Situation entity types:
automatic classification of clause-level aspect

Annemarie Friedrich[1] Alexis Palmer[2] Manfred Pinkal[1]

[1]Saarland University [2]Heidelberg University / University of North Texas

ACL 2016, Berlin, Germany

What is clause-level aspect?

aspect = how a situation is presented [Smith 1997]

[Vendler 1957, Bach 1986]:
aktionsart

state:	*The ship is in motion.*
event:	*The ship moved.*
process:	*The ship is moving.*

[Krifka et al. 1995]:
habituals / genericity

John cycles to work.

Students like coffee.

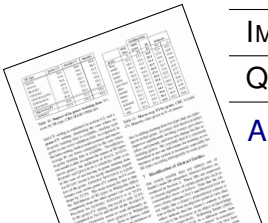
Why model these?

- understand temporal relations in discourse
- distinguish between / extract different types of knowledge

What are situation entity types?

inventory of aspectual clause types
motivated by a theory of discourse
[Smith 2003]

Situation entity types



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?

ABSTRACT ENTITIES → see paper

What we show in this paper

- first large reliably annotated **corpus** for situation entity types (40,000 clauses), 13 genres
- use of **distributional information** (Brown clusters) to make approach robust + scalable

  45% (informed baseline) - 76% (system) - 80% (humans)
- **sequence** labeling method (CRF) vs. local method (MaxEnt): small impact, depending on genre

Related work

- modeling of Vendler classes
  - *state, activity, accomplishment, achievement*
  - Italian [Zarcone & Lenci, 2008], German [Hermes et al., 2015]
  - stative vs. dynamic [Siegel & McKeown, 2000], [Friedrich & Palmer, 2014a]
  - completedness [Siegel & McKeown, 2000]
- modeling genericity
  - identifying genericity of NPs / reference to kinds
    [Reiter & Frank, 2010], [Friedrich & Pinkal, 2015b]
  - recognizing habituals
    [Mathew & Katz, 2009], [Friedrich & Pinkal, 2015a]
- labeling situation entities [Palmer et al. 2007]
  - maximum entropy model, features: pos tags, words, linguistic
  - data set: 20 texts / 4391 clauses, Brown corpus, $\kappa = 0.52$

Data collection: MASC / Wiki corpus

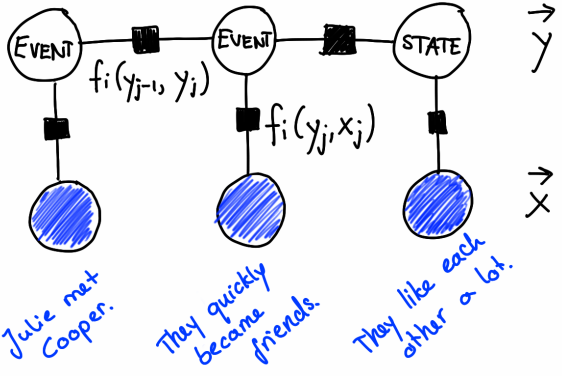
- $\approx$ 30,000 clauses from MASC [Ide et al. 2010]
  + $\approx$ 10,000 clauses from Wikipedia
- automatically segmented using SPADE [Soricut & Marcu, 2003]
- 3 annotators $\rightarrow$ majority voting $\rightarrow$ gold standard

	% in gold standard		**Fleiss' κ**
Situation entity type	**MASC**	**Wiki**	Krippendorff's diagnostics
STATE	49.8	24.3	0.67
EVENT	24.3	18.9	0.74
REPORT	4.8	0.9	0.80
GENERIC	7.3	49.7	0.68
GENERALIZING	3.8	2.5	0.43
QUESTION	3.3	0.1	0.91
IMPERATIVE	3.2	0.2	0.94
undecided	2.4	2.1	-

Conditional random field (CRF)

- text document
  = sequence of clauses
- $\vec{y}$ = sequence of situation entity type labels
- $\vec{x}$ = features representing the clauses
- λ_i = weight for feature x_i
- $f_i(y_j, x_j)$ = clause / type
  $\rightarrow$ MaxEnt
- $f_i(y_{j-1}, y_j)$ = type / type
  $\rightarrow$ CRF

$$P(\vec{y}|\vec{x}) = \frac{1}{Z(\vec{x})} exp(\sum_{j=1}^{n} \sum_{i=1}^{m} \lambda_i f_i(y_{j-1}, y_j, \vec{x}, j))$$



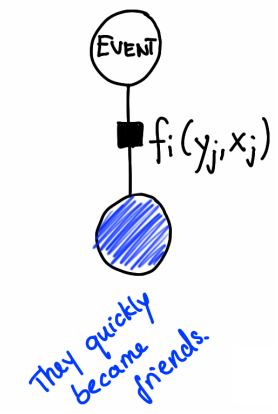
Situation entity types

Which parts of the clause are most important to distinguish the types?

[Friedrich & Palmer 2014b], [Friedrich et al. 2015], [Smith 2003]

Main verb	→ verb that heads the clause	
	Julie **likes** Cooper.	STATE
	Julie **met** Cooper.	EVENT
	Julie **teases** Cooper.	GENERALIZING SENTENCE
Main referent	→ subject of main verb	
	Julie is an owl.	STATE
	Owls are nocturnal animals.	GENERIC SENTENCE

Features for clauses



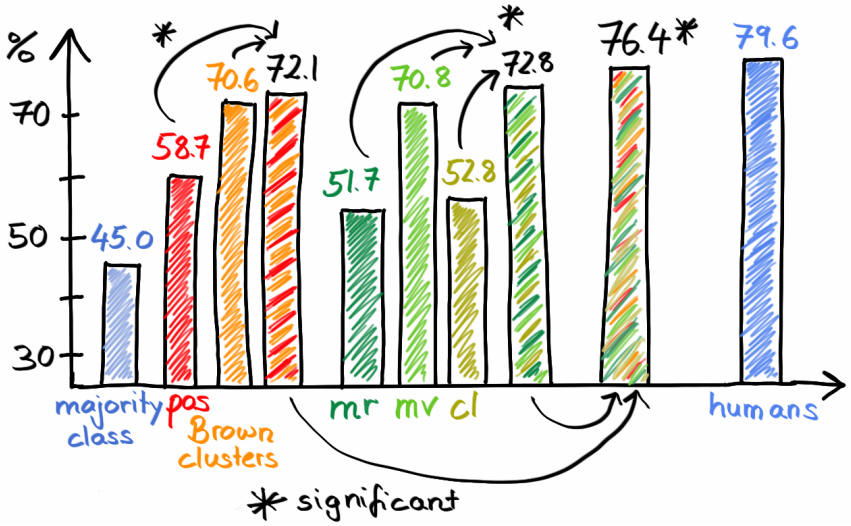
- **pos**: part of speech tags
- **bc**: Brown word clusters
  pretrained Turian et al. 2010

- **mv**: main verb
  - tense, voice, progressive, perfect, lemma, WordNet hypernyms ...
- **mr**: main referent
  - lemma, determiner type, noun type, number, person, countability, WordNet, dependency relations ...
- **cl**: clause
  - adverbs, conditional, modal, negated, ...

How well does it work?

Results: impact of different feature sets

Accuracy. Wiki+MASC dev set (80% of data), CRF, 10-fold CV.

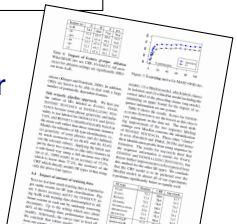


Results on heldout test set (20% of data)

Training on entire MASC+Wiki dev set.

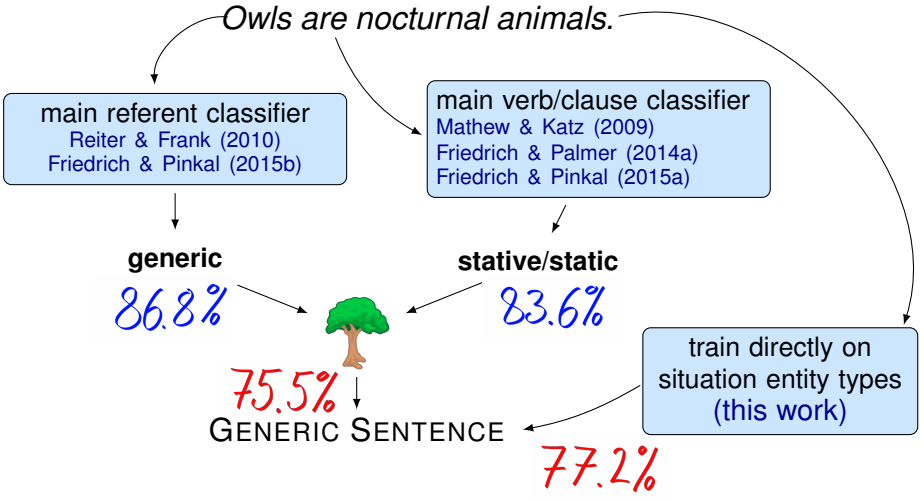
	macro-average			
feature set	**P**	**R**	**F**	**accuracy**
maj. class (STATE)	6.4	14.3	8.8	44.7
pos+Brown	67.6	60.6	63.9	69.8
mr+mv+cl	69.9	61.7	65.5	71.4
all	**73.4**	**65.5**	**69.3**	**74.7**

Ablation tests tell same story → see paper



Pipelined model?

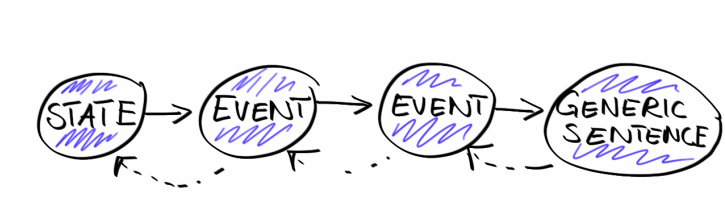
STATE, EVENT, GENERIC SENTENCE, GENERALIZING SENTENCE



Is sequential information important?

As claimed by Palmer et. al [2007]

... and if yes, when?



Maximum entropy model vs. conditional random field

SE type	MaxEnt	CRF
STATE	79.1	**80.6**
EVENT	77.5	**78.6**
REPORT	78.2	**78.9**
GENERIC	61.3	**68.3**
GENERALIZING	25.0	**29.4**
IMPERATIVE	72.3	**75.3**
QUESTION	84.4	**84.4**
macro-avg F1	68.7	**71.2**
accuracy	$*$74.1	$*$**76.4**

$*$ statistically significant

How genre-dependent is this task?

- How important is it to have in-genre training data?
  helpful, $\approx$ +5% accuracy/F_1
- Is it a good idea to add out-of-genre / domain training data?
  YES! 49.0 $\rightarrow$ 64.0 (macro-average F_1)
  system gets better at identifying infrequent types
- Statistics per type / genre $\rightarrow$ see paper

Lessons learned

- situation entity type classification task is difficult even for humans
- system performs well when comparing to human upper bound (76% vs. 80%)
- our system performs well across genres
- some types are infrequent in particular genres
  $\rightarrow$ adding out-of-domain training data helps to identify them
- a wide range of syntactic-semantic features are useful for this task
- sequential information useful for identifying 'generic contexts'

What next?

- integration of aspectual information into applications
  - temporal relation processing, argumentation mining, information extraction, translation
  - distinguishing different 'modes' of discourse (NARRATIVE, INFORMATION, REPORT, DESCRIPTION, ARGUMENTATIVE [Smith 2003])
- modification of situation entity types inventory
  - set of types by Smith [2003] possibly too coarse-grained

Thanks!

Ambika Kirkland – Christine Bocionek – Damyana Gateva – Fernando Ardente – Kleio-Isidora Mavridou – Melissa Peate Sørensen – Ruth Kühn

`http://coli.uni-saarland.de/projects/sitent`

Questions?

"Burrowing owl" by kuhnmi / CC BY 2.0

`http://coli.uni-saarland.de/projects/sitent`

References

- Emmon Bach. 1986. **The algebra of events.** Linguistics and philosophy 9(1):5–16.
- Annemarie Friedrich and Alexis Palmer. 2014a. **Automatic prediction of aspectual class of verbs in context.** In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL). Baltimore, USA.
- Annemarie Friedrich and Alexis Palmer. 2014b. **Situation entity annotation.** In Proceedings of the Linguistic Annotation Workshop VIII .
- Annemarie Friedrich, Alexis Palmer, Melissa Peate Sørensen, and Manfred Pinkal. 2015. **Annotating genericity: a survey, a scheme, and a corpus.** In Proceedings of the 9th Linguistic Annotation Workshop (LAW IX). Denver, Colorado, US.
- Annemarie Friedrich and Manfred Pinkal. 2015a. **Automatic recognition of habituals: a three-way classification of clausal aspect.** In Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP). Lisbon, Portugal.
- Annemarie Friedrich and Manfred Pinkal. 2015b. **Discourse-sensitive Automatic Identification of Generic Expressions.** In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics (ACL). Beijing, China.
- Jürgen Hermes, Michael Richter, and Claes Neuefind. 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).
- Nancy Ide, Collin Baker, Christiane Fellbaum, and Charles Fillmore. 2008. **MASC: The manually annotated sub-corpus of American English.**
- Manfred Krifka, Francis Jeffrey Pelletier, Gregory N. Carlson, Alice ter Meulen, Godehard Link, and Gennaro Chierchia. 1995. **Genericity: An Introduction.** The Generic Book, pages 1–124.

- Thomas A. Mathew and Graham E. Katz. 2009. **Supervised categorization for habitual versus episodic sentences.** In Sixth Midwest Computational Lingustics Colloquium. Indiana University, Bloomington, Indiana.
- Alexis Palmer, Elias Ponvert, Jason Baldridge, and Carlota Smith. 2007. **A sequencing model for situation entity classification.** In Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics.
- Nils Reiter and Anette Frank. 2010. **Identifying Generic Noun Phrases.** In Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics. Uppsala, Sweden.
- Eric V Siegel and Kathleen R McKeown. 2000. **Learning methods to combine linguistic indicators: Improving aspectual classification and revealing linguistic insights.** Computational Linguistics 26(4):595–628.
- Smith, Carlota. 1997. **The parameter of aspect.** Studies in Linguistics and Philosophy 43 (1991).
- Carlota S Smith. 2003. **Modes of discourse: The local structure of texts**, volume 103. Cambridge University Press.
- Radu Soricut and Daniel Marcu. 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 Language Technology.
- Joseph Turian, Lev Ratinov, and Yoshua Bengio. 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.
- Zeno Vendler. 1957. **Verbs and times.** The philosophical review pages 143–160.
- Alessandra Zarcone and Alessandro Lenci. 2008. **Computational models of event type classification in context.** In Proceedings of LREC2008.