Computerlinguistik Kolloquium

Potsdam, November 2014

Annotation and automatic classification of situation entity types

Annemarie Friedrich
joint work with Alexis Palmer
Department of Computational Linguistics
Saarland University

Situation entity types [Smith 2003]

- clauses introduce situations to a discourse
- classification of types of situation (entities)

SE type	*Example*
STATE	*Mary likes cats.*
EVENT	*Mary fed the cats.*
GENERALIZING SENTENCE	*Mary often feeds my cats.*
GENERIC SENTENCE	*Cats are always hungry.*

Modes of discourse [Smith 2003]

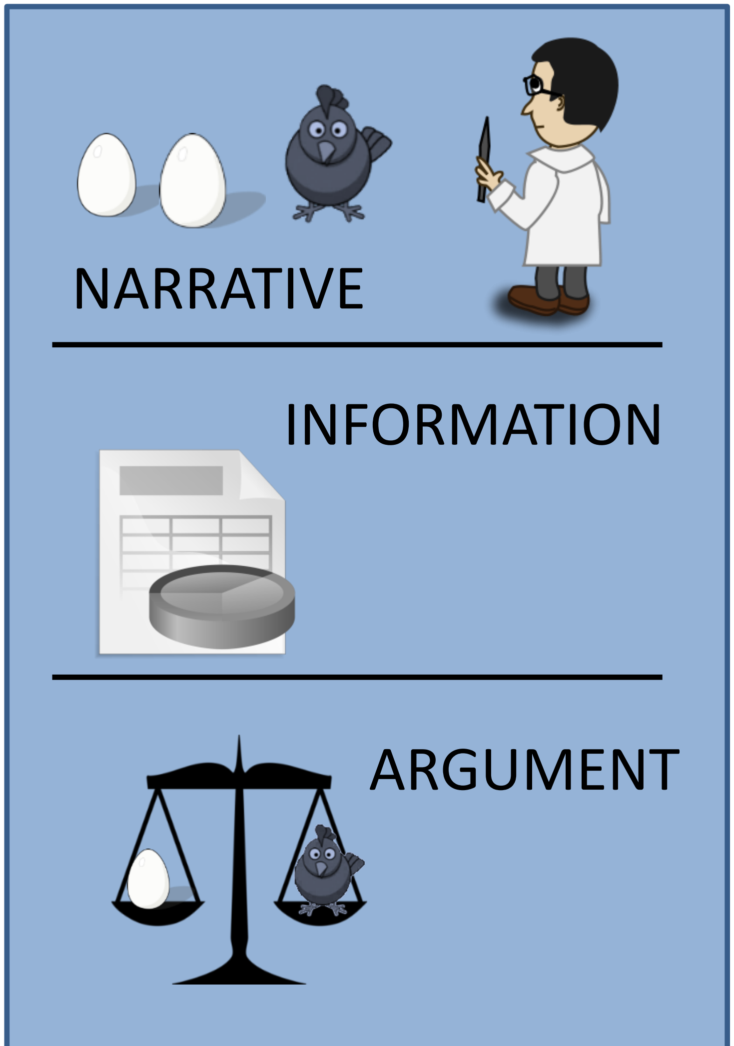
NARRATIVE

Modes of discourse [Smith 2003]

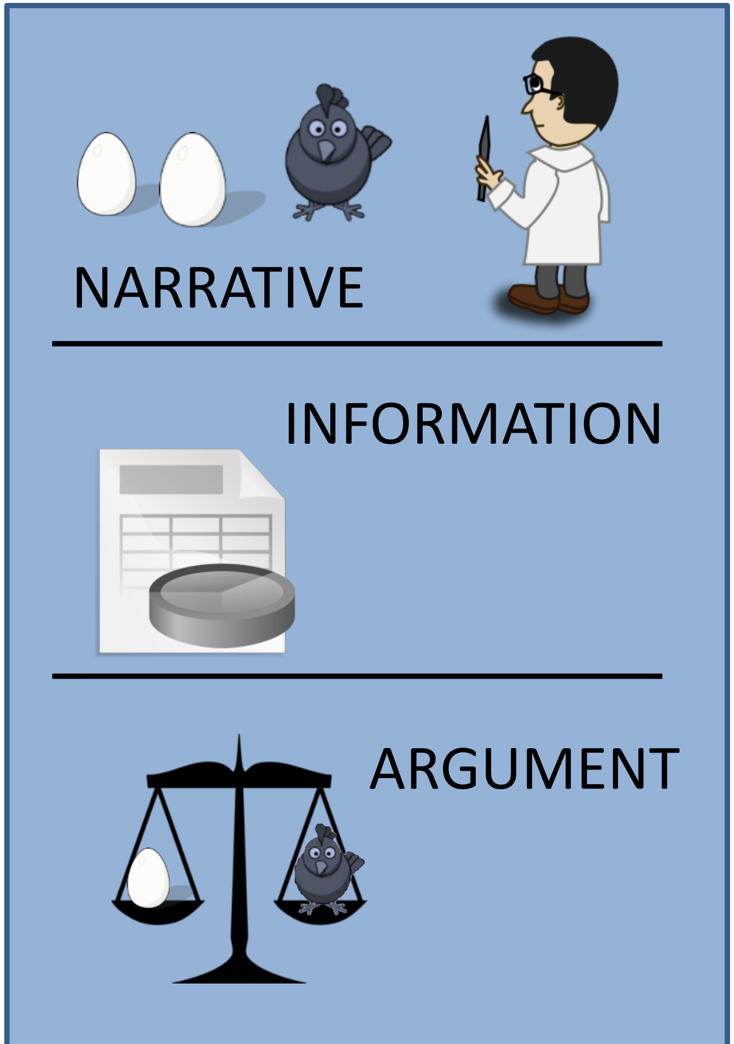
NARRATIVE

INFORMATION

Modes of discourse [Smith 2003]

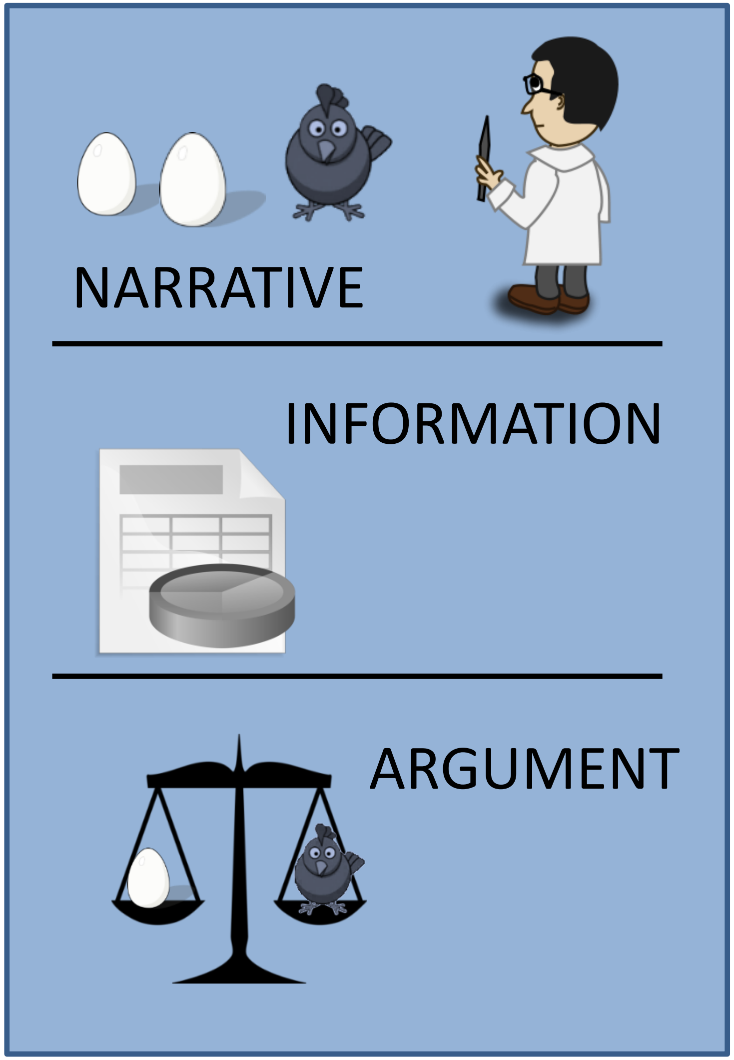


Modes of discourse [Smith 2003]



Different passages of a text can have different discourse modes.

Modes of discourse [Smith 2003]



Different passages of a text can have different discourse modes.

one text ≈ one genre

one text ≠ one discourse mode

related: Werlich's typology of texts (1975)

Modes of discourse [Smith 2003]

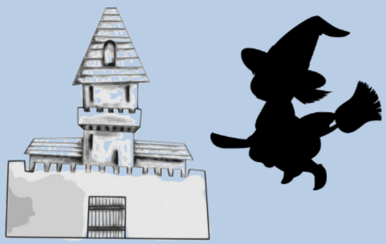
NARRATIVE

temporal progression

EVENT,
STATE

Modes of discourse [Smith 2003]

NARRATIVE



temporal progression

EVENT,
STATE

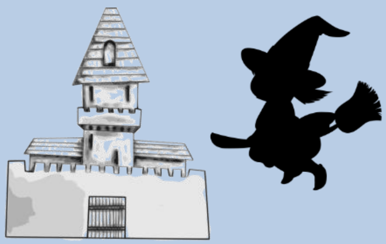
REPORT

temporal progression,
related to speech time

EVENT, STATE,
general statives

Modes of discourse [Smith 2003]

NARRATIVE



temporal progression

EVENT,
STATE

REPORT

temporal progression,
related to speech time

EVENT, STATE,
general statives

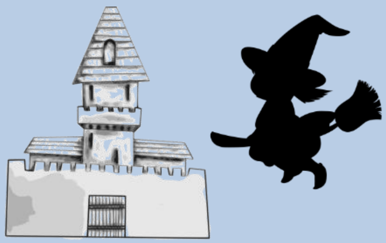
DESCRIPTION

spatial progression

EVENT, STATE,
ongoing EVENT

Modes of discourse [Smith 2003]

NARRATIVE



temporal progression

EVENT, STATE

REPORT

temporal progression, related to speech time

EVENT, STATE, general statives

DESCRIPTION

spatial progression

EVENT, STATE, ongoing EVENT

INFORMATION

general statives

atemporal, metaphoric progression

Modes of discourse [Smith 2003]

NARRATIVE

temporal progression

EVENT,
STATE

REPORT

temporal progression,
related to speech time

EVENT, STATE,
general statives

DESCRIPTION

spatial progression

EVENT, STATE,
ongoing EVENT

general
statives

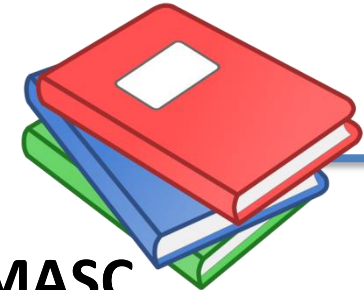
INFORMATION

ARGUMENT

FACT,
PROPOSITION,
general statives

atemporal, metaphoric progression

Overview

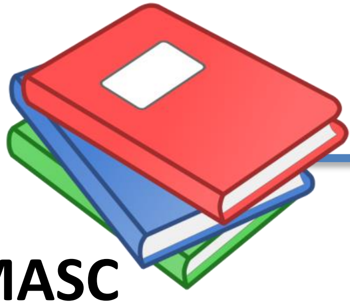


MASC

(automatic) segmentation

Situation entity (SE) types

Overview



MASC

(automatic) segmentation

Situation entity (SE) types

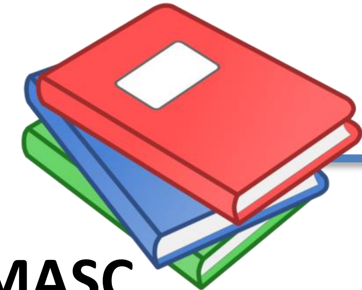
genericity of main referent

fundamental aspectual class

habituality

Feature-based annotation

Overview



MASC

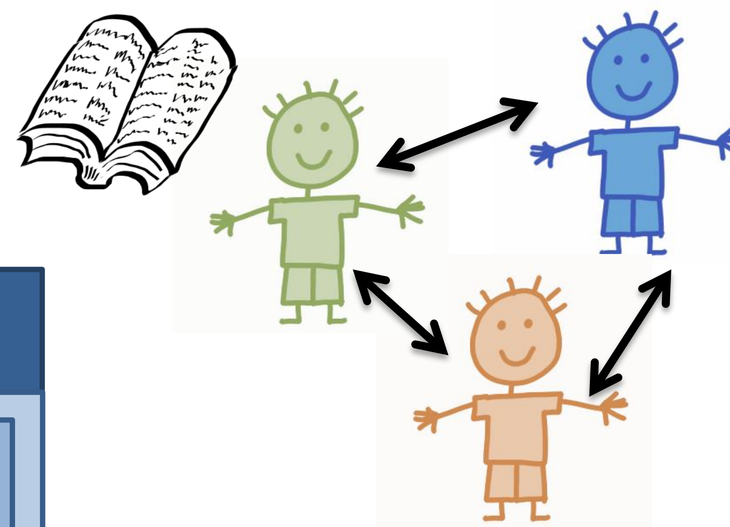
(automatic) segmentation

Situation entity (SE) types

genericity of main referent	fundamental aspectual class	habituality

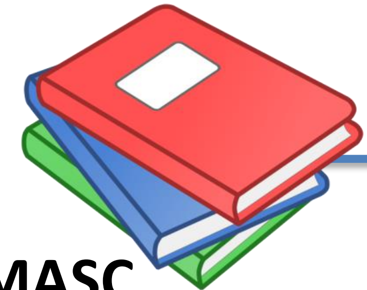
Feature-based annotation

1) Corpus annotation



inter-annotator agreement

Overview



MASC

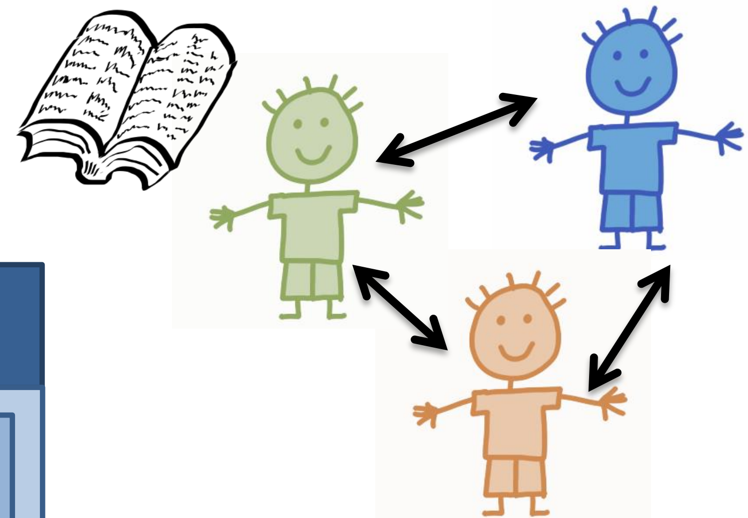
(automatic) segmentation

Situation entity (SE) types

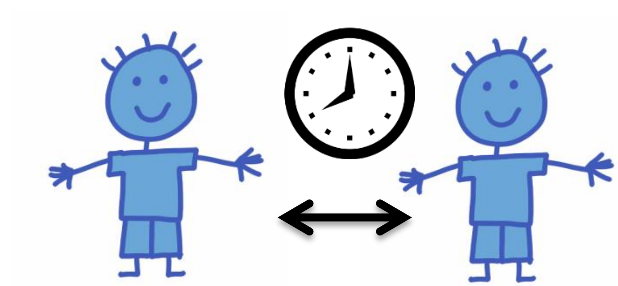
genericity of main referent	fundamental aspectual class	habituality

Feature-based annotation

1) Corpus annotation

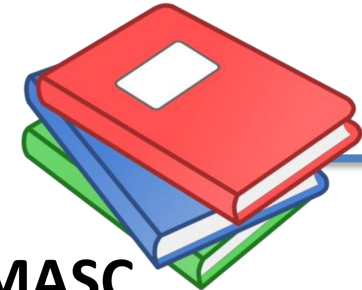


inter-annotator agreement



intra-annotator consistency

Overview



MASC

(automatic) segmentation

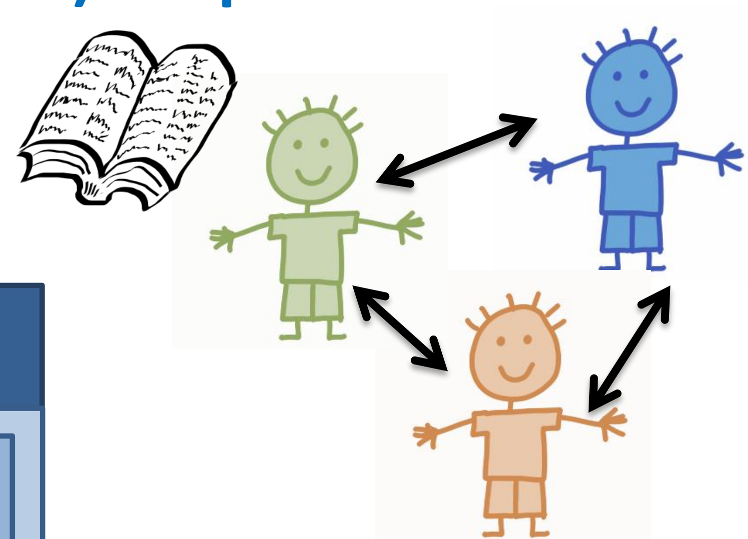
Situation entity (SE) types

genericity of main referent	fundamental aspectual class	habituality

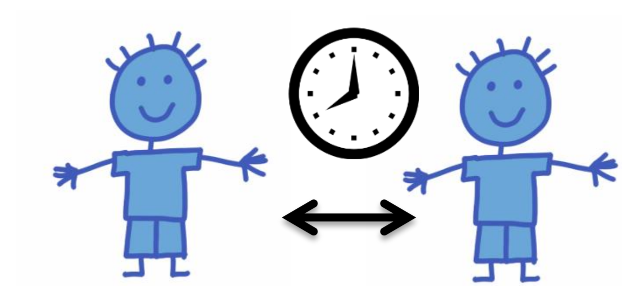
Feature-based annotation

2) automatic classification

1) Corpus annotation

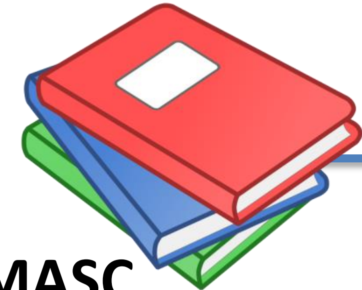


inter-annotator agreement



intra-annotator consistency

Overview



MASC

(automatic) segmentation

Situation entity (SE) types

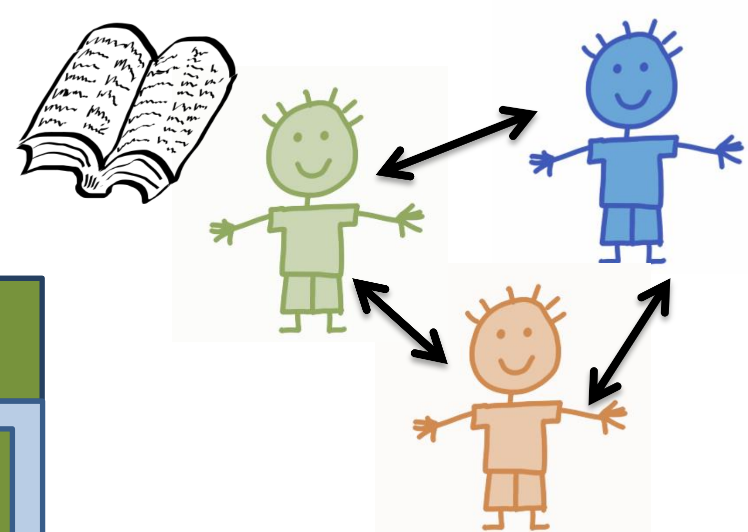
genericity of main referent	fundamental aspectual class	habituality

Feature-based annotation

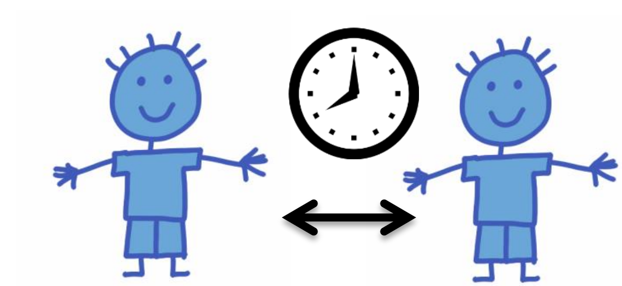
2) automatic classification

3) current status, ongoing & future work

1) Corpus annotation



inter-annotator agreement



intra-annotator consistency

Motivation of annotation study

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

borderline cases? human agreement?

Motivation of annotation study

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

Motivation of annotation study

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

foundation for analysis of the theory of Discourse Modes [Smith 2003]

Situation entity types (SE types)

EVENT **?**

Yesterday, Mary bought a cat.

Now she owns four cats.

Susie often feeds Mary's cats.

Cats are very social animals.

Situation entity types (SE types)

Yesterday, Mary bought a cat. **EVENT**

Now she owns four cats.

Susie often feeds Mary’s cats.

Cats are very social animals.

Situation entity types (SE types)

STATE ?

Yesterday, Mary bought a cat. **EVENT**

Now she owns four cats.

Susie often feeds Mary's cats.

Cats are very social animals.

Situation entity types (SE types)

Yesterday, Mary bought a cat. **EVENT**

Now she owns four cats. **STATE**

Susie often feeds Mary's cats.

Cats are very social animals.

Situation entity types (SE types)

GENERIC SENTENCE ?

Yesterday, Mary bought a cat. **EVENT**

Now she owns four cats. **STATE**

Susie often feeds Mary's cats.

Cats are very social animals.

Situation entity types (SE types)

Yesterday, Mary bought a cat. **EVENT**

Now she owns four cats. **STATE**

Susie often feeds Mary's cats.

Cats are very social animals. **GENERIC SENTENCE**

Situation entity types (SE types)

GENERALIZING SENTENCE ?

Yesterday, Mary bought a cat. **EVENT**

Now she owns four cats. **STATE**

Susie often feeds Mary's cats.

Cats are very social animals. **GENERIC SENTENCE**

Situation entity types (SE types)

Yesterday, Mary bought a cat. **EVENT**

Now she owns four cats. **STATE**

Susie often feeds Mary's cats. **GENERALIZING SENTENCE**

Cats are very social animals. **GENERIC SENTENCE**

Situation entity types (SE types)

Yesterday, Mary bought a cat. **EVENT**

Now she owns four cats. **STATE**

eventualities

Susie often feeds Mary's cats. **GENERALIZING SENTENCE**

Cats are very social animals. **GENERIC SENTENCE**

general statives

SE types: abstract entities

here: clausal complements of factive / implicative verbs

Susie **knows** **STATE**

that Mary loves her cats a lot. **FACT** objects of knowledge

SE types: abstract entities

here: clausal complements
of factive / implicative verbs

Susie **knows** **STATE**

that Mary loves her cats a lot. **FACT** objects of knowledge

Susie **believes** **STATE**

that the cats also love Mary. **PROPOSITION**

objects of belief

SE types: speech act types [Palmer et al. 2007]

Did you see my cats? **QUESTION**

Don’t forget to feed the cats! **IMPERATIVE**

Derived situation entity types

coerce **EVENTs** to **STATEs**:

negation, modality, future / perfect tense, conditionality, subjectivity

Susie will feed the cats.

Susie has not fed the cats.

If Susie has forgotten the cats, they might be hungry now.

Derived SE types

general statives are not subject to such coercion:

Susie never feeds Mary's cats. **GENERALIZING SENTENCE**

Cats might be the most popular pet. **GENERIC SENTENCE**

SE types: summary

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 that Mary fed the cats.
	PROPOSITION	I believe that Mary fed the cats.
Speech Acts	QUESTION	Does Mary like cats?
	IMPERATIVE	Don't forget to feed the cats!

Related work

- Palmer et al. [2007]:
  - first labeled data set for SEs
  - ~6000 clauses
  - no annotation manual
  - Cohen's $\kappa = 0.54$

Related work

- Palmer et al. [2007]:
  - first labeled data set for SEs
  - ~6000 clauses
  - no annotation manual
  - Cohen's κ = 0.54
- Stede & Peldzsus [2012]:
  - illocutionary status of clauses in causal relations
    ~pragmatic role, e.g. REPORT, DIRECTIVE, COMMITMENT

Data: Manually Annotated SubCorpus (MASC) of Open American National Corpus

[Ide et al. 2008]

✓ additional types of annotation available

✓ open distribution of annotations

✓ wide range of genres

MASC section	# of situations (segments)	average # tokens per segment
news	3455	9.9
jokes	2563	6.9
letters	1851	11.1

annotation status LAW 2014

Segmentation

SPADE [Soricut & Marcu 2003]

+ heuristic post-processing

+ manual correction

Segmentation

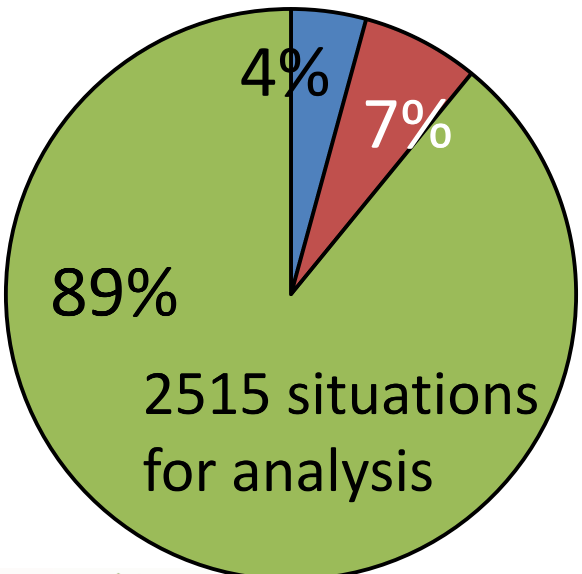
SPADE [Soricut & Marcu 2003]

+ heuristic post-processing

+ manual correction

marked as **NO SITUATION**
by at least one annotator
(e.g. headlines, names, dates)

merged to other segment
by at least one annotator



MASC news: 2823 segments

Segmentation

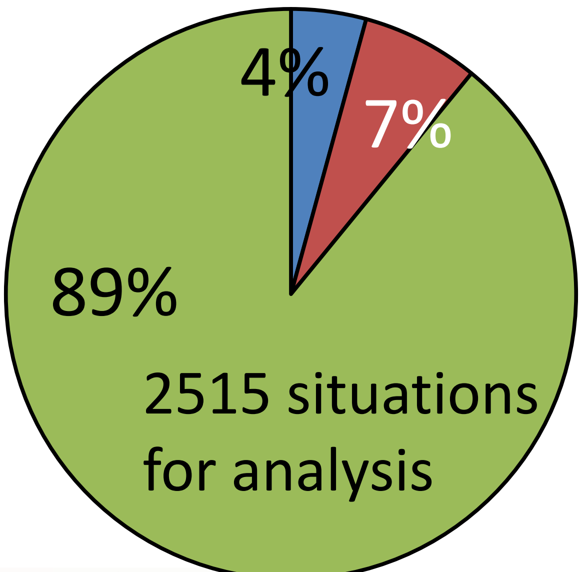
SPADE [Soricut & Marcu 2003]

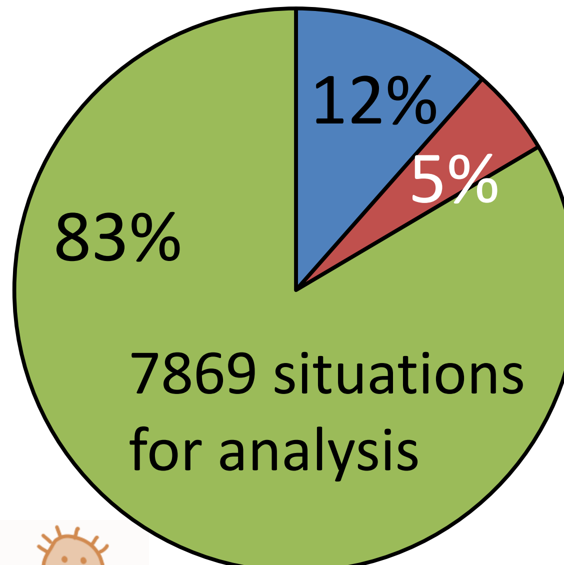
+ heuristic post-processing

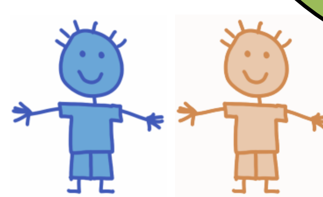
+ manual correction

marked as **NO SITUATION**
by at least one annotator
(e.g. headlines, names, dates)

merged to other segment
by at least one annotator







MASC news: 2823 segments

MASC news, jokes, letters:
9428 segments

Feature-driven annotation

1 label “easy” cases: speech acts, lexically-triggered abstract entities, other clear-cut cases

Feature-driven annotation

1 label "easy" cases: speech acts, lexically-triggered abstract entities, other clear-cut cases

2 determine **feature values**

genericity of main referent	fundamental aspectual class	habituality

Which features distinguish the SE types from each other?

Feature-driven annotation

1 label "easy" cases: speech acts, lexically-triggered abstract entities, other clear-cut cases

2 determine **feature values**

genericity of main referent	fundamental aspectual class	habituality

3 use feature values to assign

Situation entity (SE) types

Which features distinguish the SE types from each other?

Feature-driven annotation

1 label “easy” cases: speech acts, lexically-triggered abstract entities, other clear-cut cases

2 determine **feature values**

genericity of main referent	fundamental aspectual class	habituality

3 use feature values to assign

Situation entity (SE) types

+ Options for indicating uncertainty, multiple SE types / feature values.

Which features distinguish the SE types from each other?

Feature-driven annotation

1. label "easy" cases: speech acts, lexically-triggered abstract entities, other clear-cut cases
2. determine **feature values**

genericity of main referent	fundamental aspectual class	habituality

3. use feature values to assign

Situation entity (SE) types

\+ Options for indicating uncertainty, multiple SE types / feature values.

Advantages

✓ easier to convey annotation scheme

Feature-driven annotation

1. label "easy" cases: speech acts, lexically-triggered abstract entities, other clear-cut cases
2. determine **feature values**

genericity of main referent	fundamental aspectual class	habituality

3. use feature values to assign

Situation entity (SE) types

\+ Options for indicating uncertainty, multiple SE types / feature values.

Advantages

✓ easier to convey annotation scheme

✓ get partial information

Feature-driven annotation

1 label "easy" cases: speech acts, lexically-triggered abstract entities, other clear-cut cases

2 determine **feature values**

genericity of main referent	fundamental aspectual class	habituality

3 use feature values to assign

Situation entity (SE) types

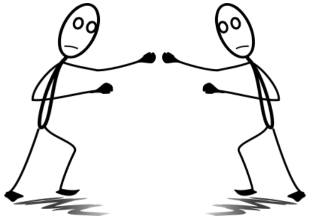
\+ Options for indicating uncertainty, multiple SE types / feature values.

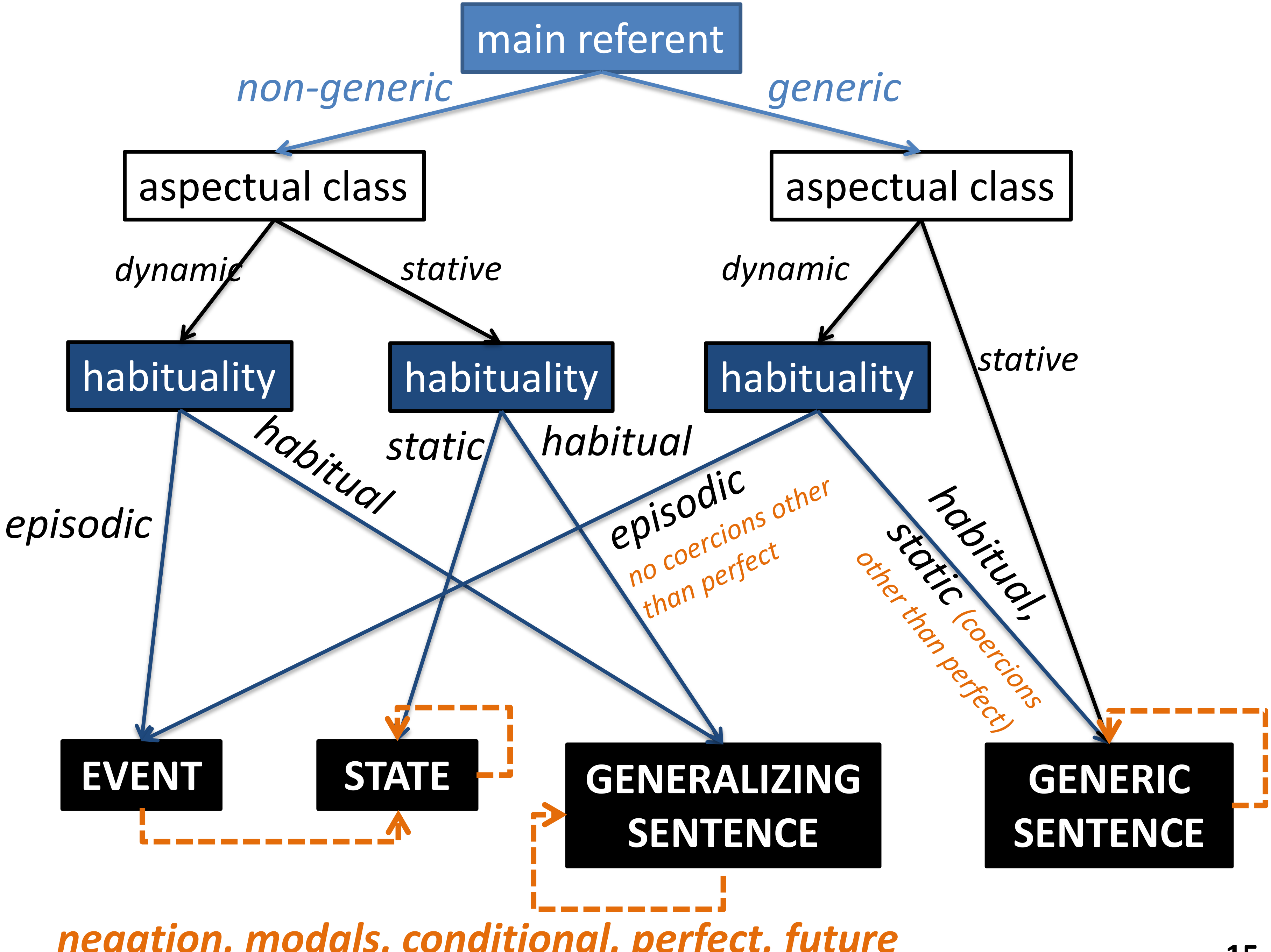
Advantages

✓ easier to convey annotation scheme

✓ get partial information

✓ analyze disagreements




main referent
non-generic
generic
aspectual class
aspectual class
dynamic
stative
dynamic
stative
habituality
habituality
habituality
episodic
habitual
static
habitual
episodic
no coercions other than perfect
habitual,
static (coercions other than perfect)
EVENT
STATE
GENERALIZING SENTENCE
GENERIC SENTENCE
negation, modals, conditional, perfect, future

Feature: genericity of main referent

What is this clause about? → usually the grammatical subject

Feature: genericity of main referent

What is this clause about? → usually the grammatical subject

NON-GENERIC

particular entity / group / company / organization / situation / process

Mary likes cats.
The cats broke the TV.
WWF protects animals.
That she didn't answer upset me.
Knitting this scarf took me two days.

Feature: genericity of main referent

What is this clause about? → usually the grammatical subject

NON-GENERIC

particular entity / group / company / organization / situation / process

Mary likes cats.
The cats broke the TV.
WWF protects animals.
That she didn't answer upset me.
Knitting this scarf took me two days.

GENERIC

kind-referring / class-referring NPs
generic concepts

Cats eat mice.
Lions in captivity have trouble to produce offspring.
Dinosaurs are extinct.
Security is an important issue.
Knitting a scarf is generally fun.

Feature: genericity of main referent

What is this clause about? → usually the grammatical subject

NON-GENERIC

particular entity / group / company / organization / situation / process

Mary likes cats.
The cats broke the TV.
WWF protects animals.
That she didn't answer upset me.
Knitting this scarf took me two days.

GENERIC

kind-referring / class-referring NPs
generic concepts

Cats eat mice.
Lions in captivity have trouble to produce offspring.
Dinosaurs are extinct.
Security is an important issue.
Knitting a scarf is generally fun.

distinguishes GENERIC SENTENCEs from other SE types
(in combination with other features)

[Friedrich & Palmer, ACL 2014]

Feature: fundamental aspectual class

feature of the entire clause,
marks main verb.

[Friedrich & Palmer, ACL 2014]

Feature: fundamental aspectual class

feature of the entire clause,
marks main verb.

distinguishes
EVENTs from STATEs

[Friedrich & Palmer, ACL 2014]

Feature: fundamental aspectual class

feature of the entire clause,
marks main verb.

distinguishes
EVENTs from STATEs

Juice **fills** the glass.
STATIVE

[Friedrich & Palmer, ACL 2014]

Feature: fundamental aspectual class

feature of the entire clause,
marks main verb.

distinguishes
EVENTs from STATEs

Juice **fills** the glass.
STATIVE

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

Feature: fundamental aspectual class

feature of the entire clause, marks main verb.

distinguishes EVENTs from STATEs

Juice **fills** the glass.
STATIVE

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

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

Feature: habituality

feature of the entire clause, marks main verb.

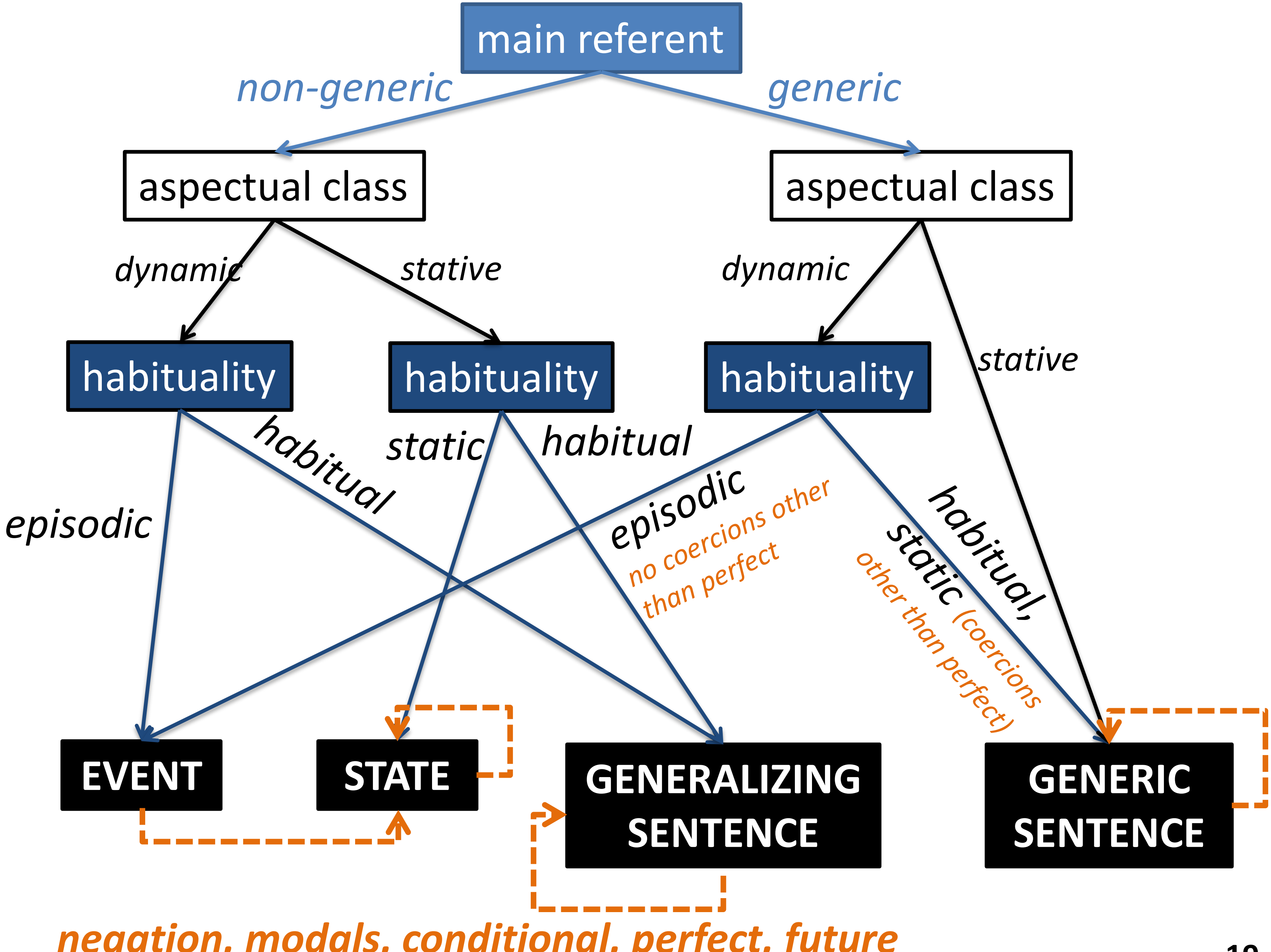
distinguishes EVENTs from general statives.

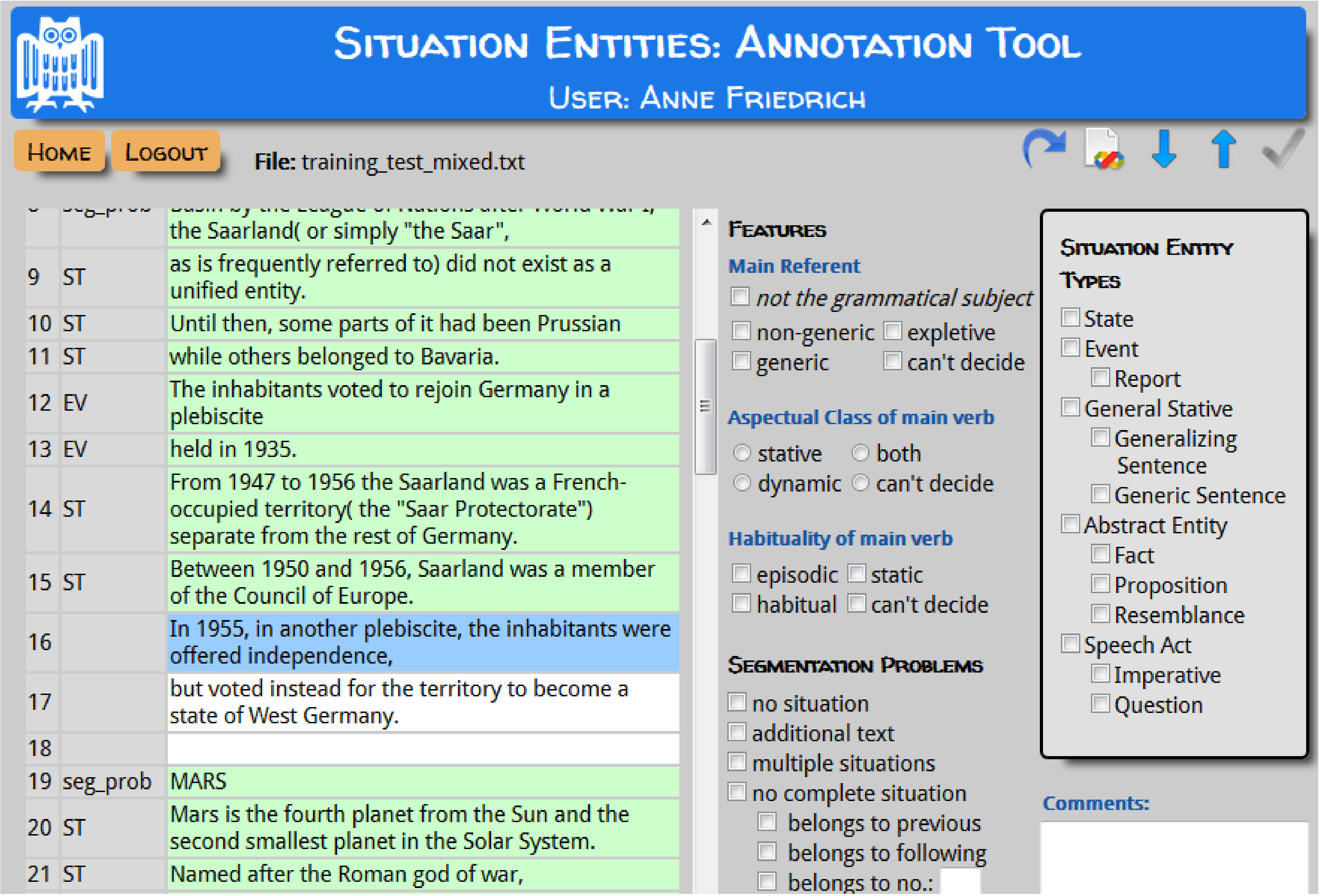
Feature: habituality

feature of the entire clause, marks main verb.

distinguishes EVENTs from general statives.

Mary fed her cats this morning.	**episodic**: one-time event
Mary feeds her cats every morning.	**habitual**: regularity
Glass breaks easily.	**habitual**: regularity
Mary owns four cats.	**static**: for STATEs


main referent
non-generic
generic
aspectual class
aspectual class
dynamic
stative
dynamic
stative
habituality
habituality
habituality
episodic
habitual
static
habitual
episodic
no coercions other than perfect
habitual, static (coercions other than perfect)
EVENT
STATE
GENERALIZING SENTENCE
GENERIC SENTENCE
negation, modals, conditional, perfect, future


Situation Entities: Annotation Tool
User: Anne Friedrich
Home
Logout
File: training_test_mixed.txt
the Saarland(or simply "the Saar",
9 ST as is frequently referred to) did not exist as a unified entity.
10 ST Until then, some parts of it had been Prussian
11 ST while others belonged to Bavaria.
12 EV The inhabitants voted to rejoin Germany in a plebiscite
13 EV held in 1935.
14 ST From 1947 to 1956 the Saarland was a French-occupied territory(the "Saar Protectorate") separate from the rest of Germany.
15 ST Between 1950 and 1956, Saarland was a member of the Council of Europe.
16 In 1955, in another plebiscite, the inhabitants were offered independence,
17 but voted instead for the territory to become a state of West Germany.
18
19 seg_prob MARS
20 ST Mars is the fourth planet from the Sun and the second smallest planet in the Solar System.
21 ST Named after the Roman god of war,
Features
Main Referent
not the grammatical subject
non-generic
expletive
generic
can't decide
Aspectual Class of main verb
stative
both
dynamic
can't decide
Habituality of main verb
episodic
static
habitual
can't decide
Segmentation Problems
no situation
additional text
multiple situations
no complete situation
belongs to previous
belongs to following
belongs to no.:
Situation Entity Types
State
Event
Report
General Stative
Generalizing Sentence
Generic Sentence
Abstract Entity
Fact
Proposition
Resemblance
Speech Act
Imperative
Question
Comments:

Features – broader perspective

corpus data for sub-tasks studied in the NLP community for which no large data sets are available

Features – broader perspective

corpus data for sub-tasks studied in the NLP community for which no large data sets are available

- automatic classification of fundamental aspectual class [Siegel & McKeown 2000, Friedrich & Palmer 2014] with the aim of improving temporal discourse processing [UzZaman et al. 2013, Bethard 2013, Costa & Branco 2012]

Features – broader perspective

corpus data for sub-tasks studied in the NLP community for which no large data sets are available

- automatic classification of fundamental aspectual class [Siegel & McKeown 2000, Friedrich & Palmer 2014] with the aim of improving temporal discourse processing [UzZaman et al. 2013, Bethard 2013, Costa & Branco 2012]
- identifying generic noun phrases [Reiter & Frank 2010]

Features – broader perspective

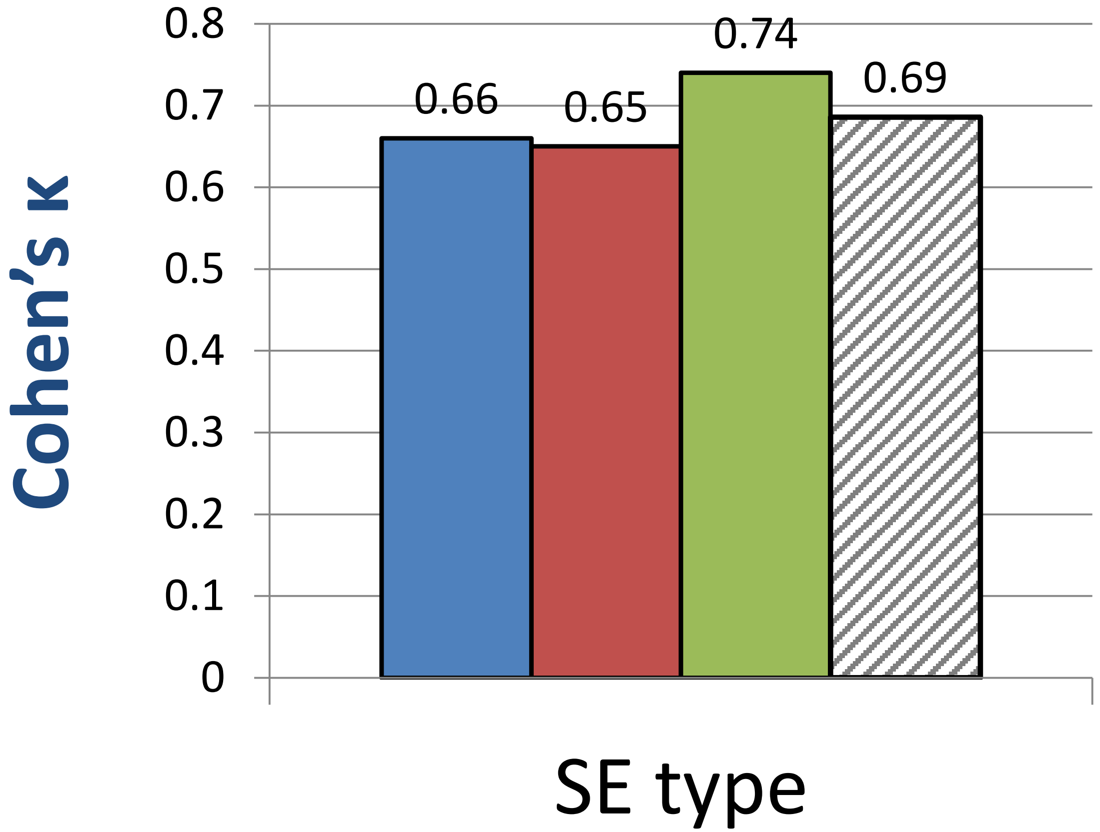
corpus data for sub-tasks studied in the NLP community for which no large data sets are available

- automatic classification of fundamental aspectual class [Siegel & McKeown 2000, Friedrich & Palmer 2014] with the aim of improving temporal discourse processing [UzZaman et al. 2013, Bethard 2013, Costa & Branco 2012]
- identifying generic noun phrases [Reiter & Frank 2010]
- identifying habitual vs. episodic sentences [Mathew & Katz 2009]

SE types: inter-annotator agreement

labels: STATE, EVENT, GENERIC SENTENCE, GENERALIZING SENTENCE

MASC: news (2823 situations)



pairs of annotators

A:B

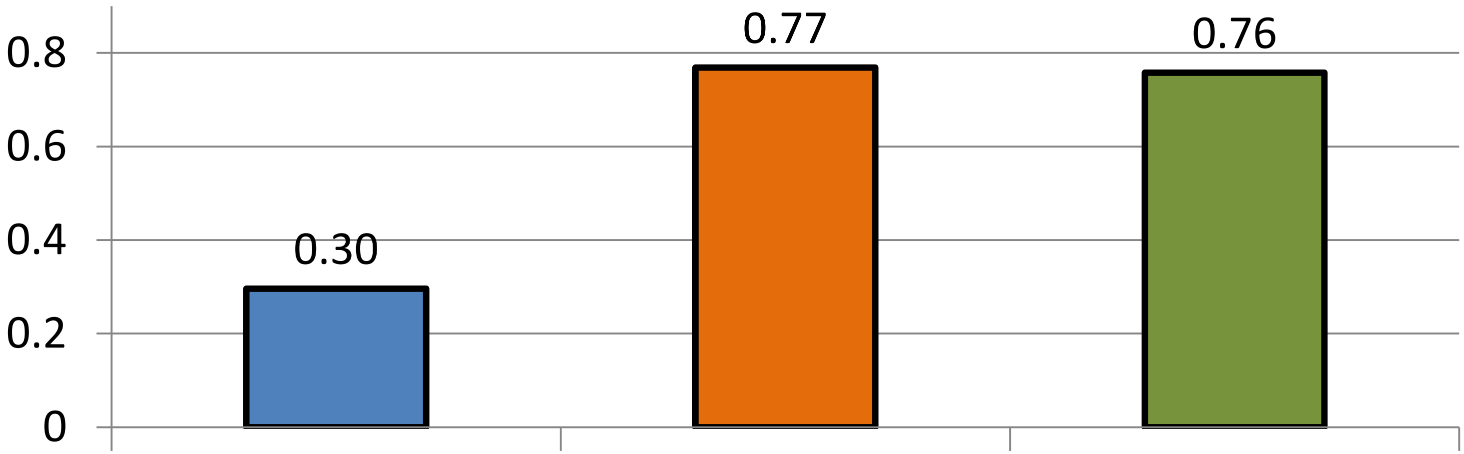
A:C

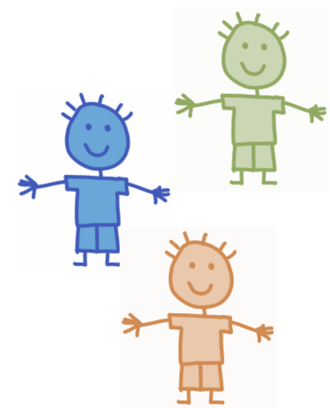
B:C

Features: inter-annotator agreement

MASC: news (2823 situations)

Fleiss' κ



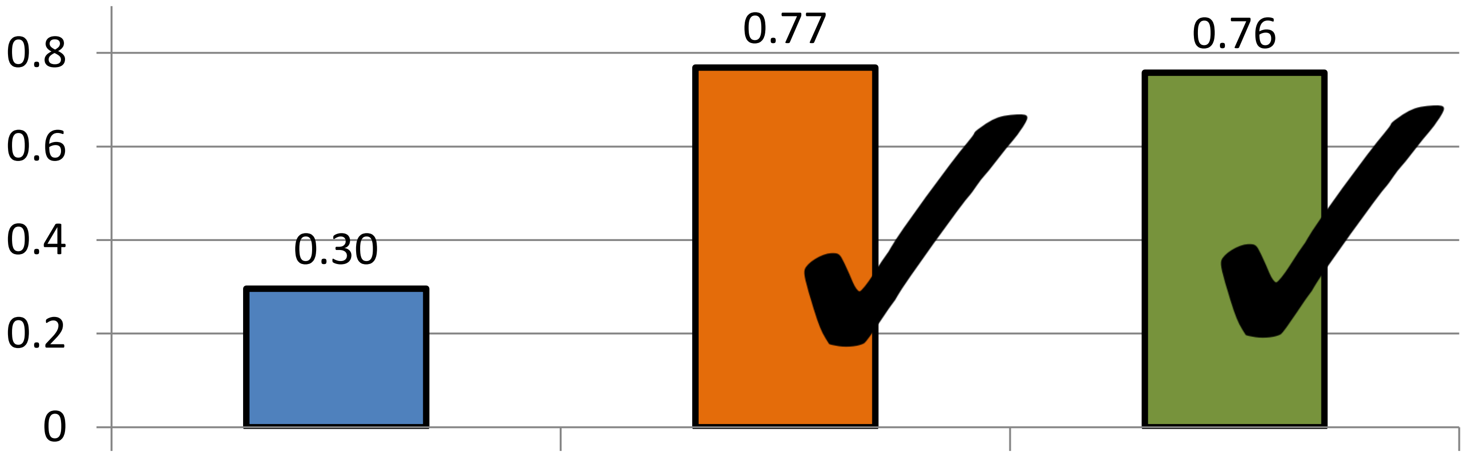


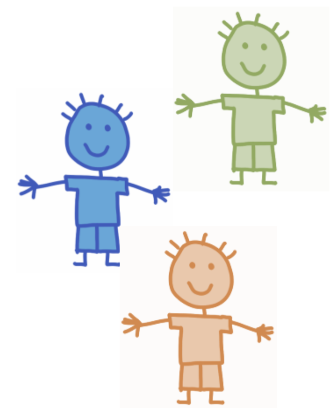
main referent	**aspectual class**	**habituality**
specific	*stative*	*episodic*
generic	*dynamic*	*habitual*
expletive	*both*	*static*

Features: inter-annotator agreement

MASC: news (2823 situations)

Fleiss' κ





main referent

specific
generic
expletive

aspectual class

stative
dynamic
both

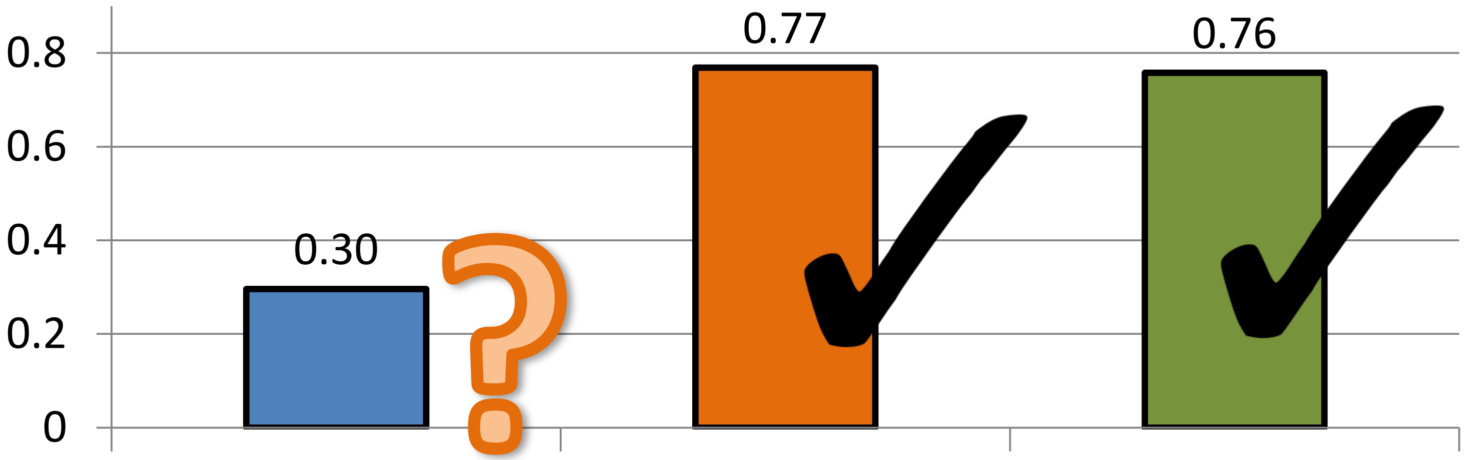
habituality

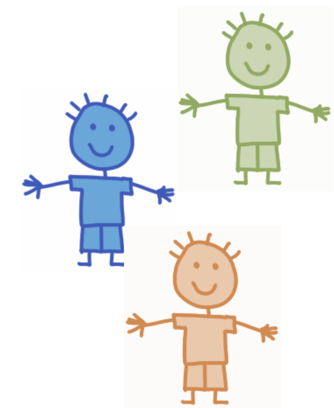
episodic
habitual
static

Features: inter-annotator agreement

MASC: news (2823 situations)

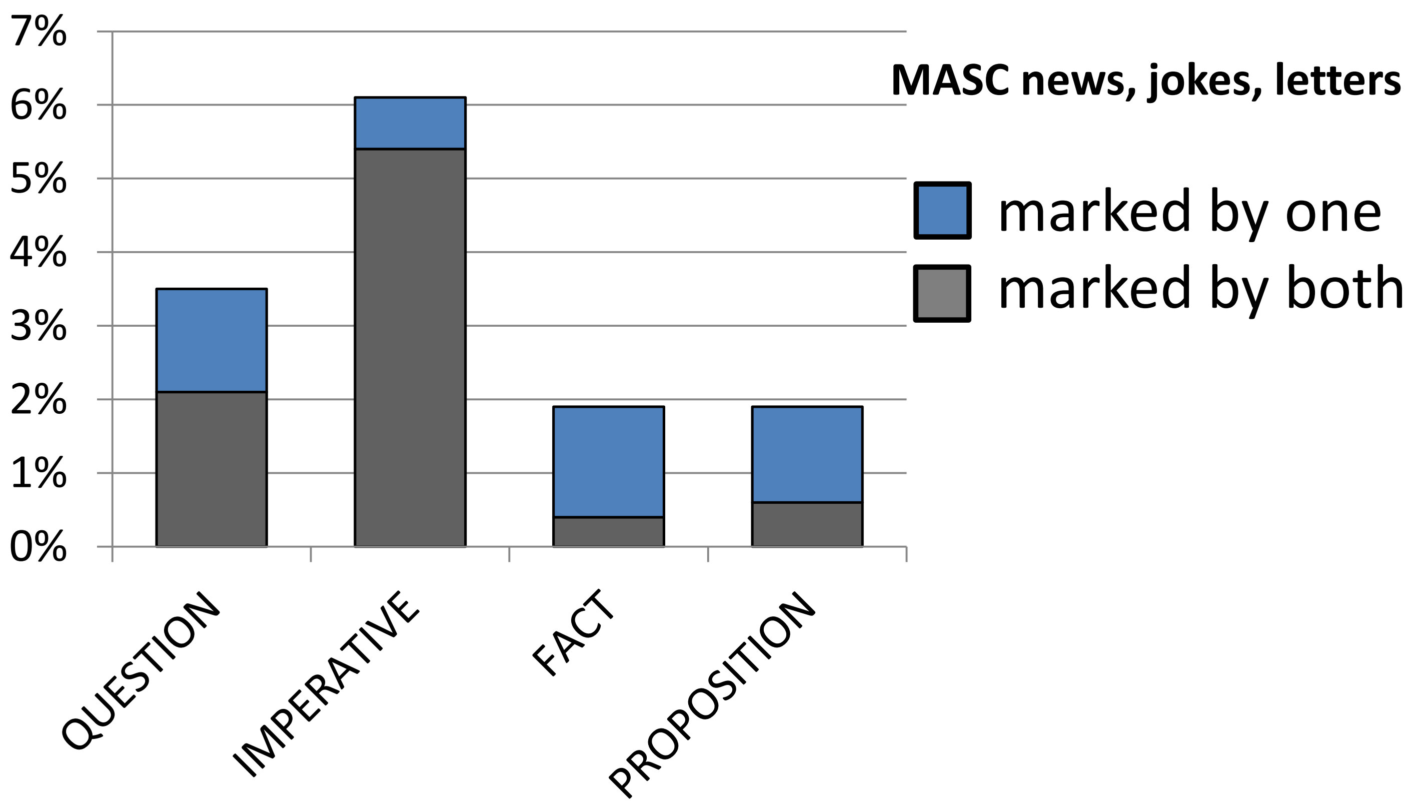
Fleiss' κ



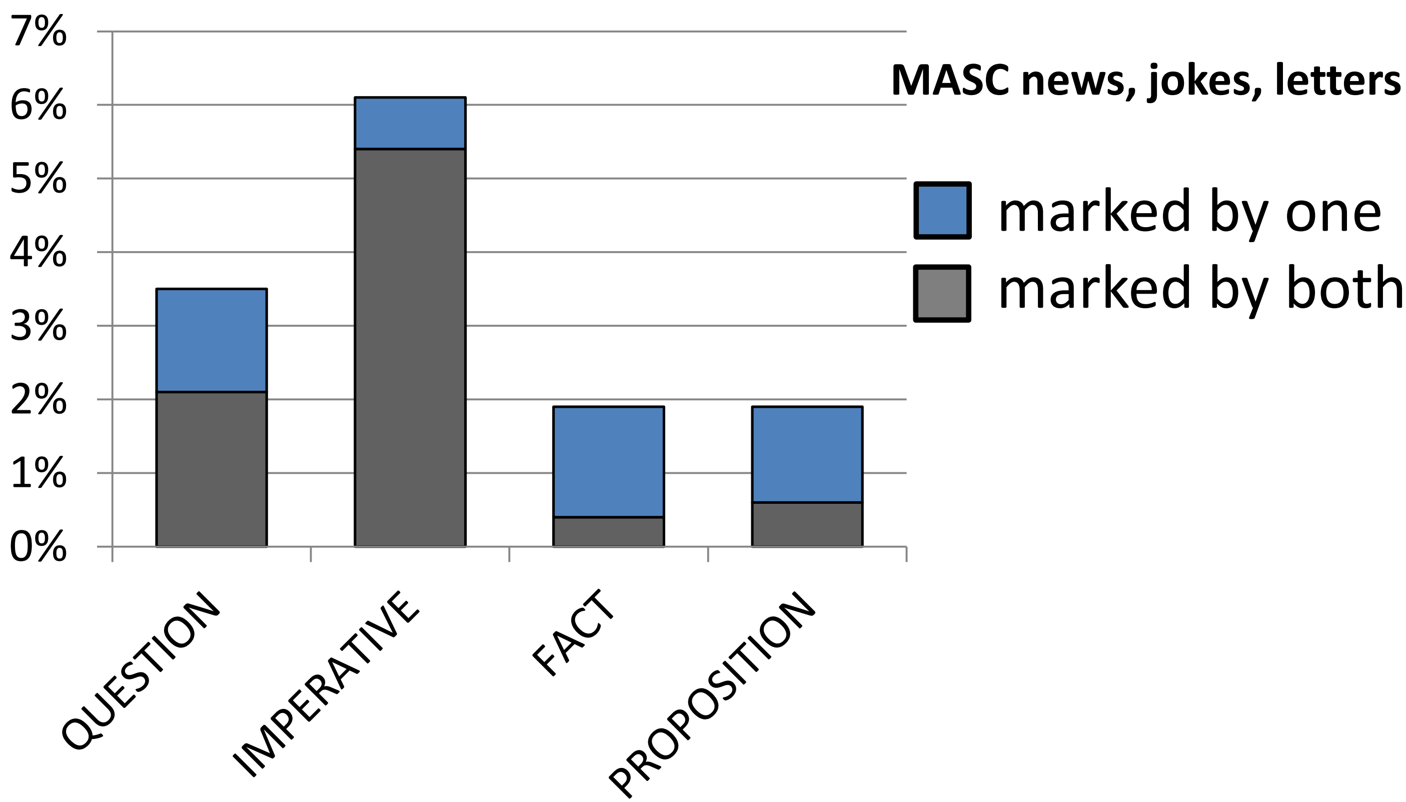


main referent	**aspectual class**	**habituality**
specific	*stative*	*episodic*
generic	*dynamic*	*habitual*
expletive	*both*	*static*

% of situations marked as speech acts / abstract entities:

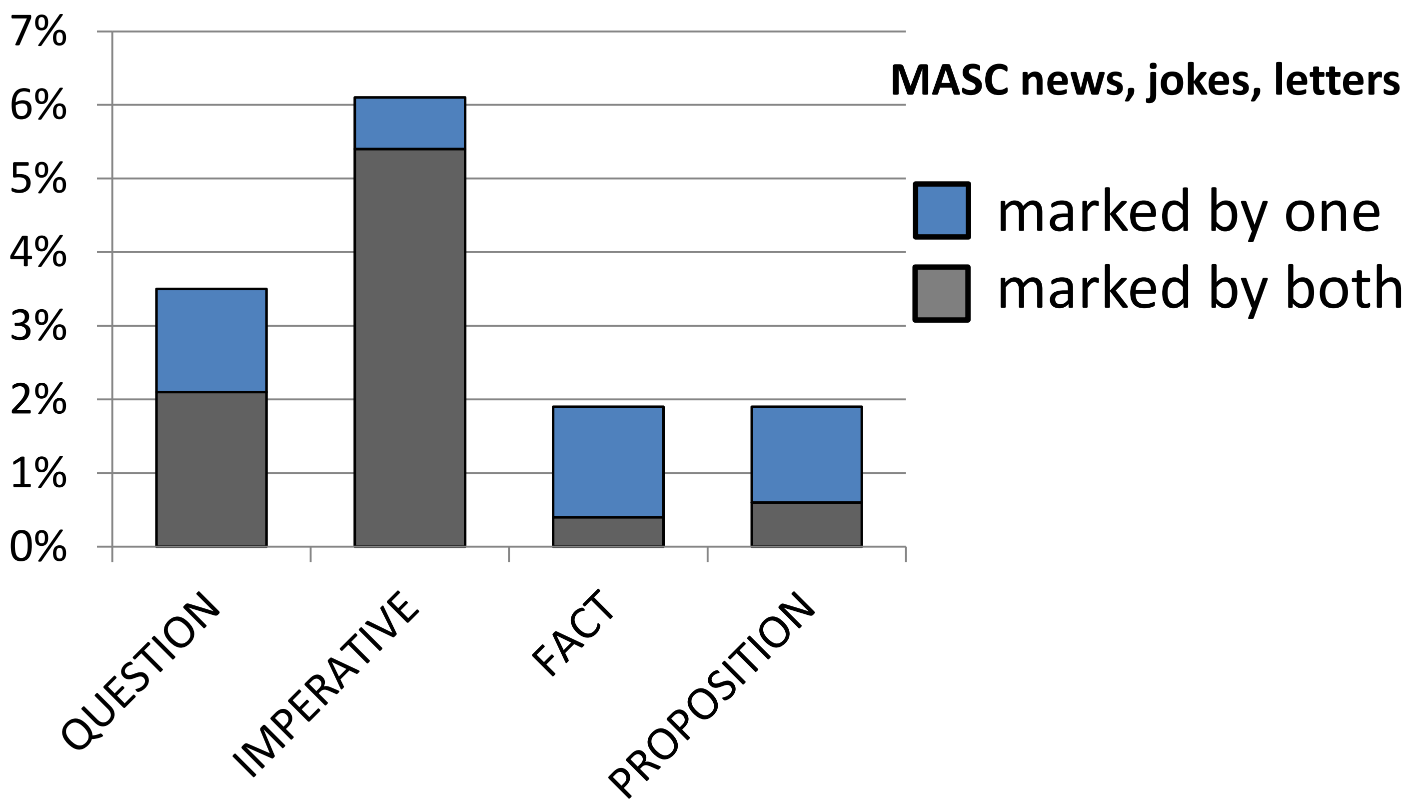


% of situations marked as speech acts / abstract entities:



indirect questions?

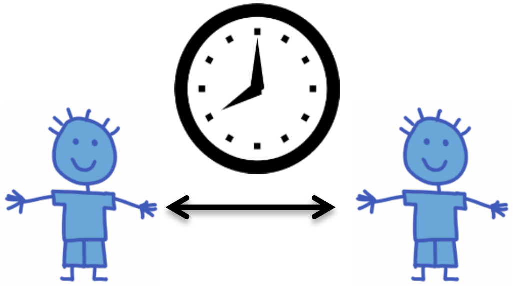
% of situations marked as speech acts / abstract entities:



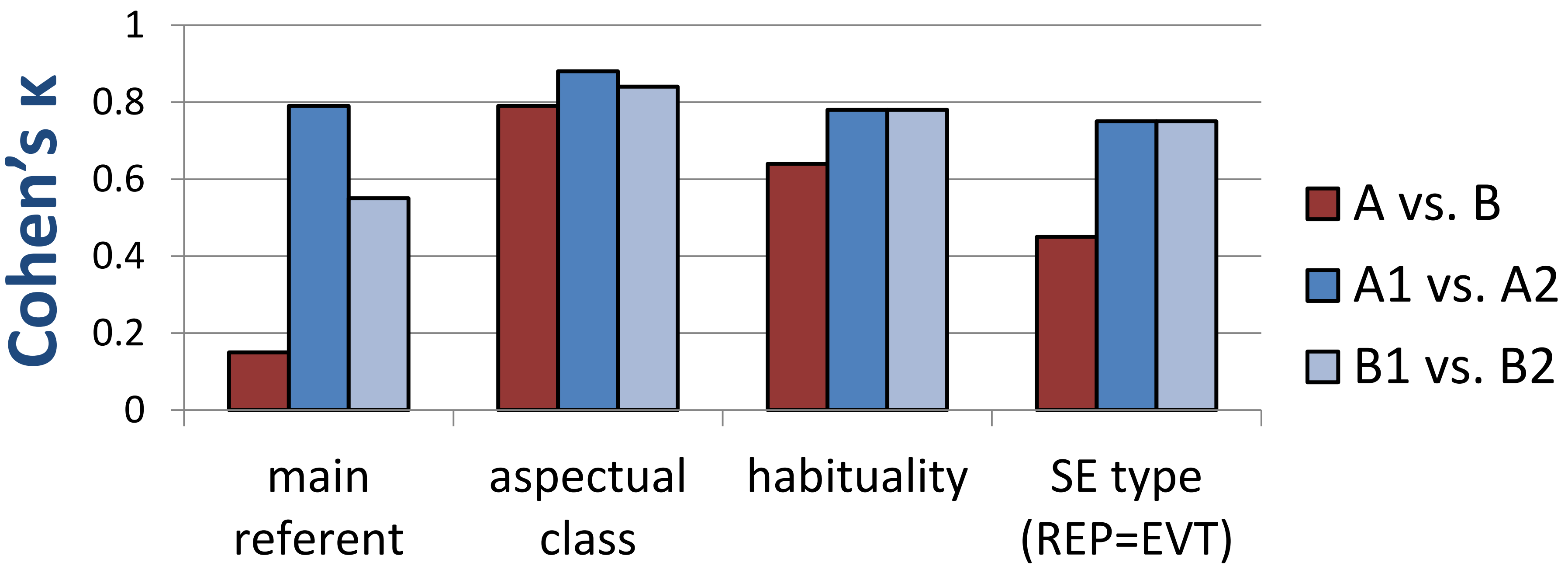
indirect questions?

no satisfying agreement yet
– lacking recall?

Intra-annotator consistency

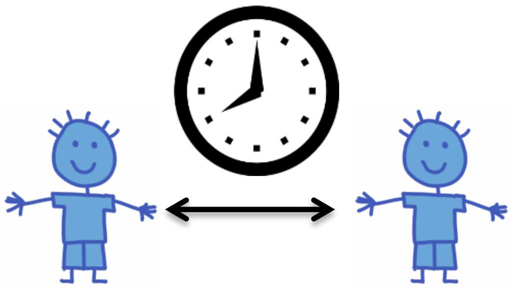


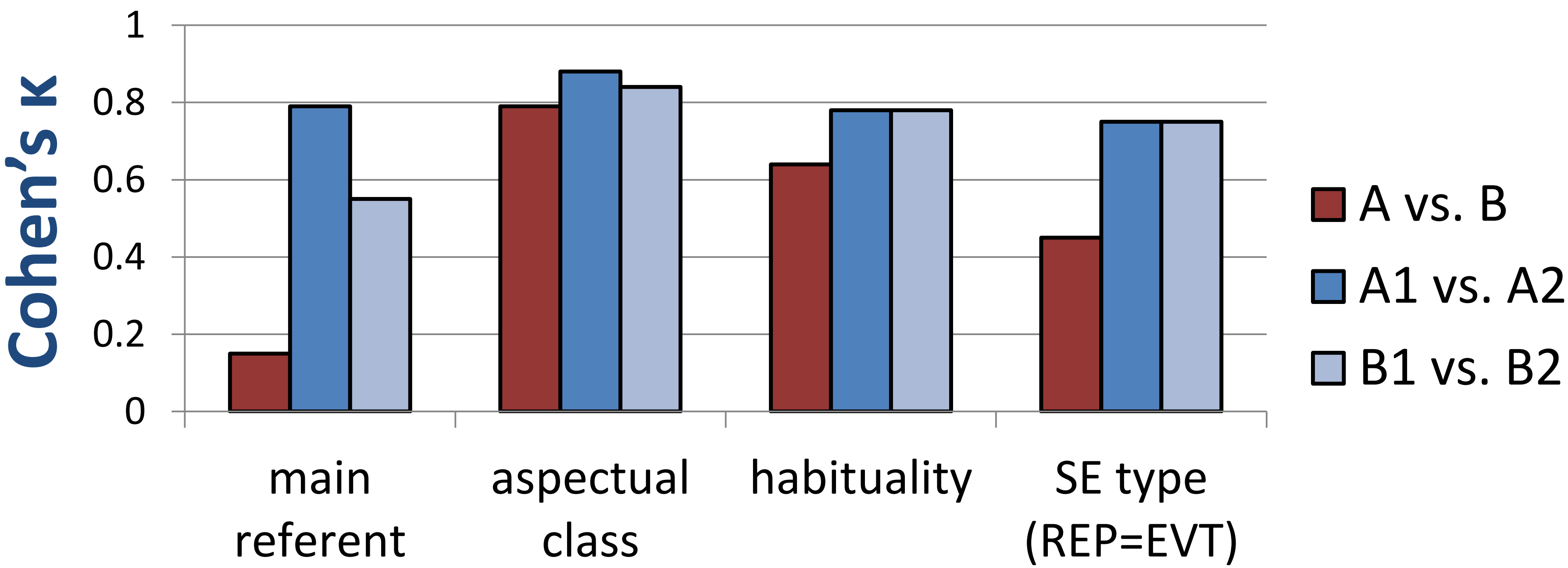
11 (5 news, 5 letters, 1 jokes) documents, 600 segments
(lowest agreements on SE type)



Intra-annotator consistency

11 (5 news, 5 letters, 1 jokes) documents, 600 segments
(lowest agreements on SE type)

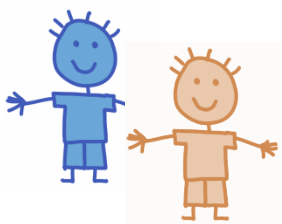




→ intra-agreement > inter-agreement
→different understanding of some cases

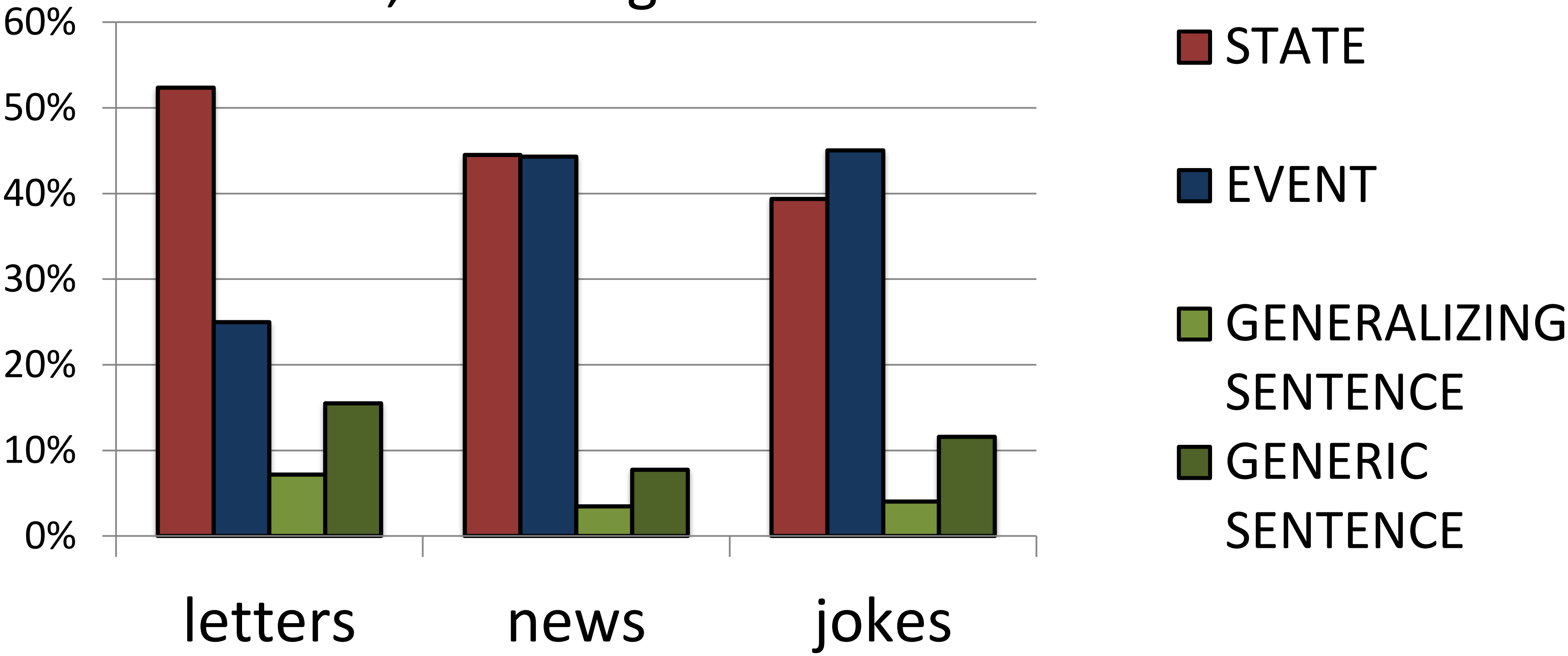
→ annotators occasionally *do* disagree with themselves
(but: hardest part of data set, total % of noise on SE type level << 20%)

Distribution of SE types: genres

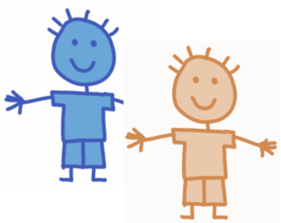


average of SE labels assigned

MASC, 9428 segments

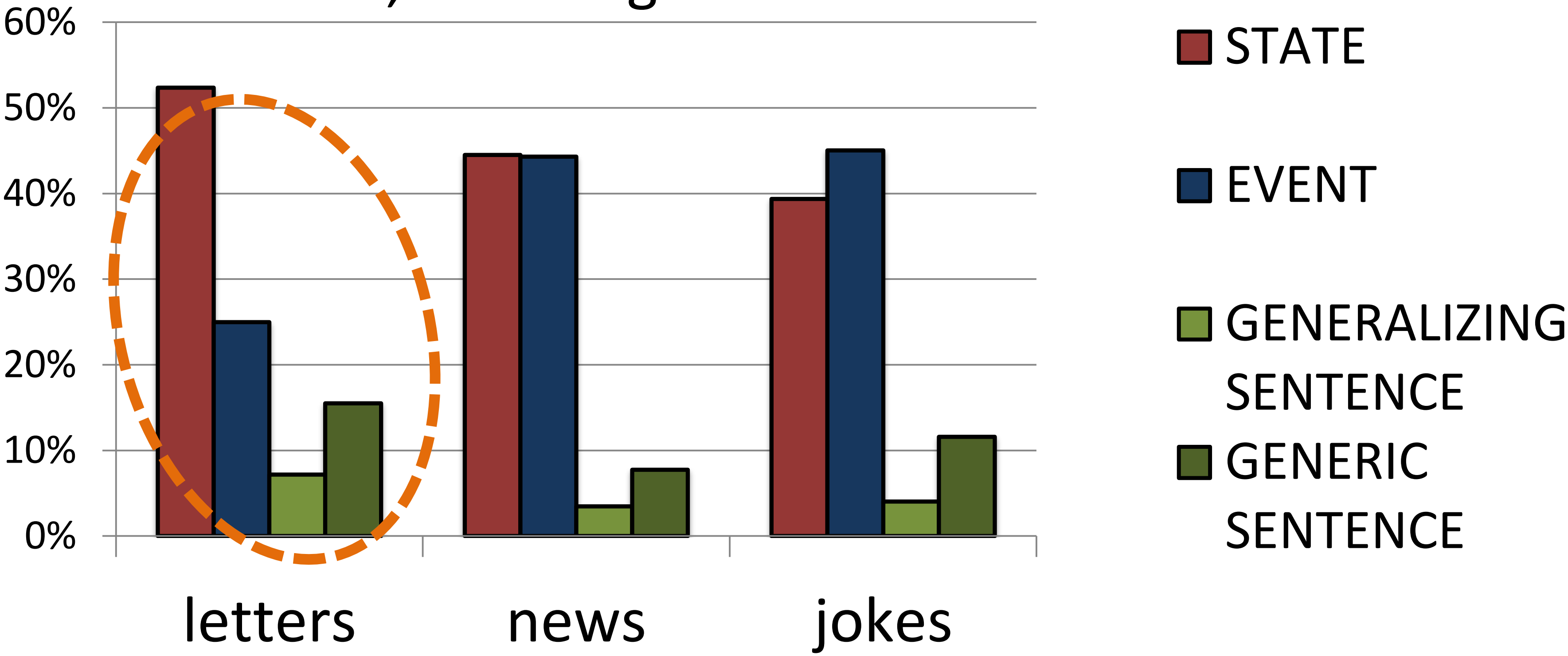


Distribution of SE types: genres



average of SE labels assigned

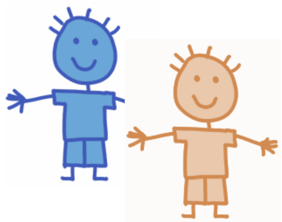
MASC, 9428 segments



letters has fewer events, more **general statives**

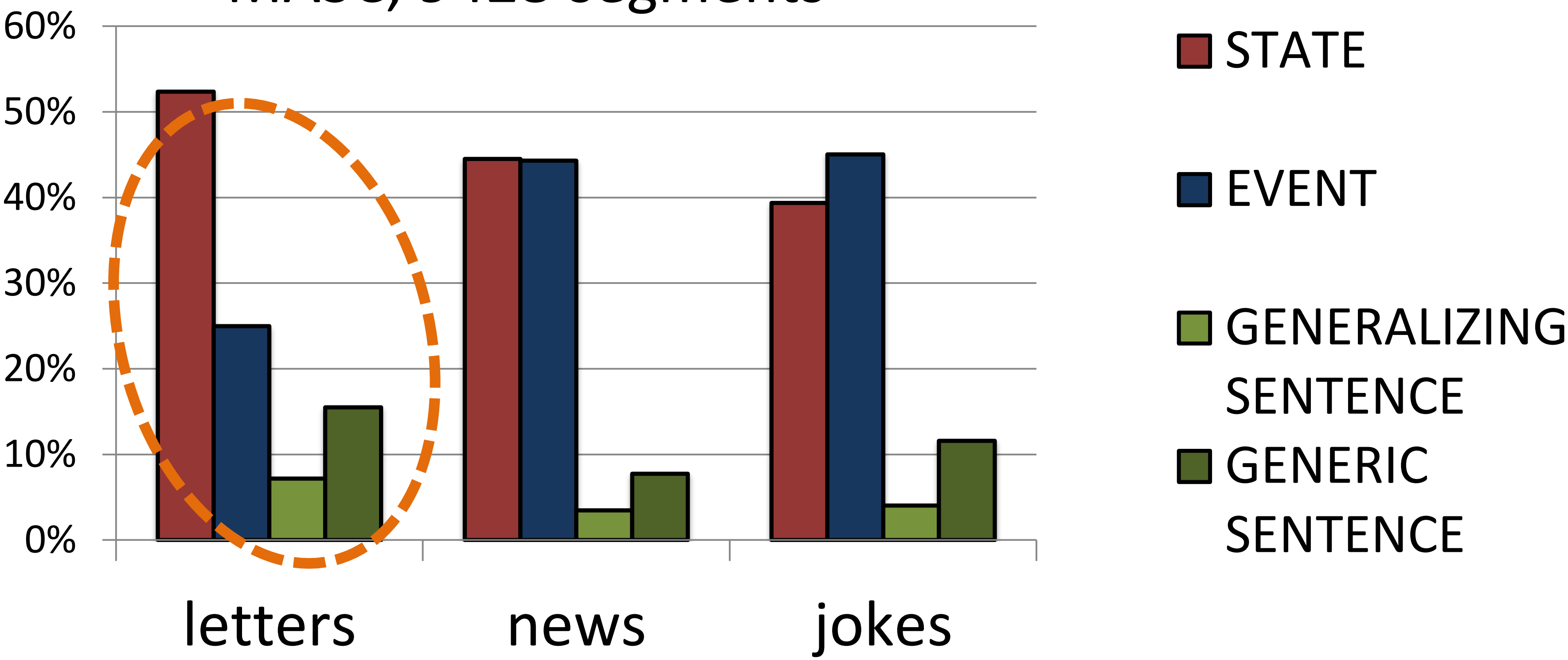
Distribution of SE types: genres

more details: [Palmer & Friedrich, 2014]



average of SE labels assigned

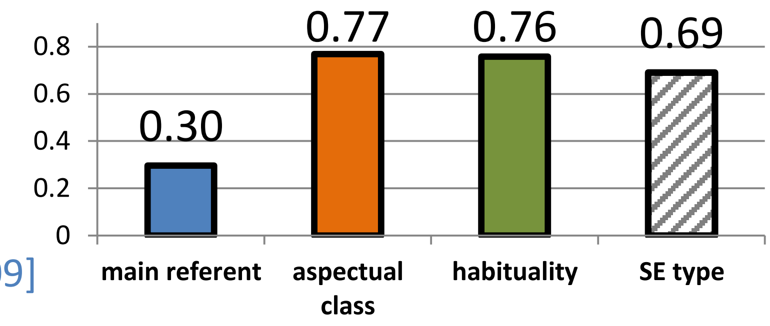
MASC, 9428 segments



➡ letters has fewer events, more **general statives**

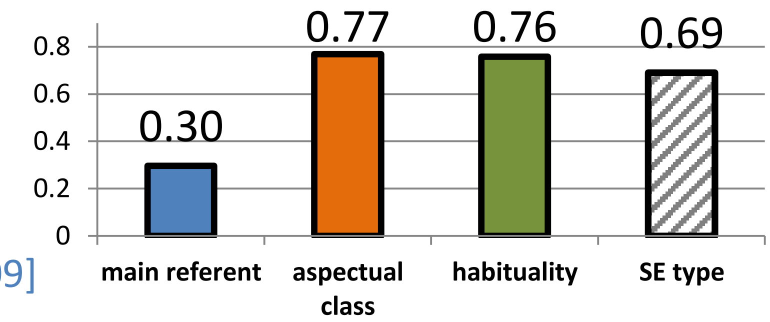
Summary:
annotation of situation entity types

- Annotation guidelines for situation entity types:
  – substantial agreement achieved for SE type, aspectual class & habituality
  – part of disagreements: hard cases
    →leverage for training
    [Plank et al. 2014, Beigman Klebanov & Beigman 2009]

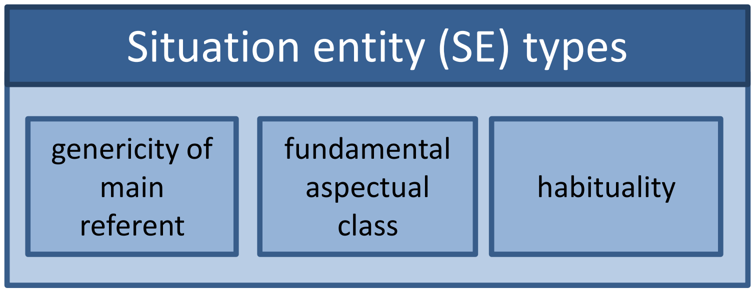


Summary:
annotation of situation entity types

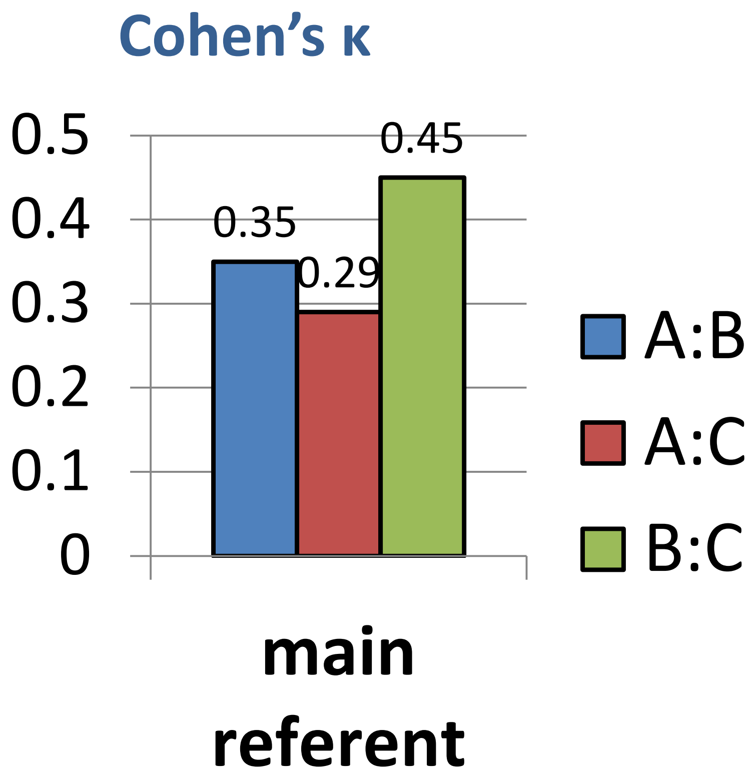
- Annotation guidelines for situation entity types:
  - substantial agreement achieved for SE type, aspectual class & habituality
  - part of disagreements: hard cases
    →leverage for training
    [Plank et al. 2014, Beigman Klebanov & Beigman 2009]



- Feature-based approach
  - helps annotators during annotation
  - analysis of disagreements
  - identify problems in guidelines
    → follow-up study on genericity

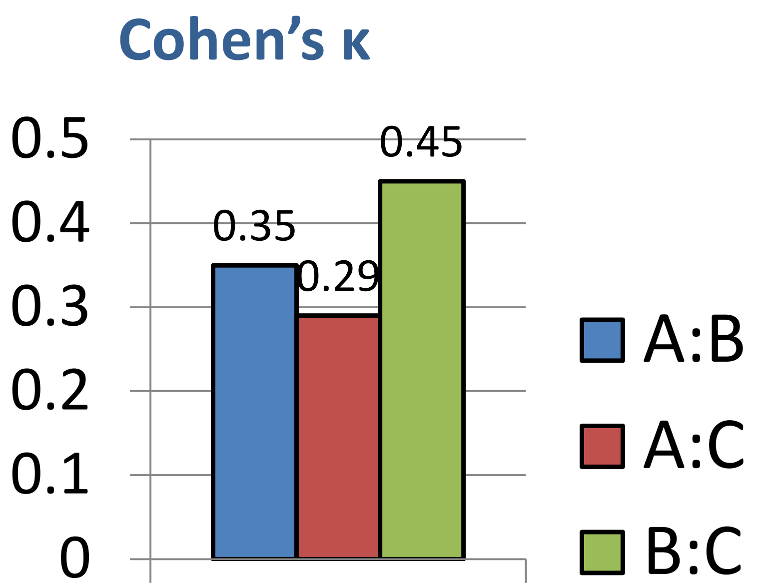


Feature: genericity of main referent (inter-annotator agreement)



specific
generic
expletive

Feature: genericity of main referent (inter-annotator agreement)

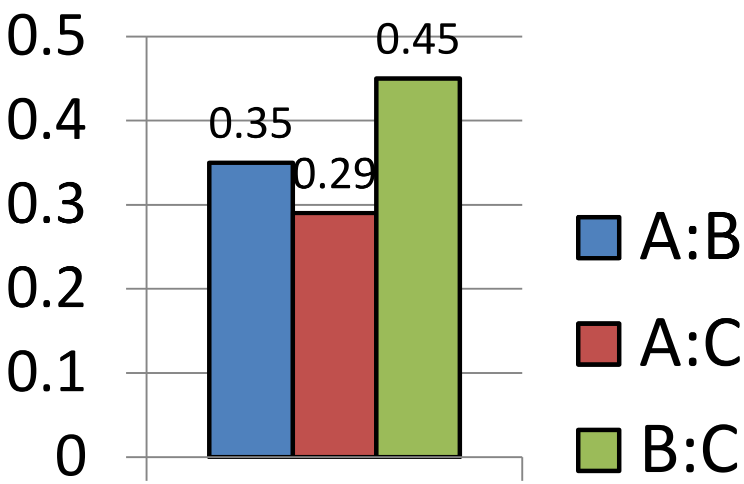


main referent

specific
generic
expletive

Feature: genericity of main referent
(inter-annotator agreement)

Cohen's κ

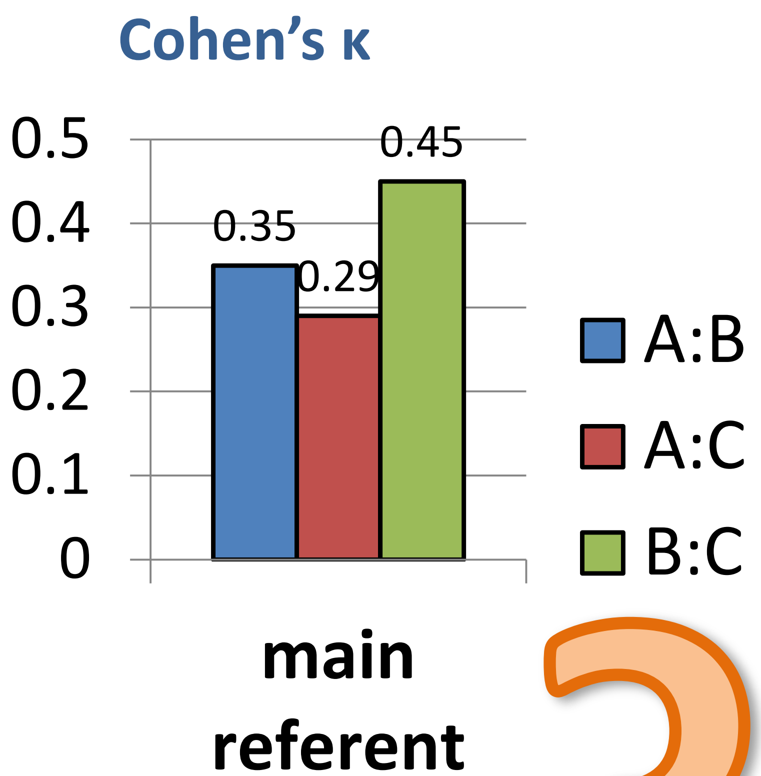


main referent

specific
generic
expletive

- **clarity** of annotation guidelines?

Feature: genericity of main referent (inter-annotator agreement)



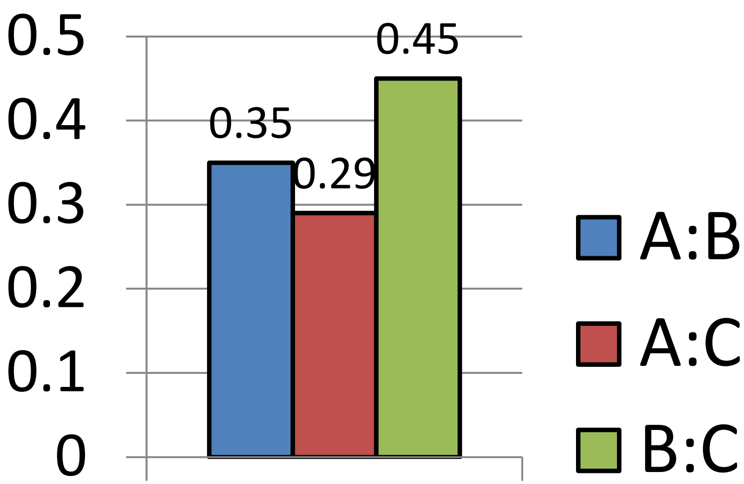
specific
generic
expletive

- **clarity** of annotation guidelines?
  - ***sparsity*** of label *generic*:

    B&C ($\kappa = 0.45$)
    – 2358 non-generic
    – 122 generic by one
    – 43 generic by both

Feature: genericity of main referent (inter-annotator agreement)

Cohen's κ



main referent

specific
generic
expletive

- **clarity** of annotation guidelines?
  - ***sparsity*** of label *generic*:

    B&C (κ = 0.45)
    - 2358 non-generic
    - 122 generic by one
    - 43 generic by both
- ***ambiguity / underspecification***

  ~ 30% of disagreements (estimate based on small qualitative analysis)

  every kid in New York

  "you" in letters

Generics follow-up study

address the issue of ***clarity***:

compared definition to existing theories [Carlson & Pelletier 1995]

& corpora (ACE 2005),

clarified definition in manual, added examples.

joint work with Melissa Peate Sorensen

Generics follow-up study

address the issue of ***clarity***:
compared definition to existing theories [Carlson & Pelletier 1995]
& corpora (ACE 2005),
clarified definition in manual, added examples.

Generic noun phrases (theory applied to subjects):
(compare to Krifka et al. 1995: "The Generic Book")

kind-referring: *The lion disappeared from Asia.*
nonspecific, referring to arbitrary member of kind:
A lion roars when it smells food.

joint work with Melissa Peate Sorensen

Generics follow-up study

- address the issue of ***sparsity***: enrich corpus with documents where we expect a higher proportion of general statives

joint work with Melissa Peate Sorensen

Generics follow-up study

- address the issue of ***sparsity***: enrich corpus with documents where we expect a higher proportion of general statives
- **Wikipedia documents:** ≈ 6100 situations, ≈ 50% marked generic

category
animals
games
gangs
history
sports
tribes

The blobfish *is a deep sea fish of the family…*
Blobfish *are typically shorter than 30cm.*

American football *is a sport*
played by two teams of eleven players.
The offense *attempts to advance an oval ball …*

Five cards *are dealt from a standard 52-card deck.*
The player *with the most piles wins.*

The Bari tribe *feels the effects as a whole.*
The Bari *trade …*

joint work with Melissa Peate Sorensen

Generics follow-up study

- address the issue of ***sparsity***: enrich corpus with documents where we expect a higher proportion of general statives
- **Wikipedia documents:** ≈ 6100 situations, ≈ 50% marked generic

category
animals
games
gangs
history
sports
tribes

The blobfish *is a deep sea fish of the family…*
Blobfish *are typically shorter than 30cm.*

inductive

[Carlson 1995]

American football *is a sport*
played by two teams of eleven players.
The offense *attempts to advance an oval ball …*

rules and regulations

Five cards *are dealt from a standard 52-card deck.*
The player *with the most piles wins.*

The Bari tribe *feels the effects as a whole.*
The Bari *trade …*

Wikipedia documents: agreement

- WikiGen corpus: 49 documents (≈ 6100 situations)
- agreement study: 14 documents (≈1800 situations), 3 annotators

Fleiss' K

main referent	**aspectual class**	**habituality**	**SE type**
0.64	0.66	0.63	0.67

substantial agreement

Generics follow-up study: lesson learned

1) Descriptions in **manual** were clarified, added more examples → third newly hired annotator learned scheme almost exclusively from manual.

Generics follow-up study: lesson learned

1) Descriptions in **manual** were clarified, added more examples → third newly hired annotator learned scheme almost exclusively from manual.
2) Selected (Wikipedia) data with more GENERIC SENTENCES

Generics follow-up study: lesson learned

1) Descriptions in **manual** were clarified, added more examples → third newly hired annotator learned scheme almost exclusively from manual.
2) Selected (Wikipedia) data with more GENERIC SENTENCES

substantial agreement

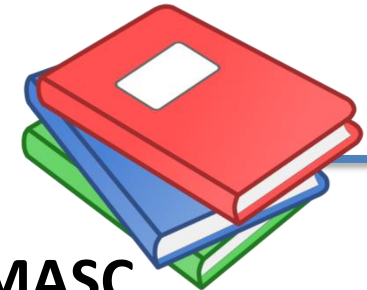
Generics follow-up study: lesson learned

1) Descriptions in **manual** were clarified, added more examples → third newly hired annotator learned scheme almost exclusively from manual.
2) Selected (Wikipedia) data with more GENERIC SENTENCES

substantial agreement

TODO: build computational model for detecting genericity of clauses

Overview



MASC

(automatic) segmentation

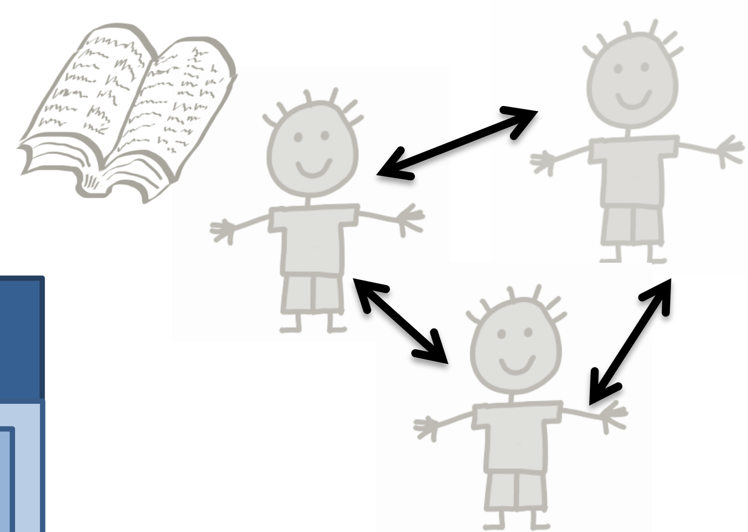
Situation entity (SE) types

genericity of main referent	fundamental aspectual class	habituality

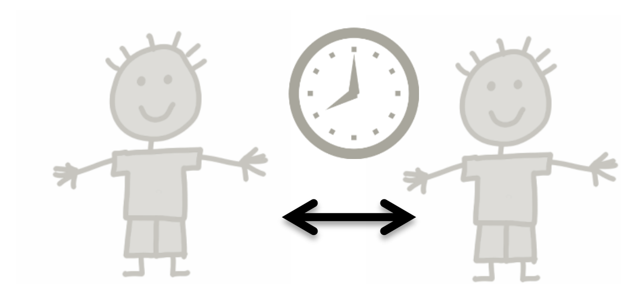
Feature-based annotation

2) automatic classification

1) Corpus annotation



inter-annotator agreement



intra-annotator consistency

Automatic prediction of aspectual class of verbs in context

[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**

Linguistic background

Vendler (1957):

time schemata of **verbs**

lexical aspect / aktionsart

states	*love, own*	**stative**
activities	*run*	**dynamic**
accomplishments	*write a letter*	
achievements	*realize*	

Linguistic background

Vendler (1957):

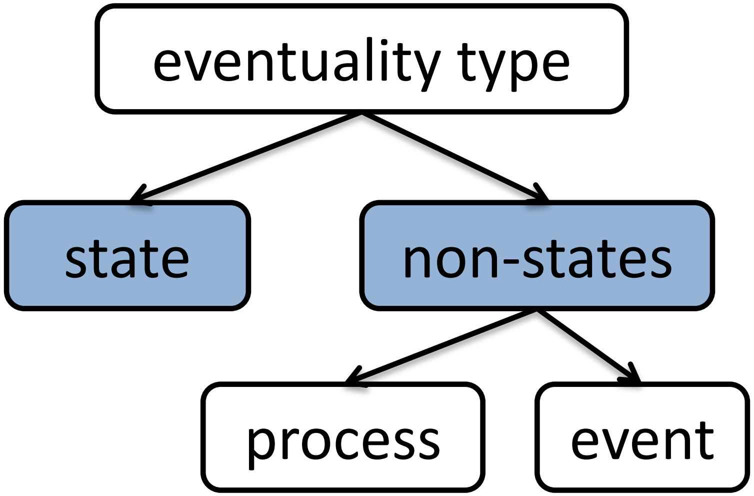
time schemata of **verbs**

lexical aspect / aktionsart

states	*love, own*	**stative**
activities	*run*	**dynamic**
accomplishments	*write a letter*	
achievements	*realize*	

Bach (1986):

time schemata of **sentences**



Task: predicting **fundamental** aspectual class

- a function of the main verb and a select group of arguments (may differ per verb)
- Siegel & McKeown (2000)

*John will **love** this cake!*	*John love cake*	**stative**
*John has **kissed** Mary.*	*John kiss Mary*	**dynamic**
*John **drives** to work.*	*John drive to work*	

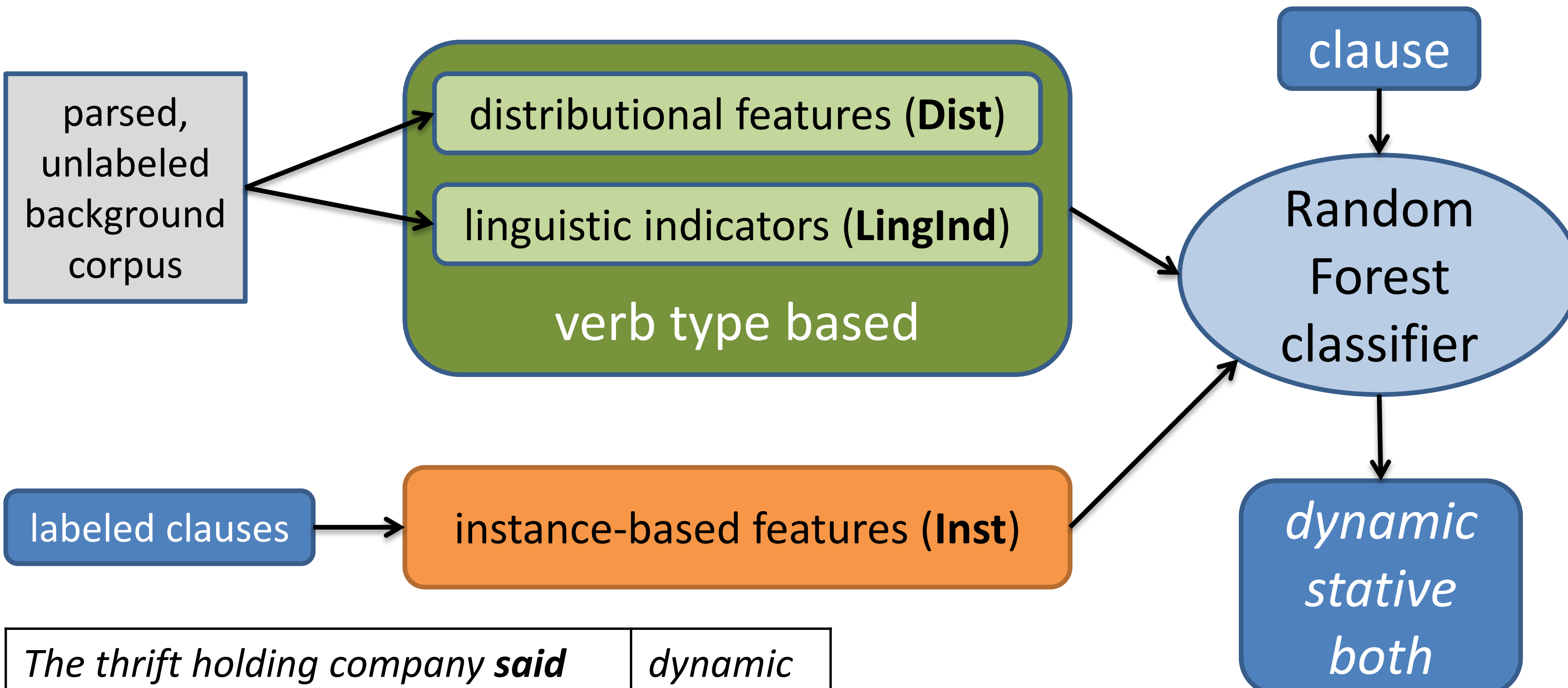
Task: predicting **fundamental** aspectual class

- a function of the main verb and a select group of arguments (may differ per verb)
- Siegel & McKeown (2000)
  - evaluation type-based
  - our work: instance-based

<table>
<tr><td>John will love this cake!</td><td>John love cake</td><td>stative</td></tr>
<tr><td>John has kissed Mary.</td><td>John kiss Mary</td><td rowspan="2">dynamic</td></tr>
<tr><td>John drives to work.</td><td>John drive to work</td></tr>
</table>

Method: Overview

- supervised three-way classification setting



*The thrift holding company **said***	*dynamic*
*it **expects** to obtain approval*	*stative*
*and **recognizes** the danger.*	*both*

Linguistic Indicators

- co-occurrence of **verb types** with certain linguistic features (Siegel & McKeown 2000)

parsed, unlabeled background corpus (GigaWord)

verb types

frequency	-
present	*says*
past	*said*
future	*will say*
perfect	*had won*
progressive	*is winning*
negated	*not/never*
particle	*up / in / …*
no subject	-

continuous adverb	*continually endlessly*
evaluation adverb	*better horribly*
manner adverb	*furiously patiently*
temporal adverb	*again finally*
in-PP	*in an hour*
for-PP	*for an hour*

tense extraction: Loaiciga et al. (2004)

Linguistic Indicators

- co-occurrence of **verb types** with certain linguistic features (Siegel & McKeown 2000)

parsed, unlabeled background corpus (GigaWord)

verb types

verb type: **fill**
feature: temporal-adverb
value: 0.0085

0.85% of the occurrences of fill are modified by one of the temporal adverbs.

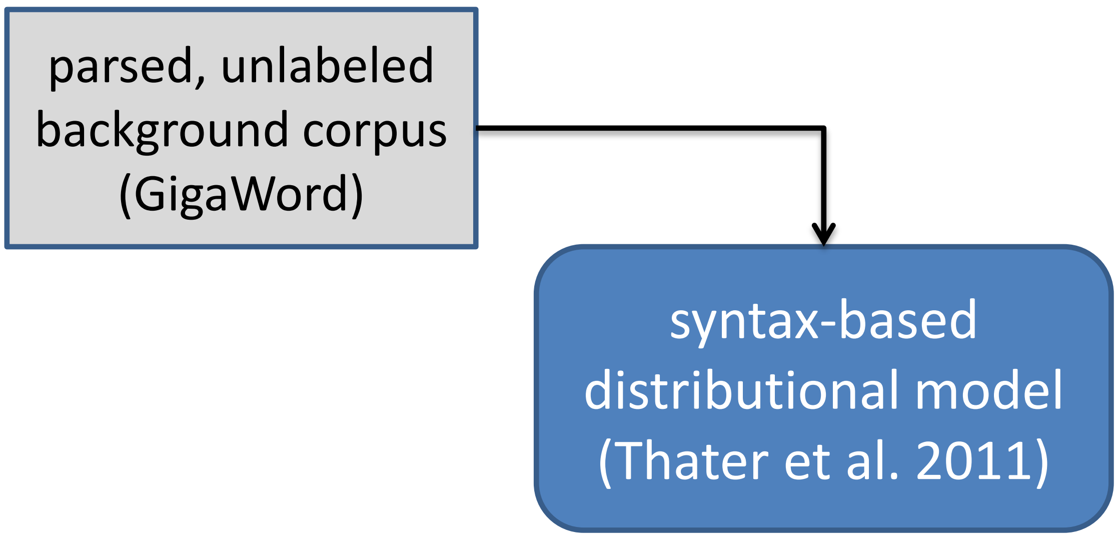
frequency	-
present	*says*
past	*said*
future	*will say*
perfect	*had won*
progressive	*is winning*
negated	*not/never*
particle	*up / in / …*
no subject	-

continuous adverb	*continually endlessly*
evaluation adverb	*better horribly*
manner adverb	*furiously patiently*
temporal adverb	*again finally*
in-PP	*in an hour*
for-PP	*for an hour*

tense extraction: Loaiciga et al. (2004)

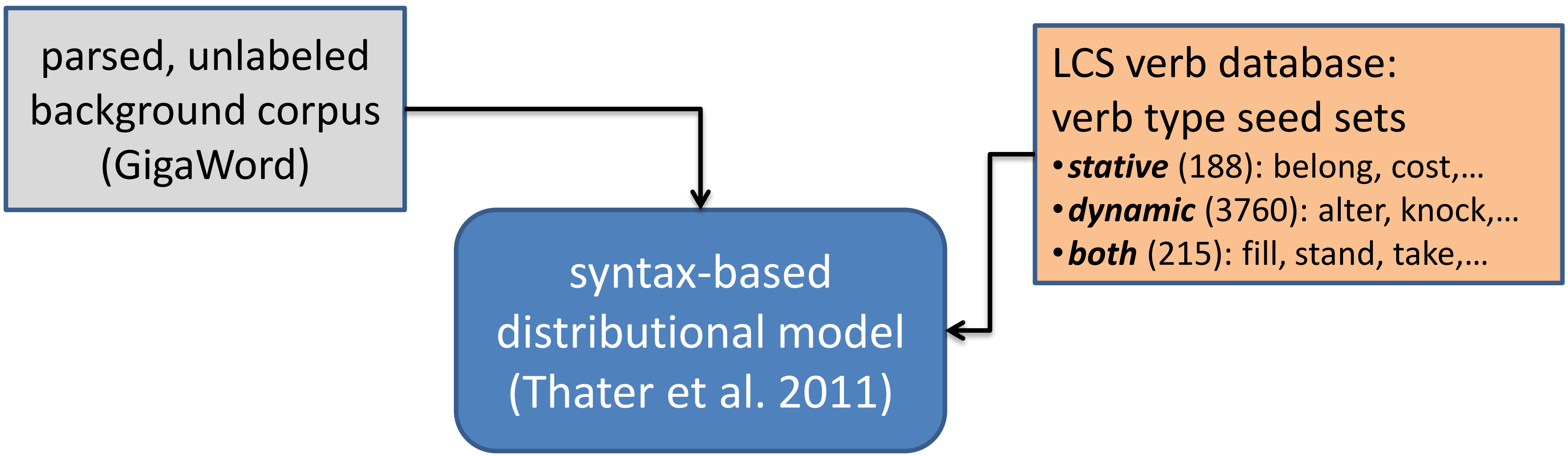
Distributional features

- average similarities with verbs in seed sets



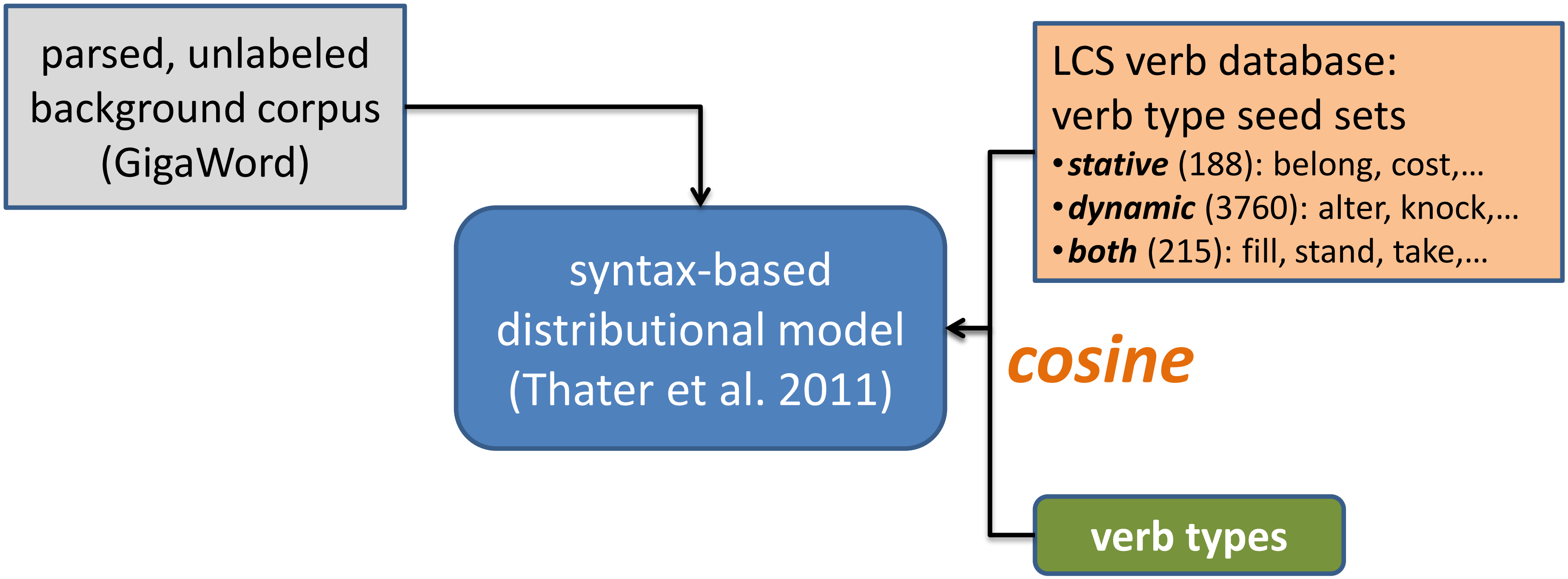
Distributional features

- average similarities with verbs in seed sets



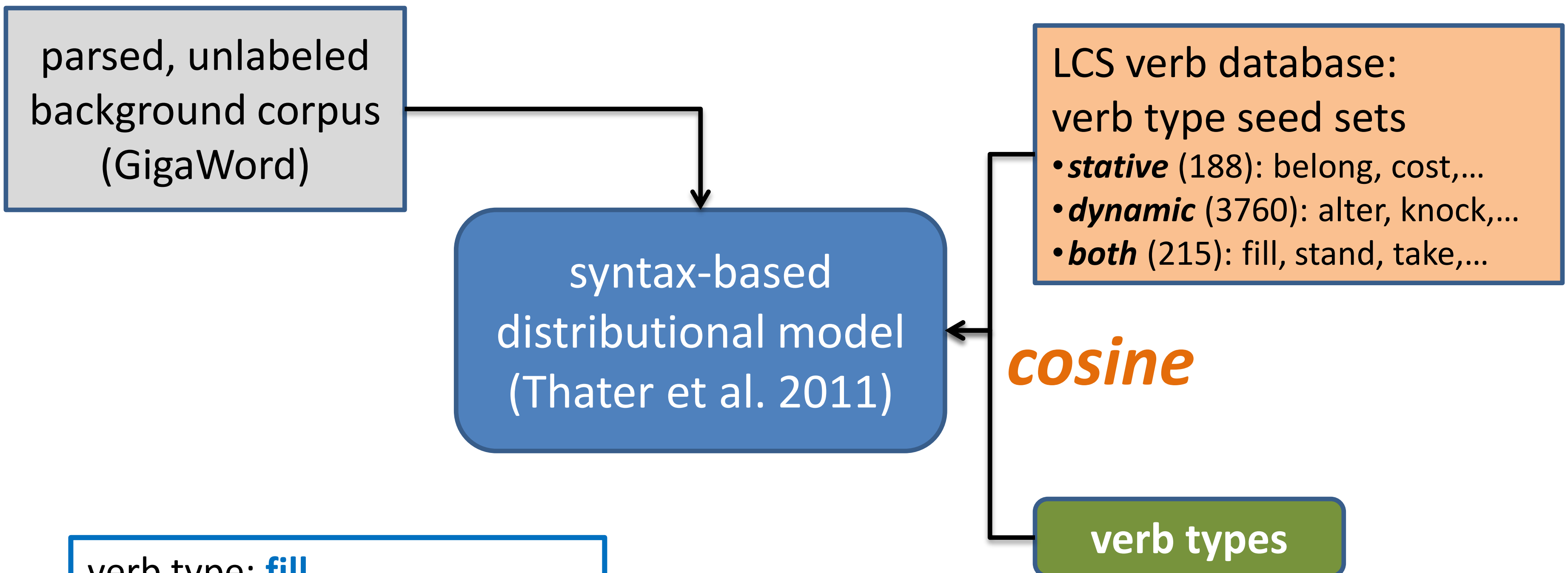
Distributional features

- average similarities with verbs in seed sets



Distributional features

- average similarities with verbs in seed sets



verb type: **fill**
avg. sim. with *stative*: 0.31
avg. sim. with *dynamic*: 0.27
avg. sim. with *both*: 0.16

Instance-based features

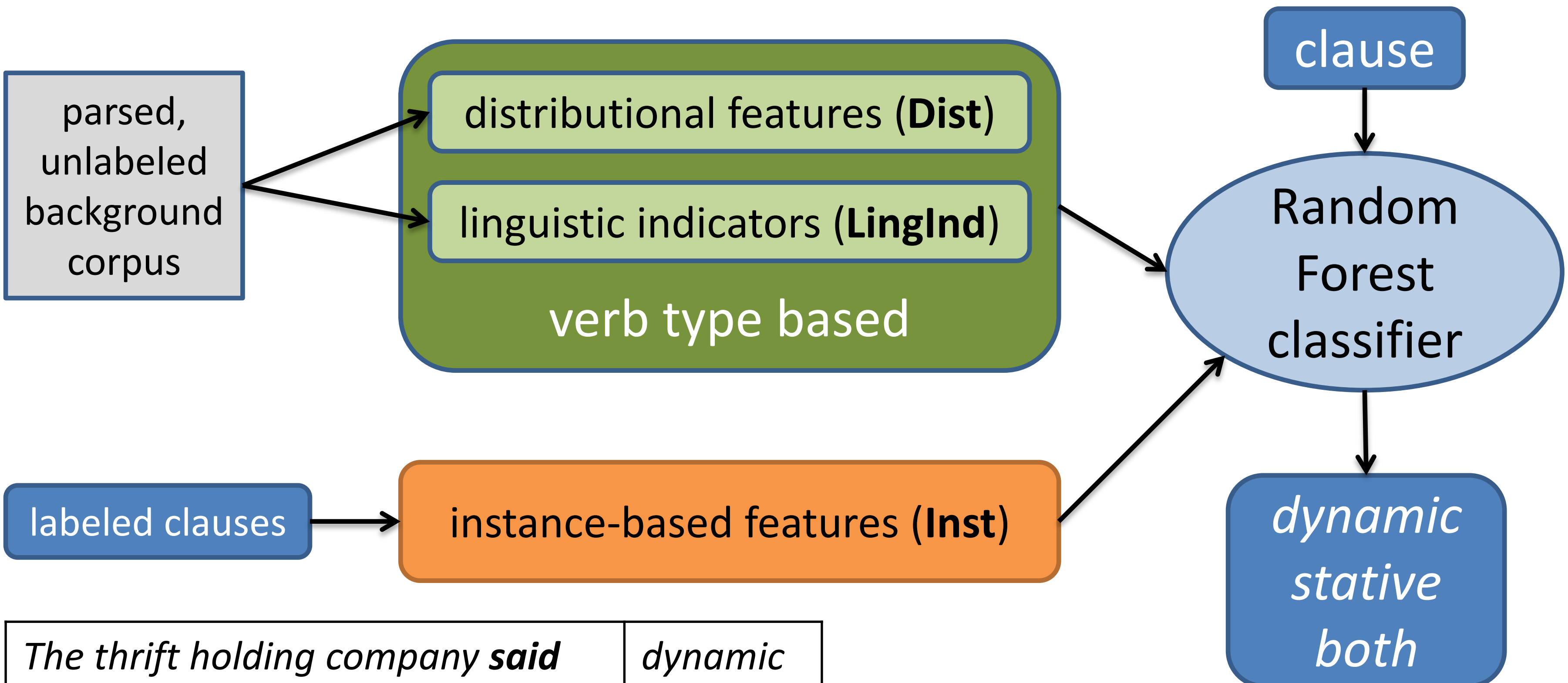
- verb-centric syntactic-semantic features

A little girl had just ***finished***
her first week of school.

tense:past	**progressive**:false
pos:VBD	**dobj**:noun.time
perfect:true	**particle**:none
voice:active	**subj**:noun.person

Method: Overview

- supervised three-way classification setting



*The thrift holding company **said***	*dynamic*
*it **expects** to obtain approval*	*stative*
*and **recognizes** the danger.*	*both*

Experiments 1&2: SEEN vs. UNSEEN verbs

Asp-MASC: 6161 clauses (complete texts) excluding be/have, 2 annotators, $\kappa = 0.7$, 10-fold cross validation

Experiments 1&2: SEEN vs. UNSEEN verbs

Asp-MASC: 6161 clauses (complete texts) excluding be/have, 2 annotators, κ = 0.7, 10-fold cross validation

SEEN verbs:

labeled training data available

Type-based features
→ same accuracy (**84%**) as only using Lemma
(= memorizing most frequent class per verb)

Experiments 1&2: SEEN vs. UNSEEN verbs

Asp-MASC: 6161 clauses (complete texts) excluding be/have, 2 annotators, κ = 0.7, 10-fold cross validation

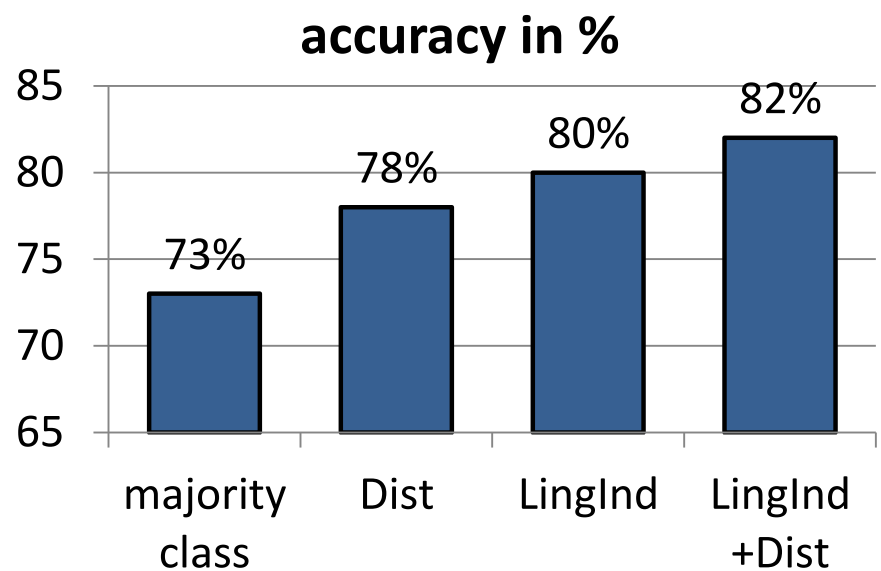
SEEN verbs:

labeled training data available

Type-based features
→ same accuracy (**84%**) as only using Lemma
(= memorizing most frequent class per verb)

UNSEEN verbs:

no labeled training data available



Experiments 1&2: SEEN vs. UNSEEN verbs

Asp-MASC: 6161 clauses (complete texts) excluding be/have, 2 annotators, κ = 0.7, 10-fold cross validation

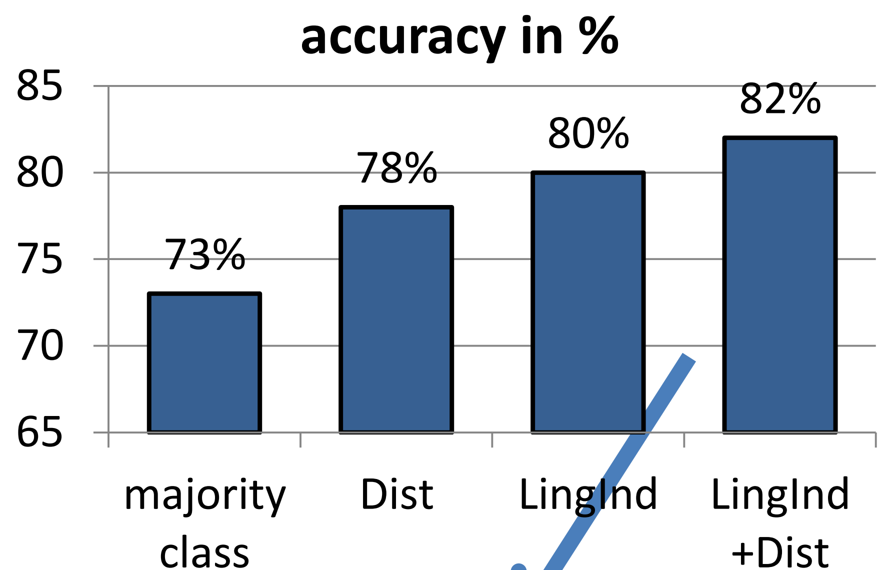
SEEN verbs:

labeled training data available

Type-based features
→ same accuracy (**84%**) as only using Lemma
(= memorizing most frequent class per verb)

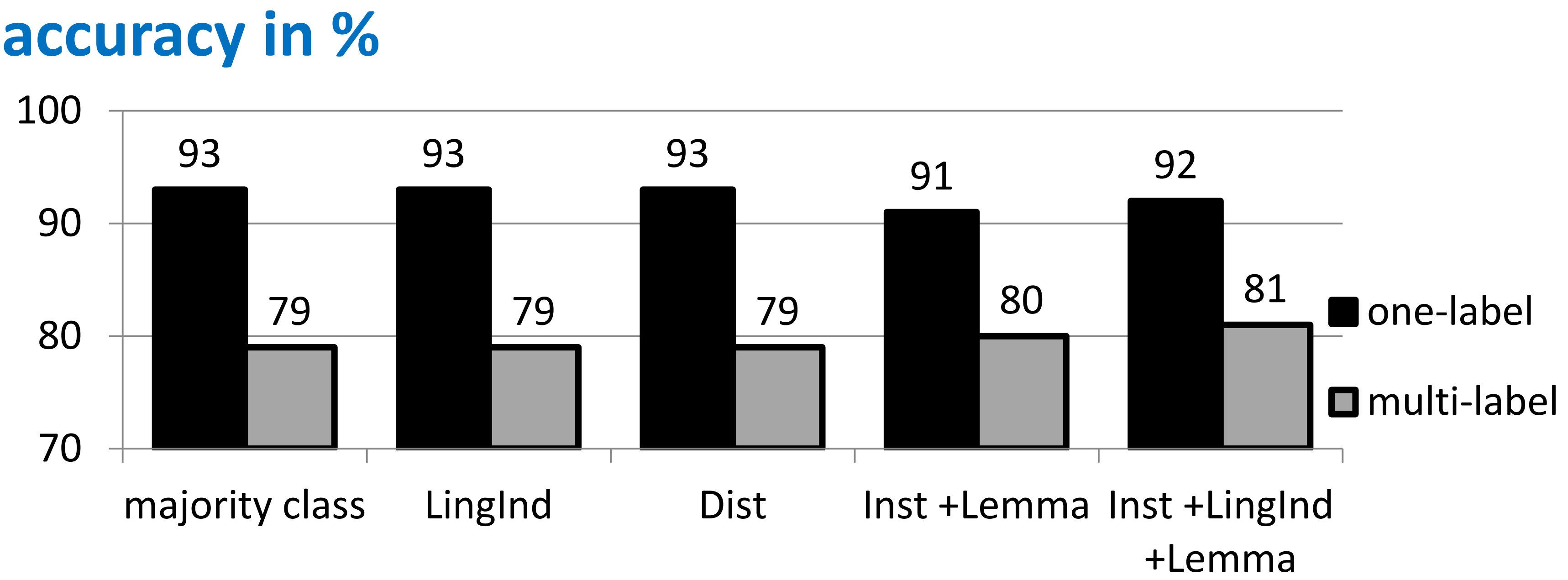
UNSEEN verbs:

no labeled training data available



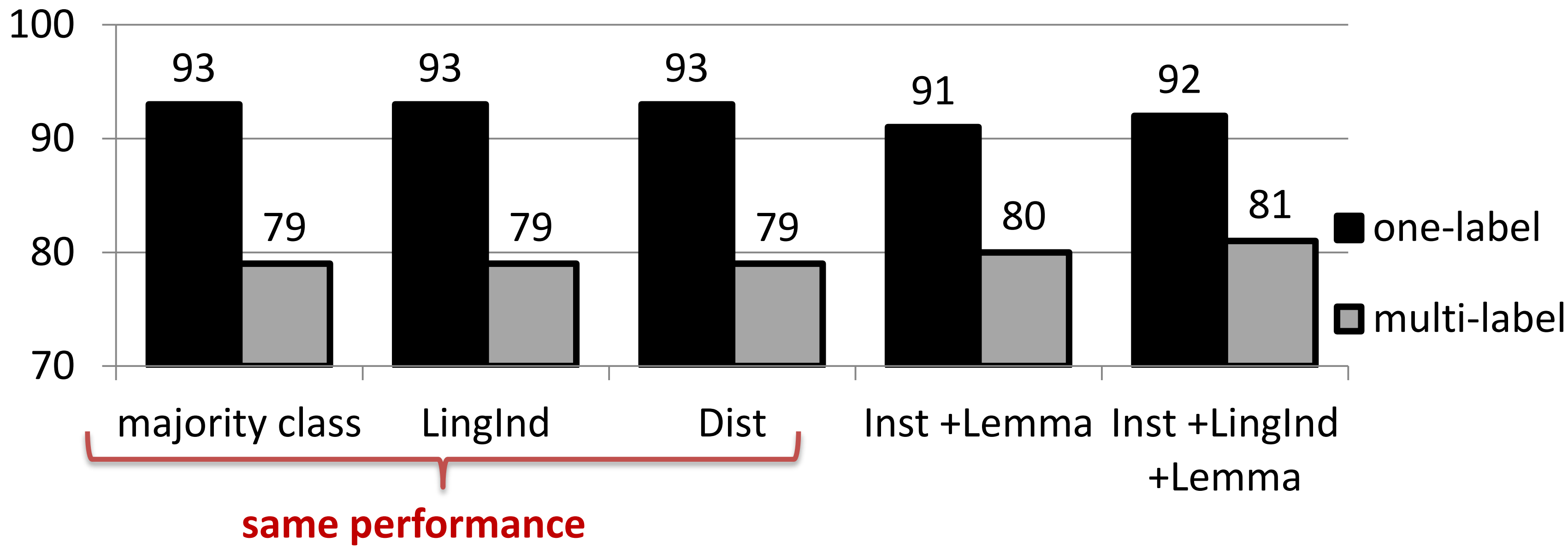
Type-based features generalize across verb types.

Experiment 3: ONE-LABEL vs. MULTI-LABEL verbs



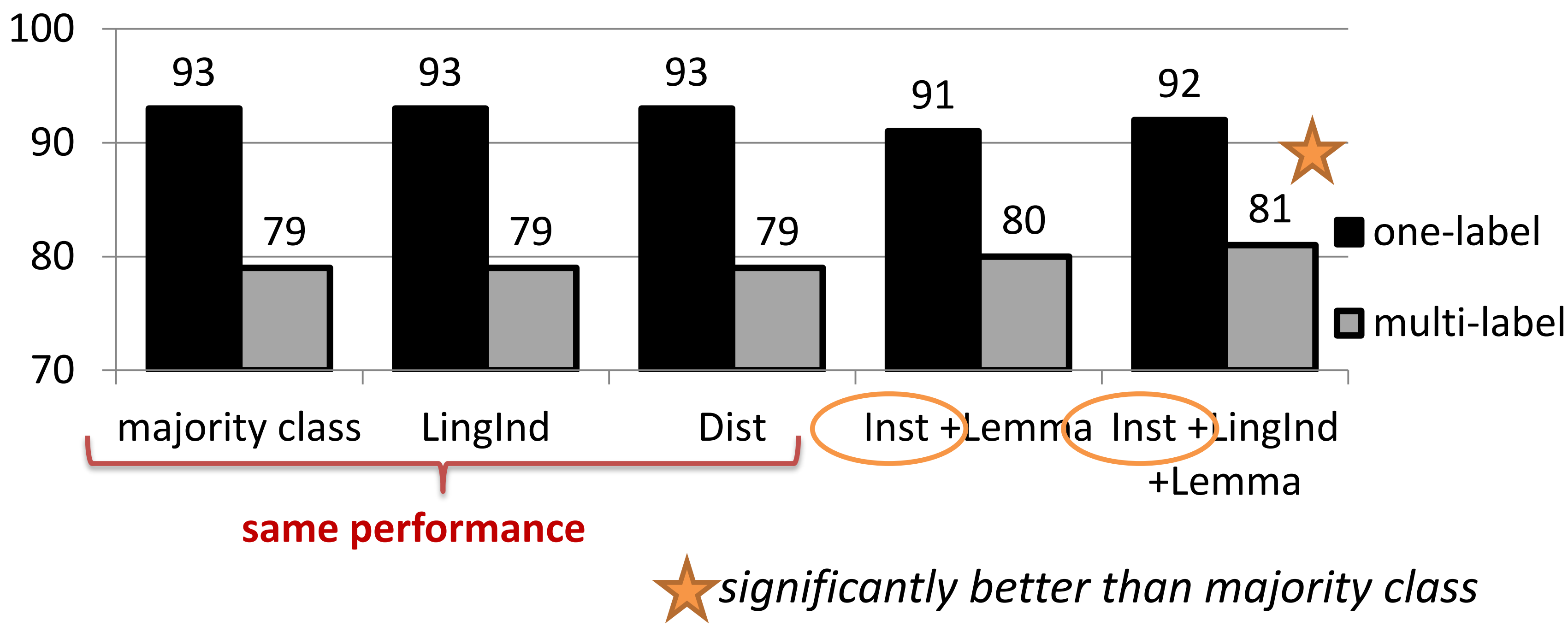
Experiment 3:
ONE-LABEL vs. MULTI-LABEL verbs

accuracy in %



Experiment 3: ONE-LABEL vs. MULTI-LABEL verbs

accuracy in %



Instance-based features are essential for classifying ambiguous verbs.

Experiment 4: INSTANCE-BASED classification

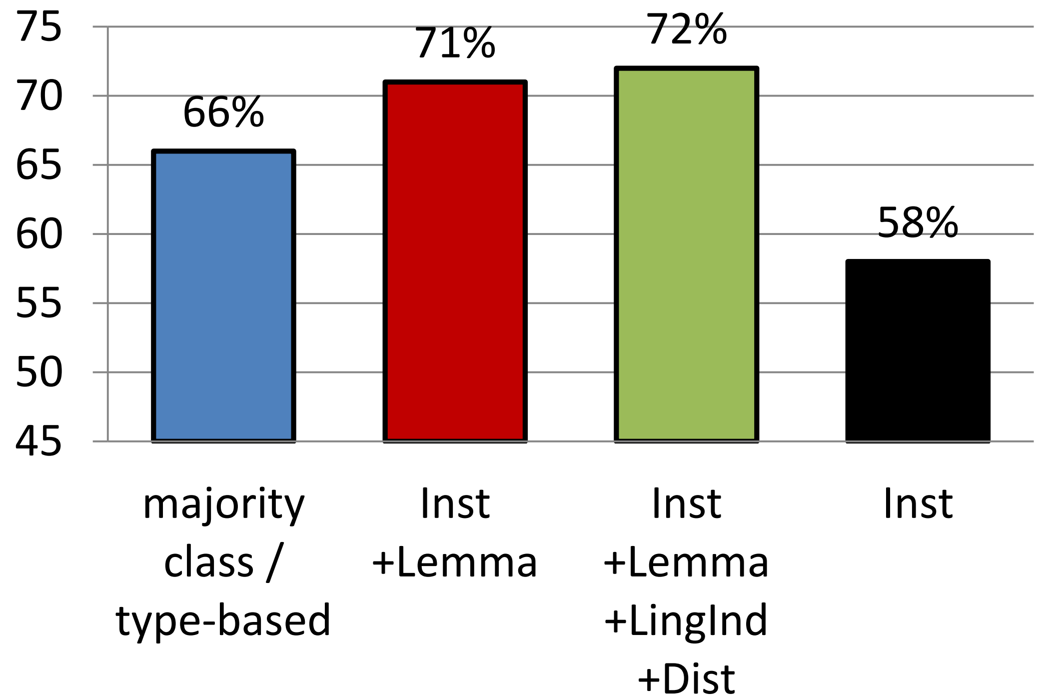
Asp-Ambig:

- 2667 sentences for 20 frequent ambiguous verbs (from Brown)
- 2 annotators, $\kappa = 0.6$

Experiment 4: INSTANCE-BASED classification

Asp-Ambig:
- 2667 sentences for 20 frequent ambiguous verbs (from Brown)
- 2 annotators, κ = 0.6

Asp-Ambig: micro-average accuracy

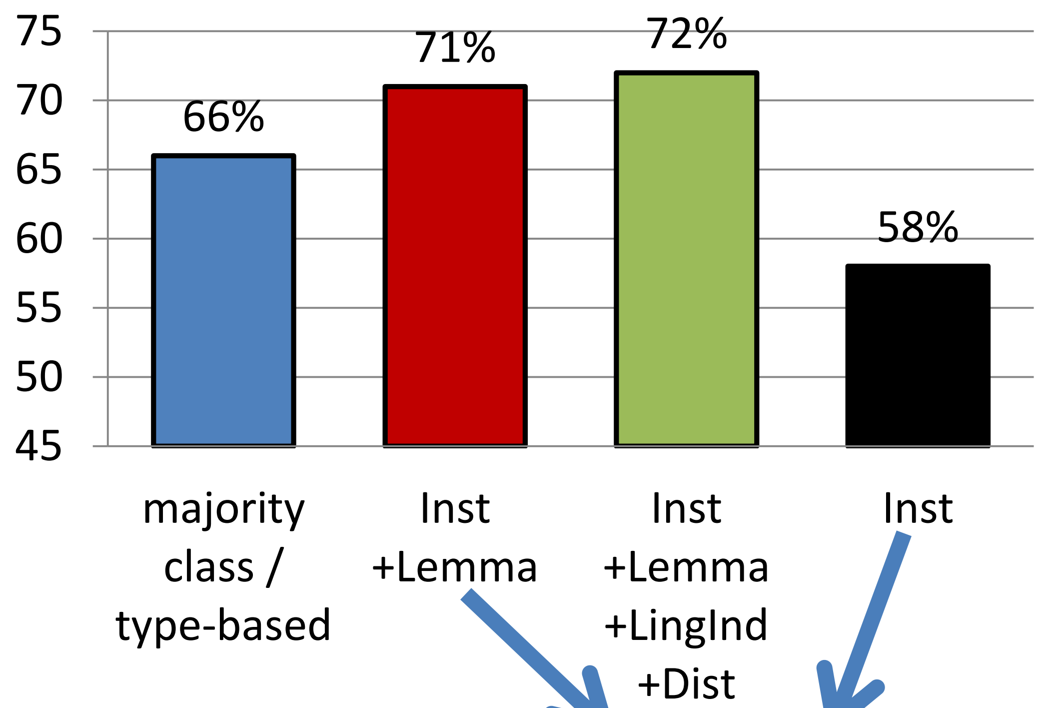


Experiment 4: INSTANCE-BASED classification

Asp-Ambig:

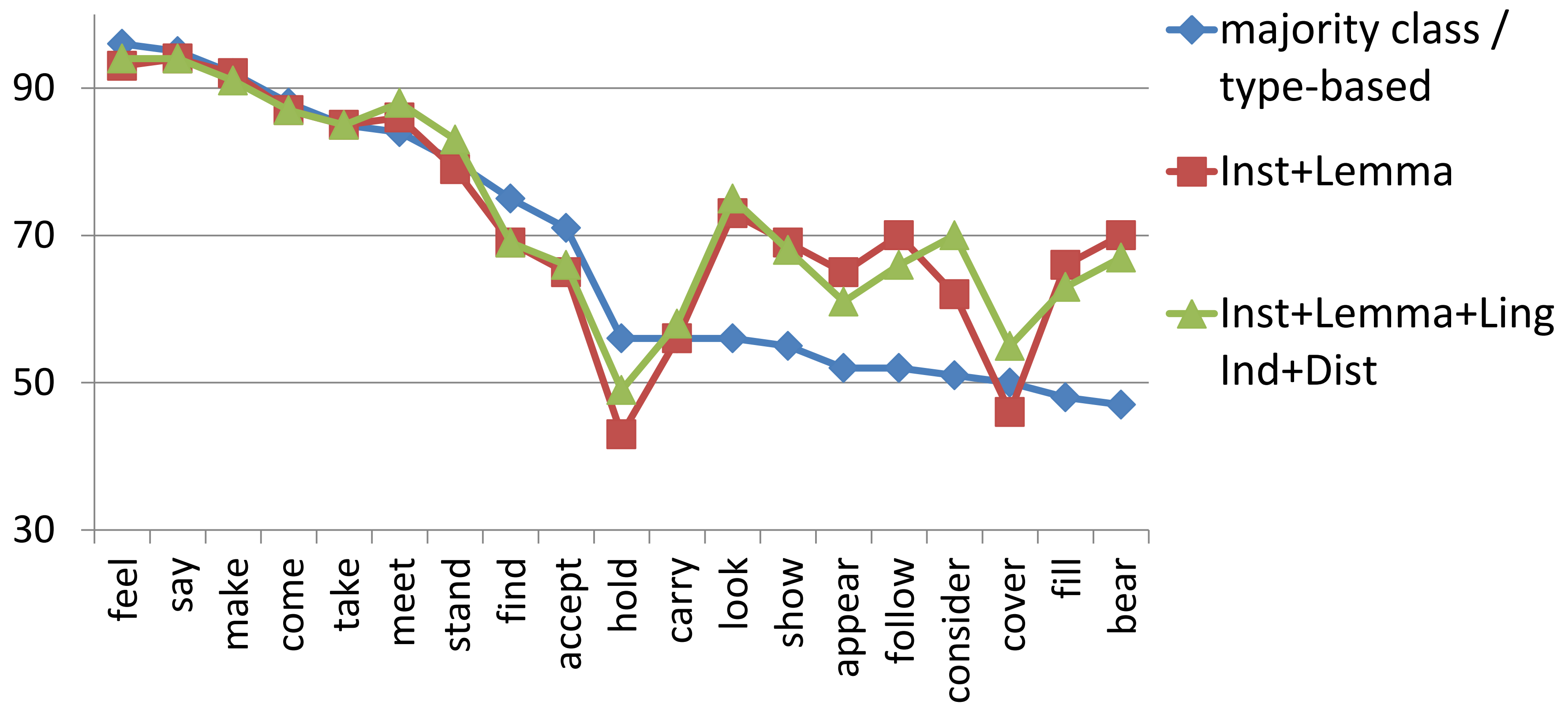
- 2667 sentences for 20 frequent ambiguous verbs (from Brown)
- 2 annotators, κ = 0.6

Asp-Ambig: micro-average accuracy

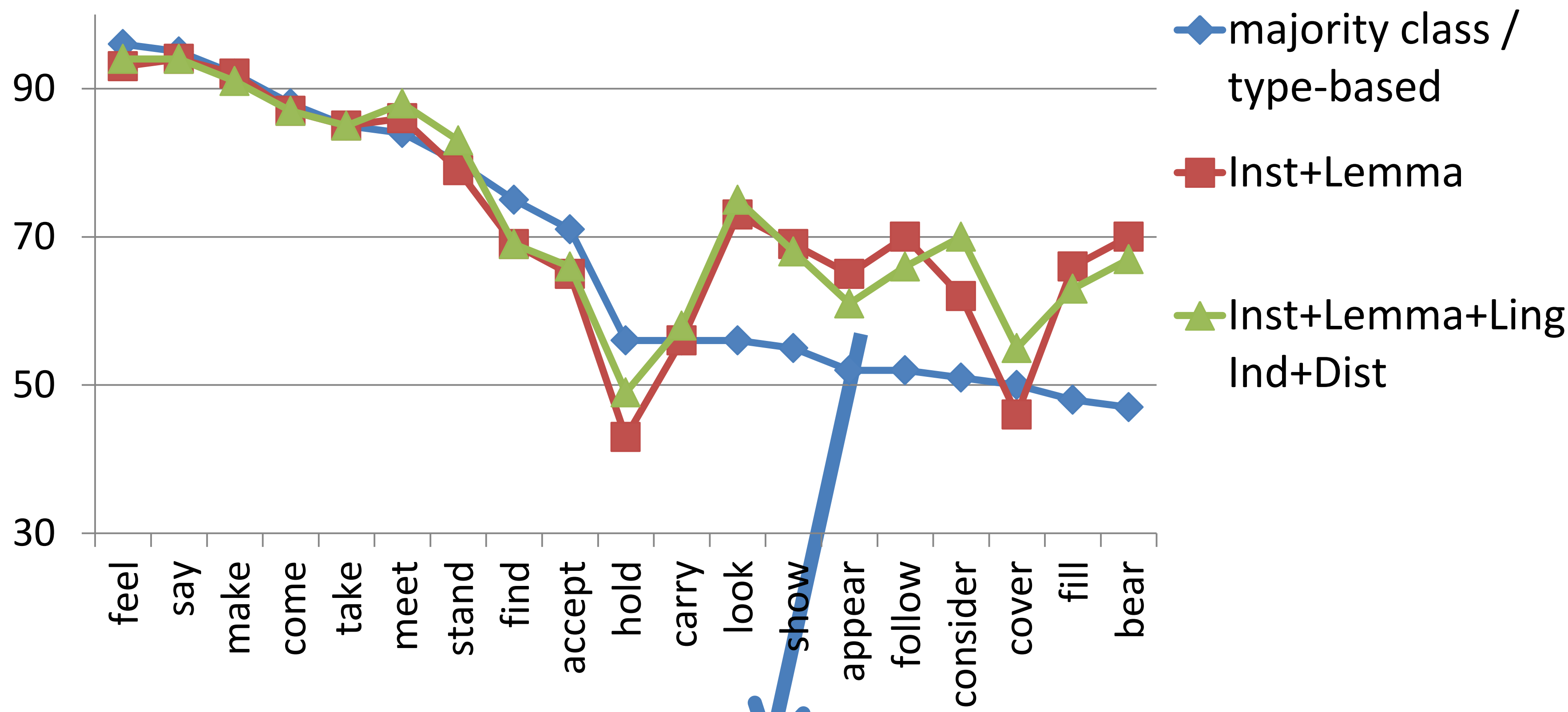


Instance-based features do not generalize across verb types.

Experiment 4: INSTANCE-BASED classification



Experiment 4: INSTANCE-BASED classification



→ The more ambiguous the verb, the more essential are instance-based features.

→ Type-based features (bias) helpful?

→ depends on verb type

Summary:
Automatic prediction of aspectual class of verbs in context

- if **no labeled training data** is available, can make type-based prediction with high accuracy.

Summary:
Automatic prediction of aspectual class of verbs in context

- if **no labeled training data** is available, can make type-based prediction with high accuracy.
- for **ambiguous verbs**: need training data & context-based features.

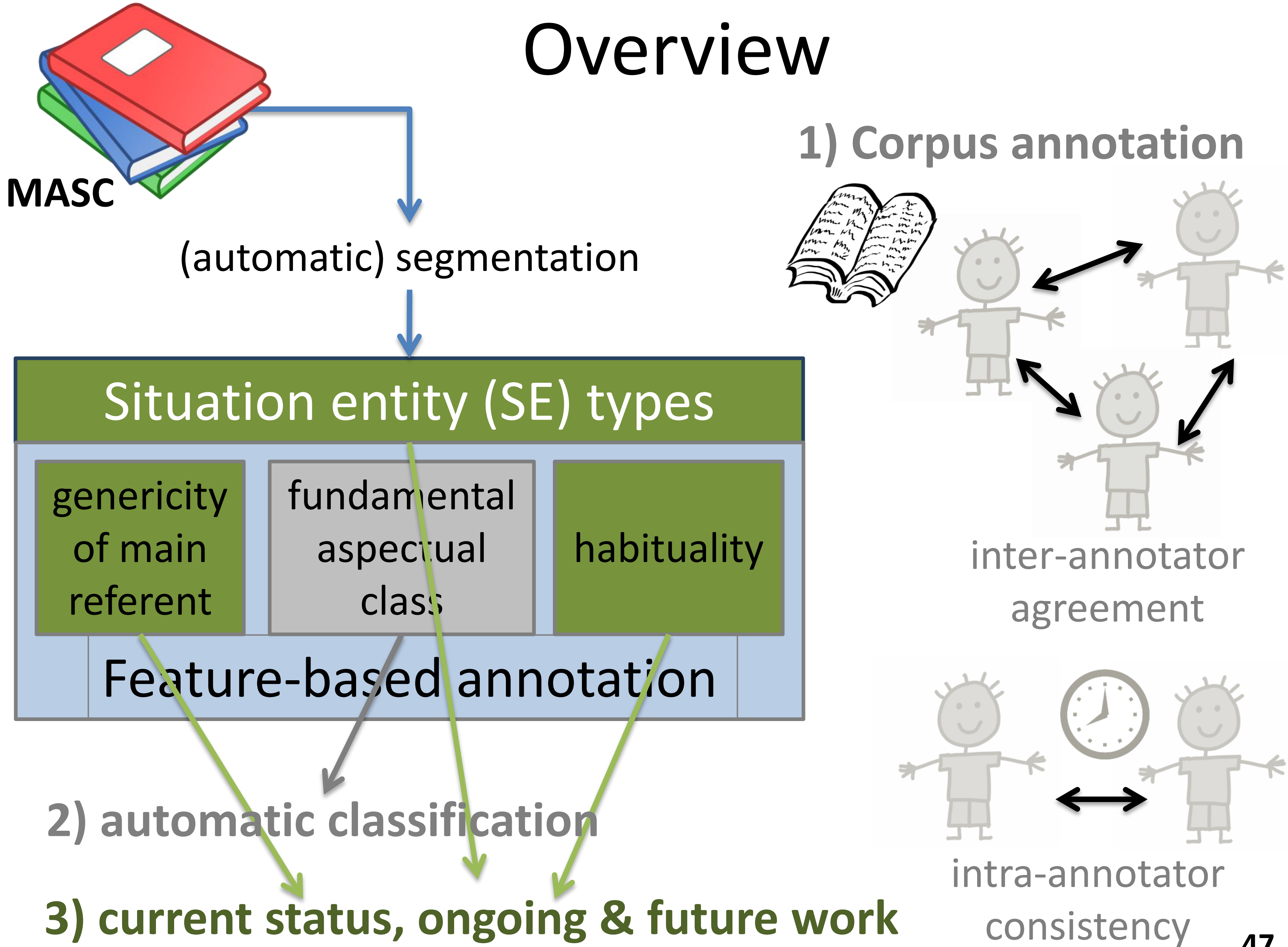
Summary:
Automatic prediction of aspectual class of verbs in context

- if **no labeled training data** is available, can make type-based prediction with high accuracy.
- for **ambiguous verbs**: need training data & context-based features.

treat different verb types differently

globally well-performing system

Overview


MASC
(automatic) segmentation
Situation entity (SE) types
genericity of main referent
fundamental aspectual class
habituality
Feature-based annotation
1) Corpus annotation
inter-annotator agreement
intra-annotator consistency
2) automatic classification
3) current status, ongoing & future work

Annotation status

Plan: gold standard via majority vote

→ label all clauses twice, have third annotator give annotations for disagreed segments (without seeing the other annotator's markup)

Annotation status

Plain: gold standard via majority vote

→ label all clauses twice, have third annotator give annotations for disagreed segments (without seeing the other annotator's markup)

corpus		# segments	2x	3x
MASC	news	3382	done	done
	essays	3357	done	done
	letters	2757	done	in progress
	jokes	4414	done	in progress
	fiction	5560	in progress	in progress
	journal	2581	in progress	in progress
	travel guides	4414	done	in progress
Wikipedia		8266	done	in progress

additional planned MASC sections: email (part), blog, non-fiction, technical

Future / Ongoing work: Automatic classification

- of **habituality**
- of the main referent's **genericity**
- of the clause's **situation entity type**

Future / Ongoing work: Automatic classification

- of **habituality**
- of the main referent's **genericity**
- of the clause's **situation entity type**

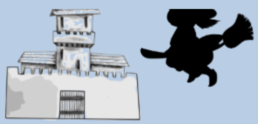
approach: combination of local features with discourse-based features

- extending upon Palmer et al. (2007)

Relevance of discourse modes

[Smith 2003]

NARRATIVE



EVENT,
STATE

REPORT

EVENT, STATE,
general statives

DESCRIPTION

EVENT,
STATE,
ongoing
EVENT

general statives

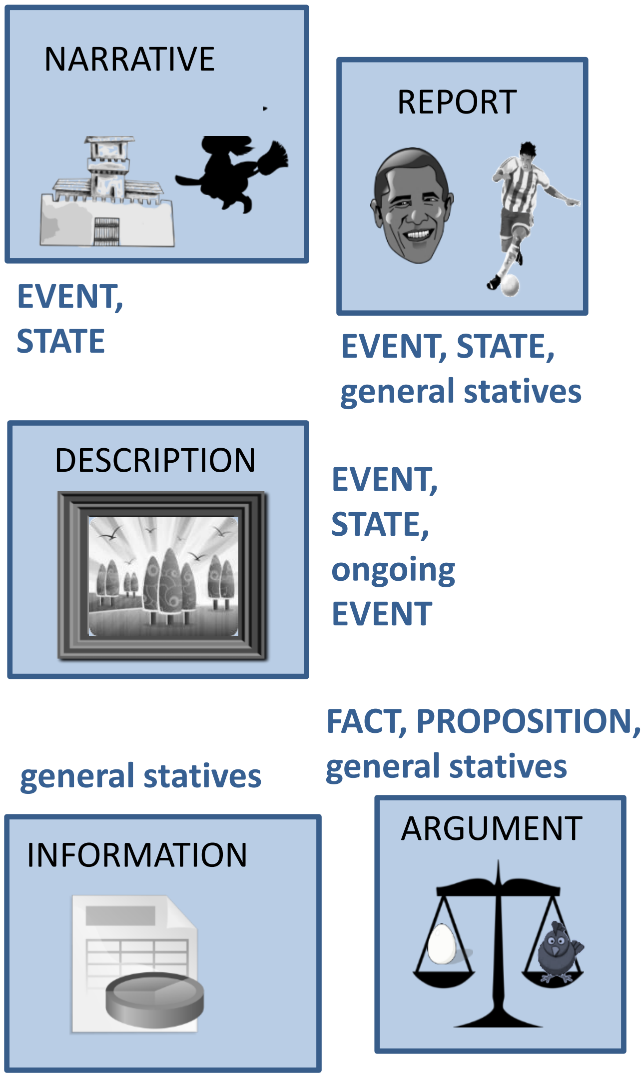
INFORMATION

FACT, PROPOSITION,
general statives

ARGUMENT

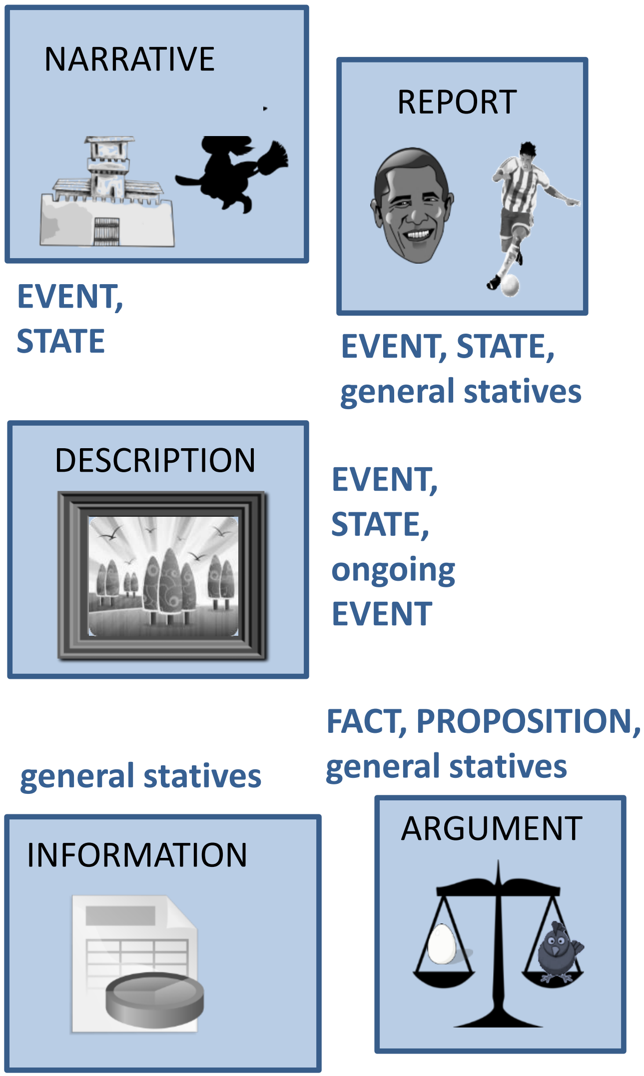
- **future work**: create **annotated corpus** for discourse modes

Relevance of discourse modes
[Smith 2003]



- **future work**: create **annotated corpus** for discourse modes
- **automatic classification** of discourse modes (using SE types & other features)

Relevance of discourse modes
[Smith 2003]



- **future work**: create **annotated corpus** for discourse modes
- **automatic classification** of discourse modes (using SE types & other features)
- 'applications'
  - temporal processing of discourse
  - genre, stylistics
  - machine translation
  - argumentation mining

Future / ongoing work

Aspectual class of light verbs

have a heart attack vs. *have a daughter*
make sense vs. *make a cake*

frequent & ambigous verbs, object matters
→need a good solution to improve overall performance
→does distributional information help?

Future / ongoing work

situation entity types

aspectual information

how speaker/writer presents a situation

use of SEs in different languages? relationships?

Future / ongoing work

situation entity types

aspectual information

how speaker/writer presents a situation

use of SEs in different languages? relationships?

MT evaluation

Can we use SE type information for evaluating translation quality? (start with related languages)

Future / ongoing work

situation entity types

aspectual information

how speaker/writer presents a situation

→ **use of SEs in different languages? relationships?**

MT evaluation

Can we use SE type information for evaluating translation quality? (start with related languages)

Situation entities in 汉语

aspectual information leads to default interpretations of time in Chinese

[Smith & Erbaugh 2005]

→ inferring temporal information

[Zhang & Xue 2014]

Future / ongoing work

situation entity types
aspectual information
how speaker/writer presents a situation

→ **use of SEs in different languages? relationships?**

MT evaluation

Can we use SE type information for evaluating translation quality?
(start with related languages)

Situation entities in 汉语

aspectual information leads to default interpretations of time in Chinese

[Smith & Erbaugh 2005]

→ inferring temporal information

[Zhang & Xue 2014]

→ develop annotation scheme

→ compare use of SE types / features vs. English http://sitent.coli.uni-saarland.de

Thanks to

Manfred Pinkal

Bonnie Webber

Andreas Peldzsus

Melissa Peate Sorensen

Ambika Kirkland

Ruth Kühn

Fernando Ardente

Christine Bocionek

We are

to hear your suggestions or ideas for collaborations.

References

Beata Beigman Klebanov and Eyal Beigman. 2009. **From annotator agreement to noise models.** Computational Linguistics, 35(4):495–503.

Steven Bethard. 2013. **ClearTK-TimeML: A minimalist approach to TempEval 2013**. In Second Joint Conference on Lexical and Computational Semantics (* SEM), volume 2, pages 10–14.

Christelle Cocco. 2012. **Discourse type clustering using pos n-gram profiles and high-dimensional embeddings.** In Proceedings of the Student Research Workshop at the 13th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2012.

Francisco Costa and António Branco. 2012. **Aspectual type and temporal relation classification.** In Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics (EACL), pages 266–275.

Annemarie Friedrich and Alexis Palmer. 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.

Nancy Ide, Christiane Fellbaum, Collin Baker, and Rebecca Passonneau. 2010. **The manually annotated subcorpus: A community resource for and by the people.** In Proceedings of the ACL 2010 conference short papers, pages 68–73.

References (ctd)

Alexis Palmer, Elias Ponvert, Jason Baldridge, and Carlota Smith. 2007. **A sequencing model for situation entity classification.** Proceedings of ACL 2007.

Nils Reiter and Anette Frank. 2010. **Identifying generic noun phrases.** In Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL).

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.

Carlota S Smith. 2003. **Modes of discourse: The local structure of texts.** 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-Volume 1, pages 149–156. Association for Computational Linguistics.

Naushad UzZaman, Hector Llorens, Leon Derczynski, Marc Verhagen, James Allen, and James Pustejovsky. 2013. Semeval-2013 task 1: **Tempeval-3: Evaluating time expressions, events, and temporaectual classification and revealing linguistic insights.** Computational Linguistics, 26(4):595–628.

References (ctd)

Alexis Palmer and Annemarie Friedrich. 2014. **Genre distinctions and discourse modes: Text types differ in their situation type distributions.** Frontiers and Connections between Argumentation Theory and Natural Language Processing. Bertinoro, Italy. *(to appear)*

Carlson, Gregory N., and Francis Jeffry Pelletier, eds. 1995. **The generic book**. University of Chicago Press.

Krifka, Manfred, Francis Jeffry Pelletier, Gregory Carlson, Alice Ter Meulen, Gennaro Chierchia, and Godehard Link. 1995**. Genericity: an introduction**. *The generic book* (1995): 1-124.

Carlson, Gregory N. "**Truth conditions of generic sentences: Two contrasting views**." *The generic book* (1995): 224-237.

Situation entity annotation

BACKUP SLIDES

Future / Ongoing work: Aspectual class of **light verbs**

- For some frequent & ambigous verbs, the object matters → need a good solution to improve overall performance
  - *have a heart attack* ←→ *have a daughter*
  - *make sense* ←→ *make a cake*
- Idea: using distributional information
- M.Sc. thesis 2015 (Liesa Heuschkel)

Future / Ongoing work: Eventuality information in **Chinese**

- boundedness of events/states leads to default interpretations of time in Chinese (Smith & Erbaugh 2005)
- potentially useful for inferring temporal information (Zhang & Xue 2014)
- Develop annotation scheme, compare use of situation entity types vs. English
- M.Sc. thesis 2015 (Bryan Zhang)

Future / Ongoing work:
Machine Translation evaluation

- situation entity types
  - aspectual information
  - how speaker/writer presents a situation
- **Question 1**: relation between SE types in different languages (e.g. German-English, French-English)
- **Question 2**: can we use SE type information for evaluating translation quality?
- M.Sc. thesis 2015 (Kleo Mavridou)

Segmentation post-processing

- merge situationless segments with appropriate neighboring segment, respecting parse trees. (automatic)

verbless segments

For every dollar
donated to Goodwill in 1998,
we helped our “graduates” earn an estimated $102.

⬇

For every dollar donated to Goodwill in 1998,
we helped our “graduates” earn an estimated $102.

Segmentation post-processing

- merge situationless segments with appropriate neighboring segment, respecting parse trees. (manual)

to-infinitives

She has learned
to make her own money.

⇩

She has learned to make her own money.

Segmentation post-processing

cases requiring manual merging

So the shift in the image of Gates has been an interesting one
for me to watch.

⇩

So the shift in the image of Gates has been an interesting one for me to watch.

- multiple situations per segment

genericity of main referent -- disagreements

- 183 cases, B&C marked *specific*, A marked *generic,* judged by authors:
  - *50 (27.3%) both readings possible*
  - *69 (37.7%) specific*
  - *22 (12%) generic*
  - *36 (19.7%) not sure given the context 5 segmentation problems*
  - *1 (0.5%) expletive*

genericity of main referent -- disagreements

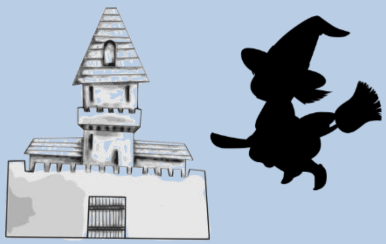
- Cases marked generic by A, which should be specific:
  - simple plurals
    - But **some regulations**, aimed at specific regional problems…
    - **Some of his fellow historians** question …
  - specific concepts
    - **a humanitarian crisis** that has festered since the Gulf War
  - you (either reading possible)
    - if **you** have the will and dedication here, **you** can learn a lot

genericity of main referent -- disagreements

- Cases marked generic by A, which should be generic:
  - Under the plan, **unsecured creditors** would receive about $92 million,…
  - **Chinese hot and sour soup** often includes bitter melon.
  - … in **many regions of the world** that boast moderate to warm climates.

Modes of discourse [Smith 2003]

NARRATIVE



temporal progression

**EVENT,
STATE**

REPORT

temporal progression,
related to speech time

**EVENT, STATE,
general statives**

DESCRIPTION

spatial progression

**EVENT, STATE,
ongoing EVENT**

**general
statives**

INFORMATION

ARGUMENT

**FACT,
PROPOSITION,
general statives**

atemporal, metaphoric progression

Feature: genericity of main referent (inter-annotator agreement)

183 clauses: B &C agree, A disagrees			
92%: B & C →specific, A → generic			
40%: misunder-standing by A	**30%**: multiple readings	**30%**: other	

As a governor, I'll make sure that every kid in New York has the same opportunity.

you in letters → generic or addressee?

annotators with different preferences:
identify ambiguous cases

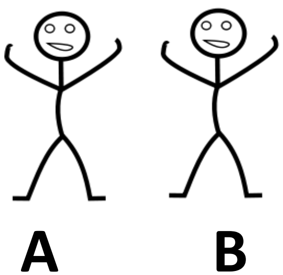
Comparing B and C: 2358 segments : specific by both

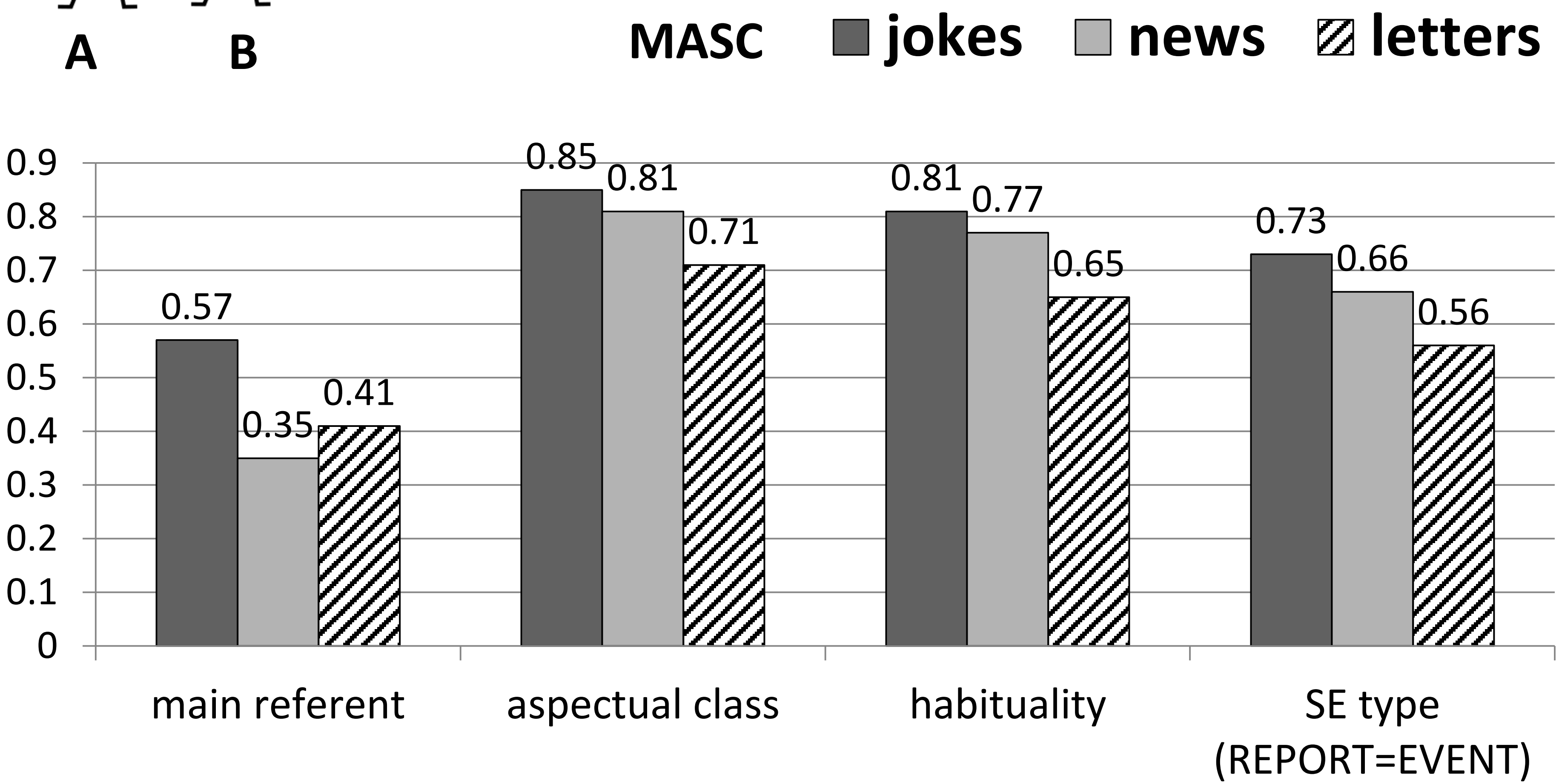
- 122 segments: generic by at least one
- 43 segments: generic by both

→ very few cases, cannot draw conclusions on reasons for low κ yet.

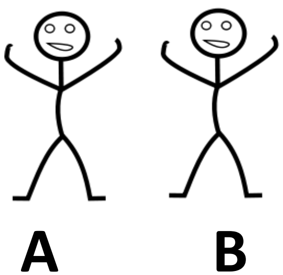
→ follow-up study with data targeting generics in progress

Inter-annotator agreement: genres

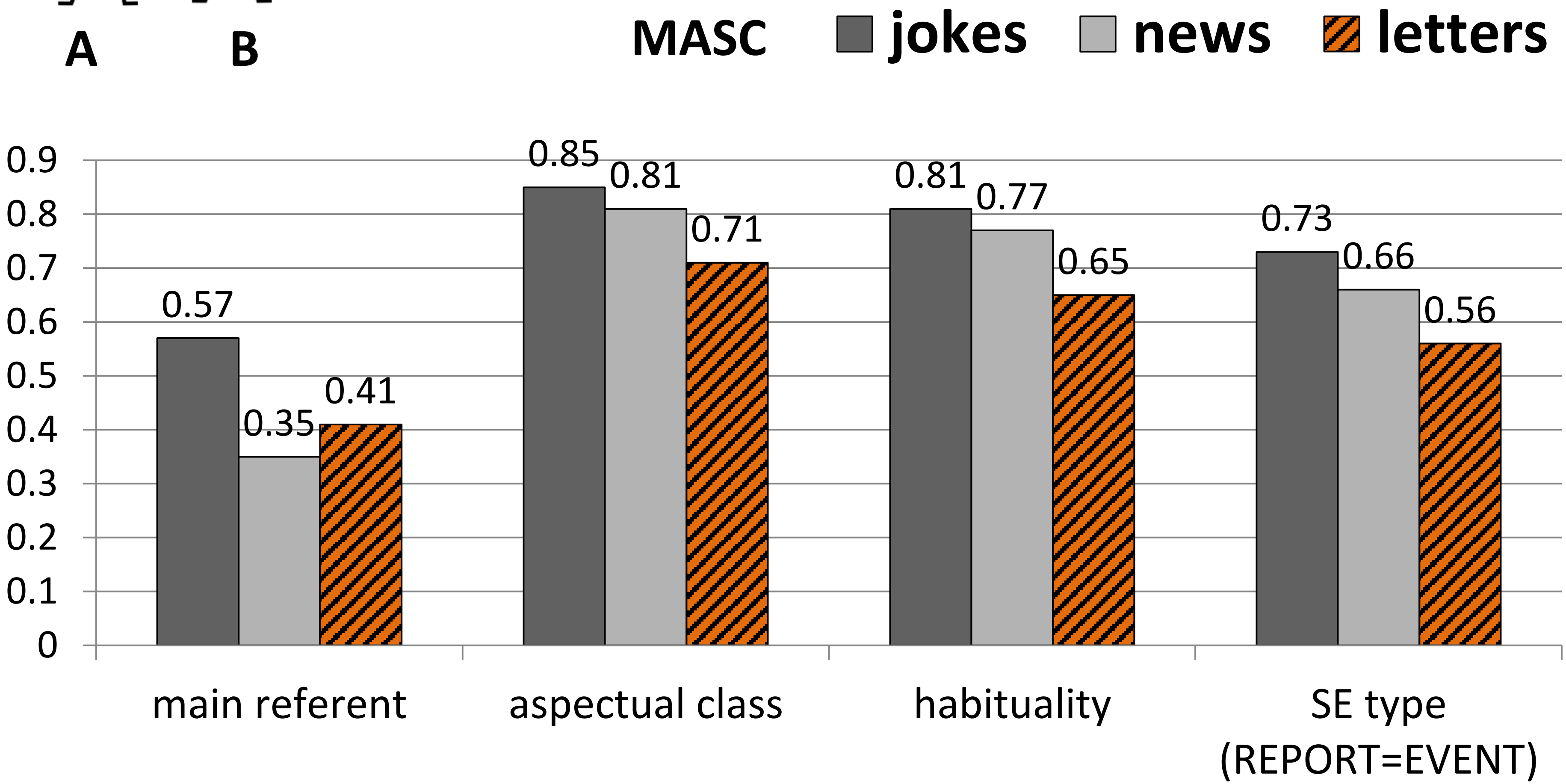




Inter-annotator agreement: genres



Why is agreement lower on letters subsection?



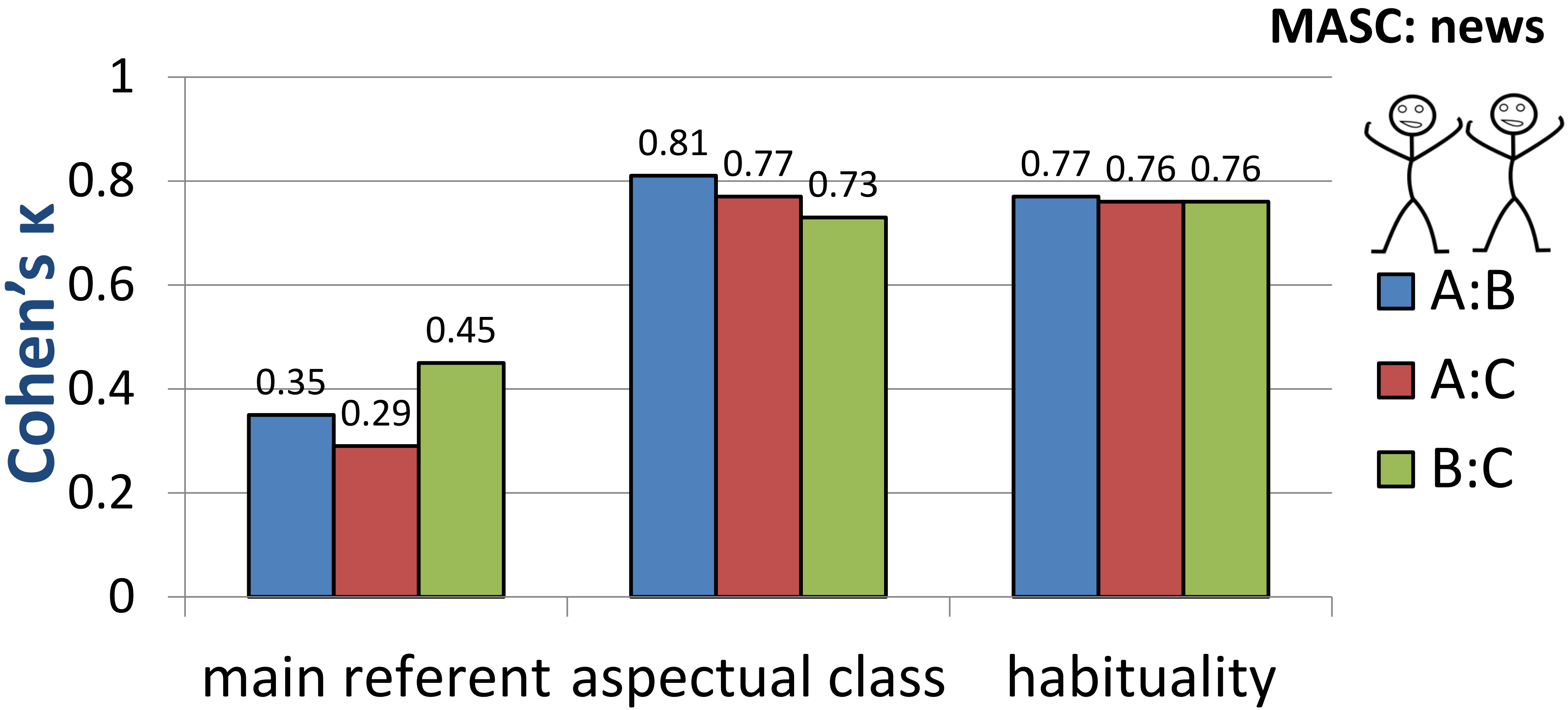
joint work with Melissa Peate Sorensen

Generics follow-up study

- address the issue of ***sparsity***: enrich corpus with documents where we expect a higher proportion of general statives
- **Wikipedia documents:** agreement study
  - (corpus total: about 8200 situations)

category	# situations	Fleiss' κ, main referent	Fleiss' κ, SE type
animals	160	0.65	0.71
games	159	0.71	0.55
gangs	255	0.23	0.58
history	485	0.10	0.57
sports	508	0.62	0.67
tribes	167	0.45	0.57
average/total	1808	0.64	0.67

Features: inter-annotator agreement



main referent	aspectual class	habituality
specific	*stative*	*episodic*
generic	*dynamic*	*habitual*
expletive	*both*	*static*

joint work with Melissa Peate Sorensen

Generics follow-up study

address the issue of ***clarity***: compared definition to existing theories & corpora, clarified definition in manual, added examples.

Generic noun phrases (theory applied to subjects):
(compare to Krifka et al. 1995: "The Generic Book")

kind-referring: *The lion disappeared from Asia.*
nonspecific, referring to arbitrary member of kind:
A lion roars when it smells food.

ACE-2005:

GEN ≈ generic
SPC ≈ non-generic
NEG → negated
No lawyer would...

USP = underspecified ≈ non-generic
non-generic nonspecific reference:
Many people will come.
mention of entities whose identity would be hard to locate:
Officials said ...

→ difficult to annotate, especially for non-subjects (story is different, see Krifka et al. 1995)
→ only 1567 GEN subjects.

Wikipedia documents: agreement

- WikiGen corpus: 49 documents (≈ 8200 situations)
- agreement study: 14 documents (≈1800 situations)

annotators	main referent	aspectual class	habituality	SE type
A1, A2	0.61	0.65	0.65	0.68
A1, A3	0.60	0.63	0.64	0.65
A2, A3	0.70	0.68	0.60	0.69
all, Fleiss' κ	0.64	0.66	0.63	0.67

Future / ongoing work

Aspectual class of light verbs

have a heart attack vs. *have a daughter*

make sense vs. *make a cake*

frequent & ambigous verbs, object matters

→ need a good solution to improve overall performance

→ does distributional information help?

situation entity types

aspectual information

how speaker/writer presents a situation

→ **use of SEs in different languages? relationships?**

Situation entities in 汉语

aspectual information leads to default interpretations of time in Chinese [Smith & Erbaugh 2005]

→ inferring temporal information [Zhang & Xue 2014]

→ develop annotation scheme

→ compare use of situation entity types vs. English

MT evaluation

Can we use SE type information for evaluating translation quality?

(start with related languages)