(o|e)\ P(v|e)\ P(s|e)$$

SO$_p$: CE in a given context maximizes $P(s, o, e)$:

$$\hat{e} = \arg\max_e P(e)\ P(o|e)\ P(s|e)$$

- probabilities estimated from corpus extracted co-occurrences (simple co-occurrences)

Probabilistic models

Established model for the task: probabilistic models of logical metonymy [Lapata et al., 2003, Lapata and Lascarides, 2003]

- logical metonymy interpretation as joint distribution $P(s, v, o, e)$

the student avoided the beer $\longrightarrow$ drinking / brewing

- two models

SOV$_p$: CE in a given context maximizes $P(s, v, o, e)$:

$$\hat{e} = \arg\max_e P(e)\ P(o|e)\ P(v|e)\ P(s|e)$$

SO$_p$: CE in a given context maximizes $P(s, o, e)$:

$$\hat{e} = \arg\max_e P(e)\ P(o|e)\ P(s|e)$$

- probabilities estimated from corpus extracted co-occurrences (simple co-occurrences)

Probabilistic models

Established model for the task: probabilistic models of logical metonymy [Lapata et al., 2003, Lapata and Lascarides, 2003]

- logical metonymy interpretation as joint distribution $P(s, v, o, e)$

the student avoided the beer $\longrightarrow$ drinking / brewing

- two models

SOV_p: CE in a given context maximizes $P(s, v, o, e)$:

$$\hat{e} = \arg\max_e P(e)\ P(o|e)\ P(v|e)\ P(s|e)$$

SO_p: CE in a given context maximizes $P(s, o, e)$:

$$\hat{e} = \arg\max_e P(e)\ P(o|e)\ P(s|e)$$

- probabilities estimated from corpus extracted co-occurrences (simple co-occurrences)

Probabilistic models

Probabilistic baseline

the student avoided the beer ⟶ drinking / brewing

B_p: CE in a given context maximizes $P(o, e)$:

$$\hat{e} = \arg\max_e P(e)\ P(o|e)$$

- given our dataset, the baseline reaches 50% accuracy, because the dataset is counterbalanced:

S	V	O	CE: high-typicality	CE: low-typicality
Braumeister	vermeiden	Bier	brauen	trinken
Student	vermeiden	Bier	trinken	brauen

Probabilistic models

Probabilistic baseline

the student avoided **the beer** ⟶ drinking / brewing

B_p: CE in a given context maximizes $P(o, e)$:

$$\hat{e} = \arg\max_e P(e)\ P(o|e)$$

- given our dataset, the baseline reaches 50% accuracy, because the dataset is counterbalanced:

S	V	O	CE high-typicality	CE low-typicality
Braumeister	vermeiden	Bier	brauen	trinken
Student	vermeiden	Bier	trinken	brauen

Probabilistic models

Probabilistic baseline

the student avoided the beer ⟶ drinking / **brewing**

B_p: CE in a given context maximizes $P(o, e)$:

$$\hat{e} = \arg\max_e P(e)\ P(o|e)$$

- given our dataset, the baseline reaches 50% accuracy, because the dataset is counterbalanced:

			CE	
S	V	O	high-typicality	low-typicality
Braumeister	vermeiden	Bier	brauen	trinken
Student	vermeiden	Bier	trinken	brauen

Probabilistic models

Probabilistic baseline

the student avoided the beer ⟶ drinking / brewing

B_p: CE in a given context maximizes $P(o, e)$:

$$\hat{e} = \arg\max_e P(e)\ P(o|e)$$

- given our dataset, the baseline reaches 50% accuracy, because the dataset is counterbalanced:

			CE	
S	V	O	high-typicality	low-typicality
Braumeister	vermeiden	Bier	brauen	trinken
Student	vermeiden	Bier	trinken	brauen

Similarity-based models

Similarity-based models

Distributional Hypothesis [Harris, 1954, Miller and Charles, 1991]

- words occurring in similar contexts → semantically similar
- meaning of a word → vector of features of its linguistic context
- semantic similarity → vector similarity

↓

A cognitive hypothesis about the form of semantic representations

- word distributional behavior → semantic content (cognitive level)
- graded category membership [Rosch, 1975], multiple sense activation [Erk, 2010]
- lexical development [Li et al., 2004], category-related deficits [Vigliocco et al., 2004], selectional preferences [Erk, 2007]
  → a multi-purpose framework: Distributional Memory

Similarity-based models

Distributional Hypothesis [Harris, 1954, Miller and Charles, 1991]

- words occurring in similar contexts → semantically similar
- meaning of a word → vector of features of its linguistic context
- semantic similarity → vector similarity

↓

A cognitive hypothesis about the form of semantic representations

- word distributional behavior → semantic content (cognitive level)
- graded category membership [Rosch, 1975], multiple sense activation [Erk, 2010]
- lexical development [Li et al., 2004], category-related deficits [Vigliocco et al., 2004], selectional preferences [Erk, 2007]
  → a multi-purpose framework: **Distributional Memory**

Similarity-based models

Distributional Hypothesis [Harris, 1954, Miller and Charles, 1991]

- words occurring in similar contexts → semantically similar
- meaning of a word → vector of features of its linguistic context
- semantic similarity → vector similarity

↓

A cognitive hypothesis about the form of semantic representations

- word distributional behavior → semantic content (cognitive level)
- graded category membership [Rosch, 1975], multiple sense activation [Erk, 2010]
- lexical development [Li et al., 2004], category-related deficits [Vigliocco et al., 2004], selectional preferences [Erk, 2007]
  → a multi-purpose framework: **Distributional Memory**

Similarity-based models

Distributional Hypothesis [Harris, 1954, Miller and Charles, 1991]

- words occurring in similar contexts → semantically similar
- meaning of a word → vector of features of its linguistic context
- semantic similarity → vector similarity

↓

A cognitive hypothesis about the form of semantic representations

- word distributional behavior → semantic content (cognitive level)
- graded category membership [Rosch, 1975], multiple sense activation [Erk, 2010]
- lexical development [Li et al., 2004], category-related deficits [Vigliocco et al., 2004], selectional preferences [Erk, 2007]
  → a multi-purpose framework: **Distributional Memory**

Similarity-based models

Distributional Hypothesis [Harris, 1954, Miller and Charles, 1991]

- words occurring in similar contexts → semantically similar
- meaning of a word → vector of features of its linguistic context
- semantic similarity → vector similarity

↓

A cognitive hypothesis about the form of semantic representations

- word distributional behavior → semantic content (cognitive level)
- graded category membership [Rosch, 1975], multiple sense activation [Erk, 2010]
- lexical development [Li et al., 2004], category-related deficits [Vigliocco et al., 2004], selectional preferences [Erk, 2007]
  → a multi-purpose framework: **Distributional Memory**

Similarity-based models

Distributional Hypothesis [Harris, 1954, Miller and Charles, 1991]

- words occurring in similar contexts → semantically similar
- meaning of a word → vector of features of its linguistic context
- semantic similarity → vector similarity

↓

A cognitive hypothesis about the form of semantic representations

- word distributional behavior → semantic content (cognitive level)
- graded category membership [Rosch, 1975], multiple sense activation [Erk, 2010]
- lexical development [Li et al., 2004], category-related deficits [Vigliocco et al., 2004], selectional preferences [Erk, 2007]
  → a multi-purpose framework: **Distributional Memory**

Similarity-based models

Distributional Hypothesis [Harris, 1954, Miller and Charles, 1991]

- words occurring in similar contexts → semantically similar
- meaning of a word → vector of features of its linguistic context
- semantic similarity → vector similarity

↓

A cognitive hypothesis about the form of semantic representations

- word distributional behavior → semantic content (cognitive level)
- graded category membership [Rosch, 1975], multiple sense activation [Erk, 2010]
- lexical development [Li et al., 2004], category-related deficits [Vigliocco et al., 2004], selectional preferences [Erk, 2007]
  → a multi-purpose framework: **Distributional Memory**

Similarity-based models: DM

Distributional Memory (DM) [Baroni and Lenci, 2010]

- multi-purpose framework in distributional semantics
- **off-line**: tensors of weighted *word-link-word* tuples, each mapped onto a score by a function $\sigma: \langle w_1 \ l \ w_2 \rangle \longrightarrow \mathbb{R}^+$
  - here, syntactic and lexicalized links (TypeDM)
- **on-line**: dependent on task, dedicated semantic space generated from the tensor (e.g. *word by link-word* space $W_1 \times LW_2$)

TypeDM for German

- 884M word SDeWaC web corpus [Faaß et al., 2010] (MATE German dependency parser [Bohnet, 2010])
  - 55M instances of simple syntactic relations (*subj_tr, subj_intr, obj, iobj, comp, nmod*)
  - 104M instances of lexicalized patterns (*noun–prep–noun, adj–noun–(of)–noun*, etc.)

Similarity-based models: DM

Distributional Memory (DM) [Baroni and Lenci, 2010]

- multi-purpose framework in distributional semantics
- **off-line**: tensors of weighted *word-link-word* tuples, each mapped onto a score by a function $\sigma : \langle w_1 \ l \ w_2 \rangle \rightarrow \mathbb{R}^+$
  - here, syntactic and lexicalized links (TypeDM)
- **on-line**: dependent on task, dedicated semantic space generated from the tensor (e.g. *word by link-word* space $W_1 \times LW_2$)

TypeDM for German

- 884M word SDEWAC web corpus [Faaß et al., 2010] (MATE German dependency parser [Bohnet, 2010])
  - 55M instances of simple syntactic relations (*subj_tr, subj_intr, obj, iobj, comp, nmod*)
  - 104M instances of lexicalized patterns (*noun–prep–noun, adj–noun–(of)–noun*, etc.)

Similarity-based models: DM

Distributional Memory (DM) [Baroni and Lenci, 2010]

- multi-purpose framework in distributional semantics
- **off-line**: tensors of weighted *word-link-word* tuples, each mapped onto a score by a function $\sigma: \langle w_1\ l\ w_2 \rangle \rightarrow \mathbb{R}^+$
  - here, syntactic and lexicalized links (TypeDM)
- **on-line**: dependent on task, dedicated semantic space generated from the tensor (e.g. *word by link-word* space $W_1 \times LW_2$)

TypeDM for German

- 884M word SDEWAC web corpus [Faaß et al., 2010] (MATE German dependency parser [Bohnet, 2010])
  - 55M instances of simple syntactic relations (*subj_tr, subj_intr, obj, iobj, comp, nmod*)
  - 104M instances of lexicalized patterns (*noun–prep–noun, adj–noun–(of)–noun*, etc.)

Similarity-based models: DM

Distributional Memory (DM) [Baroni and Lenci, 2010]

- multi-purpose framework in distributional semantics
- **off-line**: tensors of weighted *word-link-word* tuples, each mapped onto a score by a function $\sigma : \langle w_1 \; l \; w_2 \rangle \rightarrow \mathbb{R}^+$
  - here, syntactic and lexicalized links (TypeDM)
- **on-line**: dependent on task, dedicated semantic space generated from the tensor (e.g. *word by link-word* space $W_1 \times LW_2$)

TypeDM for German

- 884M word SDEWAC web corpus [Faaß et al., 2010] (MATE German dependency parser [Bohnet, 2010])
  - 55M instances of simple syntactic relations (*subj_tr*, *subj_intr*, *obj*, *iobj*, *comp*, *nmod*)
  - 104M instances of lexicalized patterns (*noun–prep–noun*, *adj–noun-(of)-noun*, etc.)

Similarity-based models: DM

Distributional Memory (DM) [Baroni and Lenci, 2010]

- multi-purpose framework in distributional semantics
- **off-line**: tensors of weighted *word-link-word* tuples, each mapped onto a score by a function $\sigma : \langle w_1 \; l \; w_2 \rangle \rightarrow \mathbb{R}^+$
  - here, syntactic and lexicalized links (TypeDM)
- **on-line**: dependent on task, dedicated semantic space generated from the tensor (e.g. *word by link-word* space $W_1 \times LW_2$)

TypeDM for German

- 884M word SDeWaC web corpus [Faaß et al., 2010] (MATE German dependency parser [Bohnet, 2010])
  - 55M instances of simple syntactic relations (*subj_tr, subj_intr, obj, iobj, comp, nmod*)
  - 104M instances of lexicalized patterns (*noun–prep–noun, adj–noun-(of)-noun*, etc.)

Similarity-based models: DM

Distributional Memory (DM) [Baroni and Lenci, 2010]

- multi-purpose framework in distributional semantics
- **off-line**: tensors of weighted *word-link-word* tuples, each mapped onto a score by a function $\sigma : \langle w_1 \; l \; w_2 \rangle \rightarrow \mathbb{R}^+$
  - here, syntactic and lexicalized links (TypeDM)
- **on-line**: dependent on task, dedicated semantic space generated from the tensor (e.g. *word by link-word* space $W_1 \times LW_2$)

TypeDM for German

- 884M word SDeWaC web corpus [Faaß et al., 2010] (MATE German dependency parser [Bohnet, 2010])
  - 55M instances of simple syntactic relations (*subj_tr, subj_intr, obj, iobj, comp, nmod*)
  - 104M instances of lexicalized patterns (*noun–prep–noun, adj–noun-(of)-noun*, etc.)

Similarity-based models

- **Beyond word-level**: compositional distributional semantics [Mitchell and Lapata, 2010, Guevara, 2011]
- **Task**: given a verb and different subjects, different impact of the subjects on the semantic expectation for expected objects
  - $\langle journalist, check \rangle \xrightarrow{patient} spelling$
  - $\langle mechanic, check \rangle \xrightarrow{patient} car$
- **Expectation Composition and Update (ECU), [Lenci, 2011]**: predict the degree of thematic fit for verb-argument combinations from TypeDM

Similarity-based models

- **Beyond word-level**: compositional distributional semantics [Mitchell and Lapata, 2010, Guevara, 2011]
- **Task**: given a verb and different subjects, different impact of the subjects on the semantic expectation for expected objects
  - $\langle journalist, check \rangle \xrightarrow{patient} spelling$
  - $\langle mechanic, check \rangle \xrightarrow{patient} car$
- **Expectation Composition and Update (ECU), [Lenci, 2011]**: predict the degree of thematic fit for verb-argument combinations from TypeDM

Similarity-based models

- **Beyond word-level**: compositional distributional semantics [Mitchell and Lapata, 2010, Guevara, 2011]
- **Task**: given a verb and different subjects, different impact of the subjects on the semantic expectation for expected objects
  - $\langle journalist, check \rangle \xrightarrow{patient} spelling$
  - $\langle mechanic, check \rangle \xrightarrow{patient} car$
- **Expectation Composition and Update (ECU), [Lenci, 2011]**: predict the degree of thematic fit for verb-argument combinations from TypeDM

Similarity-based models

- **Beyond word-level**: compositional distributional semantics [Mitchell and Lapata, 2010, Guevara, 2011]
- **Task**: given a verb and different subjects, different impact of the subjects on the semantic expectation for expected objects
  - $\langle journalist, check \rangle \xrightarrow{patient} spelling$
  - $\langle mechanic, check \rangle \xrightarrow{patient} car$
- **Expectation Composition and Update (ECU), [Lenci, 2011]**: predict the degree of thematic fit for verb-argument combinations from TypeDM

Similarity-based models: ECU

Expectation Composition and Update (ECU) [Lenci, 2011]

Compute thematic fit for *car* and *spelling* as objects of $\langle journalist, check \rangle$

1. **prototypical filler**
   1. compute expectations for the object (weighted sets of objects)
   2. compose (sum or product) and update
   3. prototype object as centroid of $W_1 \times LW_2$ vectors of the 20 most expected objects
2. **object thematic fit**: similarity of a noun to the prototype object
3. compare thematic fit of *car* and *spelling*

Similarity-based models: ECU

Expectation Composition and Update (ECU) [Lenci, 2011]

Compute thematic fit for *car* and *spelling* as objects of $\langle journalist, check \rangle$

1. **prototypical filler**
   1. compute expectations for the object (weighted sets of objects)
      - verb's expectations: $EX_V(v) = \lambda o.\sigma(\langle v\ obj^{-1}\ o \rangle)$
      - subject's expectations: $EX_S(s) = \lambda o.\sigma(\langle s\ verb\ o \rangle)$
   2. compose (sum or product) and update
      - $EX_{SV}(s, v) = \lambda o.\, EX_V(v)(o) \circ EX_S(s)(o)$
   3. prototype object as centroid of $W_1 \times LW_2$ vectors of the 20 most expected objects
2. **object thematic fit**: similarity of a noun to the prototype object
3. compare thematic fit of *car* and *spelling*

Similarity-based models: ECU

Expectation Composition and Update (ECU) [Lenci, 2011]

Compute thematic fit for *car* and *spelling* as objects of $\langle journalist, check \rangle$

1. **prototypical filler**
   1. compute expectations for the object (weighted sets of objects)
      - verb's expectations: $EX_V(v) = \lambda o.\, \sigma(\langle v\ obj^{-1}\ o \rangle)$
      - subject's expectations: $EX_S(s) = \lambda o.\, \sigma(\langle s\ verb\ o \rangle)$
   2. compose (sum or product) and update
      - $EX_{SV}(s, v) = \lambda o.\, EX_V(v)(o) \circ EX_S(s)(o)$
   3. prototype object as centroid of $W_1 \times LW_2$ vectors of the 20 most expected objects
2. **object thematic fit**: similarity of a noun to the prototype object
3. compare thematic fit of *car* and *spelling*

Similarity-based models: ECU

Expectation Composition and Update (ECU) [Lenci, 2011]

Compute thematic fit for *car* and *spelling* as objects of $\langle journalist, check \rangle$

1. **prototypical filler**
   1. compute expectations for the object (weighted sets of objects)
      - verb's expectations: $EX_V(v) = \lambda o.\, \sigma(\langle v\ obj^{-1}\ o \rangle)$
      - subject's expectations: $EX_S(s) = \lambda o.\, \sigma(\langle s\ verb\ o \rangle)$
   2. compose (sum or product) and update
      - $EX_{SV}(s, v) = \lambda o.\, EX_V(v)(o) \circ EX_S(s)(o)$
   3. prototype object as centroid of $W_1 \times LW_2$ vectors of the 20 most expected objects
2. **object thematic fit**: similarity of a noun to the prototype object
3. compare thematic fit of *car* and *spelling*

Similarity-based models: ECU

Expectation Composition and Update (ECU) [Lenci, 2011]

Compute thematic fit for *car* and *spelling* as objects of $\langle journalist, check \rangle$

1. **prototypical filler**
   1. compute expectations for the object (weighted sets of objects)
      - verb's expectations: $EX_V(v) = \lambda o.\, \sigma(\langle v\ obj^{-1}\ o \rangle)$
      - subject's expectations: $EX_S(s) = \lambda o.\, \sigma(\langle s\ verb\ o \rangle)$
   2. compose (sum or product) and update
      - $EX_{SV}(s, v) = \lambda o.\, EX_V(v)(o) \circ EX_S(s)(o)$
   3. prototype object as centroid of $W_1 \times LW_2$ vectors of the 20 most expected objects
2. **object thematic fit**: similarity of a noun to the prototype object
3. compare thematic fit of *car* and *spelling*

Similarity-based models: ECU

Expectation Composition and Update (ECU) [Lenci, 2011]

Compute thematic fit for *car* and *spelling* as objects of $\langle journalist, check \rangle$

1. **prototypical filler**
   1. compute expectations for the object (weighted sets of objects)
      - verb's expectations: $EX_V(v) = \lambda o.\, \sigma(\langle v\ obj^{-1}\ o \rangle)$
      - subject's expectations: $EX_S(s) = \lambda o.\, \sigma(\langle s\ verb\ o \rangle)$
   2. compose (sum or product) and update
      - $EX_{SV}(s, v) = \lambda o.\, EX_V(v)(o) \circ EX_S(s)(o)$
   3. prototype object as centroid of $W_1 \times LW_2$ vectors of the 20 most expected objects
2. **object thematic fit**: similarity of a noun to the prototype object
3. compare thematic fit of *car* and *spelling*

Similarity-based models: ECU for logical metonymy

ECU as a model of logical metonymy

Compute thematic fit for *drink* and *brew* as CEs of $\langle brewer, avoid, beer \rangle$

1. **prototypical CE**
   1. compute expectations for the CE (weighted sets of events)
   2. compose (sum or product) and update
   3. prototype CE
2. **event thematic fit**: similarity of an event to the prototype CE
3. compare thematic fit of *drink* and *brew*

Similarity-based models: ECU for logical metonymy

ECU as a model of logical metonymy

Compute thematic fit for *drink* and *brew* as CEs of $\langle brewer, avoid, beer \rangle$

1. **prototypical CE**
   1. compute expectations for the CE (weighted sets of events)
      - subject's expectations: $EX_S(s) = \lambda e.\sigma(\langle s\ subj\ e \rangle)$
      - object's expectations: $EX_O(o) = \lambda e.\sigma(\langle o\ obj\ e \rangle)$
      - metonymic verb's expectations: $EX_V(v) = \lambda e.\sigma(\langle v\ comp^{-1}\ e \rangle)$
   2. compose (sum or product) and update
   3. prototype CE
2. **event thematic fit**: similarity of an event to the prototype CE
3. compare thematic fit of *drink* and *brew*

Similarity-based models: ECU for logical metonymy

ECU as a model of logical metonymy

Compute thematic fit for *drink* and *brew* as CEs of $\langle brewer, avoid, beer\rangle$

1. **prototypical CE**
   1. compute expectations for the CE (weighted sets of events)
      - subject's expectations: $EX_S(s) = \lambda e.\,\sigma(\langle s\ subj\ e\rangle)$
      - object's expectations: $EX_O(o) = \lambda e.\,\sigma(\langle o\ obj\ e\rangle)$
      - metonymic verb's expectations: $EX_V(v) = \lambda e.\,\sigma(\langle v\ comp^{-1}\ e\rangle$
   2. compose (sum or product) and update
   3. prototype CE
2. **event thematic fit**: similarity of an event to the prototype CE
3. compare thematic fit of *drink* and *brew*

Similarity-based models: ECU for logical metonymy

ECU as a model of logical metonymy

Compute thematic fit for *drink* and *brew* as CEs of $\langle brewer, avoid, beer \rangle$

1. **prototypical CE**
   1. compute expectations for the CE (weighted sets of events)
      - subject's expectations: $EX_S(s) = \lambda e.\, \sigma(\langle s\ subj\ e \rangle)$
      - object's expectations: $EX_O(o) = \lambda e.\, \sigma(\langle o\ obj\ e \rangle)$
      - metonymic verb's expectations: $EX_V(v) = \lambda e.\, \sigma(\langle v\ comp^{-1}\ e \rangle$
   2. compose (sum or product) and update
   3. prototype CE
2. **event thematic fit**: similarity of an event to the prototype CE
3. compare thematic fit of *drink* and *brew*

Similarity-based models: ECU for logical metonymy

ECU as a model of logical metonymy

Compute thematic fit for *drink* and *brew* as CEs of $\langle brewer, avoid, beer \rangle$

1. **prototypical CE**
   1. compute expectations for the CE (weighted sets of events)
      - subject's expectations: $EX_S(s) = \lambda e.\, \sigma(\langle s\ subj\ e \rangle)$
      - object's expectations: $EX_O(o) = \lambda e.\, \sigma(\langle o\ obj\ e \rangle)$
      - metonymic verb's expectations: $EX_V(v) = \lambda e.\, \sigma(\langle v\ comp^{-1}\ e \rangle$
   2. compose (sum or product) and update
   3. prototype CE
2. **event thematic fit**: similarity of an event to the prototype CE
3. compare thematic fit of *drink* and *brew*

Similarity-based models: ECU for logical metonymy

ECU as a model of logical metonymy

Compute thematic fit for *drink* and *brew* as CEs of $\langle brewer, avoid, beer \rangle$

1. **prototypical CE**
   1. compute expectations for the CE (weighted sets of events)
      - subject's expectations: $EX_S(s) = \lambda e.\, \sigma(\langle s\ subj\ e \rangle)$
      - object's expectations: $EX_O(o) = \lambda e.\, \sigma(\langle o\ obj\ e \rangle)$
      - metonymic verb's expectations: $EX_V(v) = \lambda e.\, \sigma(\langle v\ comp^{-1}\ e \rangle$
   2. compose (sum or product) and update
   3. prototype CE
2. **event thematic fit**: similarity of an event to the prototype CE
3. compare thematic fit of *drink* and *brew*

Similarity-based models: ECU for logical metonymy

the student avoided the beer ⟶ drinking / brewing

SOV: composing expectations from subject, object, metonymic verb

- SOV_{Σ}: composition function is sum
- SOV_{Π}: composition function is product

SO: composing expectations from subject and object

- SO_{Σ}: composition function is sum
- SO_{Π}: composition function is product

Similarity-based models: ECU for logical metonymy

the student avoided the beer ⟶ drinking / brewing

SOV: composing expectations from subject, object, metonymic verb

- SOV_{Σ}: composition function is sum
- SOV_{Π}: composition function is product

SO: composing expectations from subject and object

- SO_{Σ}: composition function is sum
- SO_{Π}: composition function is product

Similarity-based models: ECU for logical metonymy

the student avoided **the beer** ⟶ drinking / brewing

SOV: composing expectations from subject, object, metonymic verb

- SOV_{Σ}: composition function is sum
- SOV_{Π}: composition function is product

SO: composing expectations from subject and object

- SO_{Σ}: composition function is sum
- SO_{Π}: composition function is product

Similarity-based models: ECU for logical metonymy

the student avoided the beer ⟶ drinking / brewing

B_s similarity-based baseline, expectations from object only

- given our dataset, the baseline reaches 50% accuracy, because the dataset is counterbalanced:

S	V	O	CE high-typicality	CE low-typicality
Braumeister	vermeiden	Bier	brauen	trinken
Student	vermeiden	Bier	trinken	brauen

Similarity-based models: ECU for logical metonymy

the student avoided **the beer** ⟶ drinking / brewing

B_s similarity-based baseline, expectations from object only

- given our dataset, the baseline reaches 50% accuracy, because the dataset is counterbalanced:

S	V	O	CE high-typicality	CE low-typicality
Braumeister	vermeiden	Bier	brauen	trinken
Student	vermeiden	Bier	trinken	brauen

Similarity-based models: ECU for logical metonymy

the student avoided the beer ⟶ drinking / brewing

B_s similarity-based baseline, expectations from object only

- given our dataset, the baseline reaches 50% accuracy, because the dataset is counterbalanced:

			CE	
S	V	O	high-typicality	low-typicality
Braumeister	vermeiden	Bier	brauen	trinken
Student	vermeiden	Bier	trinken	brauen

Evaluation

coverage: $\frac{(\#\text{ answered datapoints})}{(\#\text{ tot. datapoints})}$

(percentage of datapoints for which a model can make a prediction)

accuracy: $\frac{(\#\text{ correct answers})}{(\#\text{ answered datapoints})}$

(covered datapoints only, ratio of correct predictions to the number of predictions)

backoff accuracy: $coverage \times accuracy + ((1 - coverage) \times 0.5)$

(emulating a backoff procedure with baseline performance for non-covered items)

differences between models: χ^2 test

(2×2 correct - incorrect answer contingency matrix)

Evaluation

coverage: $\frac{(\#\ \text{answered datapoints})}{(\#\ \text{tot. datapoints})}$

(percentage of datapoints for which a model can make a prediction)

accuracy: $\frac{(\#\ \text{correct answers})}{(\#\ \text{answered datapoints})}$

(covered datapoints only, ratio of correct predictions to the number of predictions)

backoff accuracy: $coverage \times accuracy + ((1 - coverage) \times 0.5)$

(emulating a backoff procedure with baseline performance for non-covered items)

differences between models: χ^2 test

(2×2 correct - incorrect answer contingency matrix)

Evaluation

coverage: $\frac{(\#\text{ answered datapoints})}{(\#\text{ tot. datapoints})}$

(percentage of datapoints for which a model can make a prediction)

accuracy: $\frac{(\#\text{ correct answers})}{(\#\text{ answered datapoints})}$

(covered datapoints only, ratio of correct predictions to the number of predictions)

backoff accuracy: $\mathit{coverage} \times \mathit{accuracy} + ((1 - \mathit{coverage}) \times 0.5)$

(emulating a backoff procedure with baseline performance for non-covered items)

differences between models: χ^2 test

(2×2 correct - incorrect answer contingency matrix)

Evaluation

coverage: $\frac{(\#\ \text{answered datapoints})}{(\#\ \text{tot. datapoints})}$

(percentage of datapoints for which a model can make a prediction)

accuracy: $\frac{(\#\ \text{correct answers})}{(\#\ \text{answered datapoints})}$

(covered datapoints only, ratio of correct predictions to the number of predictions)

backoff accuracy: $coverage \times accuracy + ((1 - coverage) \times 0.5)$

(emulating a backoff procedure with baseline performance for non-covered items)

differences between models: χ^2 test

(2×2 correct - incorrect answer contingency matrix)

Results

	Probabilistic Models			Similarity-based Models				
	B_p	SOV_p	SO_p	B_s	SOV_Σ	SOV_Π	SO_Σ	SO_Π
Accuracy	0.50	0.62	0.75	0.50	0.68	0.56	0.68	0.70
Coverage	1.00	0.44	0.75	1.00	0.98	0.94	0.98	0.98
Backoff Accuracy	0.50	0.55	**0.69**	0.50	0.68	0.56	0.68	**0.70**

- both classes outperform the baselines
- similarity-based models maintain the accuracy of probabilistic models while guaranteeing higher coverage
- SO models perform better than SOV models

Results

		Probabilistic Models			Similarity-based Models				
		B_p	SOV_p	SO_p	B_s	SOV_Σ	SOV_Π	SO_Σ	SO_Π
Prob.	B_p								
	SOV_p	-							
	SO_p	*	-						
Similarity	B_s	-	-	*					
	SOV_Σ	*	-	-	*				
	SOV_Π	-	-	-	-	-			
	SO_Σ	*	-	-	*	-	-		
	SO_Π	**	*	-	**	-	*	-	

- both classes outperform the baselines
- similarity-based models maintain the accuracy of probabilistic models while guaranteeing higher coverage
- SO models perform better than SOV models

Results

- ⟨*Dieb schmuggeln/schleifen Diamant*⟩ (⟨*thief smuggle/cut diamond*⟩)
  - **prob. models:** no coverage
  - **sim. models:** events associated with both *Dieb* and *Diamant*: *stehlen* (steal), *rauben* (thieve), *holen* (*get*), *entwenden* (purloin), *erbeuten* (snatch), *verkaufen* (*sell*), *nehmen* (*take*), *klauen* (swipe)
- ⟨*Mechaniker fahren/reparieren Auto*⟩ (⟨*mechanic drive/fix car*⟩)
  - **prob. models:** wrong answer (high overall frequency of *fahren*)
  - **sim. models:** events associated with both *Mechaniker* and *Auto*: *bauen* (build), *reparieren* (fix)

Results

- ⟨*Dieb schmuggeln/schleifen Diamant*⟩ (⟨*thief smuggle/cut diamond*⟩)
  - **prob. models:** no coverage
  - **sim. models:** events associated with both *Dieb* and *Diamant*: **stehlen (steal)**, **rauben (thieve)**, *holen* (*get*), **entwenden (purloin)**, **erbeuten (snatch)**, *verkaufen* (*sell*), *nehmen* (*take*), **klauen (swipe)**
- ⟨*Mechaniker fahren/reparieren Auto*⟩ (⟨*mechanic drive/fix car*⟩)
  - **prob. models:** wrong answer (high overall frequency of *fahren*)
  - **sim. models:** events associated with both *Mechaniker* and *Auto*: *bauen* (build), *reparieren* (fix)

Results

- ⟨*Dieb schmuggeln/schleifen Diamant*⟩ (⟨*thief smuggle/cut diamond*⟩)
  - **prob. models:** no coverage
  - **sim. models:** events associated with both *Dieb* and *Diamant*: *stehlen* (steal), *rauben* (thieve), *holen* (*get*), *entwenden* (purloin), *erbeuten* (snatch), *verkaufen* (*sell*), *nehmen* (*take*), *klauen* (swipe)
- ⟨*Mechaniker fahren/reparieren Auto*⟩ (⟨*mechanic drive/fix car*⟩)
  - **prob. models:** wrong answer (high overall frequency of *fahren*)
  - **sim. models:** events associated with both *Mechaniker* and *Auto*: *bauen* (build), *reparieren* (fix)

Results

$EX_{SO}(\langle Chauffeur, Auto \rangle)$		$EX_{SO}(\langle Mechaniker, Auto \rangle)$	
fahren	*(drive)*	*bauen*	*(build)*
parken	*(park)*	*lassen*	*(let/leave)*
lassen	*(let/leave)*	*besitzen*	*(own)*
geben	*(give)*	*reparieren*	*(repair)*
sehen	*(see)*	*brauchen*	*(need)*
bringen	*(bring)*	*sehen*	*(see)*
steuern	*(steer)*	*benutzen*	*(use)*
halten	*(keep/hold)*	*stellen*	*(put)*

Table: Updated expectations in $SO_{\sqcap}$

Results

$EX_{SO}(\langle Chauffeur, Auto\rangle)$		$EX_{SO}(\langle Mechaniker, Auto\rangle)$	
fahren	**(drive)**	*bauen*	*(build)*
parken	**(park)**	*lassen*	*(let/leave)*
lassen	*(let/leave)*	*besitzen*	*(own)*
geben	*(give)*	*reparieren*	*(repair)*
sehen	*(see)*	*brauchen*	*(need)*
bringen	*(bring)*	*sehen*	*(see)*
steuern	**(steer)**	*benutzen*	*(use)*
halten	*(keep/hold)*	*stellen*	*(put)*

Table: Updated expectations in $SO_{\sqcap}$

Results

$EX_{SO}(\langle Chauffeur, Auto\rangle)$		$EX_{SO}(\langle Mechaniker, Auto\rangle)$	
fahren	*(drive)*	**bauen**	**(build)**
parken	*(park)*	*lassen*	*(let/leave)*
lassen	*(let/leave)*	*besitzen*	*(own)*
geben	*(give)*	**reparieren**	**(repair)**
sehen	*(see)*	*brauchen*	*(need)*
bringen	*(bring)*	*sehen*	*(see)*
steuern	*(steer)*	*benutzen*	*(use)*
halten	*(keep/hold)*	*stellen*	*(put)*

Table: Updated expectations in $SO_{\sqcap}$

Results

Problematic cases for both model classes:

- ⟨*Lehrerin benoten/schreiben Klausur*⟩ (⟨*teacher grade/take exam*⟩)
  - **both model classes:** *schreiben* (write) is much more frequent than *benoten* (grade)
- ⟨*Schüler lernen/schreiben Geschichte*⟩ (⟨*student study/write story*⟩)
  - **prob models:** very frequent idiomatic expression (to write history)
  - **sim. models:** *history* sense gets most informative events *erzählen* (tell), *lesen* (read), *hören* (hear), *erfinden* (invent), and *studieren* (study), *lehren* (teach)
- ⟨*Geburtstagskind einpacken/auspacken Geschenk*⟩ (⟨*birthday-boy/girl wrap/unwrap present*⟩)
  - **prob. models:** no coverage
  - **sim. models:** events associated with *Geburtstagskind*: *bekommen* (receive), *sagen* (say), *auffuttern* (eat up), *herumkommandieren* (boss around), *ausblasen* (blow out)

Results

Problematic cases for both model classes:

- ⟨*Lehrerin benoten/schreiben Klausur*⟩ (⟨*teacher grade/take exam*⟩)
  - **both model classes:** *schreiben* (write) is much more frequent than *benoten* (grade)
- ⟨*Schüler lernen/schreiben Geschichte*⟩ (⟨*student study/write story*⟩)
  - **prob models:** very frequent idiomatic expression (to write history)
  - **sim. models:** *history* sense gets most informative events *erzählen* (tell), *lesen* (read), *hören* (hear), *erfinden* (invent), and *studieren* (study), *lehren* (teach)
- ⟨*Geburtstagskind einpacken/auspacken Geschenk*⟩ (⟨*birthday-boy/girl wrap/unwrap present*⟩)
  - **prob. models:** no coverage
  - **sim. models:** events associated with *Geburtstagskind*: *bekommen* (receive), *sagen* (say), *auffuttern* (eat up), *herumkommandieren* (boss around), *ausblasen* (blow out)

Results

Problematic cases for both model classes:

- ⟨*Lehrerin benoten/schreiben Klausur*⟩ (⟨*teacher grade/take exam*⟩)
  - **both model classes:** *schreiben* (write) is much more frequent than *benoten* (grade)
- ⟨*Schüler lernen/schreiben Geschichte*⟩ (⟨*student study/write story*⟩)
  - **prob models:** very frequent idiomatic expression (to write history)
  - **sim. models:** *history* sense gets most informative events *erzählen* (tell), *lesen* (read), *hören* (hear), *erfinden* (invent), and *studieren* (study), *lehren* (teach)
- ⟨*Geburtstagskind einpacken/auspacken Geschenk*⟩ (⟨*birthday-boy/girl wrap/unwrap present*⟩)
  - **prob. models:** no coverage
  - **sim. models:** events associated with *Geburtstagskind*: *bekommen* (receive), *sagen* (say), *auffuttern* (eat up), *herumkommandieren* (boss around), *ausblasen* (blow out)

Conclusions

A contrastive study of two classes of computational models predicting CEs for logical metonymies:

- **both model classes:**
  → outperform baselines which take into account only information coming from the object
  → SO models perform better than SOV models
- **prob models:** low coverage
  → based on simple (first-order) co-occurrence (sparsity issues)
  → not the case for more complex models introducing latent variables [Prescher et al., 2000]
- **sim. models:** accuracy of probabilistic models while guaranteeing higher coverage
  → take advantage of higher-order co-occurrences

Conclusions

A contrastive study of two classes of computational models predicting CEs for logical metonymies:

- **both model classes:**
  → outperform baselines which take into account only information coming from the object
  → SO models perform better than SOV models
- **prob models:** low coverage
  → based on simple (first-order) co-occurrence (sparsity issues)
  → not the case for more complex models introducing latent variables [Prescher et al., 2000]
- **sim. models:** accuracy of probabilistic models while guaranteeing higher coverage
  → take advantage of higher-order co-occurrences

Conclusions

A contrastive study of two classes of computational models predicting CEs for logical metonymies:

- **both model classes:**
  → outperform baselines which take into account only information coming from the object
  → SO models perform better than SOV models
- **prob models:** low coverage
  → based on simple (first-order) co-occurrence (sparsity issues)
  → not the case for more complex models introducing latent variables [Prescher et al., 2000]
- **sim. models:** accuracy of probabilistic models while guaranteeing higher coverage
  → take advantage of higher-order co-occurrences

Acknowledgements

The research for this paper was funded
by the German Research Foundation (DFG)
as part of the SFB 732 - project D6
at the University of Stuttgart

Thank you!

Baroni, M. and Lenci, A. (2010).
Distributional memory: A general framework for corpus-based semantics.
Computational Linguistics, 36(4):1–49.

Bicknell, K., Elman, J. L., Hare, M., McRae, K., and Kutas, M. (2010).
Effects of event knowledge in processing verbal arguments.
Journal of Memory and Language, 63(4):489–505.

Bohnet, B. (2010).
Top accuracy and fast dependency parsing is not a contradiction.
In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 89–97, Beijing, China.

De Almeida, R. G. and Dwivedi, V. D. (2008).
Coercion without lexical decomposition: Type-shifting effects revisited.
Canadian Journal of Linguistics, 53(2/3):301–326.

Erk, K. (2007).
A simple, similarity-based model for selectional preferences.
In *Proceedings of ACL*, Prague, Czech Republic.

Erk, K. (2010).
What is word meaning, really? (and how can distributional models help us describe it?).
In *Proceedings of the workshop on Geometrical Models of Natural Language Semantics (GEMS)*, Uppsala, Sweden.

Faaß, G., Heid, U., and Schmid, H. (2010).
Design and Application of a Gold Standard for Morphological Analysis: SMOR as an Example of Morphological Evaluation.
In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta.

Ferretti, T. R., McRae, K., and Hatherell, A. (2001).
Integrating verbs, situation schemas and thematic role concept.
Journal of Memory and Language, 44:516–547.

Fodor, J. A. and Lepore, E. (1998).
The emptiness of the lexicon: Reflections on James Pustejovsky's The Generative Lexicon.
Linguistic Inquiry, 29(2):269–288.

Guevara, E. R. (2011).
Computing semantic compositionality in distributional semantics.
In *Proceedings of IWCS-2011*, Oxford, UK.

Harris, Z. S. (1954).
Distributional structure.
Word, 10(23):146–162.

Lapata, M., Keller, F., and Scheepers, C. (2003).
Intra-sentential context effects on the interpretation of logical metonymy.
Cognitive Science, 27(4):649–668.

Lapata, M. and Lascarides, A. (2003).
A probabilistic account of logical metonymy.
Computational Linguistics, 29(2):263–317.

Lenci, A. (2011).
Composing and updating verb argument expectations: A distributional semantic model.
In *Proceedings of the 2nd Workshop on Cognitive Modeling and Computational Linguistics*, pages 58–66, Portland, Oregon.

Li, P., Farkas, I., and MacWhinney, B. (2004).
Early lexical development in a self-organizing neural network.
Neural Networks, 17:1345–1362.

Miller, G. A. and Charles, W. G. (1991).
Contextual correlates of semantic similarity.
Language and Cognitive Processes, 6(1):1–28.

Mitchell, J. and Lapata, M. (2010).
Composition in distributional models of semantics.
Cognitive Science, 34(8):1388–1429.

Prescher, D., Riezler, S., and Rooth, M. (2000).
Using a Probabilistic Class-Based Lexicon for Lexical Ambiguity Resolution.
In *Proceedings of COLING 2000*, Saarbrücken, Germany.

Pustejovsky, J. (1995).
The Generative Lexicon.
MIT Press.

Rosch, E. (1975).
Cognitive representations of semantic categories.
Journal of Experimental Psychology: General, 104:192–233.

Vigliocco, G., Vinson, D. P., Lewis, W., and Garrett, M. F. (2004).
Representing the meanings of object and action words: The featural and unitary semantic space hypothesis.
Cognitive Psychology, 48(4):422–488.

Zarcone, A. and Padó, S. (2011).
Generalized event knowledge in logical metonymy resolution.
In *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*, pages 944–949, Austin, TX.

Zarcone, A., Padó, S., and Lenci, A. (2012).
Inferring covert events in logical metonymies: a probe recognition experiment.
In *Proceedings of the 34th Annual Conference of the Cognitive Science Society*, Austin, TX.