States, events, and generics:

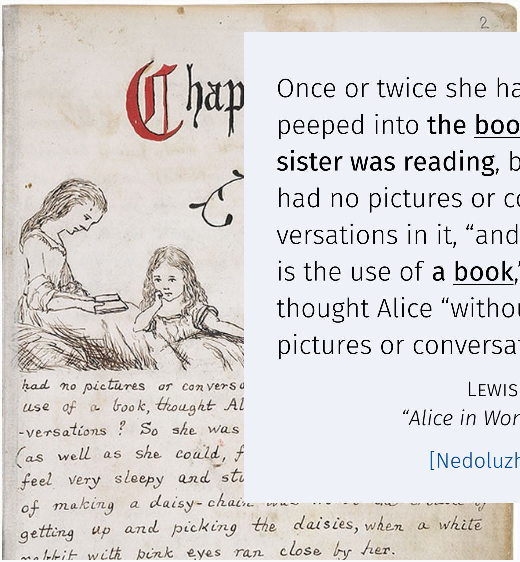
computational modeling of situation entity types

Annemarie Friedrich

Disputation • 24. Februar 2017

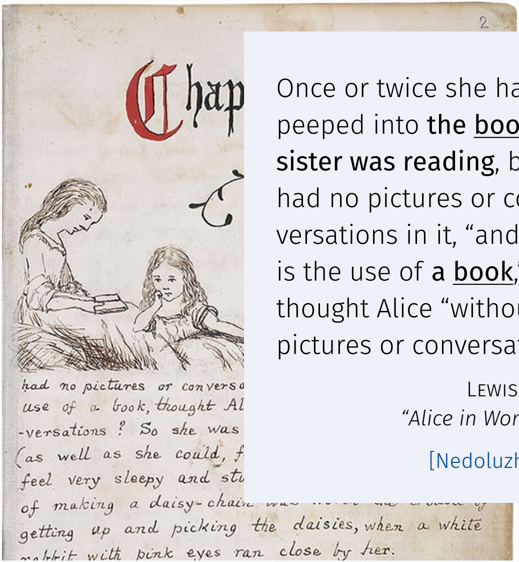
Universität des Saarlandes, Computerlinguistik

Motivation: coreference resolution



> Once or twice she had peeped into **the book her sister was reading**, but it had no pictures or conversations in it, "and what is the use of **a book**," thought Alice "without pictures or conversations?"
>
> LEWIS CARROLL: *"Alice in Wonderland"*
>
> [Nedoluzhko, 2013]

Motivation: coreference resolution



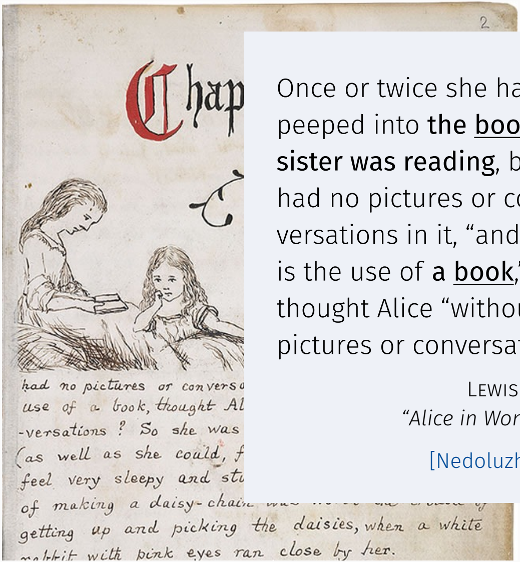
Once or twice she had peeped into **the <u>book</u> her sister was reading**, but it had no pictures or conversations in it, "and what is the use of **a <u>book</u>**," thought Alice "without pictures or conversations?"

LEWIS CARROLL:
"Alice in Wonderland"

[Nedoluzhko, 2013]

particular book

kind



Once or twice she had peeped into **the book her sister was reading**, but it had no pictures or conversations in it, "and what is the use of **a book**," thought Alice "without pictures or conversations?"

Lewis Carroll: *"Alice in Wonderland"*

[Nedoluzhko, 2013]

particular book

kind

generics

Motivation: temporal relation extraction

The rabbit took out his watch.
Alice started to her feet.

Motivation: temporal relation extraction

The rabbit took out his watch.
Alice started to her feet.

Motivation: temporal relation extraction

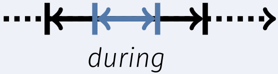
The rabbit took out his watch.
Alice started to her feet.

The rabbit was taking out his watch.
Alice started to her feet.

Motivation: temporal relation extraction

The rabbit took out his watch.
Alice started to her feet.

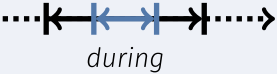
The rabbit was taking out his watch.
Alice started to her feet.



Motivation: temporal relation extraction

The rabbit took out his watch.
Alice started to her feet.

The rabbit was taking out his watch.
Alice started to her feet.



aspect

Aspect

aka aktionsart

[Vendler, 1957]
[Bach, 1986]

Aspect

aka aktionsart

[Vendler, 1957]
[Bach, 1986]

The ship moved. **event**

Aspect

aka aktionsart

[Vendler, 1957]
[Bach, 1986]

The ship moved. **event**

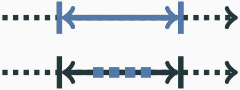
Aspect

aka aktionsart

[Vendler, 1957]
[Bach, 1986]

The ship moved. **event**

The ship was moving. **ongoing event / process**



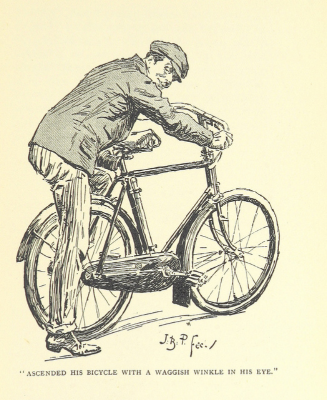
Aspect

aka aktionsart

[Vendler, 1957]
[Bach, 1986]

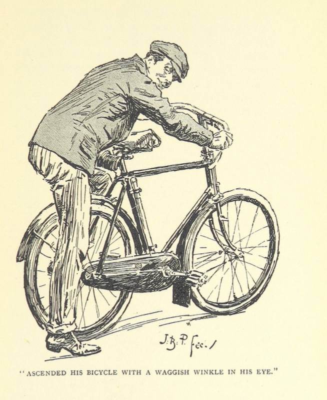
The ship moved.	**event**	
The ship was moving.	**ongoing event / process**	
The ship was in motion.	**state**	

Generics: habituals



Mike cycled to work. **episodic event**

Generics: habituals



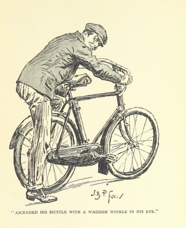
generalization over situations

[Krifka et al., 1995]

Mike cycled to work. **episodic event**

Mike cycles to work. **habitual**

Generics: reference to kinds

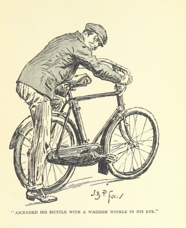


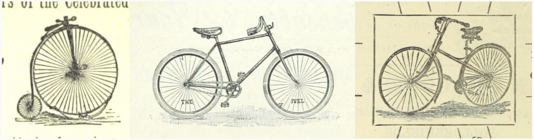
Mike's bike is blue.

particular bike

Generics: reference to kinds

[Krifka et al., 1995]





Mike's bike is blue.

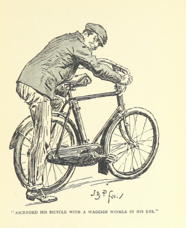
particular bike

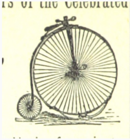
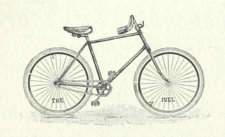
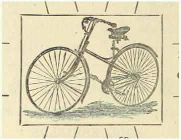
Bicycles have two wheels.

generalization over members of a kind

Generics: reference to kinds

[Krifka et al., 1995]



Mike's bike is blue.

particular bike

Bicycles have two wheels.

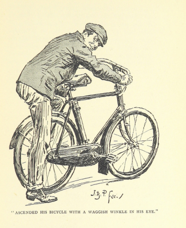
generalization over members of a kind

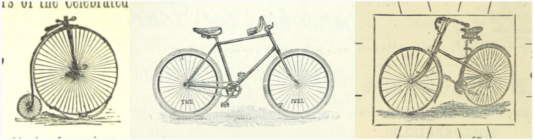
The bicycle was invented in the 19th century.

reference to kind

Generics: reference to kinds

[Krifka et al., 1995]





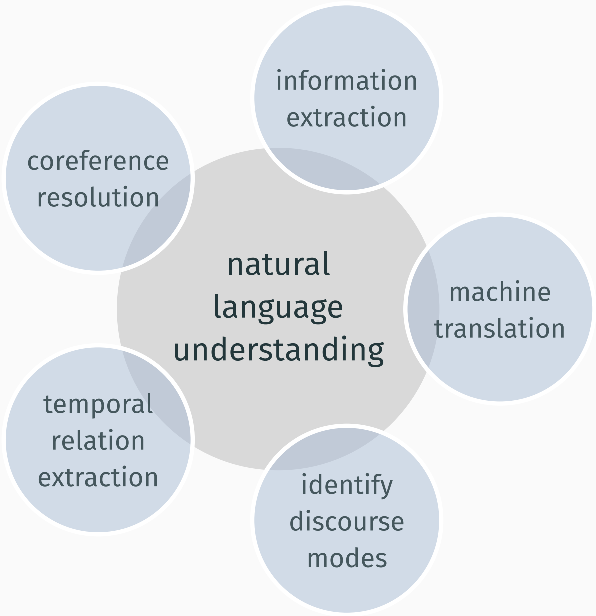
Mike's bike is blue.
particular bike

Bicycles have two wheels.
generalization over members of a kind

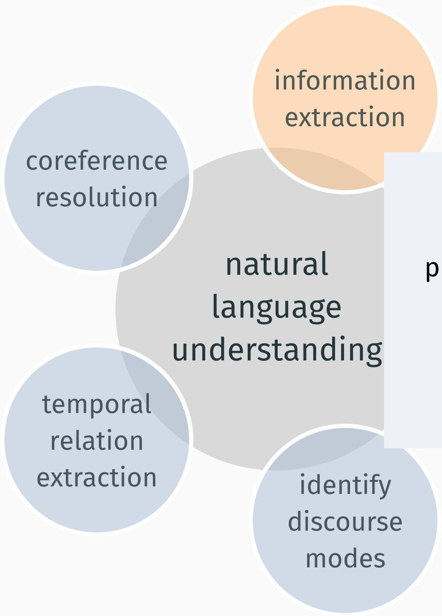
The bicycle was invented in the 19th century.
reference to kind

entailment

Why model these phenomena?



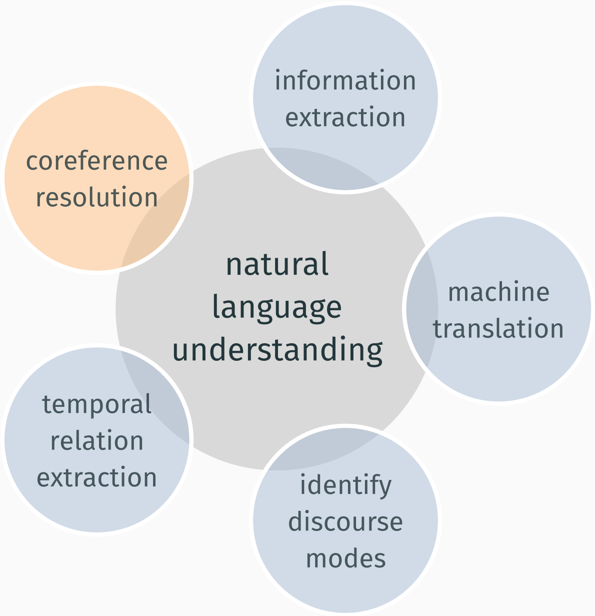
Why model these phenomena?



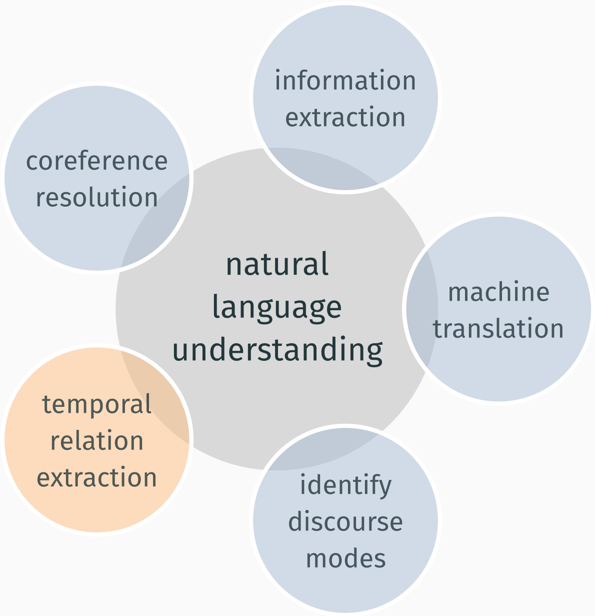
particular events,
properties of individuals
vs.
generic knowledge

[Van Durme, 2009]

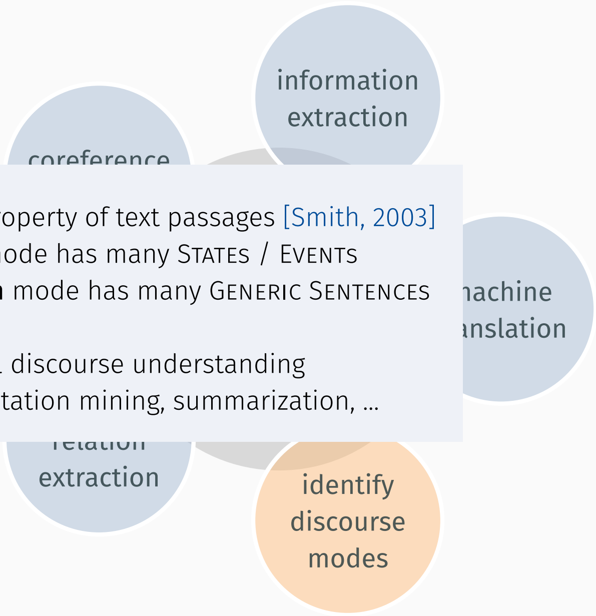
Why model these phenomena?



Why model these phenomena?



Why model these phenomena?



linguistic property of text passages [Smith, 2003]

Narrative mode has many STATES / EVENTS

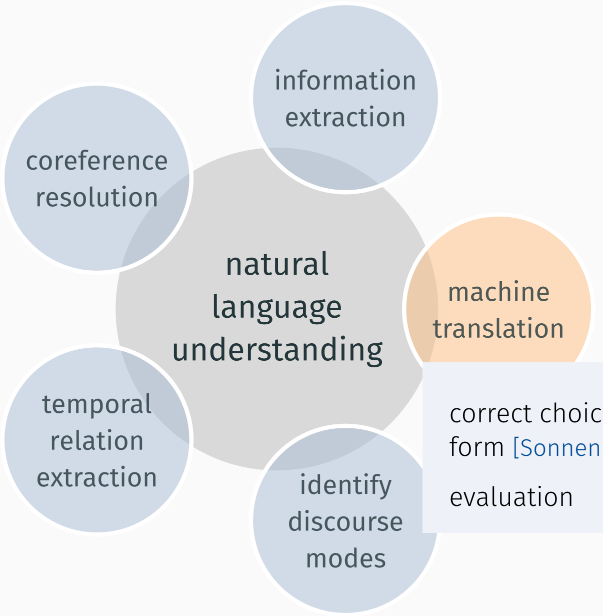
Information mode has many GENERIC SENTENCES

...

→ temporal discourse understanding

→ argumentation mining, summarization, ...

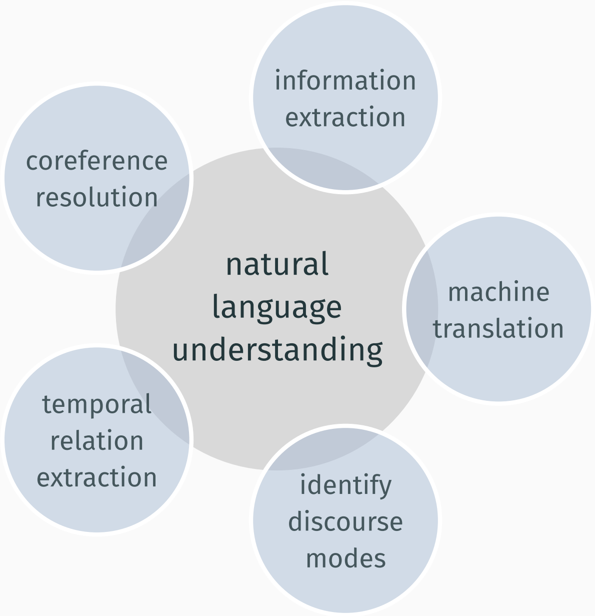
Why model these phenomena?



correct choice of aspectual form [Sonnenhauser, 2004]

evaluation

Why model these phenomena?



Overview of thesis work

generics	lexical aspect	habituals
reference to kinds	state vs. event	generalization over situations
[Friedrich & Pinkal, ACL 2015] [Friedrich et al., LAW 2015]	[Friedrich & Palmer, ACL 2014]	[Friedrich & Pinkal, EMNLP 2015]

Overview of thesis work

generics
reference to kinds

[Friedrich & Pinkal, ACL 2015]
[Friedrich et al., LAW 2015]

lexical aspect
state vs. event

[Friedrich & Palmer, ACL 2014]

habituals
generalization over situations

[Friedrich & Pinkal, EMNLP 2015]

situation entity types [Smith, 2003]

[Friedrich et al., ACL 2016], [Friedrich & Palmer, LAW 2014],
[Mavridou et al., LSDSem 2015], [Palmer & Friedrich, 2014]

Overview of thesis work

linguistic background / annotation scheme

generics
reference to kinds
[Friedrich & Pinkal, ACL 2015]
[Friedrich et al., LAW 2015]

lexical aspect
state vs. event
[Friedrich & Palmer, ACL 2014]

habituals
generalization over situations
[Friedrich & Pinkal, EMNLP 2015]

situation entity types [Smith, 2003]
[Friedrich et al., ACL 2016], [Friedrich & Palmer, LAW 2014],
[Mavridou et al., LSDSem 2015], [Palmer & Friedrich, 2014]

Overview of thesis work

corpus creation / analysis of agreement

linguistic background / annotation scheme

generics	lexical aspect	habituals
reference to kinds	state vs. event	generalization over situations
[Friedrich & Pinkal, ACL 2015] [Friedrich et al., LAW 2015]	[Friedrich & Palmer, ACL 2014]	[Friedrich & Pinkal, EMNLP 2015]

situation entity types [Smith, 2003]

[Friedrich et al., ACL 2016], [Friedrich & Palmer, LAW 2014], [Mavridou et al., LSDSem 2015], [Palmer & Friedrich, 2014]

Overview of thesis work

computational modeling

corpus creation / analysis of agreement

linguistic background / annotation scheme

generics
reference to kinds

[Friedrich & Pinkal, ACL 2015]
[Friedrich et al., LAW 2015]

lexical aspect
state vs. event

[Friedrich & Palmer, ACL 2014]

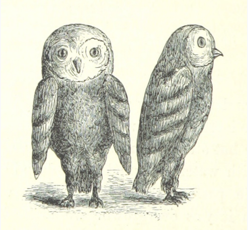
habituals
generalization over situations

[Friedrich & Pinkal, EMNLP 2015]

situation entity types [Smith, 2003]

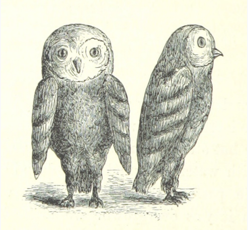
[Friedrich et al., ACL 2016], [Friedrich & Palmer, LAW 2014], [Mavridou et al., LSDSem 2015], [Palmer & Friedrich, 2014]

Situation entity types [Smith, 2003] [Palmer et al., 2007]

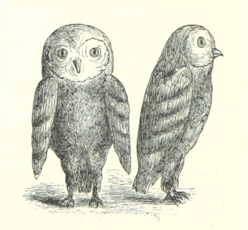


Situation entity types [Smith, 2003] [Palmer et al., 2007]

STATE	Julie likes Cooper.

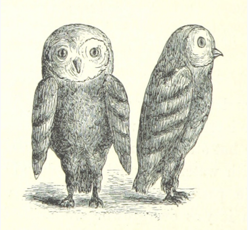


Situation entity types [Smith, 2003] [Palmer et al., 2007]



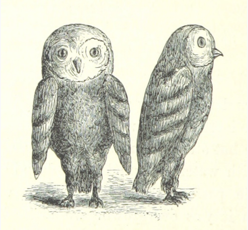
STATE	Julie likes Cooper.
EVENT	Julie met Cooper two years ago.

Situation entity types [Smith, 2003] [Palmer et al., 2007]



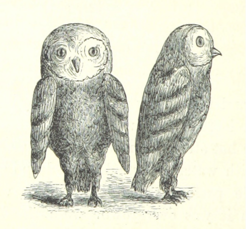
STATE	Julie likes Cooper.
EVENT	Julie met Cooper two years ago.
REPORT	..., said the zookeeper.

Situation entity types [Smith, 2003] [Palmer et al., 2007]



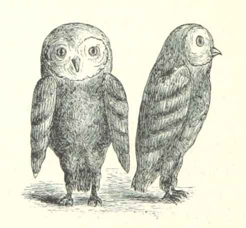
STATE	Julie likes Cooper.
EVENT	Julie met Cooper two years ago.
REPORT	..., said the zookeeper.
GENERIC SENTENCE	**Owls** are nocturnal animals.

Situation entity types [Smith, 2003] [Palmer et al., 2007]



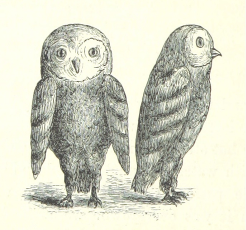
STATE	Julie likes Cooper.
EVENT	Julie met Cooper two years ago.
REPORT	..., said the zookeeper.
GENERIC SENTENCE	**Owls** are nocturnal animals.
GENERALIZING SENTENCE	Julie often teases Cooper.

Situation entity types [Smith, 2003] [Palmer et al., 2007]



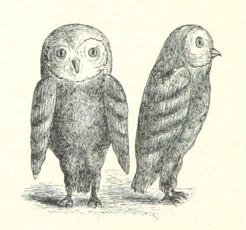
State	Julie likes Cooper.
Event	Julie met Cooper two years ago.
Report	..., said the zookeeper.
Generic Sentence	**Owls** are nocturnal animals.
Generalizing Sentence	Julie often teases Cooper.
Imperative	Catch the mouse!

Situation entity types [Smith, 2003] [Palmer et al., 2007]



STATE	Julie likes Cooper.
EVENT	Julie met Cooper two years ago.
REPORT	..., said the zookeeper.
GENERIC SENTENCE	**Owls** are nocturnal animals.
GENERALIZING SENTENCE	Julie often teases Cooper.
IMPERATIVE	Catch the mouse!
QUESTION	Why are there owls on your slides?

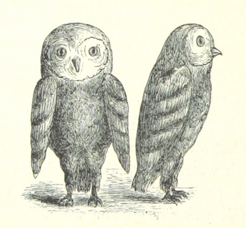
Situation entity types [Smith, 2003] [Palmer et al., 2007]



STATE	Julie likes Cooper. Julie did **not** kill the mouse.
EVENT	Julie me Cooper two years ago.
REPORT	eper.
GENERIC SE	l animals.
GENERALIZING SENTENCE	Julie often teases Cooper.
IMPERATIVE	Catch the mouse!
QUESTION	Why are there owls on your slides?

coercion to STATE:
negation, modality, future,
perfect, conditionality

Situation entity types [Smith, 2003] [Palmer et al., 2007]

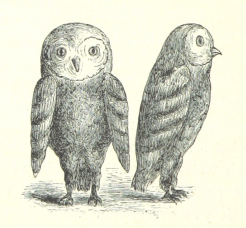


Trying one steel pen after another.

STATE	Julie likes Cooper. ...id **not** kill the mouse.
EV...	...net Cooper two years ago.
R...	...d the zoo keeper.
G...	...are nocturnal animals.
G... S...	...ften teases Cooper.
IM...	...the mouse!
Q...	...re there owls on your slides?

??

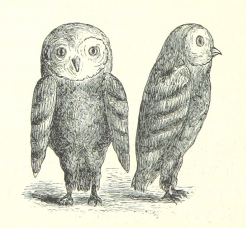
Situation entity types [Smith, 2003] [Palmer et al., 2007]



STATE	Julie $\boxed{\text{likes}}$ Coop Julie did **not** k
EVENT	Julie $\boxed{\text{met}}$ Cooper two years ago.
REPORT	..., said the zookeeper.
GENERIC SENTENCE	**Owls** are nocturnal animals.
GENERALIZING SENTENCE	Julie often teases Cooper.
IMPERATIVE	Catch the mouse!
QUESTION	Why are there owls on your slides?

lexical aspect:
dynamic or ***stative***?

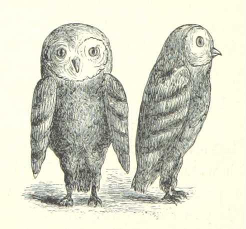
Situation entity types [Smith, 2003] [Palmer et al., 2007]



State	Julie likes Cooper. Julie did **not** kill the mouse.
Event	Julie $\boxed{\text{met}}$ Cooper two years ago.
Report	..., said the zookeeper.
Generic Sentence	**Owls** are nocturnal animals.
Generalizing Sentence	Julie often $\boxed{\text{teases}}$ Cooper.
Imperative	
Question	...ur slides?

Does something happen repeatedly?
episodic or ***habitual***?

Situation entity types [Smith, 2003] [Palmer et al., 2007]

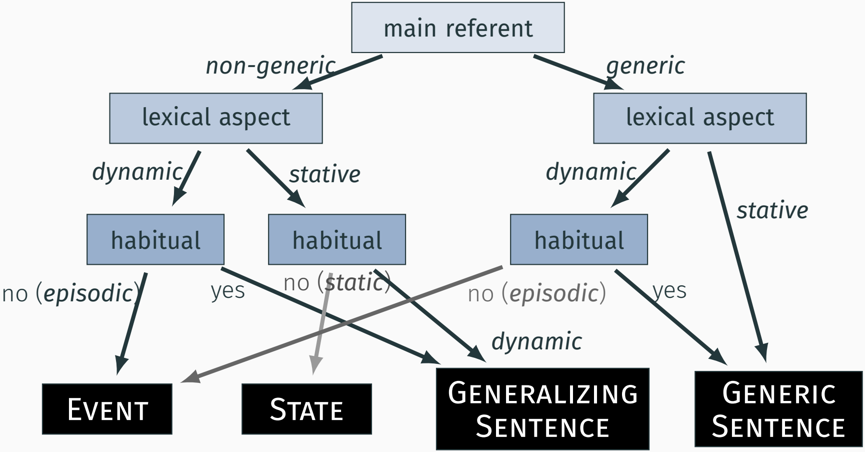


About kind/class or particular referent?
generic or *non-generic*?

State	Julie likes Cooper. Julie did **not** kill the mouse.
	years ago.
Report	..., said the zookeeper.
Generic Sentence	Owls are nocturnal animals.
Generalizing Sentence	Julie often teases Cooper.
Imperative	Catch the mouse!
Question	Why are there owls on your slides?

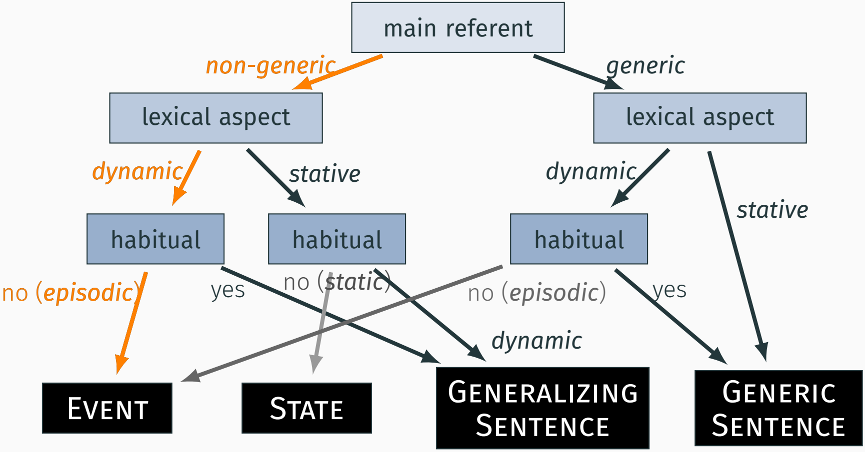
Annotation scheme

[Friedrich and Palmer, 2014] [Friedrich et al., 2015]



Annotation scheme

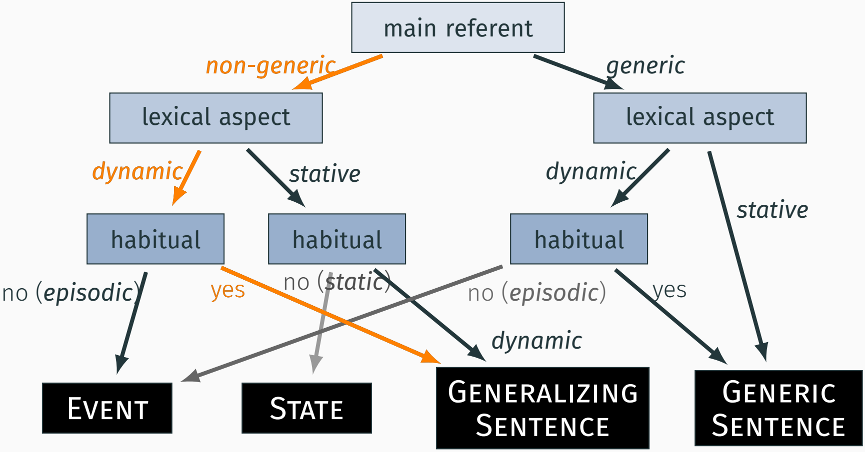
[Friedrich and Palmer, 2014] [Friedrich et al., 2015]



Mike cycled to work.

Annotation scheme

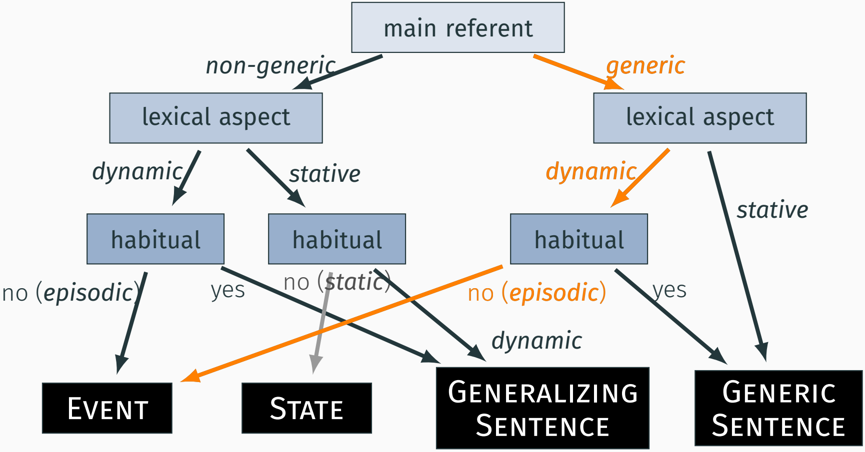
[Friedrich and Palmer, 2014] [Friedrich et al., 2015]



Mike cycles to work.

Annotation scheme

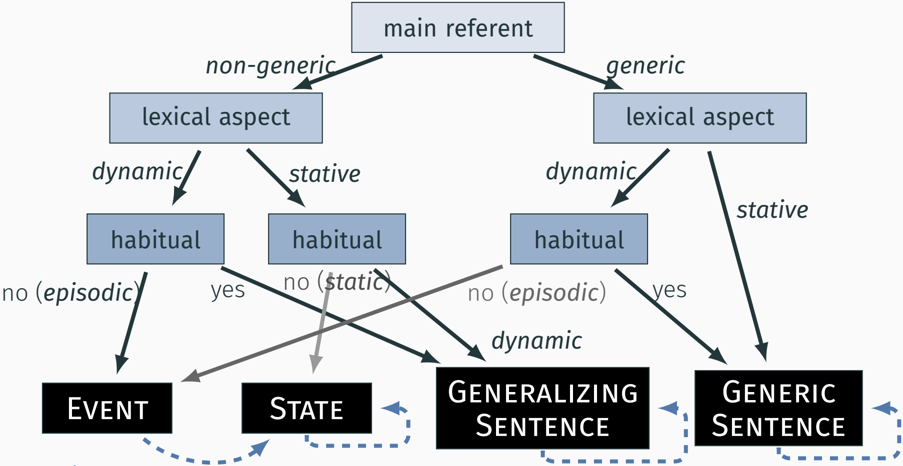
[Friedrich and Palmer, 2014] [Friedrich et al., 2015]



The bicycle was invented in the 19th century.

Annotation scheme

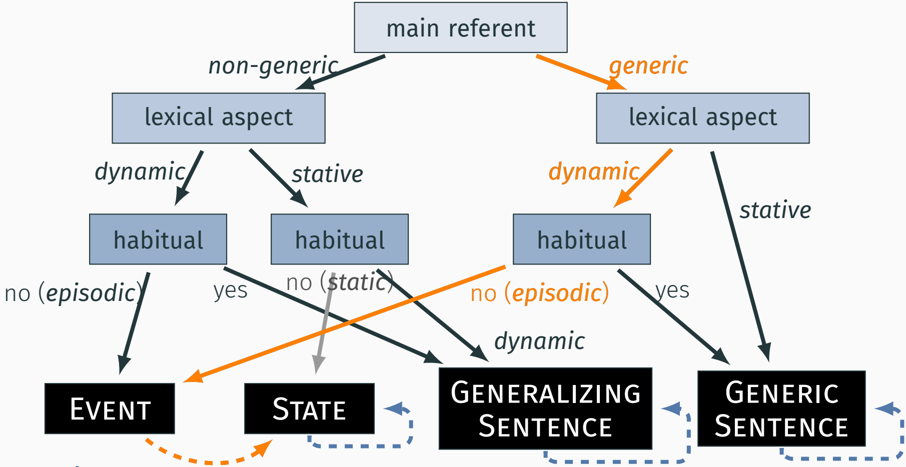
[Friedrich and Palmer, 2014] [Friedrich et al., 2015]



coercion: negation, modals, conditional, perfect, future

Annotation scheme

[Friedrich and Palmer, 2014] [Friedrich et al., 2015]



coercion: negation, modals, conditional, perfect, future

The bicycle had not yet been invented in the 18th century.

Data and annotation procedure

MASC
31,596 clauses
news, letters, fiction, journal, technical, travel, ...

Wikipedia
10,355 clauses
animals, science, sports, ethnic groups, ...

Data and annotation procedure

MASC
31,596 clauses
news, letters, fiction, journal, technical, travel, ...

Wikipedia
10,355 clauses
animals, science, sports, ethnic groups, ...

Clause segmentation
SPADE [Soricut and Marcu, 2003]
+ heuristics

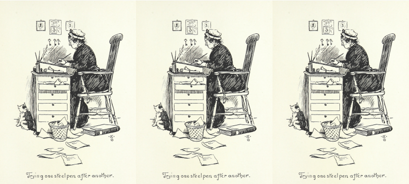
Data and annotation procedure

MASC
31,596 clauses
news, letters, fiction, journal, technical, travel, ...

Wikipedia
10,355 clauses
animals, science, sports, ethnic groups, ...

Clause segmentation
SPADE [Soricut and Marcu, 2003]
+ heuristics

manual annotation
training phase + written manual



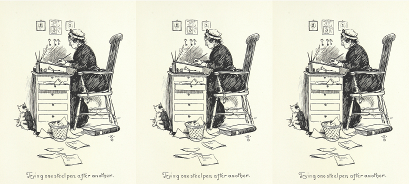
Data and annotation procedure

MASC
31,596 clauses
news, letters, fiction, journal, technical, travel, ...

Wikipedia
10,355 clauses
animals, science, sports, ethnic groups, ...

Clause segmentation
SPADE [Soricut and Marcu, 2003]
+ heuristics

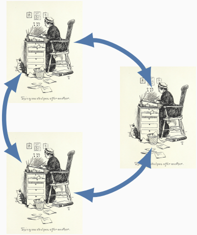
manual annotation
training phase + written manual



gold standard
majority vote over labels of 3 annotators

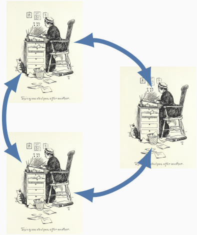
Inter-annotator agreement

Fleiss' κ:
how much agreement
beyond chance?



Inter-annotator agreement

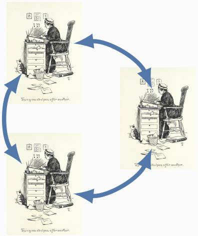
Fleiss' κ:
how much agreement beyond chance?



Annotation layer		MASC	Wiki
lexical aspect	*stative*	0.69	0.64
	dynamic		
	both		

Inter-annotator agreement

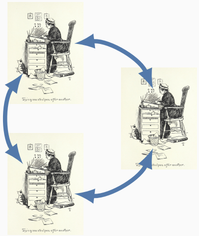
Fleiss' κ:
how much agreement beyond chance?



Annotation layer		MASC	Wiki
lexical aspect	*stative* *dynamic* *both*	0.69	0.64
main referent	*generic* *non-generic* *cannot decide*	0.69	0.65

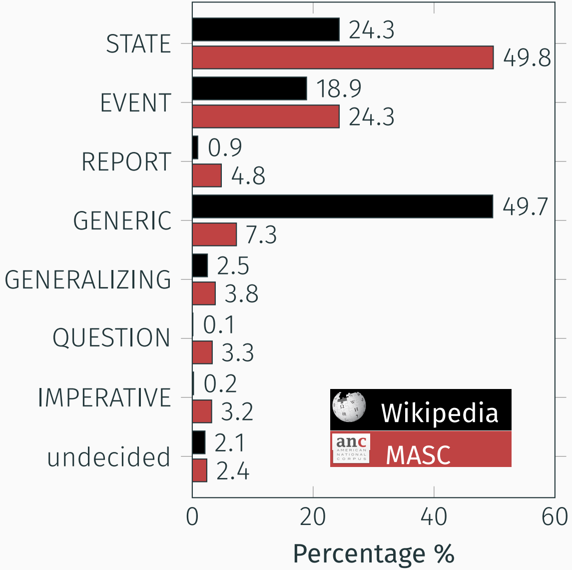
Inter-annotator agreement

Fleiss' κ:
how much agreement beyond chance?

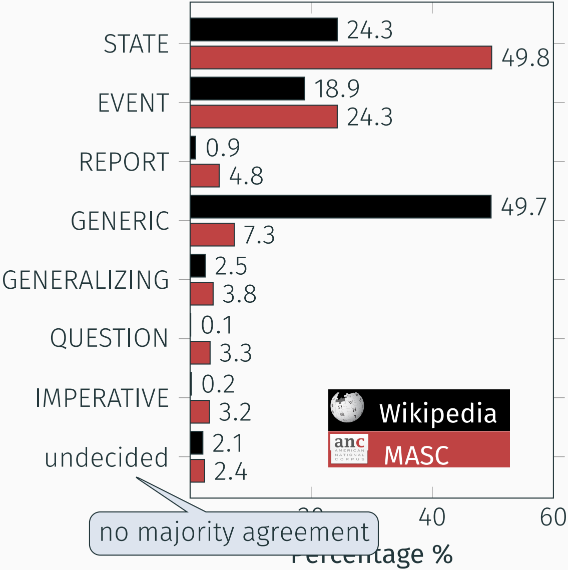


Annotation layer		MASC	Wiki
lexical aspect	*stative* *dynamic* *both*	0.69	0.64
main referent	*generic* *non-generic* *cannot decide*	0.69	0.65
habituality	*episodic* *habitual* *static* *cannot decide*	0.55	0.67

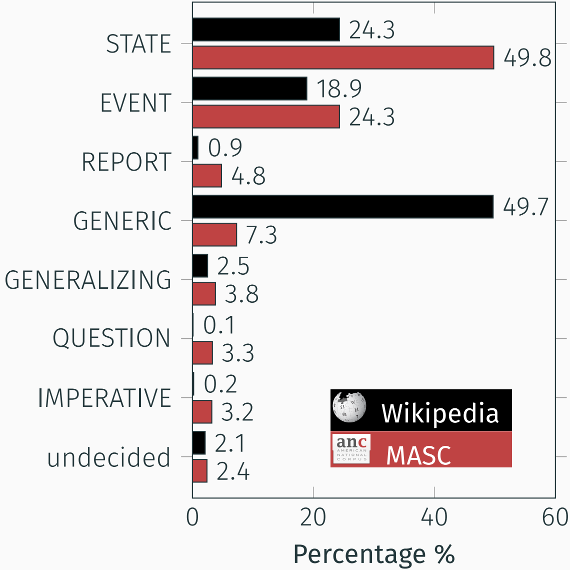
Situation entity types: distributions and agreement



Situation entity types: distributions and agreement



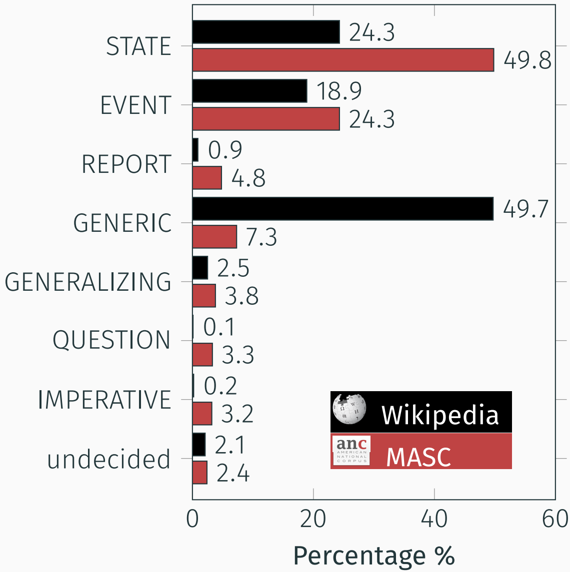
Situation entity types: distributions and agreement



Fleiss' κ [Krippendorff, 1980]

STATE	0.67
EVENT	0.74
REPORT	0.80
GENERIC	0.68
GENERALIZING	0.43
QUESTION	0.91
IMPERATIVE	0.94
undecided	

Situation entity types: distributions and agreement



Fleiss' κ [Krippendorff, 1980]

STATE	0.67
EVENT	0.74
REPORT	0.80
GENERIC	0.[illegible]
GENERALIZING	0.43
QUESTION	0.91
IMPERATIVE	0.94
undecided	

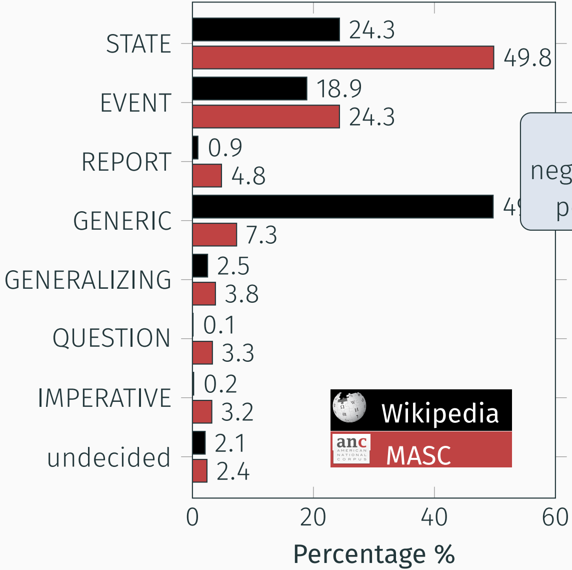
REPORT: easy

EVENT: easy

QUESTION: easy

IMPERATIVE: easy

Situation entity types: distributions and agreement

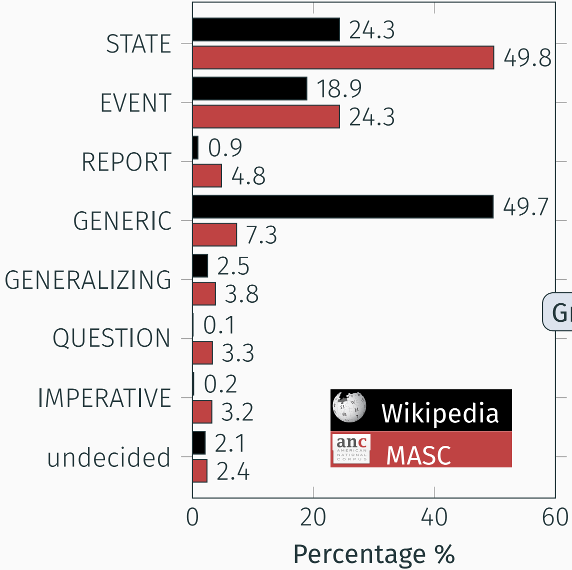


Fleiss' κ [Krippendorff, 1980]

STATE	0.67
EVENT	0.74
GENERIC	0.68
GENERALIZING	0.43
QUESTION	0.91
IMPERATIVE	0.94
undecided	

coercion to **STATE**: negation, modality, future, perfect, conditionality

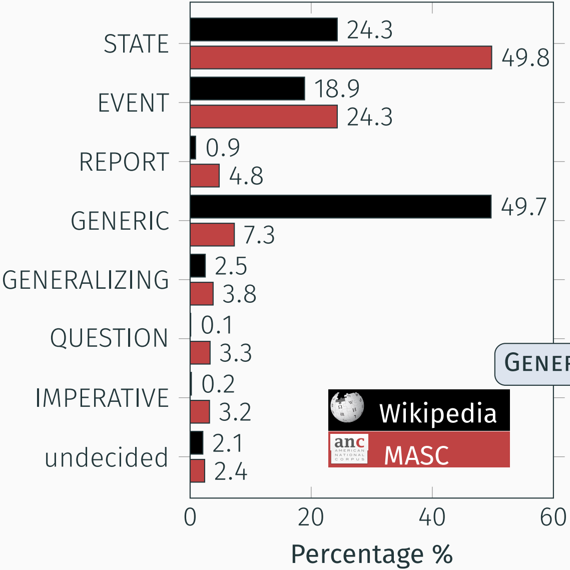
Situation entity types: distributions and agreement



Fleiss' κ [Krippendorff, 1980]	
STATE	0.67
EVENT	0.74
REPORT	0.80
GENERIC	0.68
GENERALIZING	0.43
QUESTION	0.91
IMPERATIVE	0.94
undecided	

GENERIC SENTENCES: difficult

Situation entity types: distributions and agreement



Fleiss' κ [Krippendorff, 1980]

STATE	0.67
EVENT	0.74
REPORT	0.80
GENERIC	0.68
GENERALIZING	0.43
IMPERATIVE	0.94
undecided	

GENERALIZING SENTENCES: sparse

Overview of thesis work

computational modeling

corpus creation / analysis of agreement

linguistic background / annotation scheme

generics
reference to kinds
[Friedrich & Pinkal, ACL 2015]
[Friedrich et al., LAW 2015]

lexical aspect
state vs. event
[Friedrich & Palmer, ACL 2014]

habituals
generalization over situations
[Friedrich & Pinkal, EMNLP 2015]

situation entity types [Smith, 2003]
[Friedrich et al., ACL 2016], [Friedrich & Palmer, LAW 2014],
[Mavridou et al., LSDSem 2015], [Palmer & Friedrich, 2014]

Related work in computational linguistics

- modeling of aspectual classes

Related work in computational linguistics

- modeling of aspectual classes
    - **Vendler classes** [Vendler, 1957]:
      Italian [Zarcone and Lenci, 2008], German [Hermes et al., 2015]

Related work in computational linguistics

- modeling of aspectual classes
  - Vendler classes [Vendler, 1957]:
    Italian [Zarcone and Lenci, 2008], German [Hermes et al., 2015]
  - *stative* vs. *dynamic* [Siegel and McKeown, 2000]

Related work in computational linguistics

- modeling of aspectual classes
  - **Vendler classes** [Vendler, 1957]:
    Italian [Zarcone and Lenci, 2008], German [Hermes et al., 2015]
  - ***stative*** vs. ***dynamic*** [Siegel and McKeown, 2000]
  - **completedness** [Siegel and McKeown, 2000] [Loáiciga and Grisot, 2016]

Related work in computational linguistics

- modeling of aspectual classes
  - **Vendler classes** [Vendler, 1957]:
    Italian [Zarcone and Lenci, 2008], German [Hermes et al., 2015]
  - ***stative*** vs. ***dynamic*** [Siegel and McKeown, 2000]
  - **completedness** [Siegel and McKeown, 2000] [Loáiciga and Grisot, 2016]
- **functions of tense** [Reichart and Rappoport, 2010] [Zhang and Xue, 2014]
  - (episodic/future/...) event, habitual, state, general facts, ...

Related work in computational linguistics

- modeling of aspectual classes
  - **Vendler classes** [Vendler, 1957]:
    Italian [Zarcone and Lenci, 2008], German [Hermes et al., 2015]
  - ***stative*** vs. ***dynamic*** [Siegel and McKeown, 2000]
  - **completedness** [Siegel and McKeown, 2000] [Loáiciga and Grisot, 2016]
- functions of tense [Reichart and Rappoport, 2010] [Zhang and Xue, 2014]
  - (episodic/future/...) event, habitual, state, general facts, ...
- modeling genericity
  - identifying genericity of NPs / reference to kinds [Reiter and Frank, 2010]
  - recognizing habituals [Mathew and Katz, 2009]

Related work in computational linguistics

- modeling of aspectual classes
  - Vendler classes [Vendler, 1957]:
    Italian [Zarcone and Lenci, 2008], German [Hermes et al., 2015]
  - *stative* vs. *dynamic* [Siegel and McKeown, 2000]
  - completedness [Siegel and McKeown, 2000] [Loáiciga and Grisot, 2016]
- functions of tense [Reichart and Rappoport, 2010] [Zhang and Xue, 2014]
  - (episodic/future/...) event, habitual, state, general facts, ...
- modeling genericity
  - identifying genericity of NPs / reference to kinds [Reiter and Frank, 2010]
  - recognizing habituals [Mathew and Katz, 2009]
- labeling situation entities [Palmer et al., 2007]
  - data set: 20 texts / 4391 clauses from Brown corpus

Overview of thesis work

computational modeling

corpus creation / analysis of agreement

linguistic background / annotation scheme

generics
reference to kinds

[Friedrich & Pinkal, ACL 2015]
[Friedrich et al., LAW 2015]

lexical aspect
state vs. event

[Friedrich & Palmer, ACL 2014]

habituals
generalization over situations

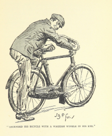
[Friedrich & Pinkal, EMNLP 2015]

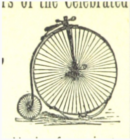
situation entity types [Smith, 2003]

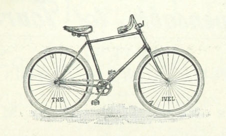
[Friedrich et al., ACL 2016], [Friedrich & Palmer, LAW 2014], [Mavridou et al., LSDSem 2015], [Palmer & Friedrich, 2014]

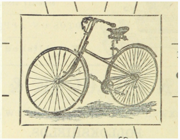
Discourse-sensitive identification of generic expressions

subject ***non-generic*** or ***generic***? [Friedrich & Pinkal, ACL 2015]









The bike is blue.

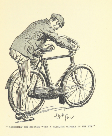
non-generic

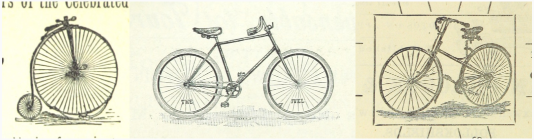
The bike was invented in the 19th century.

generic

Discourse-sensitive identification of generic expressions

subject ***non-generic*** or ***generic***? [Friedrich & Pinkal, ACL 2015]





The bike is blue.

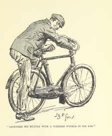
non-generic

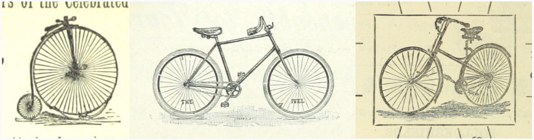
The bike was invented in the 19th century.

generic → form of NP **not sufficient for classification**

Discourse-sensitive identification of generic expressions

subject ***non-generic*** or ***generic***? [Friedrich & Pinkal, ACL 2015]





The bike is blue.

non-generic

The bike was invented in the 19th century.

generic

→ form of NP not sufficient for classification

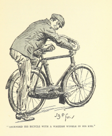
Mike keeps fixing his **bicycle**.

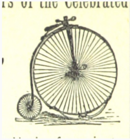
The bicycle has undergone continual adaptation and improvement.

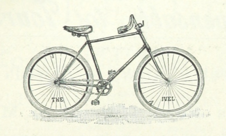
non-generic

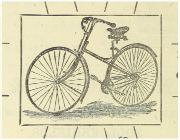
Discourse-sensitive identification of generic expressions

subject ***non-generic*** or ***generic***? [Friedrich & Pinkal, ACL 2015]









The bike is blue.

non-generic

The bike was invented in the 19th century.

generic → form of NP **not sufficient for classification**

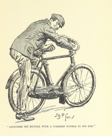
Bicycles were introduced in the 19th century in Europe.

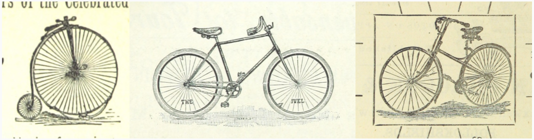
The bicycle has undergone continual adaptation and improvement.

generic

Discourse-sensitive identification of generic expressions

subject ***non-generic*** or ***generic***? [Friedrich & Pinkal, ACL 2015]





The bike is blue. The bike was invented in the 19th century.

non-generic *generic* → form of NP **not sufficient for classification**

Bicycles were introduced in the 19th century in Europe.

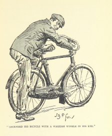
The bicycle has undergone continual adaptation and improvement.

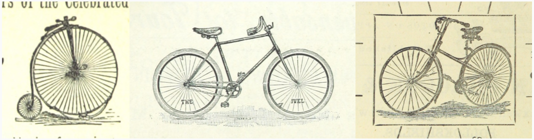
generic

→ discourse context matters

Discourse-sensitive identification of generic expressions

subject ***non-generic*** or ***generic***? [Friedrich & Pinkal, ACL 2015]





The bike is blue.

non-generic

The bike was invented in the 19th century.

generic → form of NP **not sufficient for classification**

Bicycles were introduced in the 19th century in Europe.

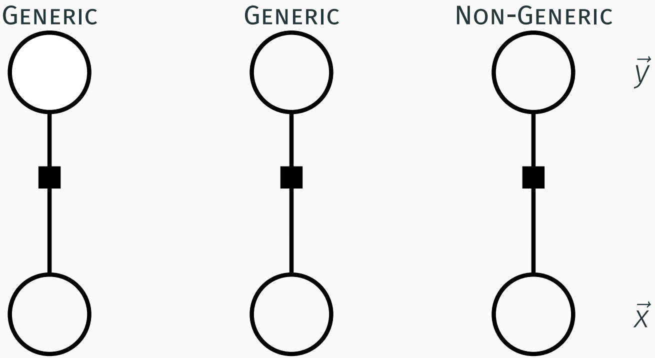
The bicycle has undergone continual adaptation and improvement.

[Reiter and Frank, 2010]

generic

→ discourse context matters

Discourse-sensitive identification of generic expressions

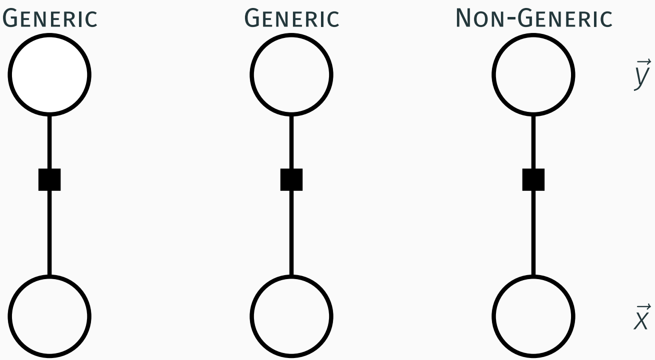


Bicycles were introduced …

The bicycle has undergone …

These innovations have continued …

Discourse-sensitive identification of generic expressions



Bicycles were introduced …

barePlural = T
simplePast = T
…

The bicycle has undergone …

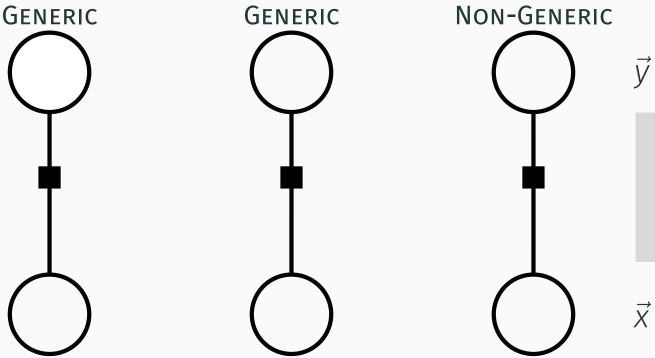
barePlural = F
perfect = T
…

These innovations have continued …

barePlural = F
countable = Y
…

syntactic-semantic features

Discourse-sensitive identification of generic expressions



maximum entropy model (MaxEnt)

$$P(y|x)$$

Bicycles were introduced ...

barePlural = T
simplePast = T
...

The bicycle has undergone ...

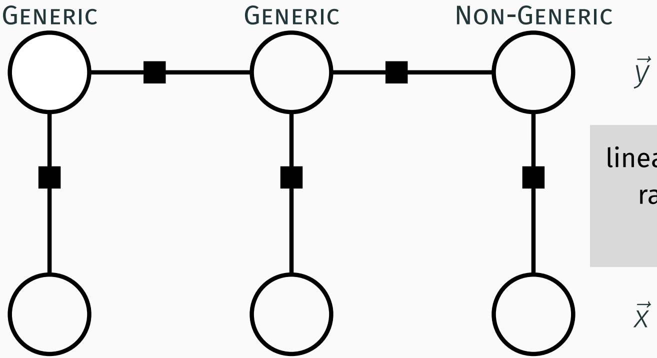
barePlural = F
perfect = T
...

These innovations have continued ...

barePlural = F
countable = Y
...

syntactic-semantic features

Discourse-sensitive identification of generic expressions



linear chain conditional random field (CRF)

$$P(\vec{y}|\vec{x})$$

Bicycles were introduced ...

barePlural = T
simplePast = T
...

The bicycle has undergone ...

barePlural = F
perfect = T
...

These innovations have continued ...

barePlural = F
countable = Y
...

syntactic-semantic features

Syntactic-semantic features

Implementation based on dkpro [Eckart de Castilho and Gurevych, 2014] and Stanford CoreNLP [Manning et al., 2014]

Syntactic-semantic features

Implementation based on dkpro [Eckart de Castilho and Gurevych, 2014] and Stanford CoreNLP [Manning et al., 2014]

The bicycle has **undergone** continual adaptation and improvement.

Syntactic-semantic features

Implementation based on dkpro [Eckart de Castilho and Gurevych, 2014] and Stanford CoreNLP [Manning et al., 2014]

The bicycle has **undergone** continual adaptation and improvement.

- main verb:
  - tense, voice, progressive, perfect, lemma, WordNet hypernyms, ...

Syntactic-semantic features

Implementation based on dkpro [Eckart de Castilho and Gurevych, 2014] and Stanford CoreNLP [Manning et al., 2014]

The bicycle has **undergone** continual adaptation and improvement.

- **main verb**:
  - tense, voice, progressive, perfect, lemma, WordNet hypernyms, ...
- **main referent** (subject):
  - lemma, determiner type, noun type, number, person, countability, WordNet senses, ...

Syntactic-semantic features

Implementation based on dkpro [Eckart de Castilho and Gurevych, 2014] and Stanford CoreNLP [Manning et al., 2014]

The bicycle has **undergone** continual adaptation and improvement.

- **main verb**:
  - tense, voice, progressive, perfect, lemma, WordNet hypernyms, ...
- **main referent** (subject):
  - lemma, determiner type, noun type, number, person, countability, WordNet senses, ...
- clause:
  - adverbs, conditional, modal, negated, ...

Syntactic-semantic features

Implementation based on dkpro [Eckart de Castilho and Gurevych, 2014] and Stanford CoreNLP [Manning et al., 2014]

The bicycle has **undergone** continual adaptation and improvement.

- **main verb**:
  - tense, voice, progressive, perfect, lemma, WordNet hypernyms, ...
- **main referent** (subject):
  - lemma, determiner type, noun type, number, person, countability, WordNet senses, ...
- **clause**:
  - adverbs, conditional, modal, negated, ...

Publicly available:
`https://github.com/annefried/sitent`

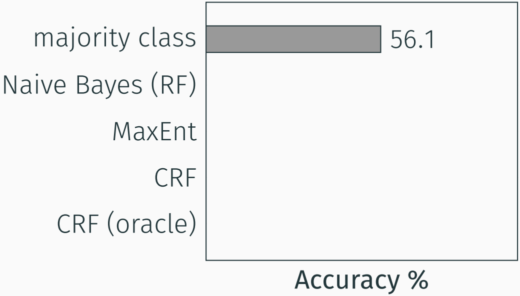
Discourse-sensitive identification of generic expressions

subject ***non-generic*** or ***generic***?

Discourse-sensitive identification of generic expressions

subject *non-generic* or *generic*?

Wikipedia
10,355 clauses
document-wise cross validation



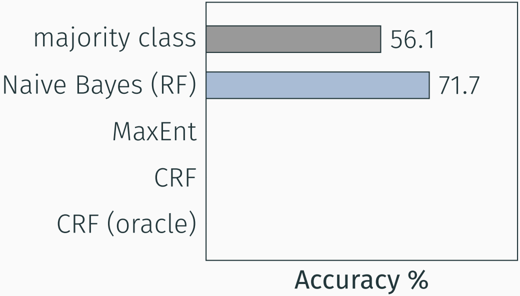
Discourse-sensitive identification of generic expressions

subject *non-generic* or *generic*?

Wikipedia

10,355 clauses

document-wise cross validation



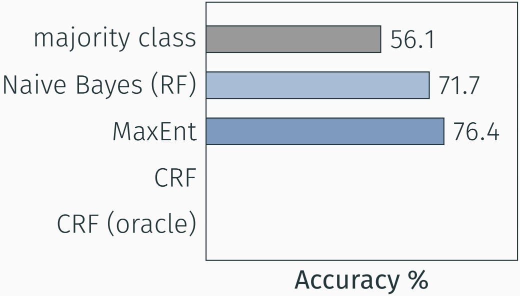
Discourse-sensitive identification of generic expressions

subject ***non-generic*** or ***generic***?

Wikipedia

10,355 clauses

document-wise cross validation



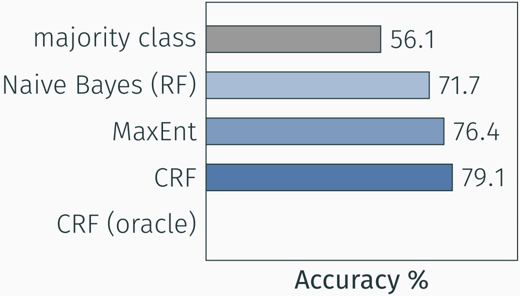
Discourse-sensitive identification of generic expressions

subject *non-generic* or *generic*?

Wikipedia

10,355 clauses

document-wise cross validation



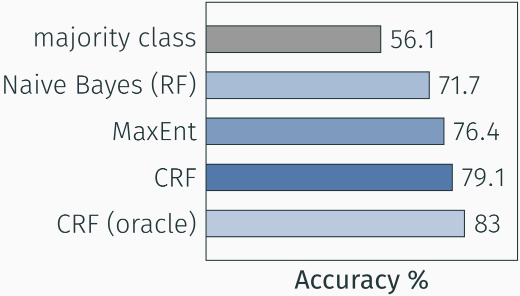
Discourse-sensitive identification of generic expressions

subject ***non-generic*** or ***generic***?

Wikipedia

10,355 clauses

document-wise cross validation



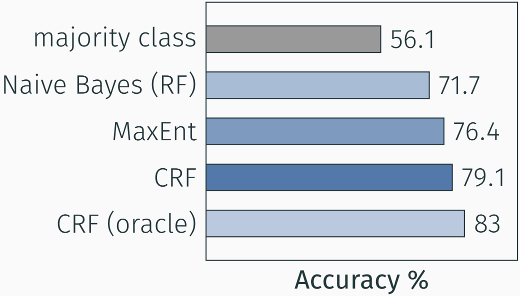
Discourse-sensitive identification of generic expressions

subject *non-generic* or *generic*?

Wikipedia

10,355 clauses

document-wise cross validation



Further findings

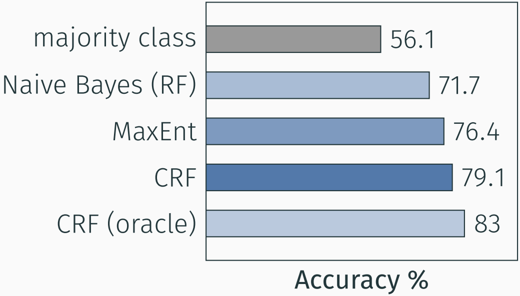
Discourse-sensitive identification of generic expressions

subject ***non-generic*** or ***generic***?

Wikipedia

10,355 clauses

document-wise cross validation



Further findings

- best results on **ACE-2** and **ACE-2005** data sets

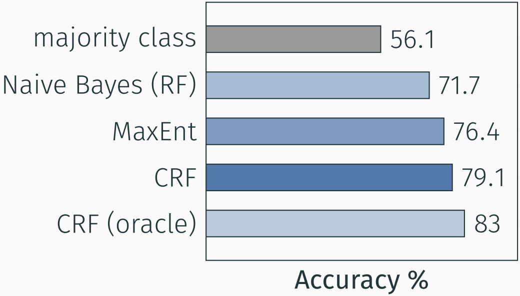
Discourse-sensitive identification of generic expressions

subject *non-generic* or *generic*?

Wikipedia

10,355 clauses

document-wise cross validation



Further findings

- best results on **ACE-2** and **ACE-2005** data sets
- features describing **clause** more important than **NP-based features**

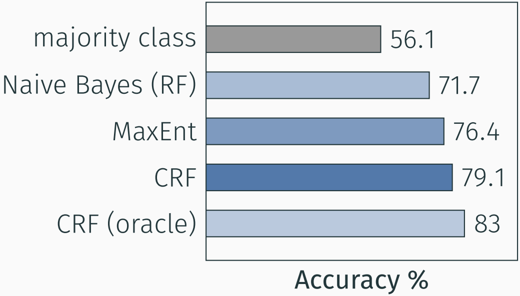
Discourse-sensitive identification of generic expressions

subject *non-generic* or *generic*?

Wikipedia

10,355 clauses

document-wise cross validation



Further findings

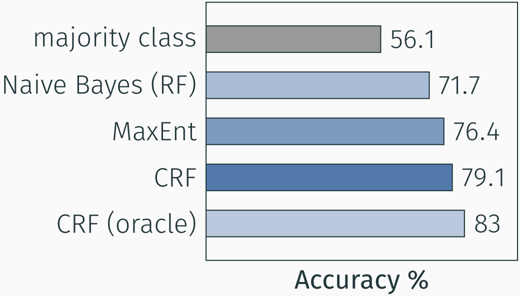
- best results on **ACE-2** and **ACE-2005** data sets
- features describing **clause** more important than **NP-based features**
- identification of **EVENTS related to kinds**

<u>Bikes</u> have two wheels. (GENERIC)
<u>The bike</u> was invented in the 19th century. (EVENT)

Discourse-sensitive identification of generic expressions

subject *non-generic* or *generic*?

Wikipedia
10,355 clauses
document-wise cross validation



Further findings

- best results on **ACE-2** and **ACE-2005** data sets
- features describing **clause** more important than **NP-based features**
- identification of **EVENTS related to kinds**

  Bikes have two wheels. (GENERIC)
  The bike was invented in the 19th century. (EVENT)

- sequence model often yields improvements when **coreference information** would be useful

Overview of thesis work

computational modeling

corpus creation / analysis of agreement

linguistic background / annotation scheme

generics
reference to kinds
[Friedrich & Pinkal, ACL 2015]
[Friedrich et al., LAW 2015]

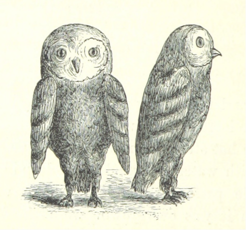
lexical aspect
state vs. event
[Friedrich & Palmer, ACL 2014]

habituals
generalization over situations
[Friedrich & Pinkal, EMNLP 2015]

situation entity types [Smith, 2003]
[Friedrich et al., ACL 2016], [Friedrich & Palmer, LAW 2014], [Mavridou et al., LSDSem 2015], [Palmer & Friedrich, 2014]

Automatic classification of situation entity types

[Smith, 2003] [Palmer et al., 2007]



STATE	Julie likes Cooper. Julie did **not** kill the mouse.
EVENT	Julie met Cooper two years ago.
REPORT	..., said the zookeeper.
GENERIC SENTENCE	**Owls** are nocturnal animals.
GENERALIZING SENTENCE	Julie often teases Cooper.
IMPERATIVE	Catch the mouse!
QUESTION	Why are there owls on your slides?

Automatic classification of situation entity types

4391 clauses from Brown corpus [Francis and Kučera, 1979]
majority class STATE (35.3%), κ = 0.52 [Palmer et al., 2007]

Automatic classification of situation entity types

4391 clauses from Brown corpus [Francis and Kučera, 1979]
majority class STATE (35.3%), κ = 0.52 [Palmer et al., 2007]

Features: words, pos tags, linguistic cues, grammatical cues

Accuracy (%)			
[Palmer et al., 2007]		[Friedrich et al., 2016]	
MaxEnt	LB	MaxEnt	CRF
50.6	53.1	55.8	60.0

Automatic classification of situation entity types

4391 clauses from Brown corpus [Francis and Kučera, 1979]
majority class STATE (35.3%), κ = 0.52 [Palmer et al., 2007]

Features: words, pos tags, linguistic cues, grammatical cues

Accuracy (%)			
[Palmer et al., 2007]		[Friedrich et al., 2016]	
MaxEnt	LB	MaxEnt	CRF
50.6	53.1	55.8	60.0

lookback features: predicted labels for previous clauses

Automatic classification of situation entity types

4391 clauses from Brown corpus [Francis and Kučera, 1979]
majority class STATE (35.3%), κ = 0.52 [Palmer et al., 2007]

Features: words, pos tags, linguistic cues, grammatical cues

Accuracy (%)			
[Palmer et al., 2007]		[Friedrich et al., 2016]	
MaxEnt	LB	MaxEnt	CRF
50.6	53.1	55.8	60.0

first true **sequence labeling** approach
for situation entity types

Automatic classification of situation entity types

4391 clauses from Brown corpus [Francis and Kučera, 1979]
majority class STATE (35.3%), κ = 0.52 [Palmer et al., 2007]

Features: words, pos tags, linguistic cues, grammatical cues

Accuracy (%)			
[Palmer et al., 2007]		[Friedrich et al., 2016]	
MaxEnt	LB	MaxEnt	CRF
50.6	53.1	55.8	60.0

Reason:
GENERIC SENTENCES
cluster together

Automatic classification of situation entity types

Features for clauses:

Automatic classification of situation entity types

Features for clauses:

- pos = part of speech tags

Automatic classification of situation entity types

Features for clauses:

- **pos** = part of speech tags
- **Brown clusters**
  [Brown et al., 1992],
  [Turian et al., 2010]:
  distributional information

Automatic classification of situation entity types

Features for clauses:

- **pos** = part of speech tags
- **Brown clusters** [Brown et al., 1992], [Turian et al., 2010]: distributional information
- **syntactic-semantic** features describe main verb, main referent (subject) and clause

Automatic classification of situation entity types

Features for clauses:

- **pos** = part of speech tags
- **Brown clusters** [Brown et al., 1992], [Turian et al., 2010]: distributional information
- **syntactic-semantic** features describe main verb, main referent (subject) and clause

most important

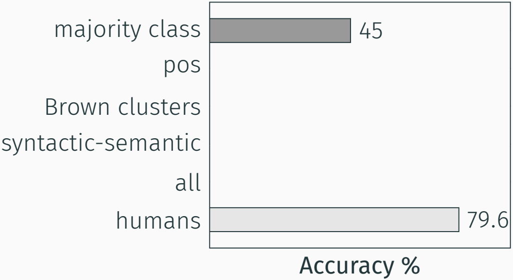
Automatic classification of situation entity types

7-way classification task
10-fold document-wise CV
dev set (80% of data)

Features for clauses:

- **pos** = part of speech tags
- **Brown clusters** [Brown et al., 1992], [Turian et al., 2010]: distributional information
- **syntactic-semantic** features describe main verb, main referent (subject) and clause

CRF (sequence model)



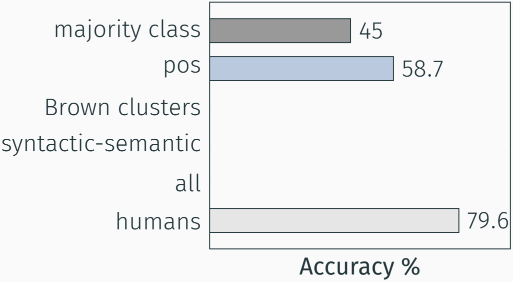
Automatic classification of situation entity types

7-way classification task
10-fold document-wise CV
dev set (80% of data)

Features for clauses:

- **pos** = part of speech tags
- **Brown clusters** [Brown et al., 1992], [Turian et al., 2010]: distributional information
- **syntactic-semantic** features describe main verb, main referent (subject) and clause

CRF (sequence model)



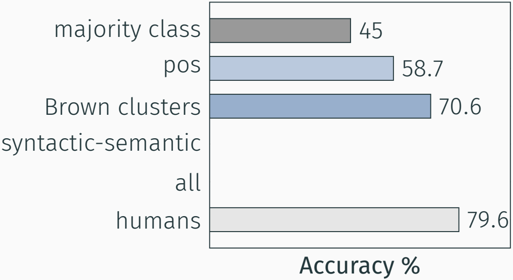
Automatic classification of situation entity types

7-way classification task
10-fold document-wise CV
dev set (80% of data)

Features for clauses:

- **pos** = part of speech tags
- **Brown clusters** [Brown et al., 1992], [Turian et al., 2010]: distributional information
- **syntactic-semantic** features describe main verb, main referent (subject) and clause

CRF (sequence model)



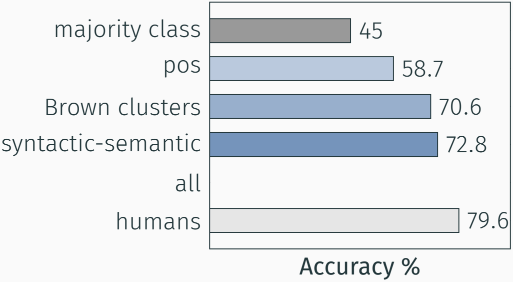
Automatic classification of situation entity types

7-way classification task
10-fold document-wise CV
dev set (80% of data)

Features for clauses:

- **pos** = part of speech tags
- **Brown clusters**
  [Brown et al., 1992],
  [Turian et al., 2010]:
  distributional information
- **syntactic-semantic** features
  describe main verb, main
  referent (subject) and clause

CRF (sequence model)



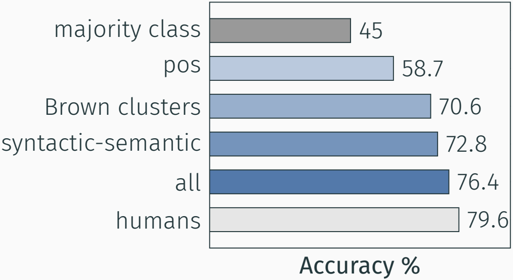
Automatic classification of situation entity types

7-way classification task
10-fold document-wise CV
dev set (80% of data)

Features for clauses:

- **pos** = part of speech tags
- **Brown clusters** [Brown et al., 1992], [Turian et al., 2010]: distributional information
- **syntactic-semantic** features describe main verb, main referent (subject) and clause

CRF (sequence model)



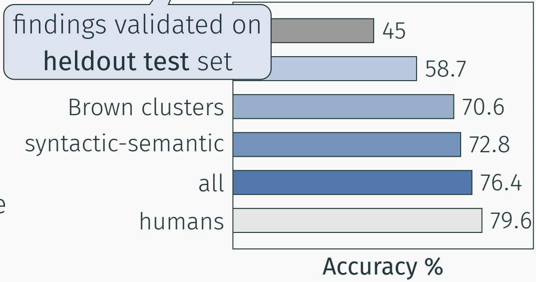
Automatic classification of situation entity types

7-way classification task
10-fold document-wise CV
dev set (80% of data)

Features for clauses:

- **pos** = part of speech tags
- **Brown clusters** [Brown et al., 1992], [Turian et al., 2010]: distributional information
- **syntactic-semantic** features describe main verb, main referent (subject) and clause

CRF (sequence model)



Automatic classification of situation entity types

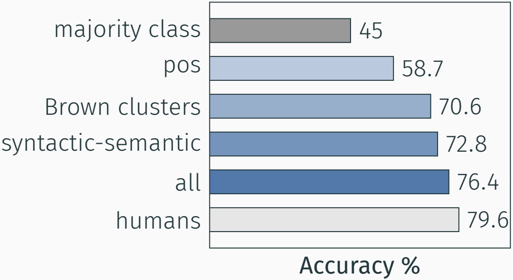
7-way classification task
10-fold document-wise CV
dev set (80% of data)

Features for clauses:

- **pos** = part of speech tags
- **Brown clusters**
  [Brown et al., 1992],
  [Turian et al., 2010]:
  distributional information
- **syntactic-semantic** features
  describe main verb, main
  referent (subject) and clause

Further findings:

CRF (sequence model)



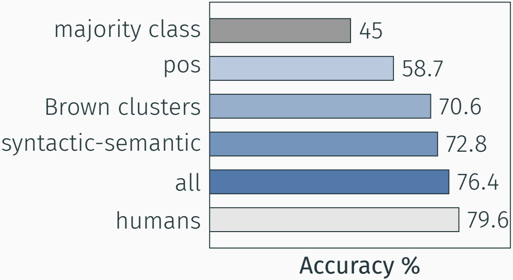
Automatic classification of situation entity types

7-way classification task
10-fold document-wise CV
dev set (80% of data)

Features for clauses:

- **pos** = part of speech tags
- **Brown clusters** [Brown et al., 1992], [Turian et al., 2010]: distributional information
- **syntactic-semantic** features describe main verb, main referent (subject) and clause

CRF (sequence model)



Further findings:

#1: model **trained directly on situation entity types** works better than pipelined model trained separately on the **subtasks**

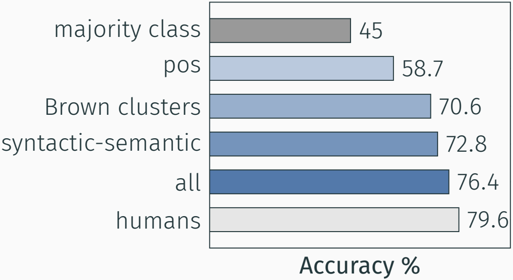
Automatic classification of situation entity types

7-way classification task
10-fold document-wise CV
dev set (80% of data)

Features for clauses:

- **pos** = part of speech tags
- **Brown clusters** [Brown et al., 1992], [Turian et al., 2010]: distributional information
- **syntactic-semantic** features describe main verb, main referent (subject) and clause

CRF (sequence model)



Further findings:

#1: model **trained directly on situation entity types** works better than pipelined model trained separately on the **subtasks**

#2: good performance across **genres**
out-of-genre training data helps for infrequent types

Overview of thesis work

computational modeling

corpus creation / analysis of agreement

linguistic background / annotation scheme

generics
reference to kinds
[Friedrich & Pinkal, ACL 2015]
[Friedrich et al., LAW 2015]

lexical aspect
state vs. event
[Friedrich & Palmer, ACL 2014]

habituals
generalization
over situations
[Friedrich & Pinkal, EMNLP 2015]

situation entity types [Smith, 2003]
[Friedrich et al., ACL 2016], [Friedrich & Palmer, LAW 2014],
[Mavridou et al., LSDSem 2015], [Palmer & Friedrich, 2014]

Automatic prediction of lexical aspectual class

She **filled** the glass with water. (***dynamic***)
The glass **is filled** with water. (***stative***)

[Friedrich & Palmer, ACL 2014]

Automatic prediction of lexical aspectual class

She **filled** the glass with water. (***dynamic***)
The glass **is filled** with water. (***stative***)

[Friedrich & Palmer, ACL 2014]

MASC (jokes, news, letters) • 7875 clauses

Random Forest classifier [Breiman, 2001] with features:

Automatic prediction of lexical aspectual class

She **filled** the glass with water. (***dynamic***)
The glass **is filled** with water. (***stative***)

[Friedrich & Palmer, ACL 2014]

anc AMERICAN NATIONAL CORPUS MASC (jokes, news, letters) • 7875 clauses

Random Forest classifier [Breiman, 2001] with features:

- **linguistic indicators**[Siegel and McKeown, 2000]:
  tendency of verb type "fill" to occur with progressive? etc.
  estimated over GigaWord [Graff et al., 2003]

Automatic prediction of lexical aspectual class

She **filled** the glass with water. (***dynamic***)
The glass **is filled** with water. (***stative***)

[Friedrich & Palmer, ACL 2014]

anc AMERICAN NATIONAL CORPUS

MASC (jokes, news, letters) • 7875 clauses

Random Forest classifier [Breiman, 2001] with features:

- **linguistic indicators**[Siegel and McKeown, 2000]:
  tendency of verb type "fill" to occur with progressive? etc.
  estimated over GigaWord [Graff et al., 2003]
- **contextual features**: subject of "fill" = "she"

Automatic prediction of lexical aspectual class

She **filled** the glass with water. (***dynamic***)
The glass **is filled** with water. (***stative***)

[Friedrich & Palmer, ACL 2014]

MASC (jokes, news, letters) • 7875 clauses

Random Forest classifier [Breiman, 2001] with features:

- **linguistic indicators**[Siegel and McKeown, 2000]:
  tendency of verb type "fill" to occur with progressive? etc.
  estimated over GigaWord [Graff et al., 2003]
- **contextual features**: subject of "fill" = "she"

Finding #1: linguistic indicators generalize across verb types

Automatic prediction of lexical aspectual class

She **filled** the glass with water. (***dynamic***)
The glass **is filled** with water. (***stative***)

[Friedrich & Palmer, ACL 2014]

MASC (jokes, news, letters) • 7875 clauses

Random Forest classifier [Breiman, 2001] with features:

- **linguistic indicators**[Siegel and McKeown, 2000]:
  tendency of verb type "fill" to occur with progressive? etc.
  estimated over GigaWord [Graff et al., 2003]
- **contextual features**: subject of "fill" = "she"

Finding #1: linguistic indicators generalize across verb types

10-fold cross validation • verbs in test folds ∩ verbs in train fold = ∅

Automatic prediction of lexical aspectual class

She **filled** the glass with water. (***dynamic***)
The glass **is filled** with water. (***stative***)

[Friedrich & Palmer, ACL 2014]

MASC (jokes, news, letters) • 7875 clauses

Random Forest classifier [Breiman, 2001] with features:

- **linguistic indicators**[Siegel and McKeown, 2000]:
  tendency of verb type "fill" to occur with progressive? etc.
  estimated over GigaWord [Graff et al., 2003]
- **contextual features**: subject of "fill" = "she"

Finding #1: **linguistic indicators generalize across verb types**

10-fold cross validation • verbs in test folds ∩ verbs in train fold = ∅

Finding #2: **contextual features help for ambiguous verb types**

Overview of thesis work

computational modeling

corpus creation / analysis of agreement

linguistic background / annotation scheme

generics
reference to kinds
[Friedrich & Pinkal, ACL 2015]
[Friedrich et al., LAW 2015]

lexical aspect
state vs. event
[Friedrich & Palmer, ACL 2014]

habituals
generalization over situations
[Friedrich & Pinkal, EMNLP 2015]

situation entity types [Smith, 2003]
[Friedrich et al., ACL 2016], [Friedrich & Palmer, LAW 2014],
[Mavridou et al., LSDSem 2015], [Palmer & Friedrich, 2014]

Automatic recognition of habituals

John bought a bike.	(***episodic***)
John cycles to work.	(***habitual***)

Automatic recognition of habituals

John bought a bike.	(***episodic***)	[Mathew and Katz, 2009]
John cycles to work.	(***habitual***)	

Automatic recognition of habituals

John bought a bike. (***episodic***)
John cycles to work. (***habitual***)
[Mathew and Katz, 2009]

John likes coffee.

stative

Automatic recognition of habituals

John bought a bike. (***episodic***)
John cycles to work. (***habitual***)
[Mathew and Katz, 2009]

John likes coffee.
Bill can cycle.

coercion to STATE:
negation, modality, future,
perfect, conditionality

Automatic recognition of habituals

John bought a bike.	(*episodic*)	[Mathew and Katz, 2009]
John cycles to work.	(*habitual*)	
John likes coffee.	(*static*)	[Friedrich & Pinkal, EMNLP 2015]
Bill can cycle.		

Automatic recognition of habituals

John bought a bike.	(*episodic*)	[Mathew and Katz, 2009]
John cycles to work.	(*habitual*)	
John likes coffee. Bill can cycle.	(*static*)	[Friedrich & Pinkal, EMNLP 2015]

Wikipedia (10,355 clauses) • Random Forest classifiers [Breiman, 2001]

Automatic recognition of habituals

John bought a bike.	(*episodic*)	[Mathew and Katz, 2009]
John cycles to work.	(*habitual*)	
John likes coffee. Bill can cycle.	(*static*)	[Friedrich & Pinkal, EMNLP 2015]

Wikipedia (10,355 clauses) • Random Forest classifiers [Breiman, 2001]

Findings / contributions:

Automatic recognition of habituals

John bought a bike.	(*episodic*)	[Mathew and Katz, 2009]
John cycles to work.	(*habitual*)	
John likes coffee. Bill can cycle.	(*static*)	[Friedrich & Pinkal, EMNLP 2015]

Wikipedia (10,355 clauses) • Random Forest classifiers [Breiman, 2001]

Findings / contributions:

#1: recognizing ***habituals*** in free text requires a **three-way distinction**

Automatic recognition of habituals

John bought a bike.	(*episodic*)	[Mathew and Katz, 2009]
John cycles to work.	(*habitual*)	
John likes coffee.	(*static*)	[Friedrich & Pinkal, EMNLP 2015]
Bill can cycle.		

Wikipedia (10,355 clauses) • Random Forest classifiers [Breiman, 2001]

Findings / contributions:

- #1: recognizing ***habituals*** in free text requires a **three-way distinction**
- #2: **contextual** and **verb type-based features** are complementary

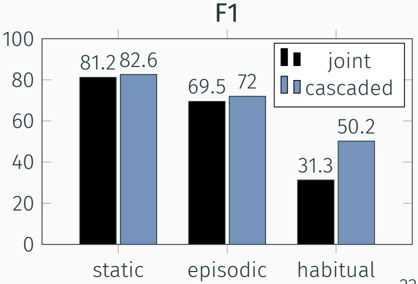
Automatic recognition of habituals

John bought a bike.	(*episodic*)	[Mathew and Katz, 2009]
John cycles to work.	(*habitual*)	
John likes coffee. Bill can cycle.	(*static*)	[Friedrich & Pinkal, EMNLP 2015]

Wikipedia (10,355 clauses) • Random Forest classifiers [Breiman, 2001]

Findings / contributions:

- #1: recognizing ***habituals*** in free text requires a **three-way distinction**
- #2: **contextual** and **verb type-based features** are complementary
- #3: filtering out ***static*** clauses first is beneficial (**cascaded** model)



Conclusion / contributions

computational modeling

corpus creation / analysis of agreement

linguistic background / annotation scheme

generics
reference to kinds

[Friedrich & Pinkal, ACL 2015]
[Friedrich et al., LAW 2015]

lexical aspect
state vs. event

[Friedrich & Palmer, ACL 2014]

habituals
generalization over situations

[Friedrich & Pinkal, EMNLP 2015]

situation entity types [Smith, 2003]

[Friedrich et al., ACL 2016], [Friedrich & Palmer, LAW 2014], [Mavridou et al., LSDSem 2015], [Palmer & Friedrich, 2014]

Conclusion / contributions

computational modeling

corpus creation / analysis of agreement

n scheme

first large reliably annotated **corpus**
for situation entity types
(40,000 clauses, 13 genres)
and related aspectual + semantic phenomena

[Friedrich & Pinkal, ACL 2015]
[Friedrich et al., LAW 2015]

[Friedrich & Palmer, ACL 2014]

habituals
generalization
over situations

[Friedrich & Pinkal, EMNLP 2015]

situation entity types [Smith, 2003]

[Friedrich et al., ACL 2016], [Friedrich & Palmer, LAW 2014],
[Mavridou et al., LSDSem 2015], [Palmer & Friedrich, 2014]

Conclusion / contributions

computational modeling

computational models for aspectual distinctions
outperform prior approaches in each case;
implementation publicly available

linguistic background / annotation scheme

generics
reference to kinds
[Friedrich & Pinkal, ACL 2015]
[Friedrich et al., LAW 2015]

lexical aspect
state vs. event
[Friedrich & Palmer, ACL 2014]

habituals
generalization
over situations
[Friedrich & Pinkal, EMNLP 2015]

situation entity types [Smith, 2003]
[Friedrich et al., ACL 2016], [Friedrich & Palmer, LAW 2014],
[Mavridou et al., LSDSem 2015], [Palmer & Friedrich, 2014]

Conclusion / contributions / directions for future work

- Situation entity types can be annotated with **reasonable agreement** if broken down into related sub-tasks [Friedrich & Palmer, LAW 2014]

Conclusion / contributions / directions for future work

- Situation entity types can be annotated with **reasonable agreement** if broken down into related sub-tasks [Friedrich & Palmer, LAW 2014]
    - **crowdsource** relevant annotations?

Conclusion / contributions / directions for future work

- Situation entity types can be annotated with **reasonable agreement** if broken down into related sub-tasks [Friedrich & Palmer, LAW 2014]
    - **crowdsource** relevant annotations?
- Inventory of situation entity types **cross-linguistically applicable**, but different implementation required [Mavridou et al., LSDSem 2015]

Conclusion / contributions / directions for future work

- Situation entity types can be annotated with **reasonable agreement** if broken down into related sub-tasks [Friedrich & Palmer, LAW 2014]
    - **crowdsource** relevant annotations?
- Inventory of situation entity types **cross-linguistically applicable**, but different implementation required [Mavridou et al., LSDSem 2015]
    - English Perfect vs. German Perfekt

Conclusion / contributions / directions for future work

- Situation entity types can be annotated with **reasonable agreement** if broken down into related sub-tasks [Friedrich & Palmer, LAW 2014]
    - **crowdsource** relevant annotations?
- Inventory of situation entity types **cross-linguistically applicable**, but different implementation required [Mavridou et al., LSDSem 2015]
    - English Perfect vs. German Perfekt
    - lexical choice: *she is startled* (STATE) vs. *sie erschrickt* (EVENT)

Conclusion / contributions / directions for future work

- Situation entity types can be annotated with **reasonable agreement** if broken down into related sub-tasks [Friedrich & Palmer, LAW 2014]
    - **crowdsource** relevant annotations?
- Inventory of situation entity types **cross-linguistically applicable**, but different implementation required [Mavridou et al., LSDSem 2015]
    - English Perfect vs. German Perfekt
    - lexical choice: *she is startled* (STATE) vs. *sie erschrickt* (EVENT)
- Semantic theory about **generics** [Krifka et al., 1995] works well in some **genres** (e.g., encyclopedic), less well in others (e.g., essays) [Friedrich et al., LAW 2015]

Conclusion / contributions / directions for future work

- Situation entity types can be annotated with **reasonable agreement** if broken down into related sub-tasks [Friedrich & Palmer, LAW 2014]
  - **crowdsource** relevant annotations?
- Inventory of situation entity types **cross-linguistically applicable**, but different implementation required [Mavridou et al., LSDSem 2015]
  - English Perfect vs. German Perfekt
  - lexical choice: *she is startled* (STATE) vs. *sie erschrickt* (EVENT)
- Semantic theory about **generics** [Krifka et al., 1995] works well in some **genres** (e.g., encyclopedic), less well in others (e.g., essays) [Friedrich et al., LAW 2015]
  - is there a way to annotate / model the "**underspecified**" cases?
    Students at Saarland university eat at the mensa.

Conclusion / contributions / directions for future work

- Situation entity types can be annotated with **reasonable agreement** if broken down into related sub-tasks [Friedrich & Palmer, LAW 2014]
  - **crowdsource** relevant annotations?
- Inventory of situation entity types **cross-linguistically applicable**, but different implementation required [Mavridou et al., LSDSem 2015]
  - English Perfect vs. German Perfekt
  - lexical choice: *she is startled* (STATE) vs. *sie erschrickt* (EVENT)
- Semantic theory about **generics** [Krifka et al., 1995] works well in some **genres** (e.g., encyclopedic), less well in others (e.g., essays) [Friedrich et al., LAW 2015]
  - is there a way to annotate / model the “**underspecified**” cases?
    Students at Saarland university eat at the mensa.

⇒ **modification of situation entity types inventory**?

Conclusion / contributions / directions for future work

- Situation entity types can be annotated with **reasonable agreement** if broken down into related sub-tasks [Friedrich & Palmer, LAW 2014]
  - **crowdsource** relevant annotations?
- Inventory of situation entity types **cross-linguistically applicable**, but different implementation required [Mavridou et al., LSDSem 2015]
  - English Perfect vs. German Perfekt
  - lexical choice: *she is startled* (STATE) vs. *sie erschrickt* (EVENT)
- Semantic theory about **generics** [Krifka et al., 1995] works well in some **genres** (e.g., encyclopedic), less well in others (e.g., essays) [Friedrich et al., LAW 2015]
  - is there a way to annotate / model the "**underspecified**" cases?
    Students at Saarland university eat at the mensa.

⇒ **modification of situation entity types inventory**?

- current set possibly too coarse-grained for many NLP applications

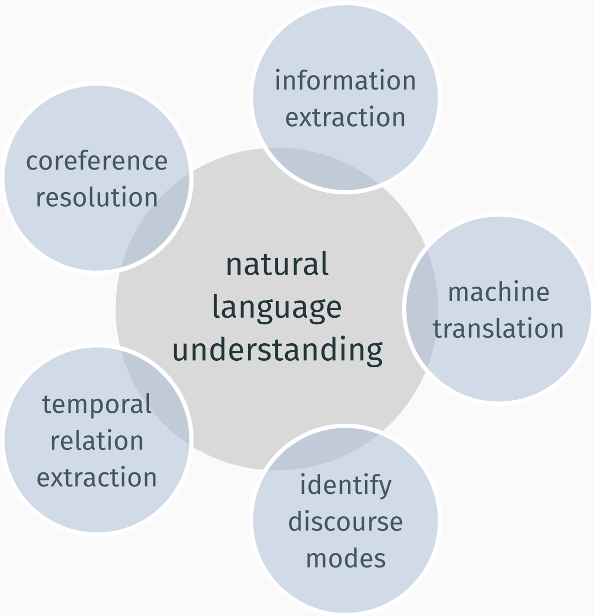
Conclusion / contributions / directions for future work

- Situation entity types can be annotated with **reasonable agreement** if broken down into related sub-tasks [Friedrich & Palmer, LAW 2014]
    - **crowdsource** relevant annotations?
- Inventory of situation entity types **cross-linguistically applicable**, but different implementation required [Mavridou et al., LSDSem 2015]
    - English Perfect vs. German Perfekt
    - lexical choice: *she is startled* (STATE) vs. *sie erschrickt* (EVENT)
- Semantic theory about **generics** [Krifka et al., 1995] works well in some **genres** (e.g., encyclopedic), less well in others (e.g., essays) [Friedrich et al., LAW 2015]
    - is there a way to annotate / model the "**underspecified**" cases?
      Students at Saarland university eat at the mensa.

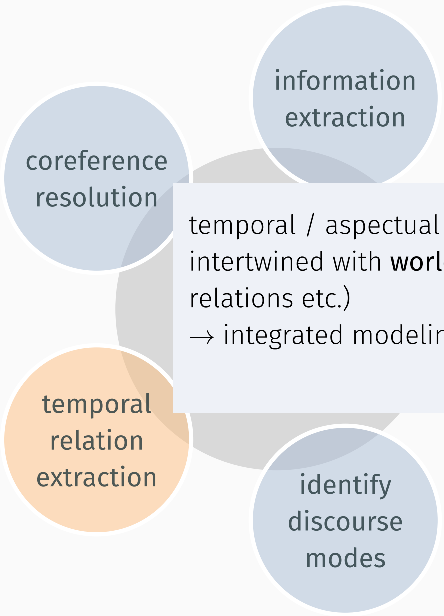
⇒ **modification of situation entity types inventory**?

- current set possibly too coarse-grained for many NLP applications
- distinguish among the different types of STATES
    - *John is tall* vs. *John is hungry* vs. *John can swim*

Directions for future work



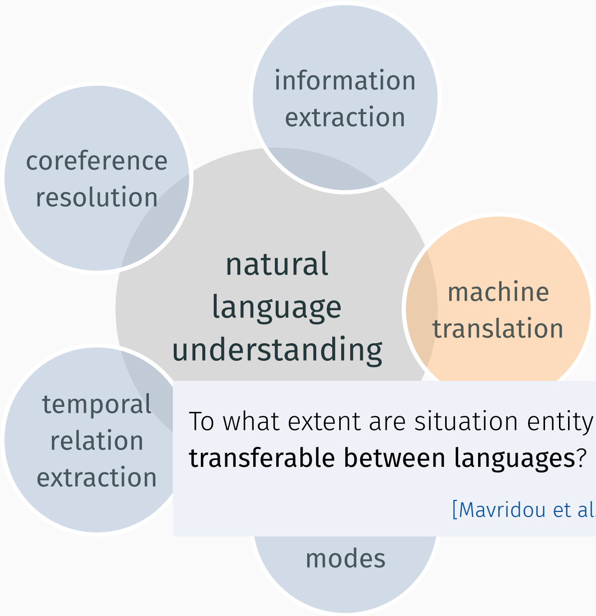
Directions for future work



temporal / aspectual interpretation always intertwined with **world knowledge** (causal relations etc.)
→ integrated modeling?

[Steedman, 2012]

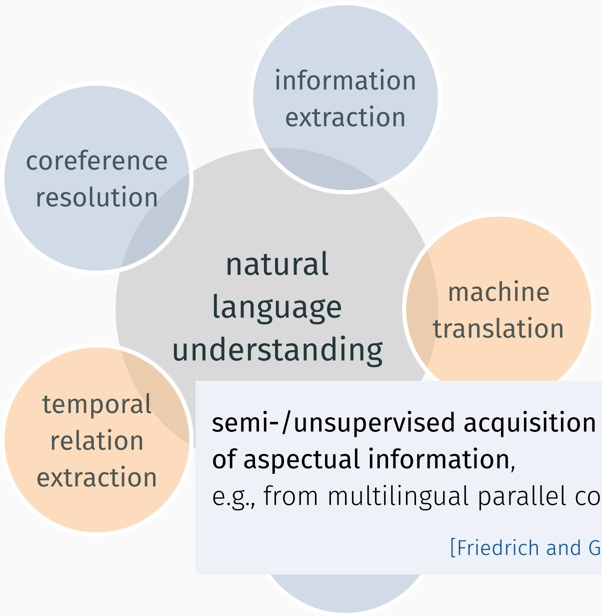
Directions for future work



To what extent are situation entity types **transferable between languages**?

[Mavridou et al., LSDSem 2015]

Directions for future work



semi-/unsupervised acquisition of aspectual information,
e.g., from multilingual parallel corpora?

[Friedrich and Gateva, 2017]

States, events, and generics: computational modeling of situation entity types

computational modeling

corpus creation / analysis of agreement

linguistic background / annotation scheme

lexical aspect

state vs. event

[Friedrich & Palmer, ACL 2014]

habituals

generalization over situations

[Friedrich & Pinkal, EMNLP 2015]

generics

reference to kinds

[Friedrich & Pinkal, ACL 2015]
[Friedrich et al., LAW 2015]

situation entity types [Smith, 2003]

[Friedrich et al., ACL 2016], [Friedrich & Palmer, LAW 2014],
[Mavridou et al., LSDSem 2015], [Palmer & Friedrich, 2014]

References I

Bach, E. (1986).
The algebra of events.
Linguistics and philosophy, 9(1):5–16.

Breiman, L. (2001).
Random forests.
Machine learning, 45(1):5–32.

Brown, P. F., Desouza, P. V., Mercer, R. L., Pietra, V. J. D., and Lai, J. C. (1992).
Class-based n-gram models of natural language.
Computational linguistics, 18(4):467–479.

References II

Eckart de Castilho, R. and Gurevych, I. (2014).
A broad-coverage collection of portable NLP components for building shareable analysis pipelines.
In *Proceedings of the Workshop on Open Infrastructures and Analysis Frameworks for HLT*, pages 1–11, Dublin, Ireland.

Francis, W. N. and Kučera, H. (1979).
Brown corpus manual.
Brown University.
http://clu.uni.no/icame/brown/bcm.html.

Friedrich, A. and Gateva, D. (2017).
Classification of telicity using cross-linguistic annotation projection.
under submission.

References III

Friedrich, A. and Palmer, A. (2014).
Situation entity annotation.
In *Proceedings of the 8th Linguistic Annotation Workshop (LAW VIII)*, Dublin, Ireland.

Friedrich, A., Palmer, A., and Pinkal, M. (2016).
Situation entity types: automatic classification of clause-level aspect.
In *In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL)*, Berlin, Germany.

Friedrich, A., Palmer, A., Sørensen, M. P., and Pinkal, M. (2015).
Annotating genericity: a survey, a scheme, and a corpus.
In *Proceedings of the 9th Linguistic Annotation Workshop (LAW IX)*, Denver, Colorado, USA.

References IV

Graff, D., Kong, J., Chen, K., and Maeda, K. (2003).
English Gigaword.
Linguistic Data Consortium, Philadelphia.

Hermes, J., Richter, M., and Neuefeind, C. (2015).
Automatic Induction of German Aspectual Verb Classes in a Distributional Framework.
In *Proceedings of the International Conference of the German Society for Computational Linguistics and Language Technology (GSCL)*, pages 122–129, Duisburg-Essen, Germany.

References V

Krifka, M., Pelletier, F. J., Carlson, G. N., ter Meulen, A., Link, G., and Chierchia, G. (1995).
Genericity: An Introduction.
In Carlson, G. N. and Pelletier, F. J., editors, *The Generic Book*, Studies in Communication, Media, and Public Opinion, pages 1–124.
University Of Chicago Press.

Krippendorff, K. (1980).
Content analysis: An introduction to its methodology.
Sage.

Loáiciga, S. and Grisot, C. (2016).
Predicting and using a pragmatic component of lexical aspect.
Linguistic Issues in Language Technology, Special issue on Modality in Natural Language Understanding, 13.

References VI

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S. J., and McClosky, D. (2014).
The Stanford CoreNLP Natural Language Processing Toolkit.
In *Proceedings of the 52nd annual meeting of the Association for Computational Linguistics (ACL) System Demonstrations*, pages 55–60, Baltimore, Maryland.

Mathew, T. A. and Katz, E. G. (2009).
Supervised Categorization of Habitual and Episodic Sentences.
In *Sixth Midwest Computational Linguistics Colloquium*, Bloomington, Indiana: Indiana University.

References VII

Nedoluzhko, A. (2013).
Generic noun phrases and annotation of coreference and bridging relations in the Prague Dependency Treebank.
In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse (LAW-VII)*, pages 103–111, Sofia, Bulgaria.

Palmer, A., Ponvert, E., Baldridge, J., and Smith, C. (2007).
A sequencing model for situation entity classification.
In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 896–903, Prague, Czech Republic.

References VIII

Reichart, R. and Rappoport, A. (2010).
Tense sense disambiguation: a new syntactic polysemy task.
In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 325–334. Association for Computational Linguistics.

Reiter, N. and Frank, A. (2010).
Identifying generic noun phrases.
In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 40–49, Sweden, Uppsala.

Siegel, E. V. and McKeown, K. R. (2000).
Learning methods to combine linguistic indicators: Improving aspectual classification and revealing linguistic insights.
Computational Linguistics, 26(4):595–628.

References IX

Smith, C. S. (2003).
Modes of discourse: The local structure of texts.
Cambridge University Press.

Sonnenhauser, B. (2004).
Proceedings of the 2nd Workshop on Text Meaning and Interpretation, chapter Underspecification of 'meaning': The case of Russian imperfective aspect.

Soricut, R. and Marcu, D. (2003).
Sentence level discourse parsing using syntactic and lexical information.
In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human*

References X

Language Technology (NAACL-HLT), pages 149–156, Edmonton, Canada.

Steedman, M. (2012).
Computational linguistics.
In Binnick, R. I., editor, *The Oxford handbook of tense and aspect.* Oxford University Press.

Turian, J., Ratinov, L., and Bengio, Y. (2010).
Word representations: a simple and general method for semi-supervised learning.
In *Proceedings of the 48th annual meeting of the Association for Computational Linguistics (ACL)*, pages 384–394, Uppsala, Sweden.

References XI

Van Durme, B. D. (2009).
Extracting implicit knowledge from text.
PhD thesis, University of Rochester.

Vendler, Z. (1957).
Verbs and times.
The philosophical review, pages 143–160.

Zarcone, A. and Lenci, A. (2008).
Computational models for event type classification in context.
In *Proceedings of The International Conference on Language Resources and Evaluation (LREC)*, Marrakech, Morocco.

References XII

Zhang, Y. and Xue, N. (2014).
Automatic Inference of the Tense of Chinese Events using Implicit Linguistic Information.
In *Proceedings of The Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar.