UNIVERSITÄT DES SAARLANDES

CLUSTER OF EXCELLENCE

Automatic identification of generic expressions

Annemarie Friedrich

afried@coli.uni-saarland.de

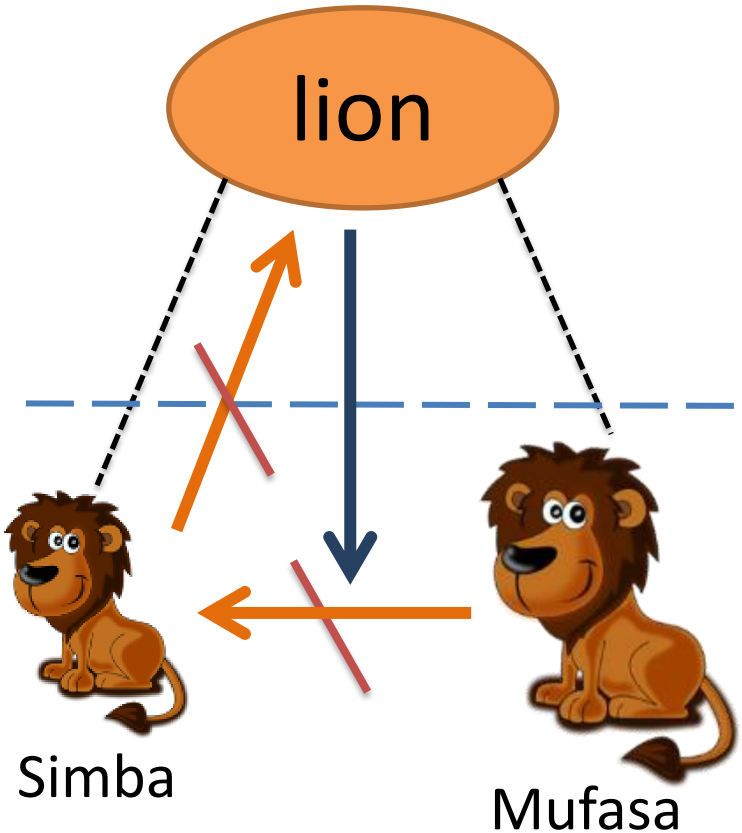
joint work with

Alexis Palmer, Melissa Peate Sørensen and Manfred Pinkal

Computational Linguistics, Universität des Saarlandes

JUNE 2015. IBM T.J. WATSON RESEARCH CENTER, YORKTOWN HEIGHTS, NY.

Generic vs. non-generic expressions

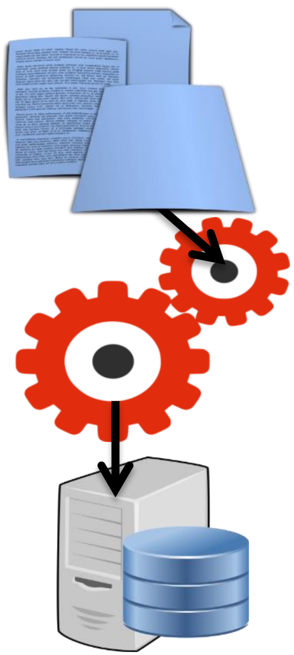


different
entailment properties

Lions are dangerous.

Mufasa is dangerous.
Simba is dangerous.

Identifying generic expressions: why?



knowledge extraction from text

natural language understanding

How? Discourse context matters

Previous work (Reiter & Frank 2010):
classification of noun phrases (in isolated sentences)

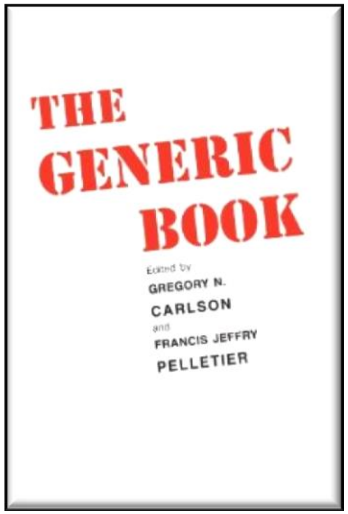
a) Sugar maples also have a tendency to color unevenly in fall. ***(generic)***
b) The recent year's growth twigs are green and turn dark brown. ***(generic)***

Discourse-sensitive approach (Friedrich & Pinkal 2015):
sequence labeling task
classification of (subject) noun phrases & clauses

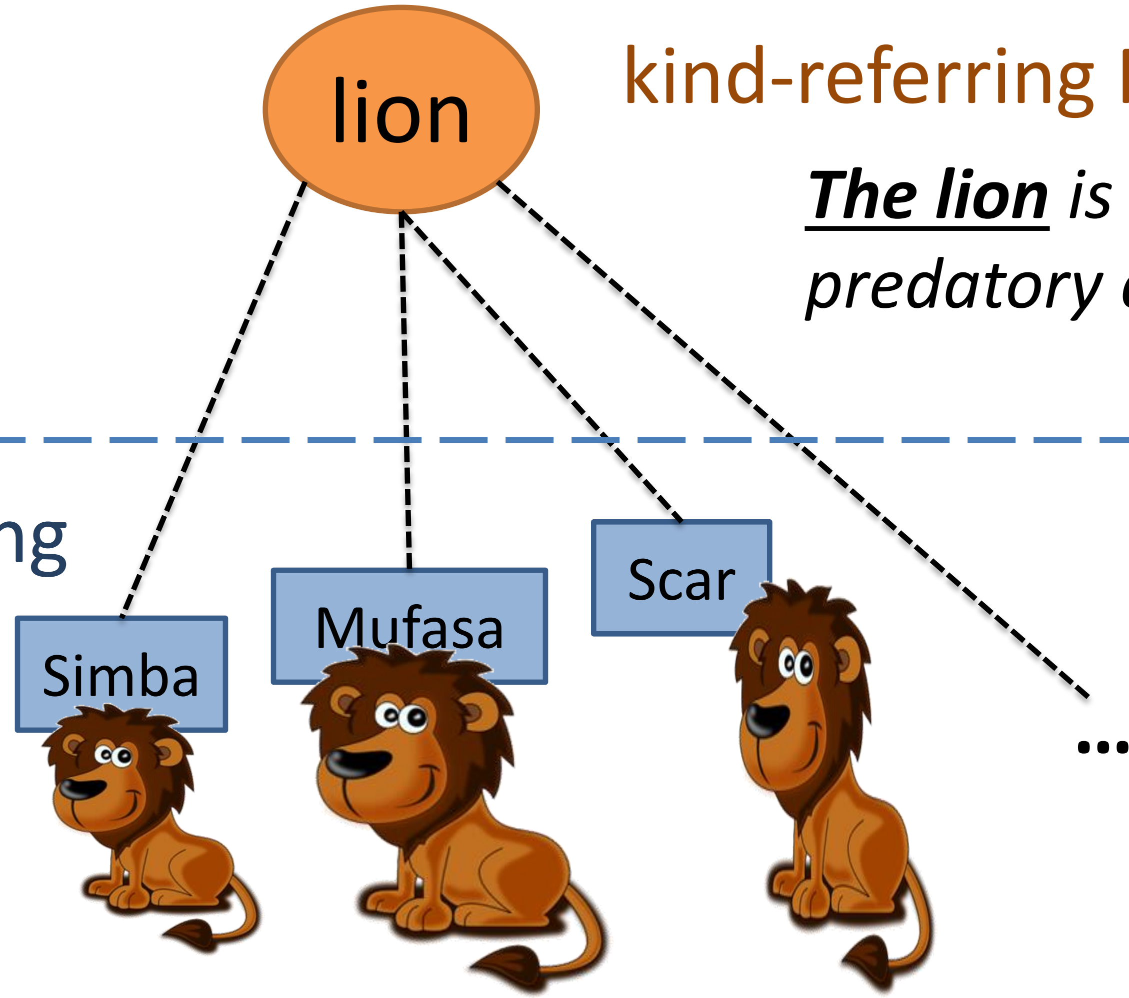
Overview of talk

- terminology: generics
- corpora & annotation
- computational model
- results

Terminology: reference to kinds



Krifka et al. (1995): Genericity: An Introduction.



kind-referring NPs

<u>The lion</u> *is a predatory cat.*

non-kind-referring NPs

<u>Simba</u> *escaped from the zoo.*

NP-level: reference to kinds

form of NP not sufficient

	kind-referring	**non-kind-referring**
definite NPs	*The lion is a predatory cat.*	*The cat chased the mouse.*
indefinite NPs	*Lions eat meat.*	*Dogs were barking outside.*
quantified NPs	*Some (type of) dinosaur is extinct.*	*Some dogs were barking outside.*
proper names	*Panthera leo persica was first described by the Austrian zoologist Meyer.*	*John likes ice cream.*

clause / context matters

Terminology: clause-level genericity

characterizing sentences

generalizations over situations

	lexically characterizing sentences	**habitual sentences**
kind-referring subject	*Lions have manes.*	*Lions eat meat.*
non-kind-referring subject	*John is tall.*	*John drives to work.*

Terminology: clause-level genericity

characterizing sentences

generalizations over situations

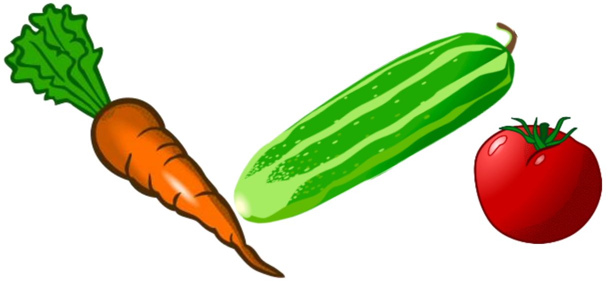
	lexically characterizing sentences	**habitual sentences**
kind-referring subject	*Lions have manes.*	*Lions eat meat.*
non-kind-referring subject	*John is tall.*	*John drives to work.*

Generic sentences

- The term is used for various (NP- and clause-level) phenomena.
- Generic sentences are not rendered false by the existence of counter-examples.

Lions eat meat.

The lion in our zoo is weird, though, it only eats vegetables.



Related Work

- **ACE corpora**
  - Automatic Content Extraction (2002-2008)
  - largest corpora annotated with NP-level genericity to date, ~40k NPs
- Reiter & Frank (ACL 2010):

  “Identification of generic noun phrases”
  - use a variety of NP-based and clause-based features
  - Bayesian network (Weka)

ACE entity class annotations

ACE-2:

generic = "any member of the set in question"

specific = "some particular, identifiable member of that set"

ACE-2005:

GEN = kind-referring

SPC = non-kind-referring

NEG = negatively quantified NPs

There are no confirmed suspects yet.

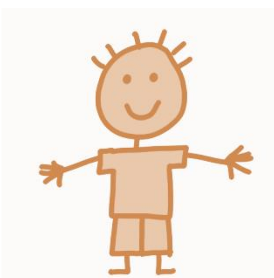
USP = underspecified: ambiguous cases

There are new opportunities for women in New Delhi.

and mentions of entities "whose identity would be difficult to locate": *Officials reported …*

ACE-2005: agreement study

533 documents



adjudication

final corpus

news, broadcast news, broadcast conversation, forum and weblog texts

annotations available from LDC

agreement study:

exactly-matching entity mention spans (~90%)

		annotator 2			
		SPC	**USP**	**GEN**	**NEG**
annotator 1	**SPC**	28168	1575	684	3
	USP	1142	1954	963	2
	GEN	757	1261	1707	10
	NEG	8	5	7	71

Cohen's κ = 0.53

confusion of SPC/GEN with USP is high

Problems of the ACE annotation guidelines

- **predicative uses** are marked
  - *John is a nice person. (specific)*
  - *John seems to be a nice person. (generic)*
- **noun modifiers** in compounds (9.5% of all mentions) are marked as generic: *subway system*

non-referential

- guidelines mix **genericity** and **specificity**

  (specificity = speaker has a particular referent in mind)
  - *Officials reported...*
  - this is not underspecified: it is not generic, but nonspecific

WikiGenerics corpus

102 Wikipedia texts

about animals, sports, politics, science, biographies, …

10279 clauses, aim: balanced corpus (many generics)

Annotation scheme

motivated by **semantic theory** (Krifka et al. 1995)

study references to and statement about kinds

(Task NP, Task Cl, Task Cl+NP)

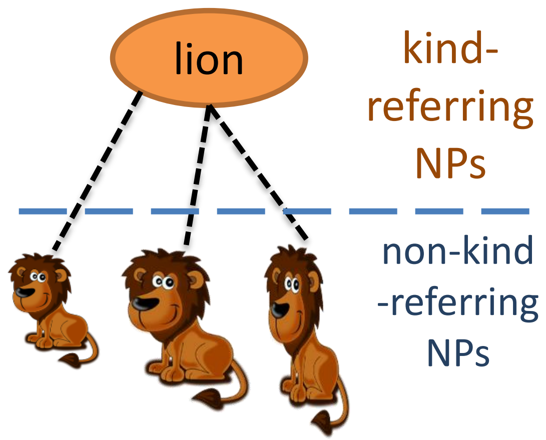
(other aspects of genericity → future work)

contribution of clauses to **discourse**:

characterizing statements ≠ particular events or states

→ relevant for processing temporal structure of discourse

Task NP: genericity of subject



generic: references to kind / class

The lion is a predatory cat.

Lions have manes.

A lion may eat up to 30kg in one sitting.

non-generic: references particular individual(s)

Simba flees into exile.

A lion must have eaten the rabbit. (nonspecific)

Lions are in this cage.

Task Cl: genericity of clause

generic: characterizing statements about kinds

subject must be **generic**.

The lion is a predatory cat.

Lions eat up to 30kg in one sitting. (habitual)

non-generic: statements about particular individuals or particular events.

John is a nice guy.

John cycles to work. (habitual)

Task Cl+NP: clause and subject

clause / subject	generic	non-generic
generic	*Lions have manes.* *Lions eat meat.*	*The blobfish was voted the "World's Ugliest Animal".* *Dinosaurs died out.*
non-generic	-- --	*John is a nice guy.* *John cycles to work.*

habitual vs. other types of clauses: future work!

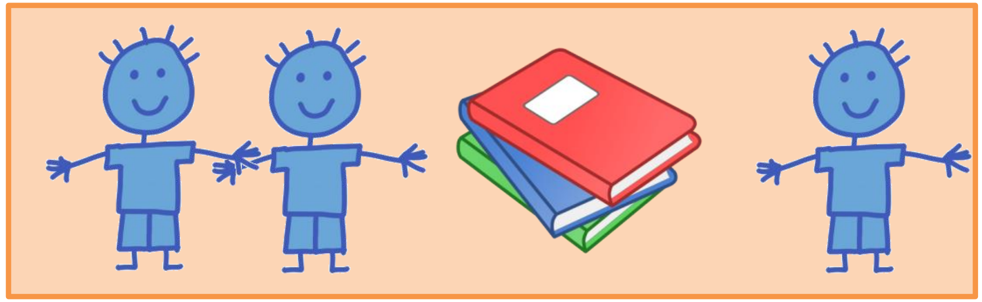
Annotation process

SPADE system

segmentation into clauses

subjects are
not pre-marked

1) <u>Lions</u> are big cats
2) and eat meat.



3 annotators label all texts

majority vote

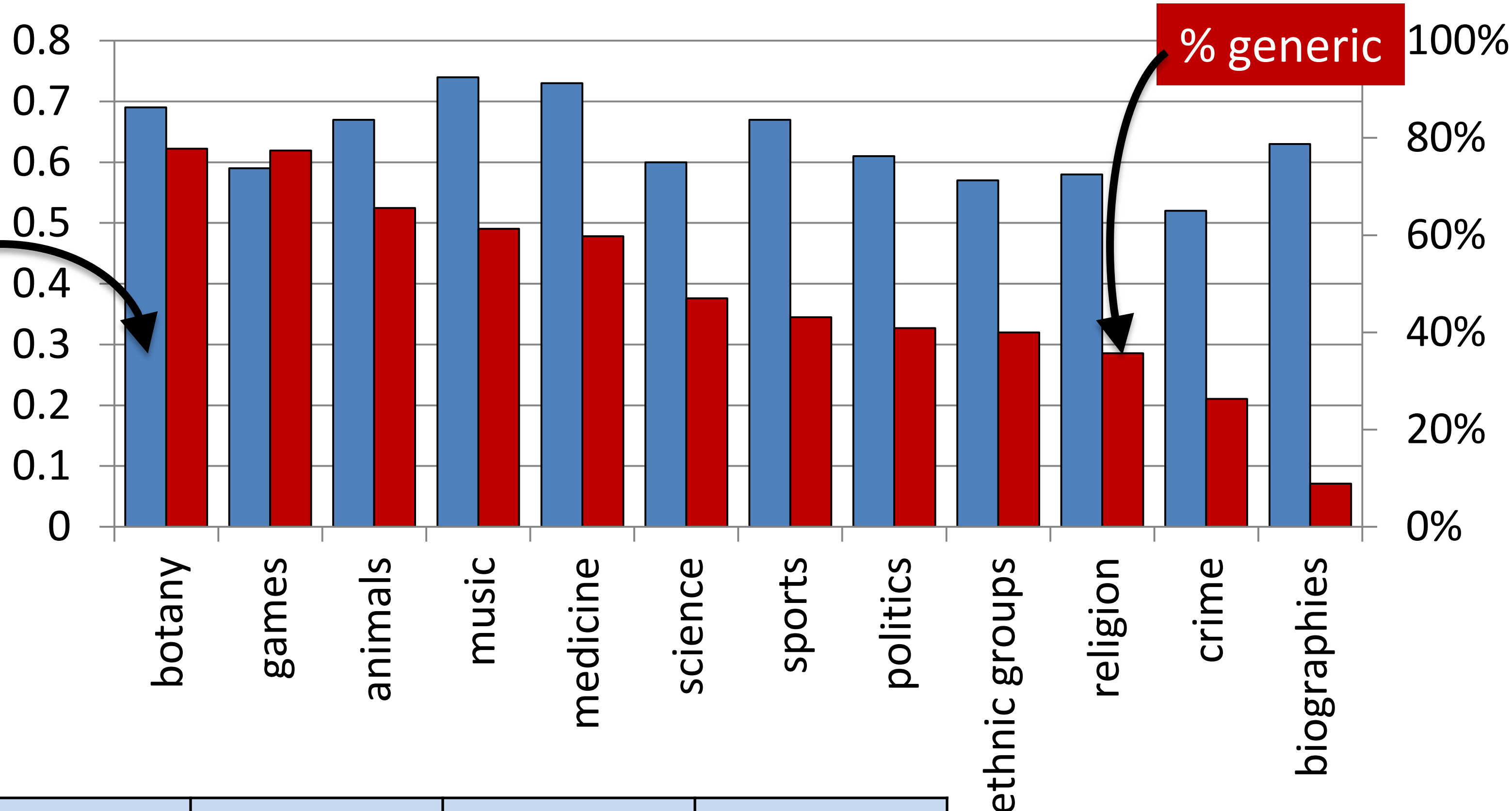
gold standard

Inter-annotator agreement: WikiGenerics

MASC, Cl+NP

Fleiss' κ

marked completely by 3 annotators



Task NP	Task CL	Task Cl+NP	% generic
0.69	0.72	0.68	50.1%

balanced corpus, substantial agreement

Computational model

Sugar maples also have a tendency to color unevenly in fall.

barePlural=true 1
determinerType=def : 0
tense=present : 1
voice=active : 1
…

The recent year's growth twigs are green.

barePlural=true : 0
determinerType=def : 1
…
currentLabel=GEN and previousLabel=GEN : 1
…

CRF

GENERIC

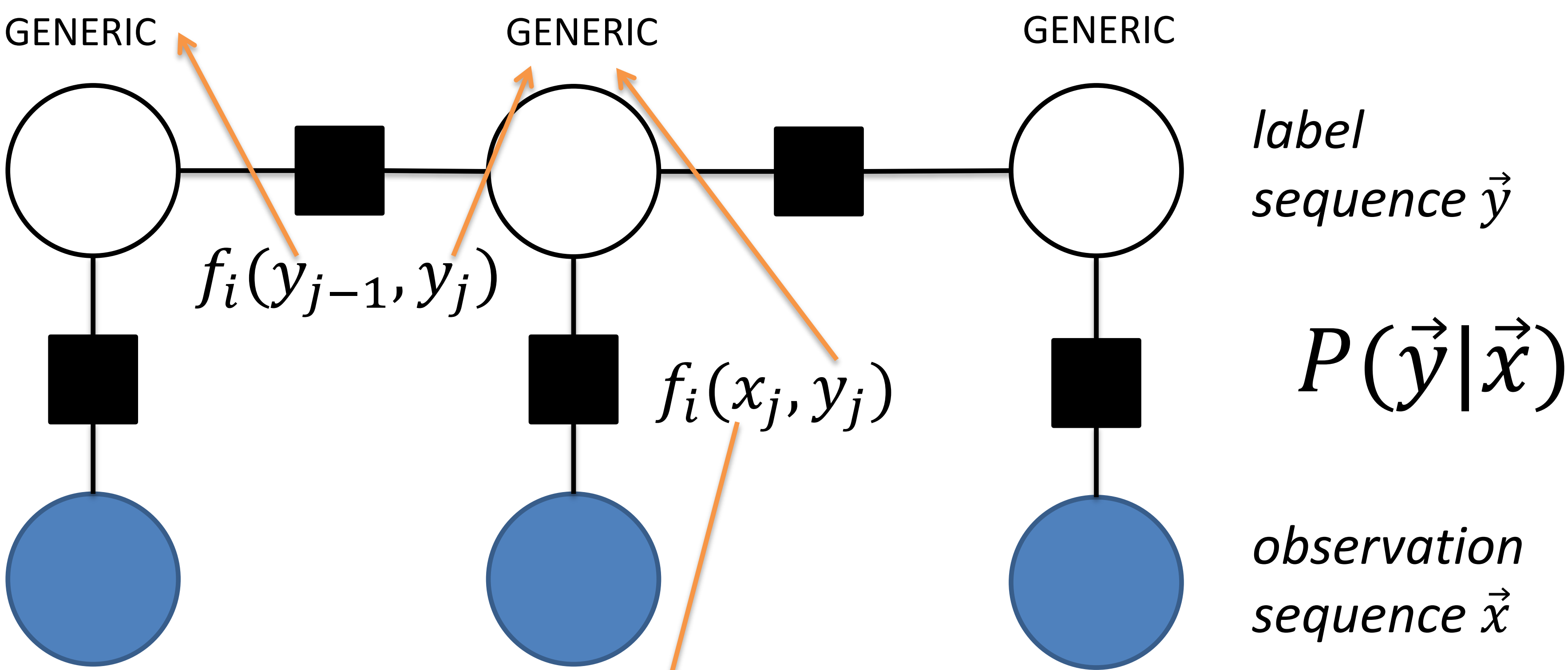
GENERIC

sequence of clauses (entire document)

features: indicator functions

sequence of labels

Linear-chain Conditional Random Field



<u>Acer saccharum</u> is a deciduous tree.

<u>Sugar maples</u> also have a tendency to color unevenly in fall.

<u>The recent year's growth twigs</u> are green.

Linear-chain Conditional Random Field

Probability of label sequence $\vec{y}$ given observation sequence $\vec{x}$

sum over observations in $\vec{x}$

sum over feature functions

$$P(\vec{y}|\vec{x}) = \frac{1}{Z(\vec{x})} \exp\left\{\sum_{j=1}^{n}\left[\sum_{i} \lambda_i f_i(y_{j-1}, y_j) + \sum_{k} \lambda_k f_k(x_j, y_j)\right]\right\}$$

normalization over scores for all possible label sequences with length $|\vec{x}|$

weights for feature functions
(independent of position in sequence j)

Discriminative training (maximum likelihood, CRF++ toolkit uses L-BGFS)

Features [see Reiter & Frank 2010]

extracted from dependency parses (Stanford parser)

NP-based features	
number	sg, pl
person	1,2,3
countability	Celex: count, uncount,…
noun type	common, proper, pronoun
determiner type	def, indef, demon
part-of-speech	POS of head
bare plural	true, false
WordNet based features	senses, lexical filename,…

Clause-based features	
dependency relations	between (subject) head and governor etc.
tense	past, present, future
progressive	true, false
perfective	true, falce
voice	active, passive
part-of-speech	POS of head
temporal modifier	true, false
number of modifiers	numeric
predicate	lemma of head
adjunct-degree	positive, comparative, superlative

Results on ACE data: NP-level

ACE-2: SPC, GEN		
System	F1	Acc.
majority class	46.5	86.8
Bayes Net (R&F)	69.8	80.4
CRF (unigram)	71.3	88.5*
CRF (bigram)	**72.4**	**88.9***
CRF (bigram, gold)	*76.0*	*90.1*

*difference statistically significant

ACE-2005 (subject mentions) SPC, GEN, USP		
System	F1	Acc.
majority class	28.6	75.1
Bayes Net (R&F)	52.7	72.5
CRF (unigram)	53.6	77.7*
CRF (bigram)	**53.7**	**77.8**
CRF (bigram, gold)	*58.6*	*79.6**

Our model outperforms previous work.
Gold information → discourse helps.
Few generic instances, problems in annotation guidelines.

WikiGenerics Task NP: genericity of subject

<u>The lion</u> is a predatory cat. *(generic)*

<u>Simba</u> had to flee. *(non-generic)*

System	Macro-avg. F1	Accuracy
majority class	35.9	56.1
Bayes Net (R&F)	72.3	71.7
CRF (unigram)	75.9	76.4
CRF (bigram)	**78.8**	**79.1**
CRF (bigram, gold)	*82.7*	*83.0*

discourse information

WikiGenerics Task Cl: genericity of clause

The lion is a predatory cat. *(generic)*
Simba had to flee. *(non-generic)*

System	Macro-avg. F1	Accuracy
majority class	35.1	43.7
Bayes Net (R&F)	73.7	73.5
CRF (unigram)	77.4	77.4
CRF (bigram)	**80.7**	**80.7**
CRF (bigram, gold)	*82.8*	*82.8*

discourse information

WikiGenerics Task Cl+NP: three-way task

(CLAUSE_subject)

The lion is a predatory cat. *(GEN_gen)*

Simba had to flee. *(NON-GEN_non-gen)*

The blobfish was voted the most ugly animal of the world. *(NON-GEN_gen)*

System	Macro-avg. F1	Accuracy
majority class	22.4	50.4
Bayes Net (R&F)	56.4	65.2
CRF (unigram)	63.4	74.0
CRF (bigram)	**65.8**	**77.4**
CRF (bigram, gold)	*69.0*	*80.6*

discourse information

Feature set ablation

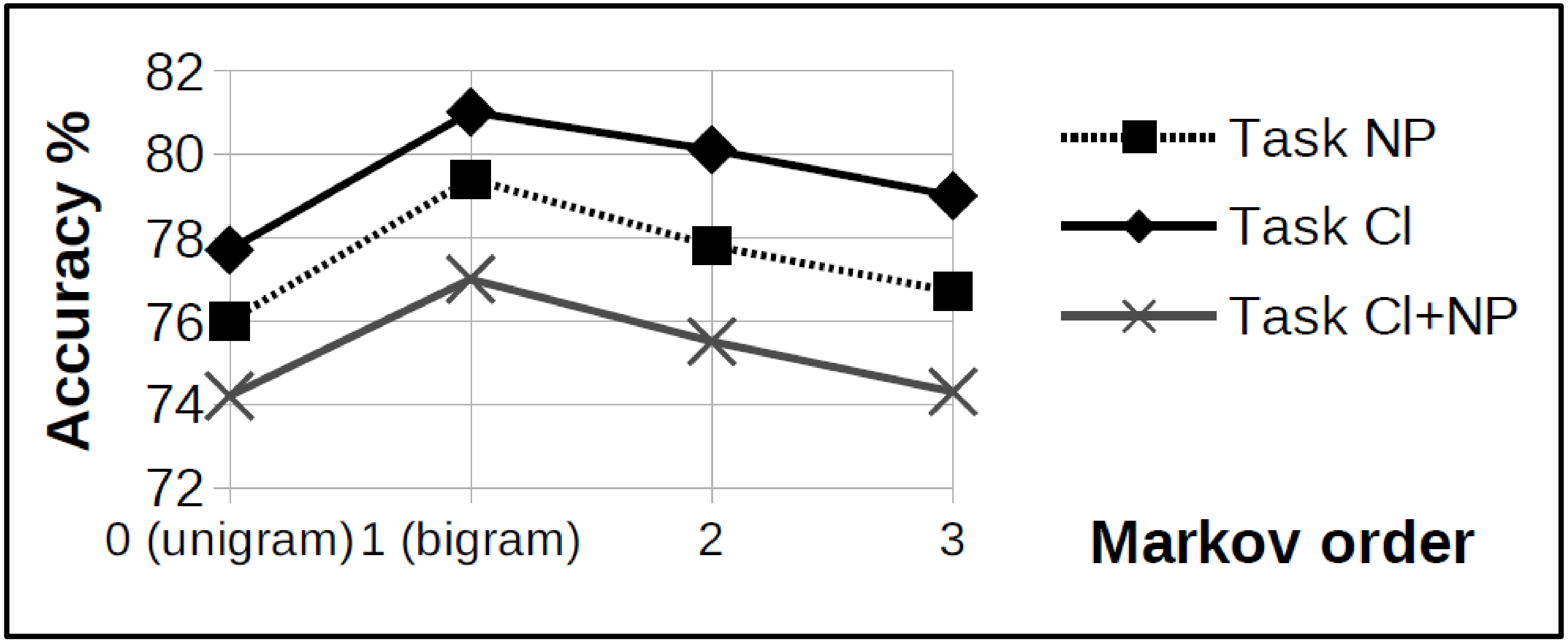
	Accuracy		
System	Task NP	Task Cl	Task Cl+NP
CRF (bigram)	79.1	80.7	77.4
- clause features only	76.0	78.8	74.3
- NP features only	74.1	71.7	70.0

→ It strongly depends on clause whether an NP is interpreted as generic or not.

Markov order

What happens if we take more than the immediately preceding label into account?

Mallet toolkit



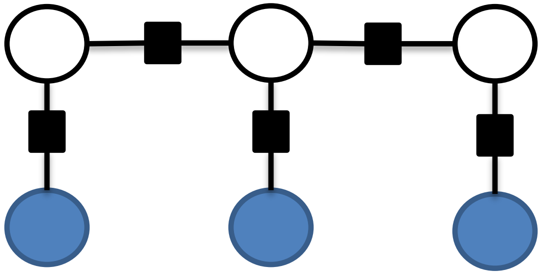
- using only the preceding label is optimal
- labels of non-adjacent clauses ***do*** influence each (score is optimized for entire sequence)

Conclusions

We classify NPs <u>and</u> clauses with regard to their genericity.

WikiGenerics corpus
balanced
substantial agreement

CRF finds **optimal label sequence** for clauses of a document, combining information from clause and surrounding labels



discourse information matters!

Future work

- Genericity of NPs other than the subject
  - annotation + automatic classification
  - *Cats chase mice.*

- Related linguistic phenomena
  - habitual vs. episodic sentences
    - *John cycled to work today.*
    - *John cycles to work.*
- Integration into applications

Thank you

Alexis Palmer

Melissa Peate Sørensen

Manfred Pinkal

Questions?

www.coli.uni-saarland.de/projects/sitent

UNIVERSITÄT DES SAARLANDES

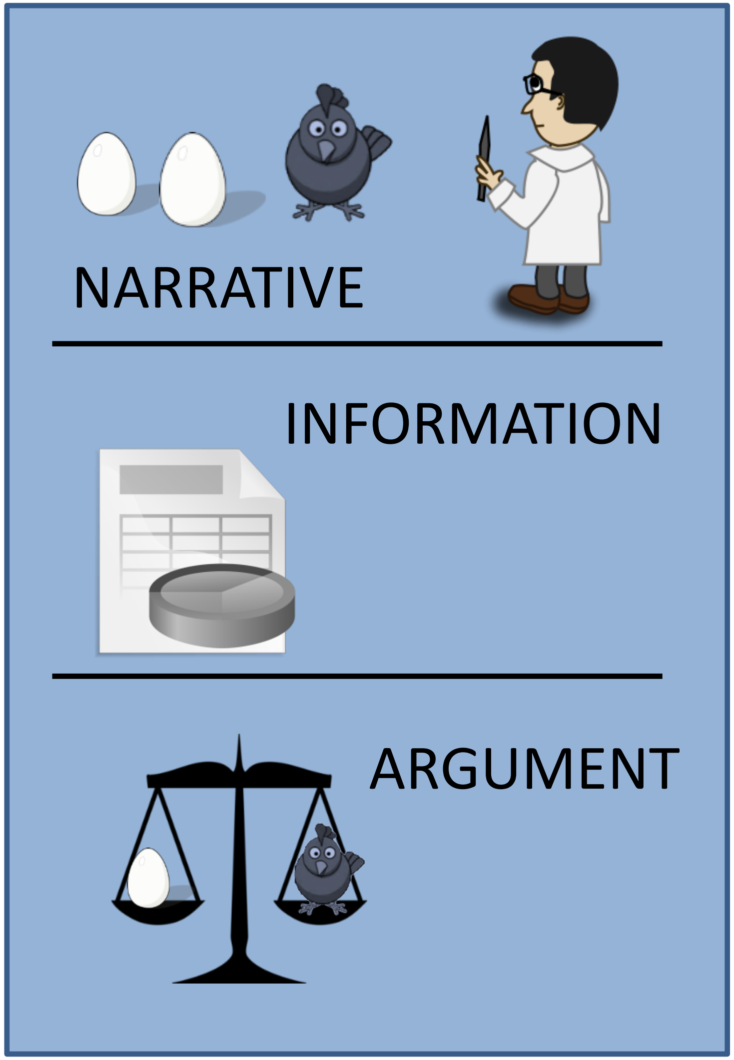
CLUSTER OF EXCELLENCE

The bigger picture

Annotation and automatic classification of situation entity types

Discourse modes [Smith 2003]

Modes of discourse [Smith 2003]



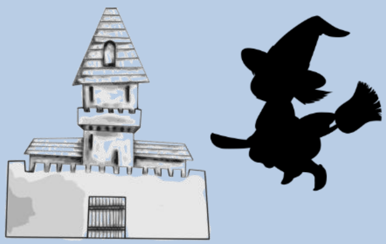
Different passages of a text can have different discourse modes.

one text ≈ one genre

one text ≠ one discourse mode

Modes of discourse [Smith 2003]

NARRATIVE



temporal progression

REPORT

temporal progression

DESCRIPTION

temporal / spatial progression

INFORMATION

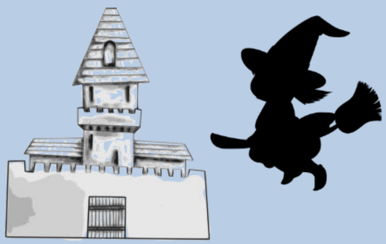
metaphorical progression

ARGUMENT

metaphorical progression

Modes of discourse [Smith 2003]:
Situation entity types

NARRATIVE



EVENT, STATE

REPORT

EVENT, STATE, general statives

DESCRIPTION

EVENT, STATE, ongoing EVENT

general statives

INFORMATION

ARGUMENT

FACT, PROPOSITION, general statives

Situation entity types

Eventualities	STATE	*Mary likes cats.*
	EVENT	*Mary fed the cats.*
	- REPORT	*..., Mary said.*
General Statives	GENERALIZING SENTENCE	*Mary often feeds my cats.*
	GENERIC SENTENCE	*Cats are always hungry.*
Abstract Entities	FACT	*I know <u>that Mary fed the cats</u>.*
	PROPOSITION	*I believe <u>that Mary fed the cats</u>.*
Speech Acts	QUESTION	*Does Mary like cats?*
	IMPERATIVE	*Don't forget to feed the cats!*

Motivation

Annotation of large data set (MASC)

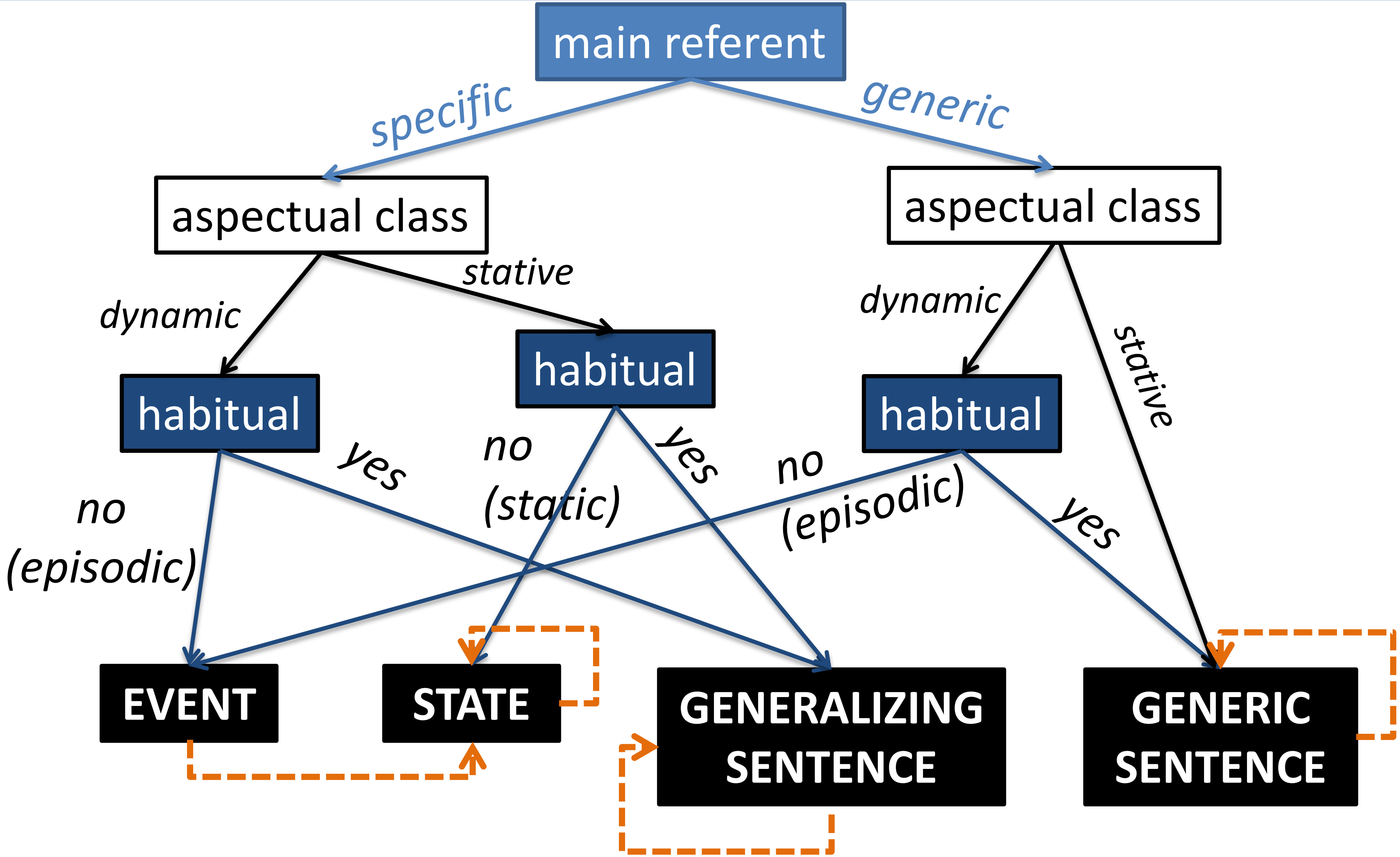
computational modeling

assess the applicability of SE type classification as described by Smith [2003]

borderline cases? human agreement?

training, development, evaluation of automatic systems for classifying SEs and related tasks: improve temporal discourse processing

Features & SE types



negation, modals, conditional, perfect, future

Feature: fundamental aspectual class

[Friedrich & Palmer, ACL 2014]

Juice **fills** the glass.
STATIVE

The glass **was filled** with juice.
BOTH readings possible

She **filled** the glass with juice. **DYNAMIC**

Summary

- **generics**: distinguish GENERIC SENTENCES from other situation entity types
  [Friedrich & Pinkal 2015]
- **lexical aspectual class**: distinguish STATES and EVENTS [Friedrich & Palmer 2014]
- **habituals**: work in progress
- **full classification task**: work in progress

References

ACE corpora: https://www.ldc.upenn.edu/collaborations/past-projects/ace

Friedrich, A. & Palmer, A (2014). **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

Friedrich, A. & Pinkal, M. (2015). **Discourse-sensitive Automatic Identification of Generic Expressions**. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics (ACL)*. Beijing, China. *(to appear)*

Krifka, M. et al. (1995). **Genericity: an introduction**. *The Generic Book*, 1-124. University of Chicago Press.

Ide, N., Baker, C., Fellbaum, C., & Fillmore, C. (2008). **MASC: The manually annotated sub-corpus of American English.** In *In Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC*).

Reiter, N., & Frank, A. (2010, July**). Identifying generic noun phrases.** In *Proceedings of ACL (*pp. 40-49). Association for Computational Linguistics.

Soricut, R., & Marcu, D. (2003). **Sentence level discourse parsing using syntactic and lexical information.** ACL-HLT. (pp. 149-156). Association for Computational Linguistics.