CLUSTER OF EXCELLENCE

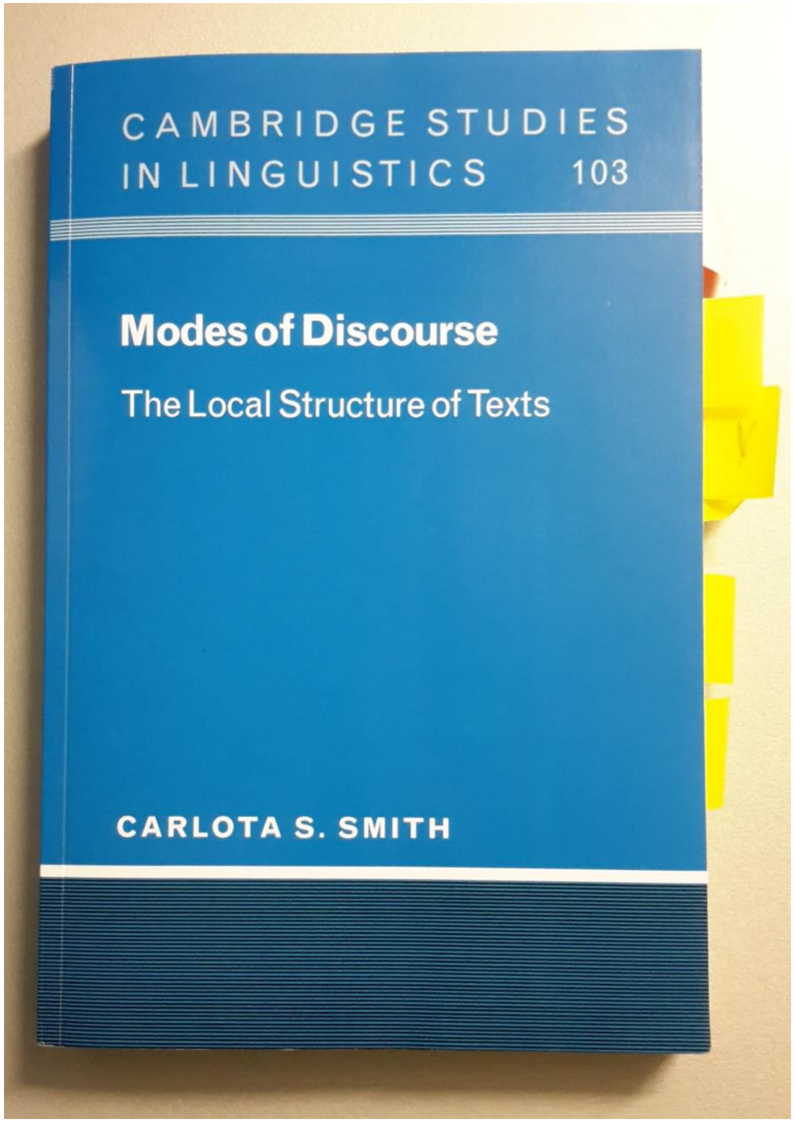
UNIVERSITÄT DES SAARLANDES

Annotation and automatic classification of situation entity types

Kolloquium Übersetzungswissenschaft -- Saarbrücken, January 2016

Annemarie Friedrich, Saarland University
joint work with Alexis Palmer and Manfred Pinkal

Carlota Smith: Modes of discourse (2003)



Thanks!

Alexis Palmer

Manfred Pinkal

Melissa Peate Sorensen

Liesa Heuschkel

Kleio-Isidora Mavridou

Christine Bocionek

Fernando Ardente

Ruth Kühn

Ambika Kirkland

Damyana Gateva

Discourse modes

Prof. Dr. Origin at Saarland University came into his office one morning and was very surprised by the results of an experiment he had started the day before. He called in his assistants to inspect the hen and the egg that were the subject of his experiments…

The chicken or the egg causality dilemma is commonly stated as "which came first, the chicken or the egg?" To ancient philosophers, the question about the first chicken or egg also evoked the questions of how life and the universe in general began. …

In my opinion, the results of Prof. Dr. Origin's group are highly interesting, but they do by no means solve the philosophical question of how life and the universe began. I believe that much more research is needed, and that the field of biology alone will not be able to answer this question.

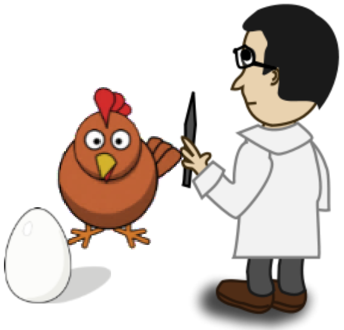
Discourse modes

Prof. Dr. Origin at Saarland University came into his office one morning and was very surprised by the results of an experiment he had started the day before. He called in his assistants to inspect the hen and the egg that were the subject of his experiments…

NARRATIVE

The chicken or the egg causality dilemma is commonly stated as "which came first, the chicken or the egg?" To ancient philosophers, the question about the first chicken or egg also evoked the questions of how life and the universe in general began. …

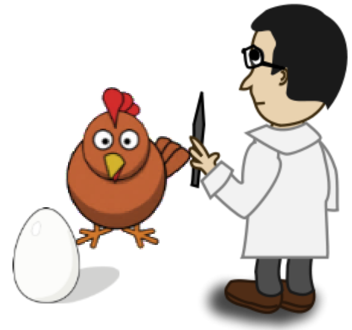
In my opinion, the results of Prof. Dr. Origin’s group are highly interesting, but they do by no means solve the philosophical question of how life and the universe began. I believe that much more research is needed, and that the field of biology alone will not be able to answer this question.



Discourse modes

Prof. Dr. Origin at Saarland University came into his office one morning and was very surprised by the results of an experiment he had started the day before. He called in his assistants to inspect the hen and the eggs that were the subject of his experiments...

NARRATIVE



The chicken or the egg causality dilemma is commonly stated as "which came first, the chicken or the egg?" To ancient philosophers, the question about the first chicken or egg also evoked the questions of how life and the universe in general began. ...

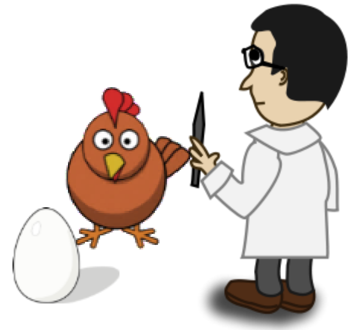
INFORMATION

In my opinion, the results of Prof. Dr. Origin's group are highly interesting, but they do by no means solve the philosophical question of how life and the universe began. I believe that much more research is needed, and that the field of biology alone will not be able to answer this question.

Discourse modes

Prof. Dr. Origin at Saarland University came into his office one morning and was very surprised by the results of an experiment he had started the day before. He called in his assistants to inspect the hen and the egg that were the subject of his experiments…

NARRATIVE

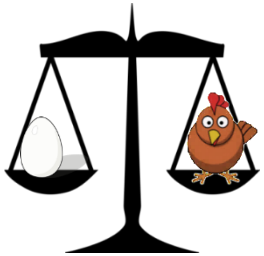


The chicken or the egg causality dilemma is commonly stated as "which came first, the chicken or the egg?" To ancient philosophers, the question about the first chicken or egg also evoked the questions of how life and the universe in general began. …

INFORMATION

In my opinion, the results of Prof. Dr. Origin’s group are highly interesting, but they do by no means solve the philosophical question of how life and the universe began. I believe that much more research is needed, and that the field of biology alone will not be able to answer this question.

ARGUMENT
COMMENTARY



Discourse modes

Prof. Dr. Origin at Saarland University came into his office one morning and was very surprised by the results of an experiment he had started the day before. He called in his assistants to inspect the hen and the egg that were the subject of his experiments...

NARRATIVE

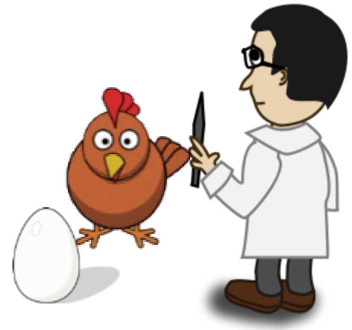
The chicken or the egg causality dilemma is commonly stated as "which came first, the chicken or the egg?" To ancient philosophers, the question about the first chicken or egg also evoked the questions of how life and the universe in general began. ...

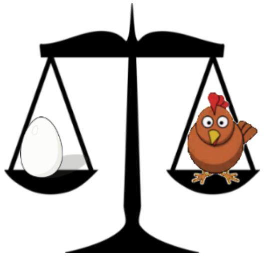
INFORMATION

In my opinion, the results of Prof. Dr. Origin's group are highly interesting, but they do by no means solve the philosophical question of how life and the universe began. I believe that much more research is needed, and that the field of biology alone will not be able to answer this question.

ARGUMENT

COMMENTARY





one text

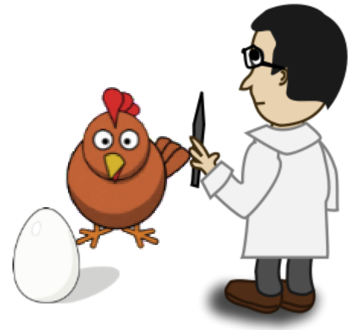
≈ one genre

one passage

≈ one discourse mode

Discourse modes & situation entity types

Prof. Dr. Origin at Saarland University <u>came into his office</u> one morning and <u>was very surprised</u> by the results of an experiment he <u>had started</u> the day before. He <u>called</u> in his assistants to inspect the hen and the egg that <u>were the subject</u> of his experiments…

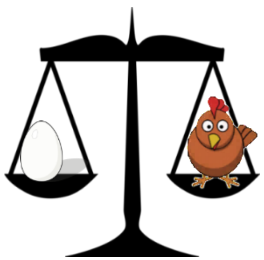


NARRATIVE

<u>The chicken or the egg causality dilemma is commonly stated as</u> "which came first, the chicken or the egg?" <u>To ancient philosophers, the question about</u> the first chicken or egg also <u>evoked the questions</u> of how life and the universe in general began. …

INFORMATION

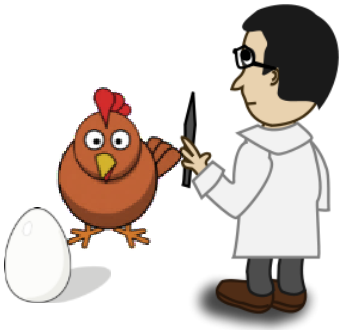
In my opinion, the results of Prof. Dr. Origin's group <u>are highly interesting</u>, but they <u>do by no means solve</u> the philosophical question of how life and the universe began. <u>I believe</u> that <u>much more research is needed</u>, and that the field of biology alone <u>will not be able to answer</u> this question.



ARGUMENT
COMMENTARY

Discourse modes & situation entity types

Prof. Dr. Origin at Saarland University came into his office one morning and was very surprised by the results of an experiment he had started the day before. He called in his assistants to inspect the hen and the egg that were the subject of his experiments…



NARRATIVE

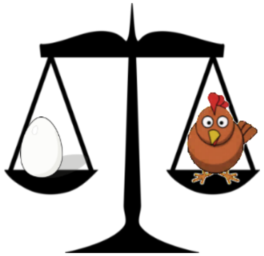
STATE

EVENT

The chicken or the egg causality dilemma is commonly stated as "which came first, the chicken or the egg?" To ancient philosophers, the question about the first chicken or egg also evoked the questions of how life and the universe in general began. …

INFORMATION

In my opinion, the results of Prof. Dr. Origin's group are highly interesting, but they do by no means solve the philosophical question of how life and the universe began. I believe that much more research is needed, and that the field of biology alone will not be able to answer this question.



ARGUMENT

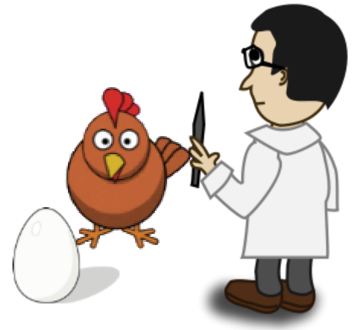
COMMENTARY

Discourse modes & situation entity types

Prof. Dr. Origin at Saarland University came into his office one morning and was very surprised by the results of an experiment he had started the day before. He called in his assistants to inspect the hen and the egg that were the subject of his experiments…

The chicken or the egg causality dilemma is commonly stated as "which came first, the chicken or the egg?" To ancient philosophers, the question about the first chicken or egg also evoked the questions of how life and the universe in general began. …

In my opinion, the results of Prof. Dr. Origin's group are highly interesting, but they do by no means solve the philosophical question of how life and the universe began. I believe that much more research is needed, and that the field of biology alone will not be able to answer this question.



NARRATIVE

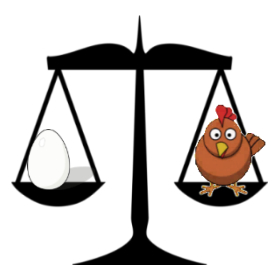
STATE

EVENT

INFORMATION

GENERIC SENTENCE

GENERALIZING SENTENCE

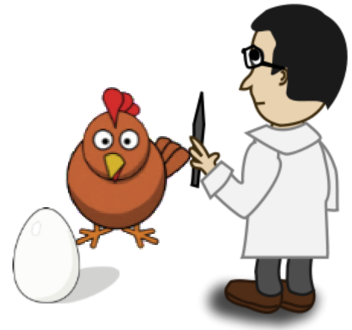


ARGUMENT

COMMENTARY

Discourse modes & situation entity types

Prof. Dr. Origin at Saarland University came into his office one morning and was very surprised by the results of an experiment he had started the day before. He called in his assistants to inspect the hen and the egg that were the subject of his experiments…



NARRATIVE

State

Event

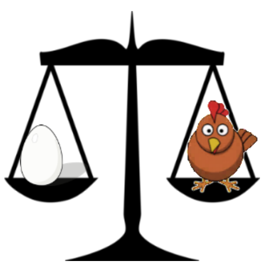
The chicken or the egg causality dilemma is commonly stated as "which came first, the chicken or the egg?" To ancient philosophers, the question about the first chicken or egg also evoked the questions of how life and the universe in general began. …

INFORMATION

Generic Sentence

Generalizing Sentence

In my opinion, the results of Prof. Dr. Origin's group are highly interesting, but they do by no means solve the philosophical question of how life and the universe began. I believe that much more research is needed, and that the field of biology alone will not be able to answer this question.



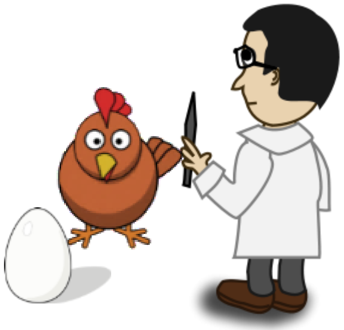
ARGUMENT

COMMENTARY

State, Event, Abstract Entities, Generic / Generalizing Sentences

Discourse modes & type of progression

Prof. Dr. Origin at Saarland University came into his office one morning and was very surprised by the results of an experiment he had started the day before. He called in his assistants to inspect the hen and the egg that were the subject of his experiments…

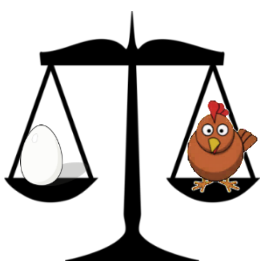


NARRATIVE

The chicken or the egg causality dilemma is commonly stated as "which came first, the chicken or the egg?" To ancient philosophers, the question about the first chicken or egg also evoked the questions of how life and the universe in general began. …

INFORMATION

In my opinion, the results of Prof. Dr. Origin's group are highly interesting, but they do by no means solve the philosophical question of how life and the universe began. I believe that much more research is needed, and that the field of biology alone will not be able to answer this question.



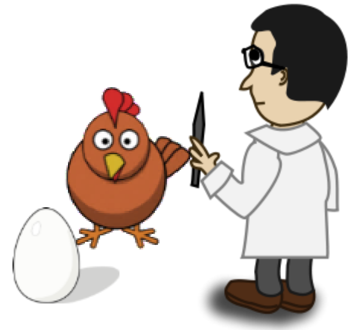
ARGUMENT
COMMENTARY

Discourse modes & type of progression

Prof. Dr. Origin at Saarland University came into his office one morning and was very surprised by the results of an experiment he had started the day before. He called in his assistants to inspect the hen and the egg that were the subject of his experiments…

The chicken or the egg causality dilemma is commonly stated as "which came first, the chicken or the egg?" To ancient philosophers, the question about the first chicken or egg also evoked the questions of how life and the universe in general began. …

In my opinion, the results of Prof. Dr. Origin's group are highly interesting, but they do by no means solve the philosophical question of how life and the universe began. I believe that much more research is needed, and that the field of biology alone will not be able to answer this question.

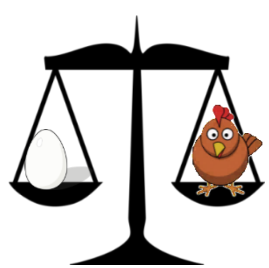


NARRATIVE

temporal

situations related to one another

INFORMATION



ARGUMENT

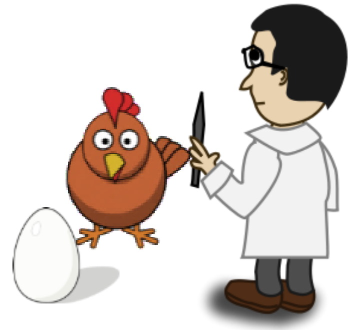
COMMENTARY

Discourse modes & type of progression

Prof. Dr. Origin at Saarland University came into his office one morning and was very surprised by the results of an experiment he had started the day before. He called in his assistants to inspect the hen and the egg that were the subject of his experiments…

The chicken or the egg causality dilemma is commonly stated as "which came first, the chicken or the egg?" To ancient philosophers, the question about the first chicken or egg also evoked the questions of how life and the universe in general began. …

In my opinion, the results of Prof. Dr. Origin's group are highly interesting, but they do by no means solve the philosophical question of how life and the universe began. I believe that much more research is needed, and that the field of biology alone will not be able to answer this question.

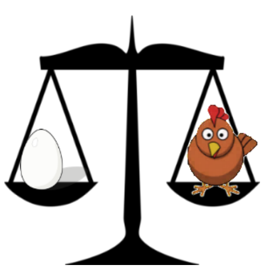


NARRATIVE

temporal
situations related to one another

INFORMATION

metaphorical
through domain



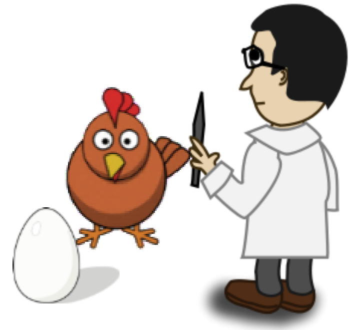
ARGUMENT
COMMENTARY

Discourse modes & type of progression

Prof. Dr. Origin at Saarland University came into his office one morning and was very surprised by the results of an experiment he had started the day before. He called in his assistants to inspect the hen and the egg that were the subject of his experiments…

The chicken or the egg causality dilemma is commonly stated as "which came first, the chicken or the egg?" To ancient philosophers, the question about the first chicken or egg also evoked the questions of how life and the universe in general began. …

In my opinion, the results of Prof. Dr. Origin’s group are highly interesting, but they do by no means solve the philosophical question of how life and the universe began. I believe that much more research is needed, and that the field of biology alone will not be able to answer this question.

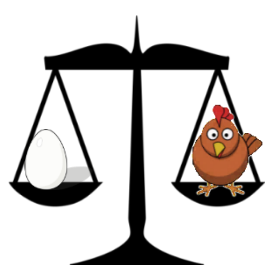


NARRATIVE

temporal
situations related to one another

INFORMATION

metaphorical
through domain

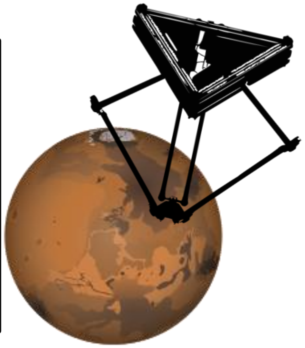


ARGUMENT
COMMENTARY

metaphorical

Additional discourse modes [Smith 2003]

On Monday, NASA **announced** that signs of liquid water **have been found** on Mars. The Mars Reconnaissance Orbiter spacecraft **found** evidence of the liquid on the Martian surface, in long dark spots on the Red Planet thought to be formed because of water flow.



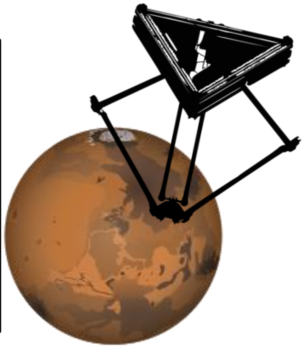
REPORT

STATE, EVENT

temporal progression related to speech time.

Additional discourse modes [Smith 2003]

On Monday, NASA **announced** that signs of liquid water **have been found** on Mars. The Mars Reconnaissance Orbiter spacecraft **found** evidence of the liquid on the Martian surface, in long dark spots on the Red Planet thought to be formed because of water flow.



REPORT

STATE, EVENT

temporal progression related to speech time.

The sand-hills here run down to the sea, and end in two spits of rock jutting out opposite each other, till you lose sight of them in the water. One is called the North Spit, and one the South.

DESCRIPTION

STATE, on-going EVENT

metaphorical progression through scene

Discourse modes: related theories

Discourse modes: related theories

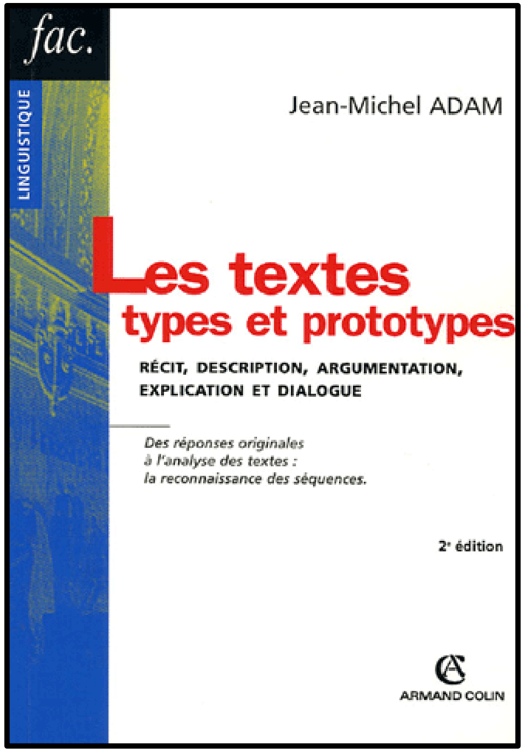
Egon Werlich, 1989

text types
narration, description, exposition, argumentation, instruction

Discourse modes: related theories

Egon Werlich, 1989

text types
narration, description, exposition, argumentation, instruction



Jean-Michel Adam, 2005

typical sequences
narrative, argumentative, descriptive, explicative, dialogued

Discourse modes: relevance for NLP

Discourse modes: relevance for NLP

- temporal discourse processing
  - knowing a passage's discourse mode is a necessary prerequisite for interpreting tense [Smith 2005]

Discourse modes: relevance for NLP

- temporal discourse processing
  - knowing a passage's discourse mode is a necessary prerequisite for interpreting tense [Smith 2005]
- automatic summarization, information extraction
  - focus on information in particular passages depending on the mode; user-specific summarization

Discourse modes: relevance for NLP

- temporal discourse processing
  - knowing a passage’s discourse mode is a necessary prerequisite for interpreting tense [Smith 2005]
- automatic summarization, information extraction
  - focus on information in particular passages depending on the mode; user-specific summarization
- argumentation mining
  - narrow the search space for claims by focusing on argumentative passages

Discourse modes: relevance for NLP

- temporal discourse processing
  - knowing a passage's discourse mode is a necessary prerequisite for interpreting tense [Smith 2005]
- automatic summarization, information extraction
  - focus on information in particular passages depending on the mode; user-specific summarization
- argumentation mining
  - narrow the search space for claims by focusing on argumentative passages
- genre distinctions
  - literary studies

Situation entity types

situations / eventualities
≈ evoked by finite clauses

1. Yesterday, Mary bought a cat.

2. Now she owns four cats.

3. Susie often feeds Mary's cats.

4. Cats are very social animals.

Situation entity types

situations / eventualities
≈ evoked by finite clauses

STATE

1. Yesterday, Mary bought a cat.
2. Now she owns four cats.
3. Susie often feeds Mary's cats.
4. Cats are very social animals.

Situation entity types

situations / eventualities
≈ evoked by finite clauses

1. Yesterday, Mary bought a cat.
2. Now she owns four cats. STATE
3. Susie often feeds Mary's cats.
4. Cats are very social animals.

Situation entity types

situations / eventualities
≈ evoked by finite clauses

EVENT

1. Yesterday, Mary bought a cat.

2. Now she owns four cats. STATE

3. Susie often feeds Mary's cats.

4. Cats are very social animals.

Situation entity types

situations / eventualities
≈ evoked by finite clauses

1. Yesterday, Mary bought a cat. EVENT
2. Now she owns four cats. STATE
3. Susie often feeds Mary's cats.
4. Cats are very social animals.

Situation entity types

situations / eventualities
≈ evoked by finite clauses

GENERIC SENTENCE

1. Yesterday, Mary bought a cat. EVENT
2. Now she owns four cats. STATE
3. Susie often feeds Mary's cats.
4. Cats are very social animals.

Situation entity types

situations / eventualities
≈ evoked by finite clauses

1. Yesterday, Mary bought a cat. EVENT
2. Now she owns four cats. STATE
3. Susie often feeds Mary's cats.
4. Cats are very social animals. GENERIC SENTENCE

Situation entity types

situations / eventualities
≈ evoked by finite clauses

GENERALIZING SENTENCE

1. Yesterday, Mary bought a cat. EVENT
2. Now she owns four cats. STATE
3. Susie often feeds Mary's cats.
4. Cats are very social animals. GENERIC SENTENCE

Situation entity types

situations / eventualities
≈ evoked by finite clauses

1. Yesterday, Mary bought a cat. EVENT
2. Now she owns four cats. STATE
3. Susie often feeds Mary's cats. GENERALIZING SENTENCE
4. Cats are very social animals. GENERIC SENTENCE

More situation entity types

ABSTRACT ENTITIES
here: clausal complements

frequent in
ARGUMENT/COMMENTARY
discourse mode

Susie **knows** — STATE

that Mary loves her cats a lot. — FACT

object of knowledge

More situation entity types

ABSTRACT ENTITIES
here: clausal complements

frequent in
ARGUMENT/COMMENTARY
discourse mode

Susie **knows** — STATE

that Mary loves her cats a lot. — FACT

object of knowledge

Susie **believes** — STATE

that the cats also love Mary. — PROPOSITION

object of belief

More situation entity types

ABSTRACT ENTITIES
here: clausal complements

frequent in ARGUMENT/COMMENTARY discourse mode

Susie **knows** — STATE

that Mary loves her cats a lot. — FACT

object of knowledge

Susie **believes** — STATE

that the cats also love Mary. — PROPOSITION

object of belief

Have you seen my cats? — QUESTION

Don’t forget to feed the cats! — IMPERATIVE

[Palmer et al. 2007]

Situation entity 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 <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!

Situation entity types: summary

Eventualities	STATE	Mary likes cats.
	EVENT	Mary fed the cats.
	- REPORT	..., Mary said.
General Statives	GENERALIZING SENTENCE	cats.
	GENERI SENTE	
Abstract	FACT	
	PROPOS	
	QUE	
	IMPERATIVE	Don't forget to feed the cats!

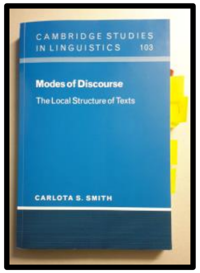
Writer / speaker chooses how to present things:

The ship was in motion. STATE

The ship moved. EVENT

Carlota Smith: The Parameter of Aspect (1997).

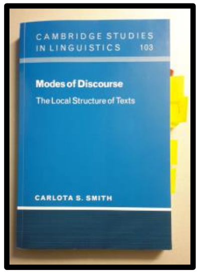
Situation entity annotation



Carlota Smith: Modes of Discourse (2003).

Many examples, but no formal definition of the different situation entity types.

Situation entity annotation



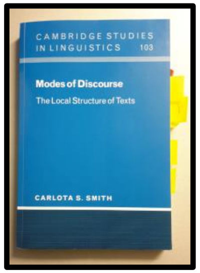
Carlota Smith: Modes of Discourse (2003).

Many examples, but no formal definition of the different situation entity types.

Alexis Palmer, Elias Ponvert, Jason Baldridge, and Carlota Smith.
A sequence model for situation entity classification. ACL 2007.

- first labeled data set for SEs, ~6000 clauses
- no annotation manual, Cohen's κ = 0.54

Situation entity annotation



Carlota Smith: Modes of Discourse (2003).

Many examples, but no formal definition of the different situation entity types.

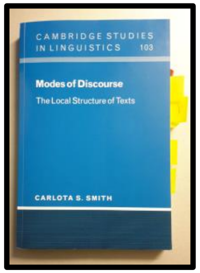
Alexis Palmer, Elias Ponvert, Jason Baldridge, and Carlota Smith.
A sequence model for situation entity classification. ACL 2007.

- first labeled data set for SEs, ~6000 clauses
- no annotati

What are the **most important differences** between Smith's situation entity types?

Annemarie Friedrich and Alexis Palmer.
Situation entity annotation. LAW 2014.

Situation entity annotation



Carlota Smith: Modes of Discourse (2003).

Many examples, but no formal definition of the different situation entity types.

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

- first labeled data set for SEs, ~6000 clauses
- no annotati

What are the **most important differences** between Smith's situation entity types?

convey **annotation scheme + guidelines** to annotators

Annemarie Friedrich and Alexis Palmer. **Situation entity annotation**. LAW 2014.

Situation entity types: feature-based annotation

What are the **main differences** between the different 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 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!

Situation entity types: feature-based annotation

What are the **main differences** between the different situation entity types?

Does the verb express an **event** or a **state**?

aspectual class

		Mary likes cats.
		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!

Situation entity types: feature-based annotation

What are the **main differences** between the different situation entity types?

Does the verb express an **event** or a **state**?

aspectual class

Does something happen repeatedly or once?

habituality

		Mary likes cats
		Mary fed the c
	- REPORT	..., Mary said.
General Statives	GENERALIZING SENTENCE	Mary often feeds m ats.
	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!

Situation entity types: feature-based annotation

What are the **main differences** between the different situation entity types?

Does the verb express an **event** or a **state**?

aspectual class

Does something happen repeatedly or once?

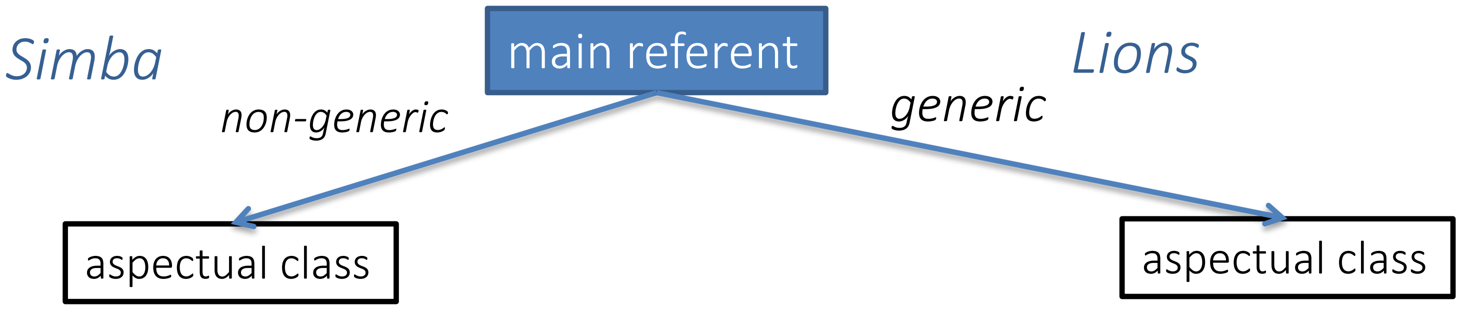
habituality

Does the sentence talk about a **particular referent** or a **kind/class**?

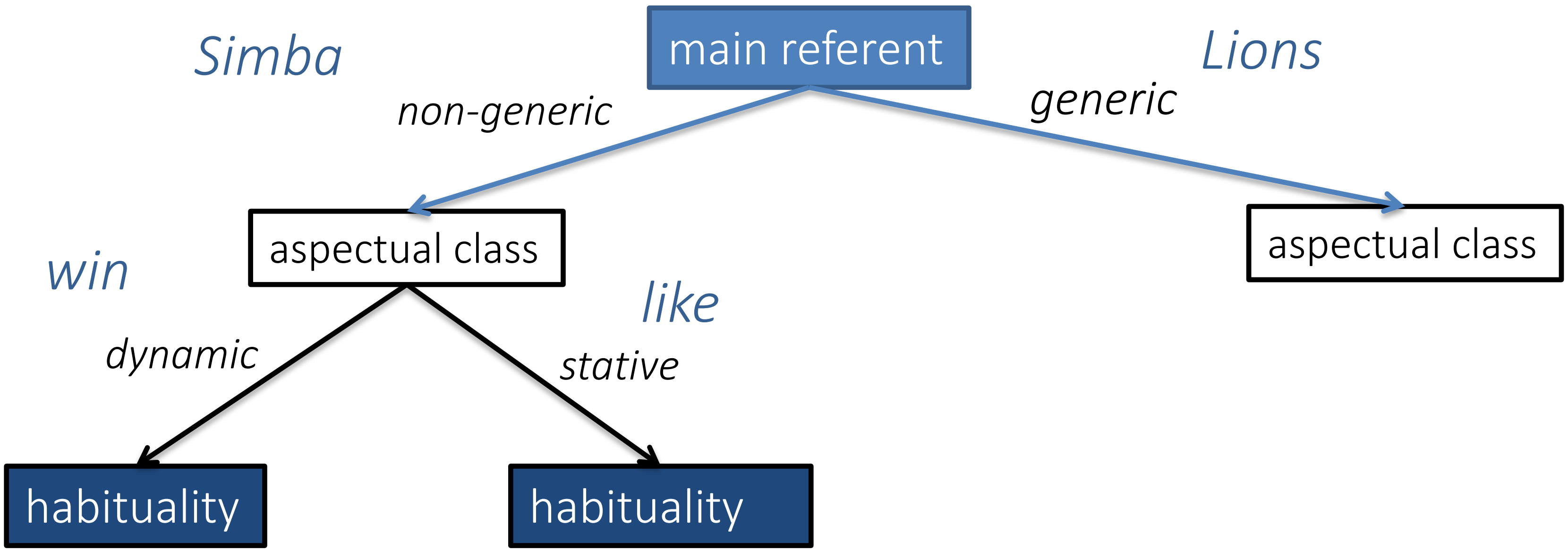
genericity

		Mary likes cats
		Mary fed the c
	- REPORT	..., Mary said.
		Mary often feeds m cats.
		Cats are always hungry.
		I know that Mary fed the cats.
Entities	PROPOSITION	I believe that Mary fed the cats.
Speech Acts	QUESTION	Does Mary like cats?
	IMPERATIVE	Don't forget to feed the cats!

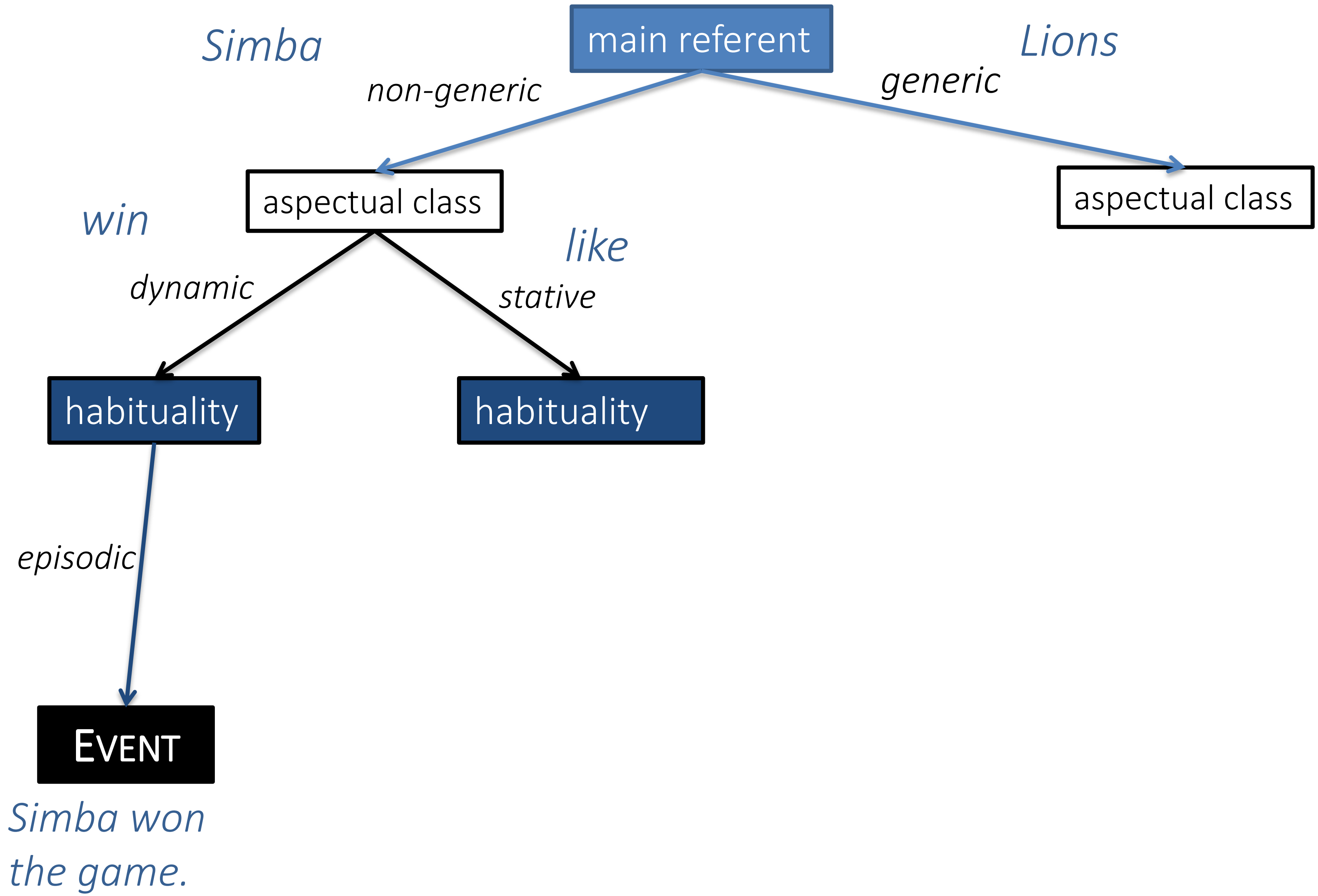
A decision tree for labeling situation entities



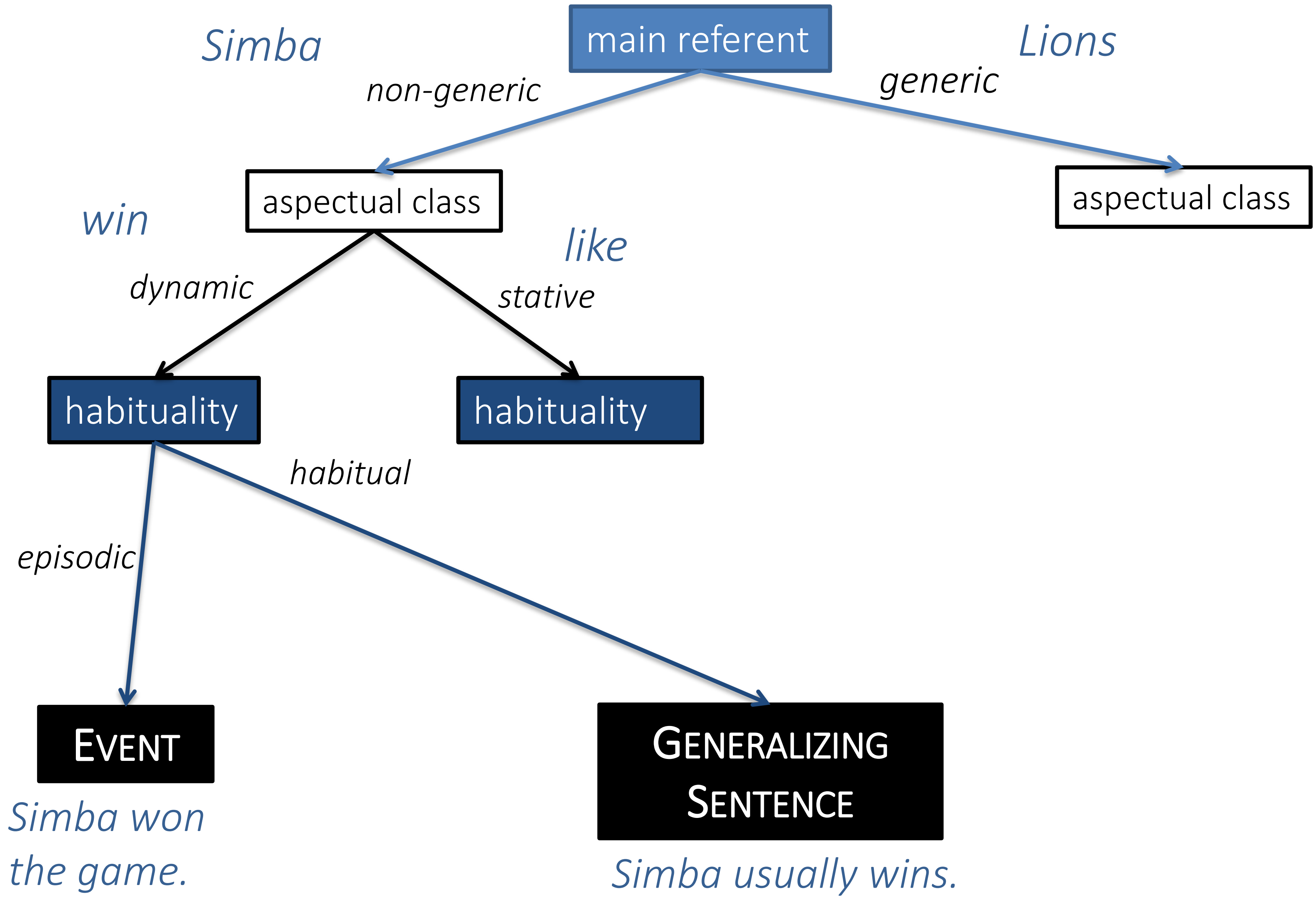
A decision tree for labeling situation entities



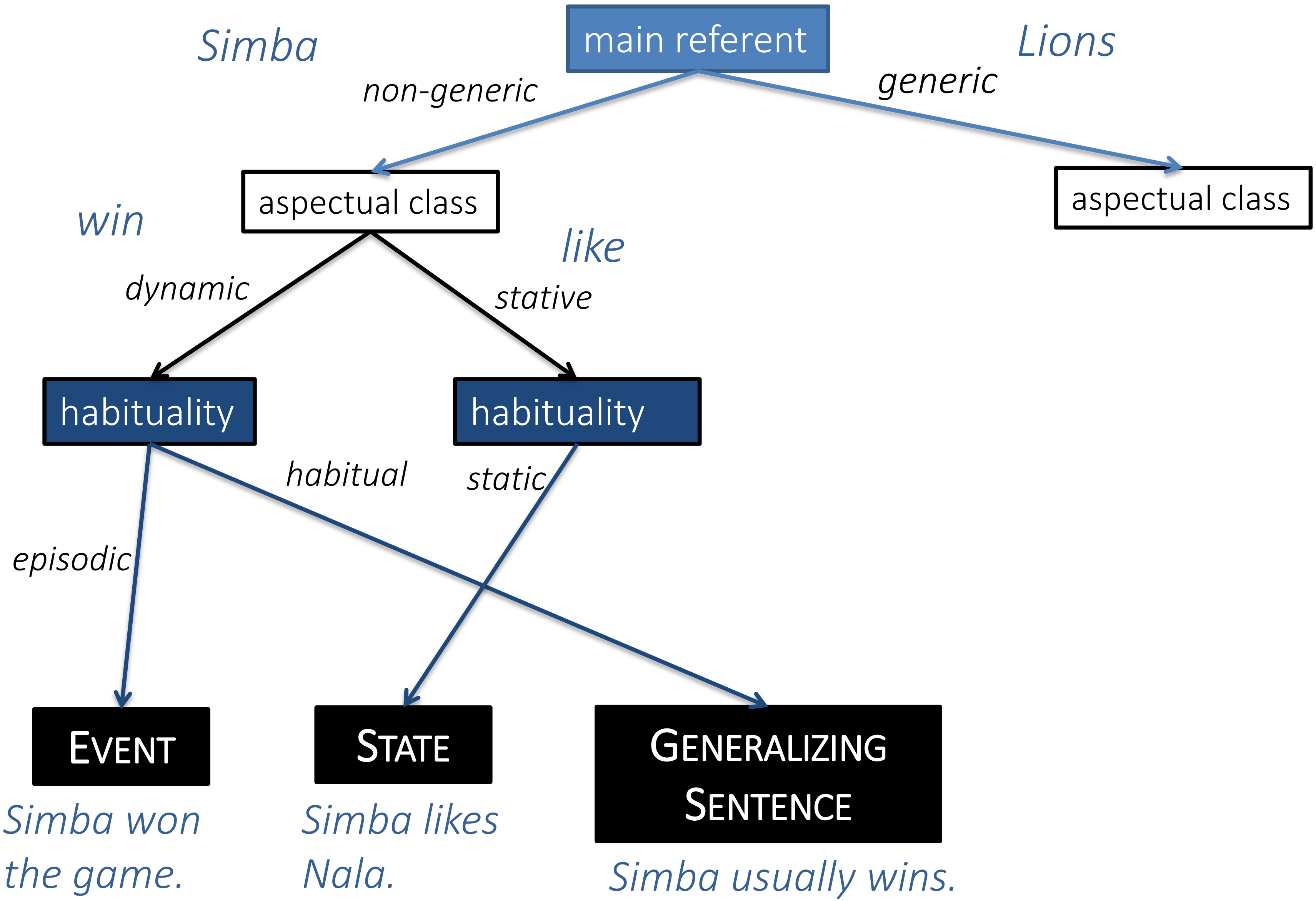
A decision tree for labeling situation entities



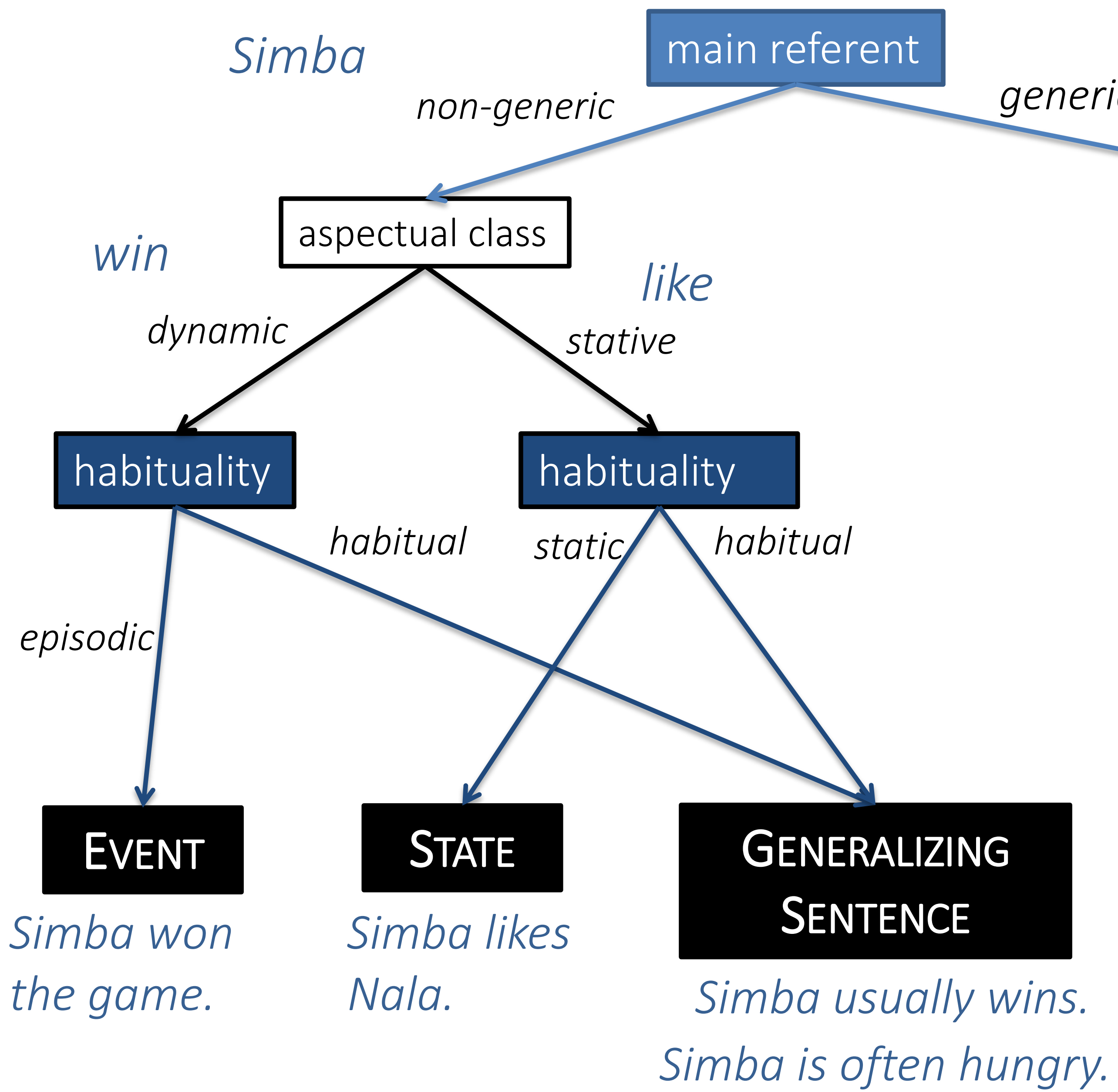
A decision tree for labeling situation entities



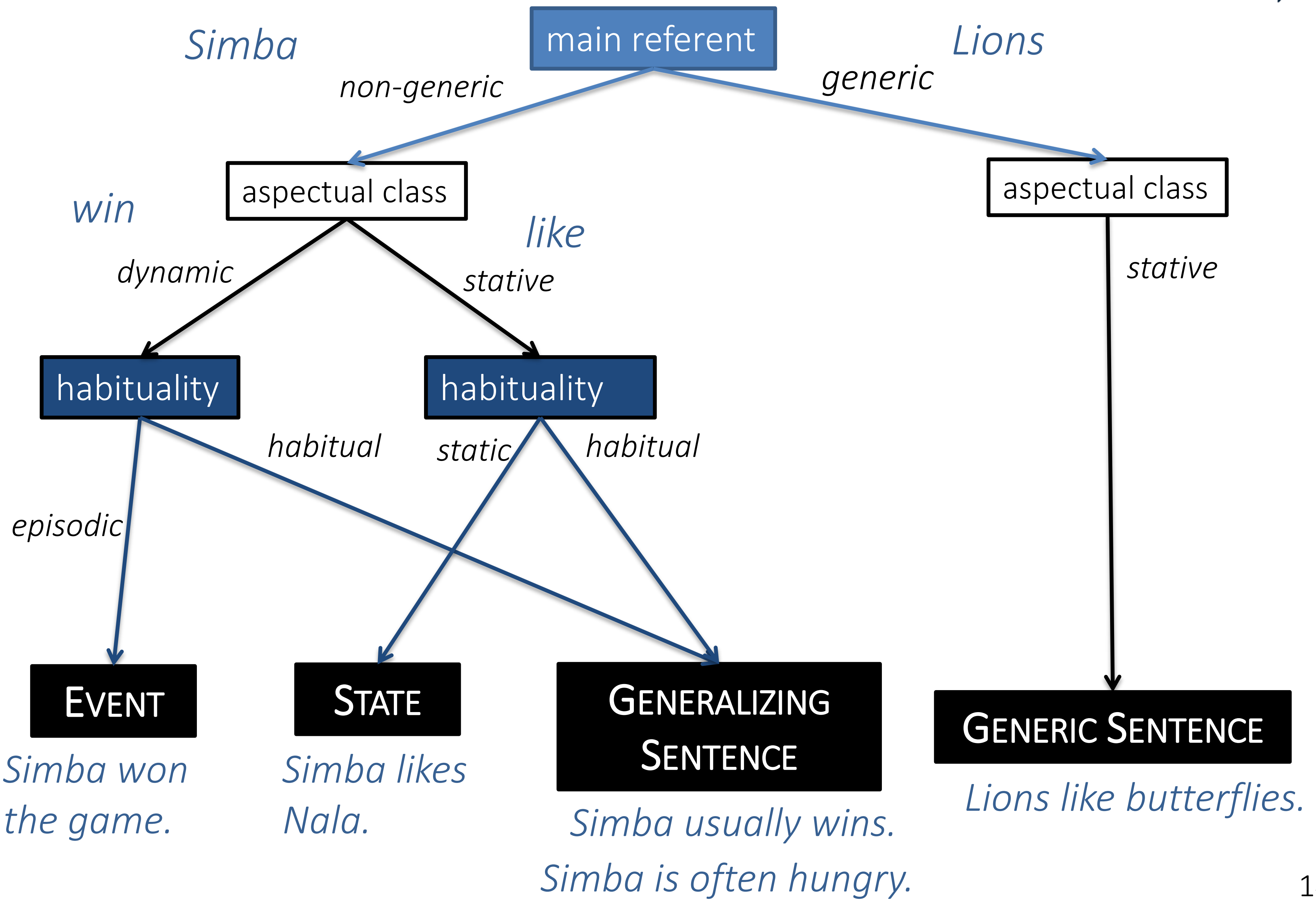
A decision tree for labeling situation entities



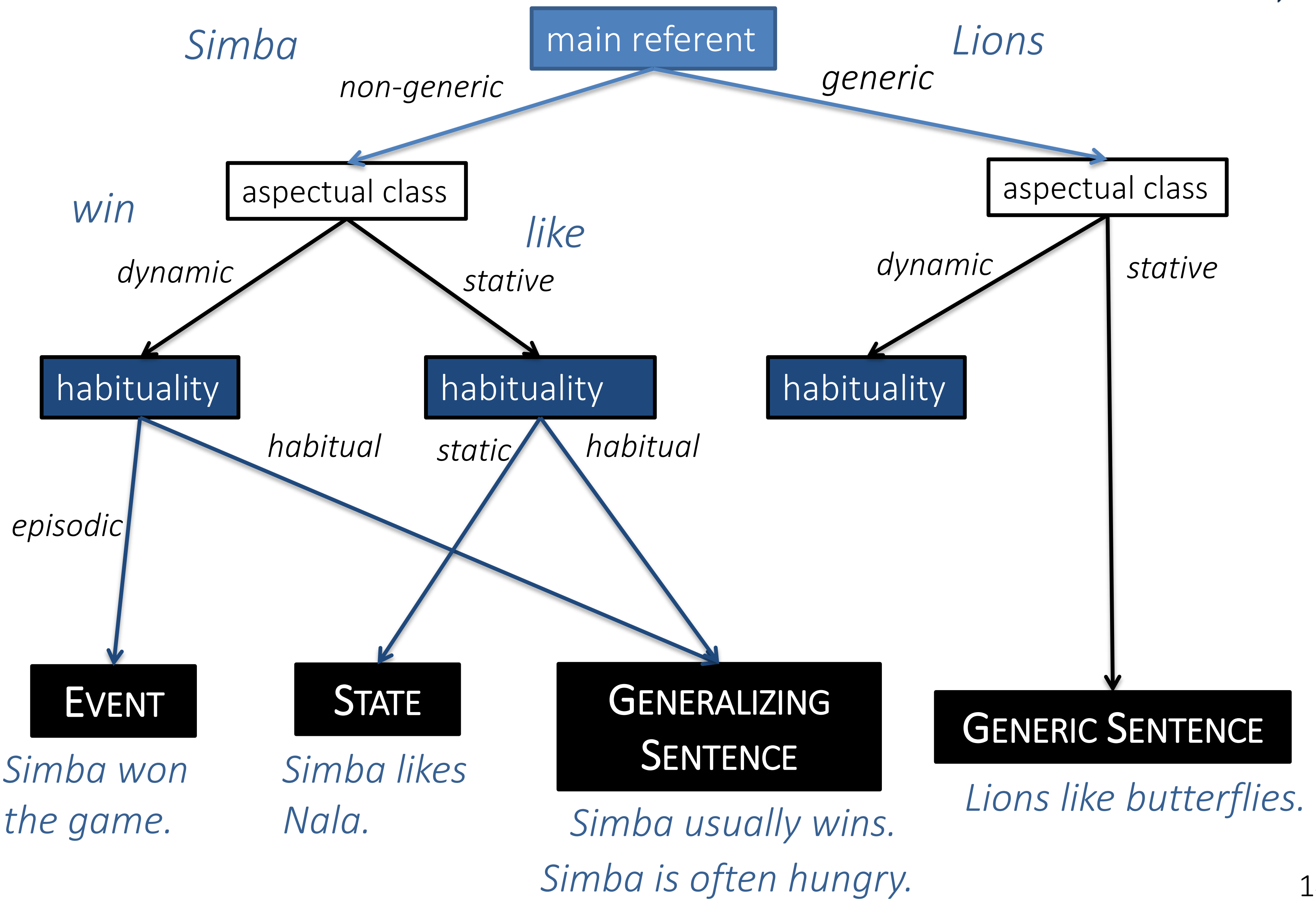
A decision tree for labeling situation entities



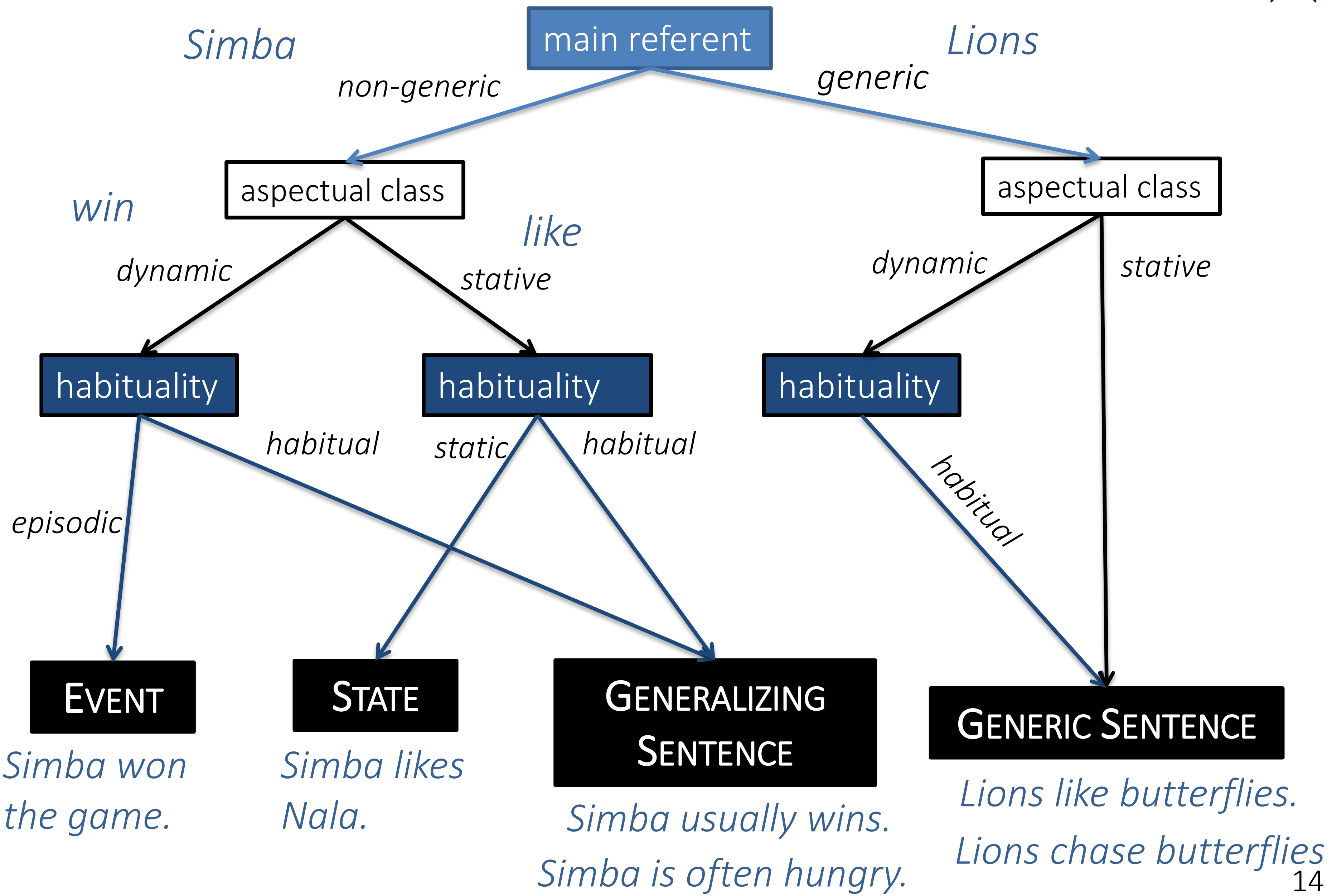
A decision tree for labeling situation entities



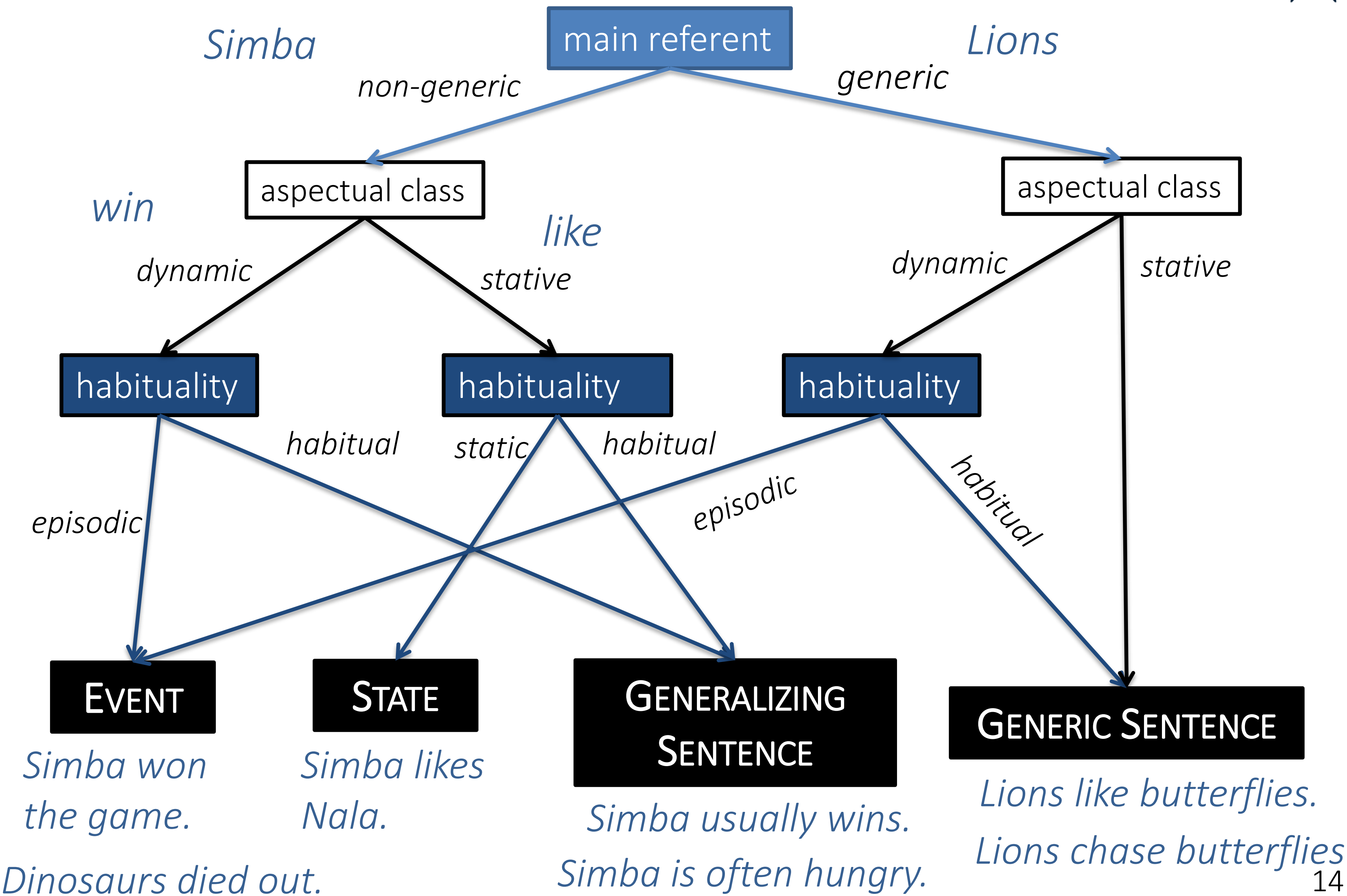
A decision tree for labeling situation entities



A decision tree for labeling situation entities



A decision tree for labeling situation entities



Situation entity types: coercion

some linguistic phenomena coerce EVENTs to STATEs:

negation, modality, future / perfect, conditionality, subjectivity

Situation entity types: coercion

some linguistic phenomena coerce EVENTs to STATEs:

negation, modality, future / perfect, conditionality, subjectivity

Susie will feed the cats.

Situation entity types: coercion

some linguistic phenomena coerce **EVENTs** to **STATEs**:

negation, modality, future / perfect, conditionality, subjectivity

Susie will feed the cats.
Susie has not fed the cats.

Situation entity types: coercion

some linguistic phenomena coerce **EVENTs** to **STATEs**:

negation, modality, future / perfect,
conditionality, subjectivity

Susie will feed the cats.
Susie has not fed the cats.
If Susie has forgotten the cats,
they might be hungry now.

Situation entity types: coercion

some linguistic phenomena coerce **EVENTs** to **STATEs**:

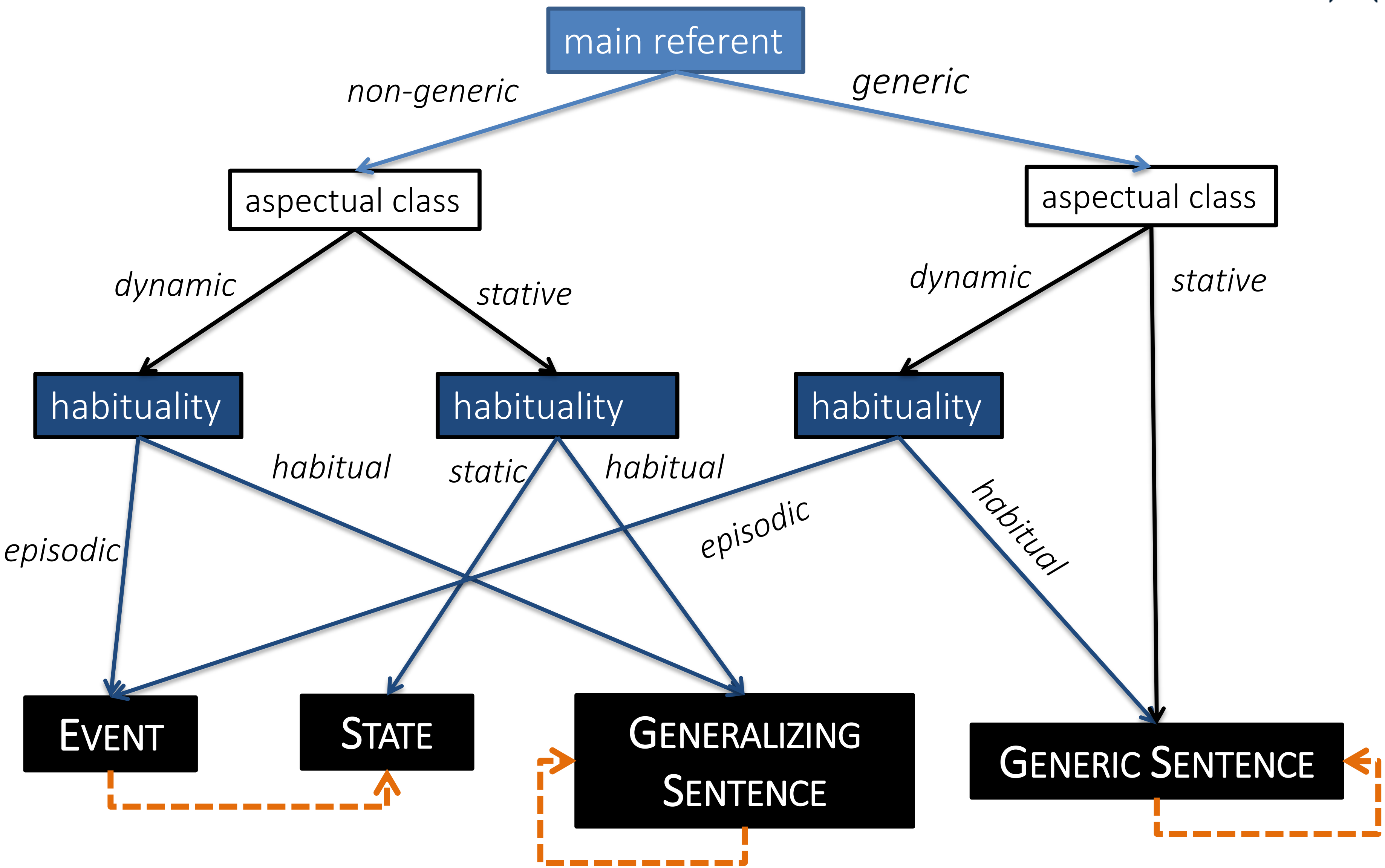
negation, modality, future / perfect,
conditionality, subjectivity

Susie will feed the cats.
Susie has not fed the cats.
If Susie has forgotten the cats,
they might be hungry now.

does not apply to general statives:

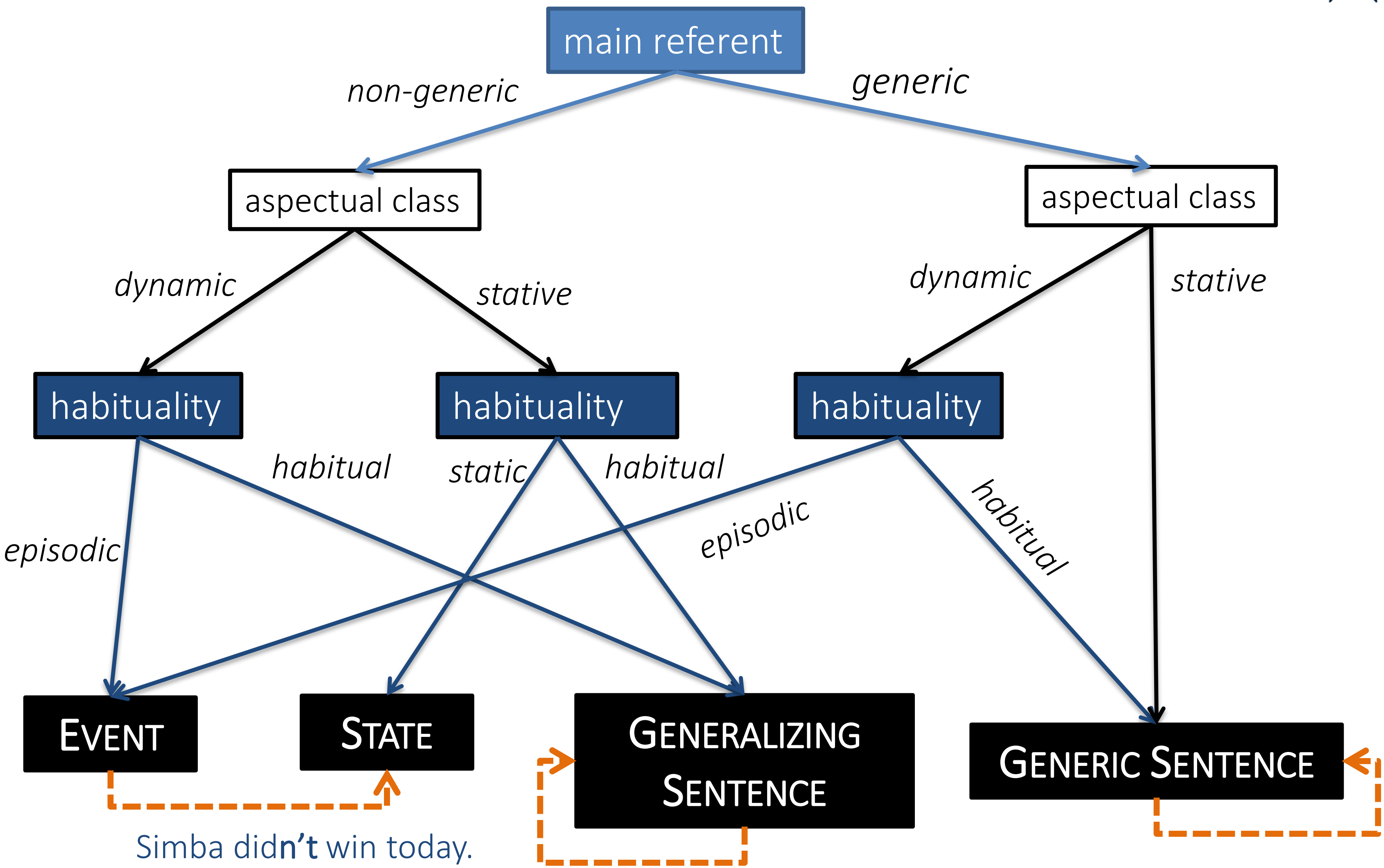
Susie never feeds Mary's cats.	GENERALIZING SENTENCE
Cats might be the most popular pet.	GENERIC SENTENCE

A decision tree for labeling situation entities



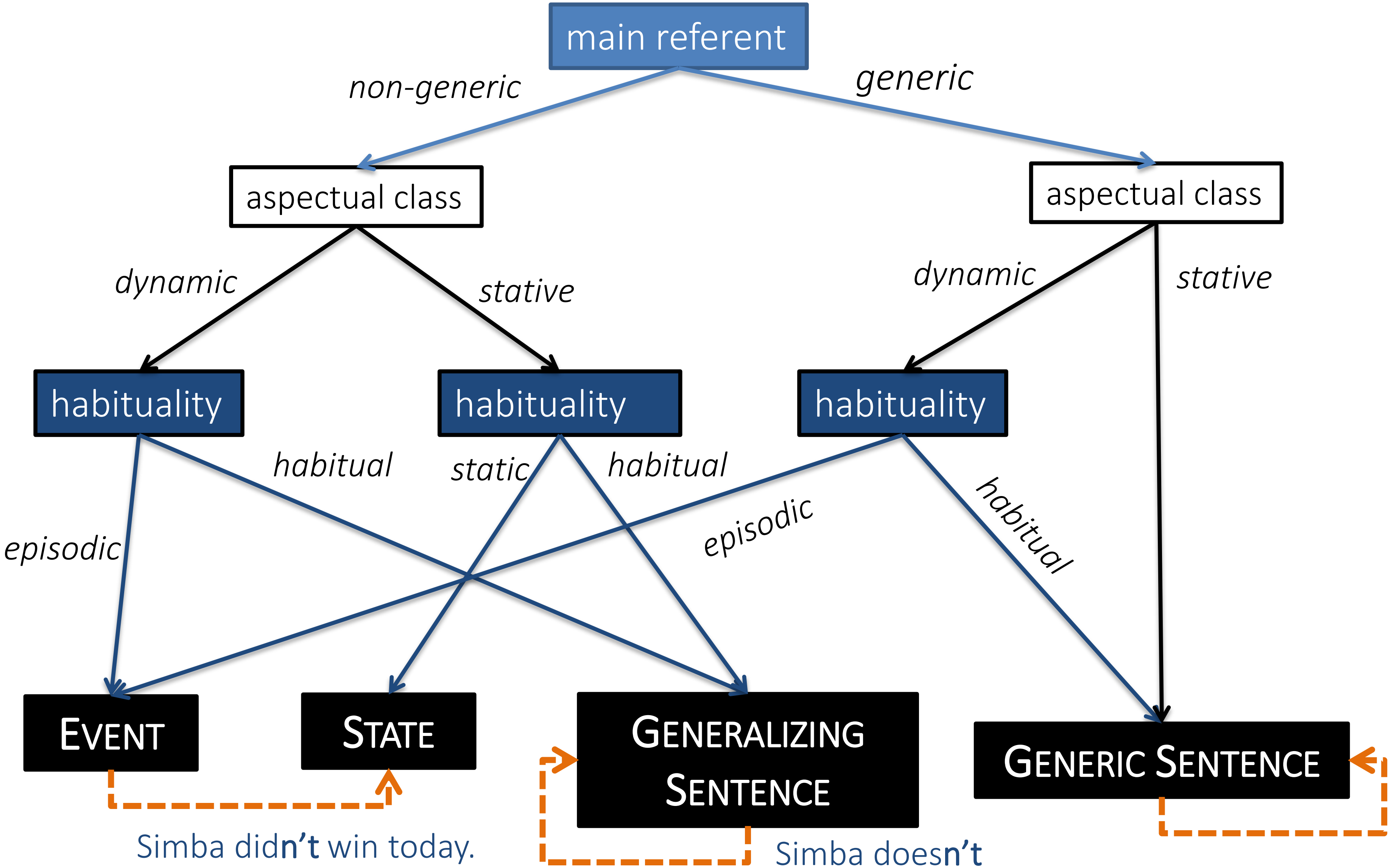
negation, modals, conditional, perfect, future

A decision tree for labeling situation entities



negation, modals, conditional, perfect, future

A decision tree for labeling situation entities

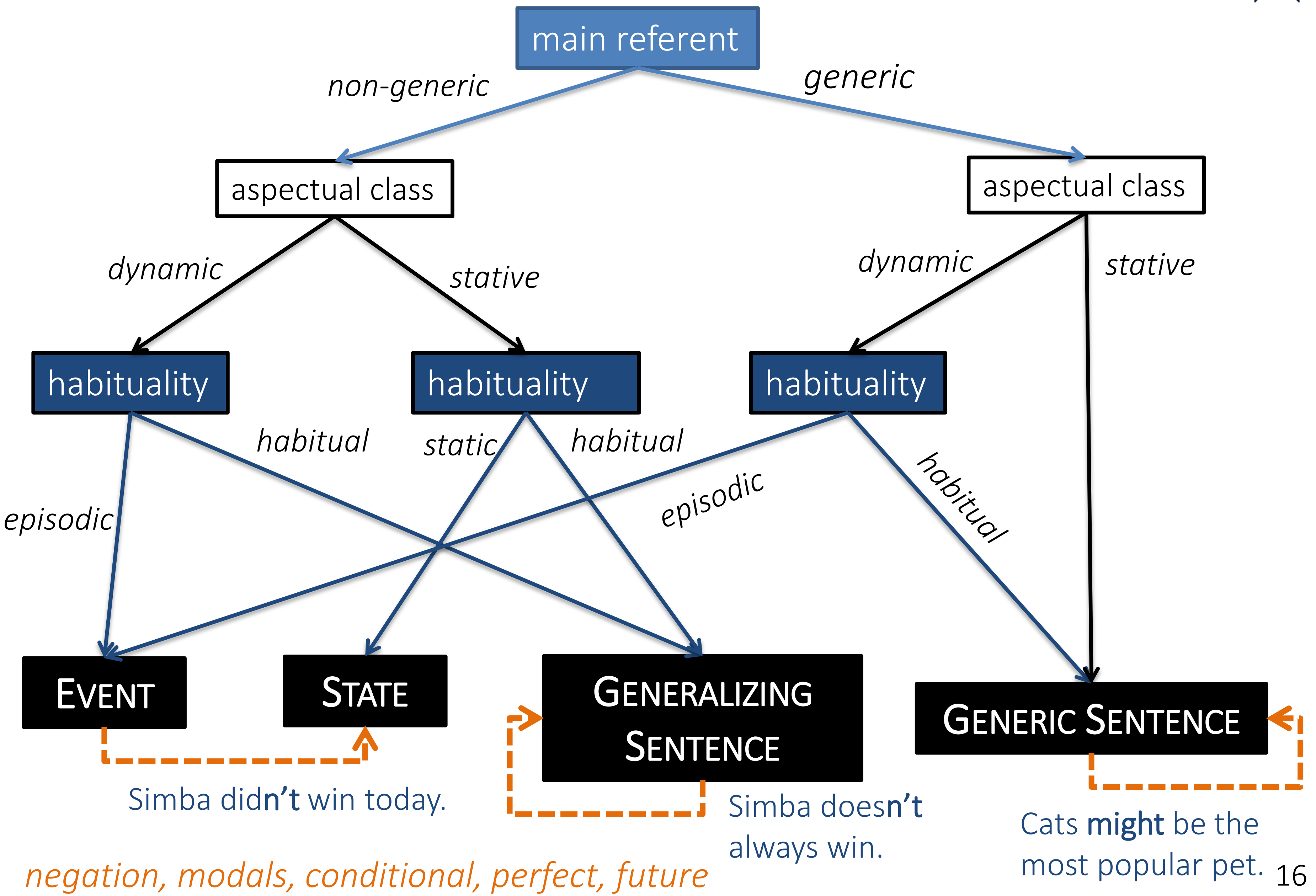


Simba did**n't** win today.

Simba does**n't** always win.

negation, modals, conditional, perfect, future

A decision tree for labeling situation entities



Data sets and annotation procedure

Data sets and annotation procedure

MASC
30,000 clauses
essays, letters, fiction, technical, travel, news …

Data sets and annotation procedure

MASC
30,000 clauses
essays, letters, fiction, technical, travel, news …

Wikipedia
10,000 clauses
botany, animals, sports, biographies, science, …

Data sets and annotation procedure

MASC
30,000 clauses
essays, letters, fiction, technical, travel, news …

Wikipedia
10,000 clauses
botany, animals, sports, biographies, science, …

segmentation into clauses (SPADE)

Data sets and annotation procedure

MASC
30,000 clauses
essays, letters, fiction, technical, travel, news …

Wikipedia
10,000 clauses
botany, animals, sports, biographies, science, …

segmentation into clauses (SPADE)

Annotators label
- situation entity type
- genericity of main referent
- lexical aspectual class of main verb
- habituality of main verb

Data sets and annotation procedure

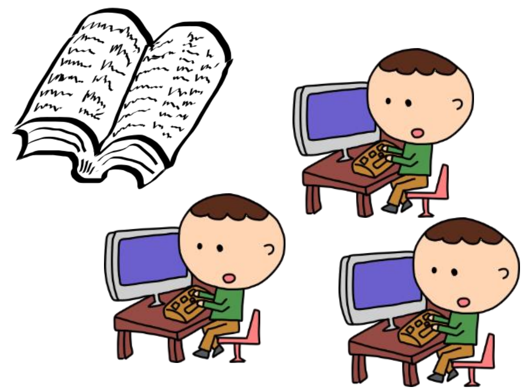
MASC
30,000 clauses
essays, letters, fiction, technical, travel, news …

Wikipedia
10,000 clauses
botany, animals, sports, biographies, science, …

segmentation into clauses (SPADE)

Annotators label
- situation entity type
- genericity of main referent
- lexical aspectual class of main verb
- habituality of main verb

training phase
+ manual



Data sets and annotation procedure

MASC
30,000 clauses
essays, letters, fiction, technical, travel, news …

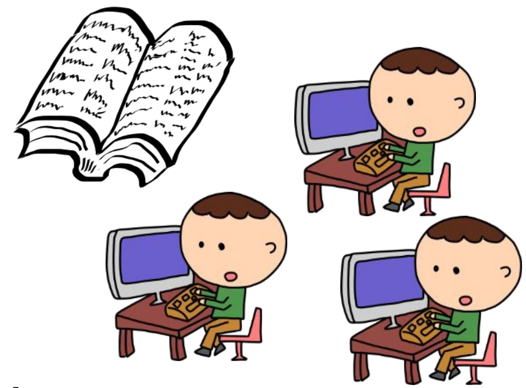
Wikipedia
10,000 clauses
botany, animals, sports, biographies, science, …

segmentation into clauses (SPADE)

Annotators label

- situation entity type
- genericity of main referent
- lexical aspectual class of main verb
- habituality of main verb

training phase
+ manual



gold standard = majority vote over labels of 3 annotators

Data sets and annotation procedure

MASC
30,000 clauses
essays, letters, fiction, technical, travel, news …

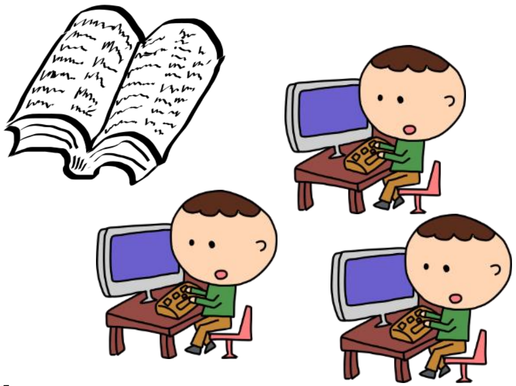
Wikipedia
10,000 clauses
botany, animals, sports, biographies, science, …

segmentation into clauses (SPADE)

Annotators label

- situation entity type
- genericity of main referent
- lexical aspectual class of main verb
- habituality of main verb

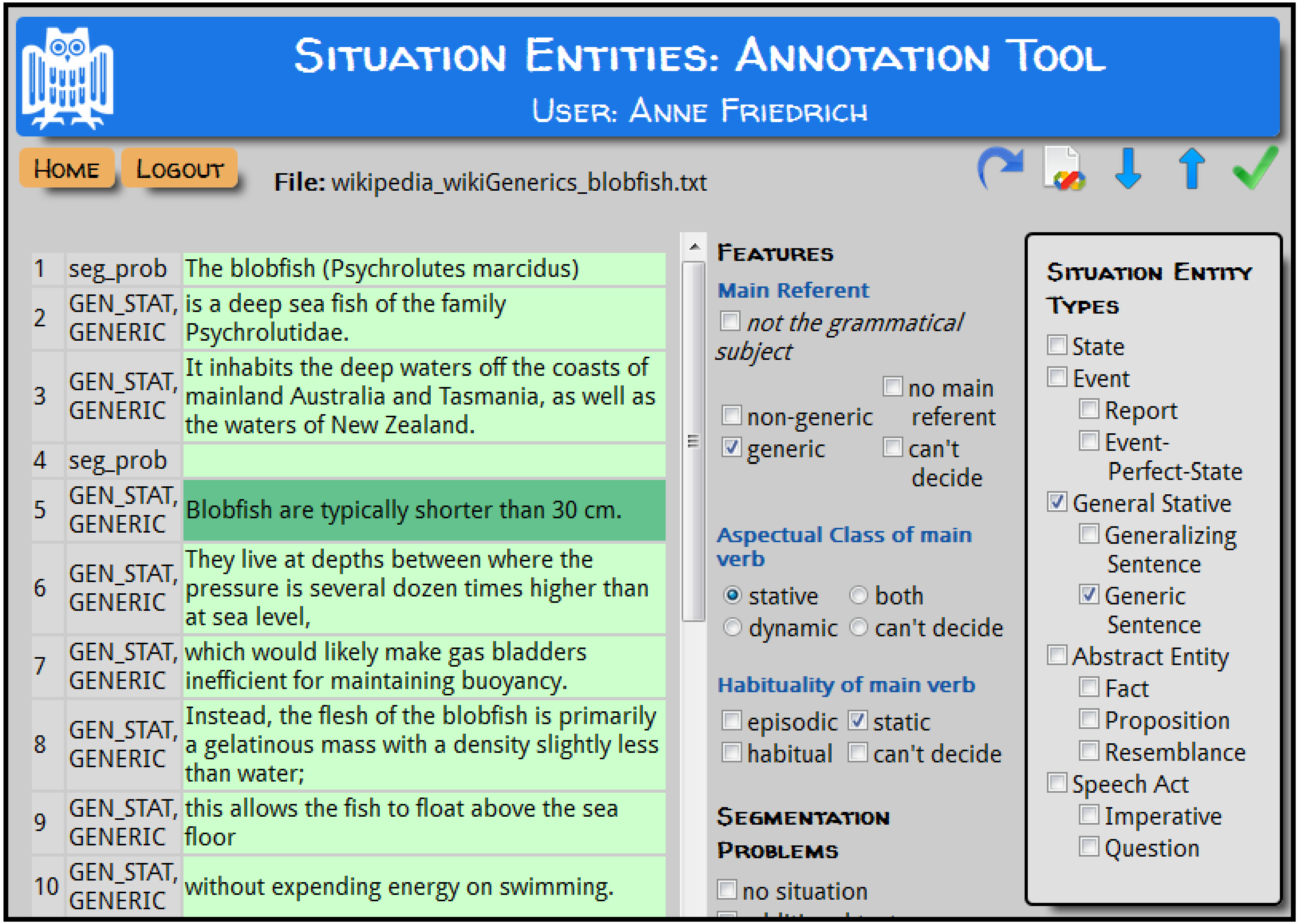
training phase + manual



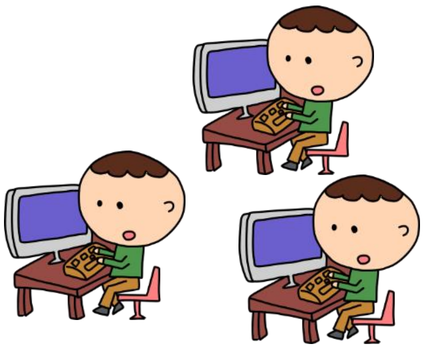
gold standard = majority vote over labels of 3 annotators

(about 10% of segments marked as "NO SITUATION") 

Annotation of situation entity types and features

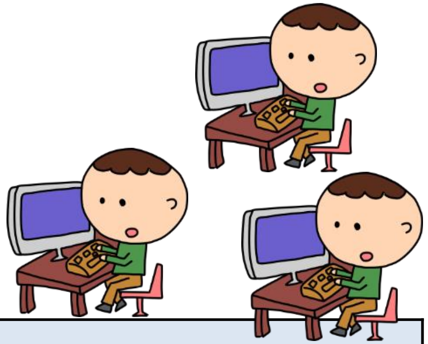


Inter-annotator agreement



Fleiss' κ: features			
Feature	labels	MASC	Wikipedia
aspectual class	stative, dynamic, both	0.69	0.64
main referent	generic, non-generic, cannot decide	0.55	0.67
habituality	episodic, static, habitual, cannot decide	0.72	0.65

Inter-annotator agreement



Fleiss' κ

Krippendorff's diagnostics: situation entity types

Fleiss' κ		
CATEGORY	MASC	Wikipedia
all categories	**0.64**	**0.63**
STATE	0.64	0.57
EVENT	0.72	0.72
REPORT	0.83	0.28
GENERIC SENTENCE	0.43	0.70
GENERALIZING SENTENCE	0.45	0.35
ABSTRACT ENTITY	0.40	0.19
QUESTION	0.85	0.85
IMPERATIVE	0.91	0.85

Situation entity types: relevance for NLP

- identifying the discourse modes of a text passage

Situation entity types: relevance for NLP

- identifying the discourse modes of a text passage
- corpus data and computational models for sub-tasks studied in the NLP community for which no large data sets are available

Situation entity types: relevance for NLP

- identifying the discourse modes of a text passage
- corpus data and computational models 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]

Situation entity types: relevance for NLP

- identifying the discourse modes of a text passage
- corpus data and computational models 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 2013]

Situation entity types: relevance for NLP

- identifying the discourse modes of a text passage
- corpus data and computational models 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 2013]
  - identifying habitual vs. episodic sentences [Mathew & Katz 2009]

Computational modeling of situation entity types

[ACL 2014]

[ACL 2015, LAW 2015]

is the main referent generic?

lexical aspectual class

[EMNLP 2015]

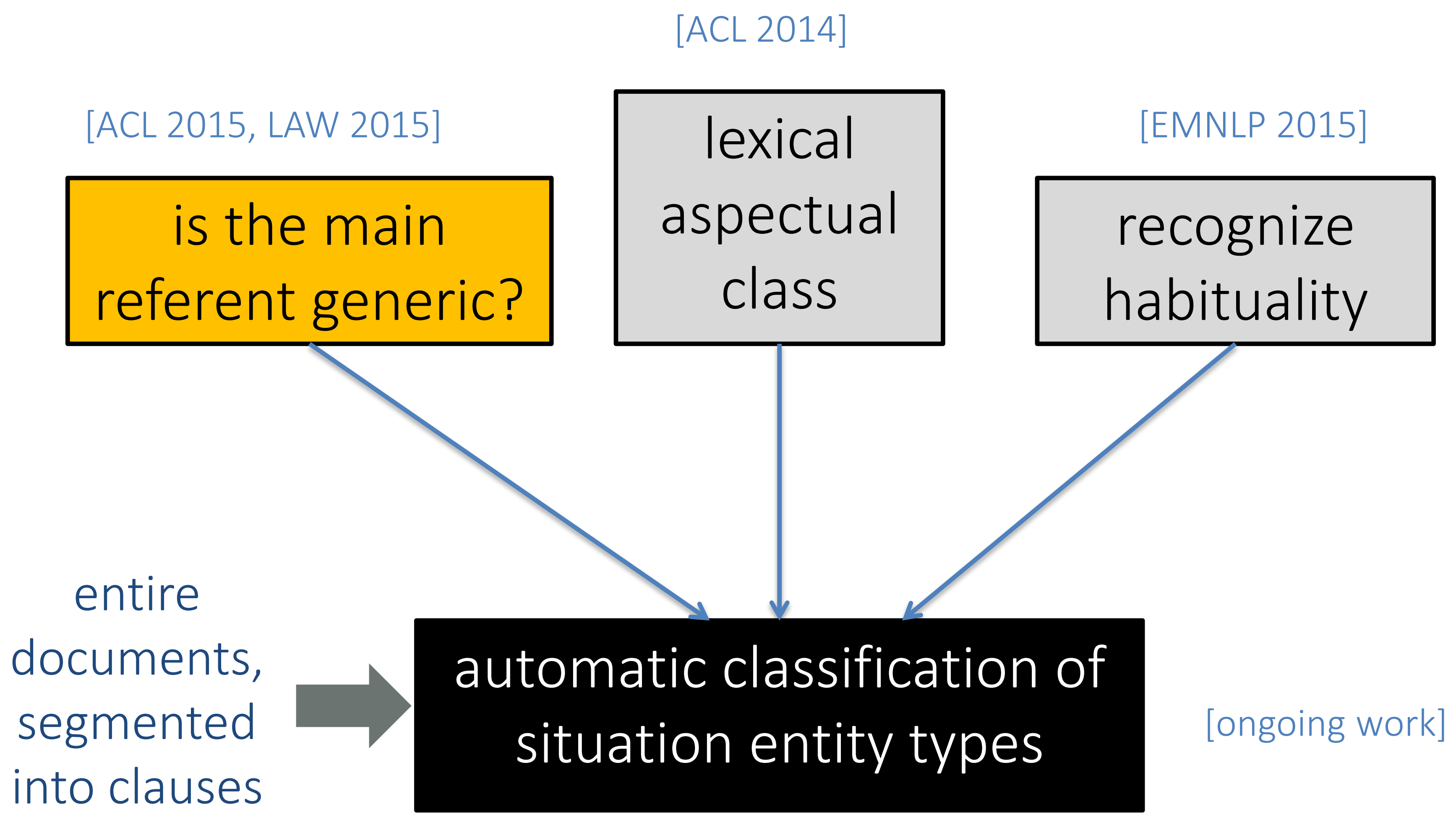
recognize habituality

entire documents, segmented into clauses

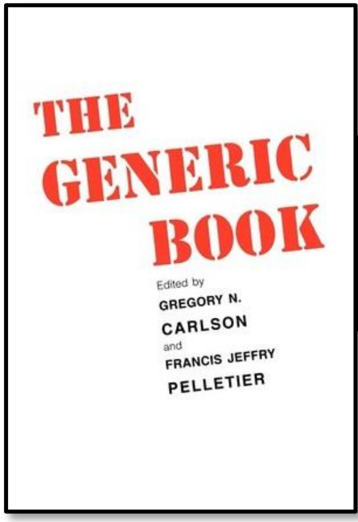
automatic classification of situation entity types

[ongoing work]

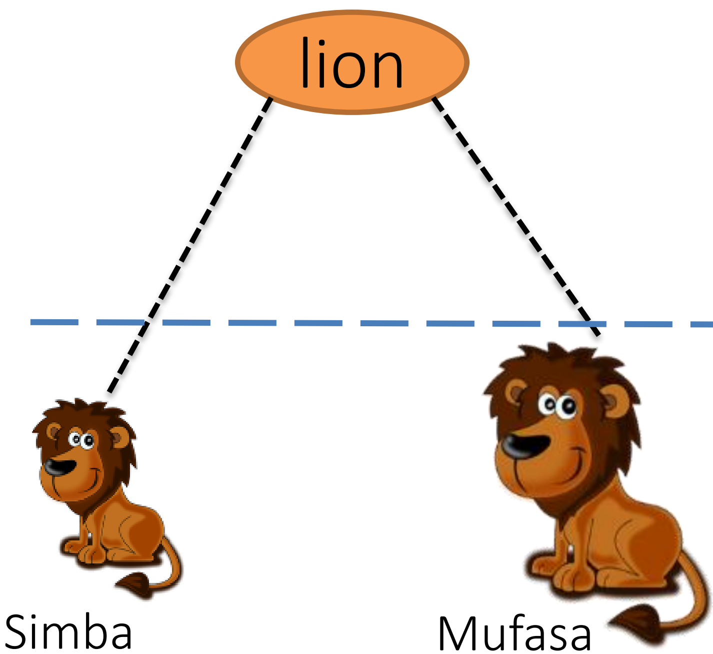
Computational modeling of situation entity types



Genericity



Krifka, Manfred, et al.
Introduction to genericity.
In *The Generic Book* (1995).



different
entailment properties

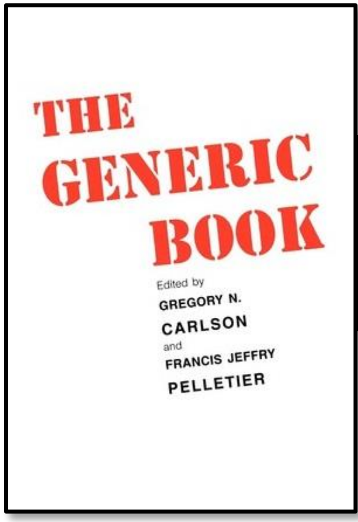
Lions are dangerous.

kind-referring
generic

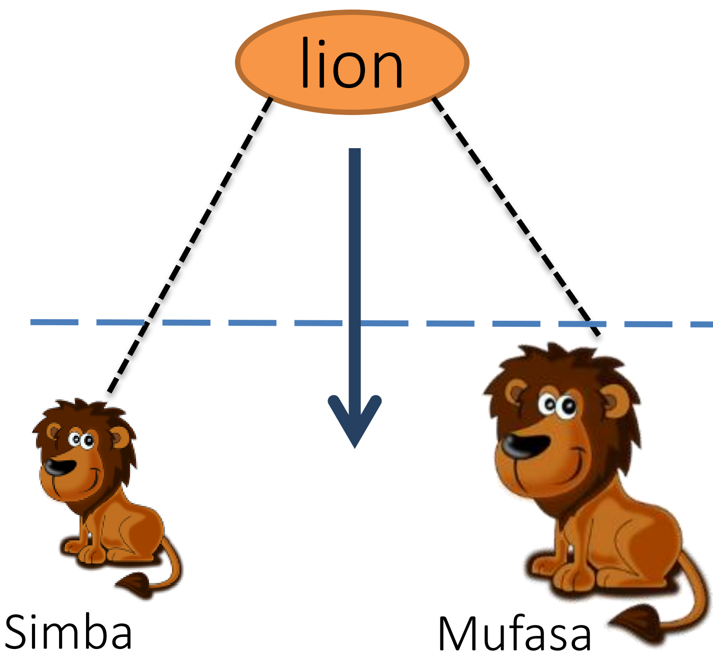
Mufasa is dangerous.
Simba is dangerous.

non-generic

Genericity



Krifka, Manfred, et al.
Introduction to genericity.
In *The Generic Book* (1995).



different
entailment properties

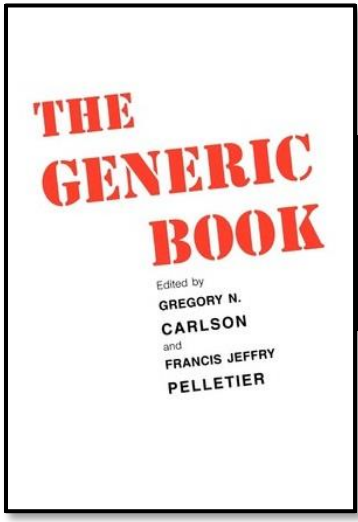
<u>Lions</u> are dangerous.

kind-referring
generic

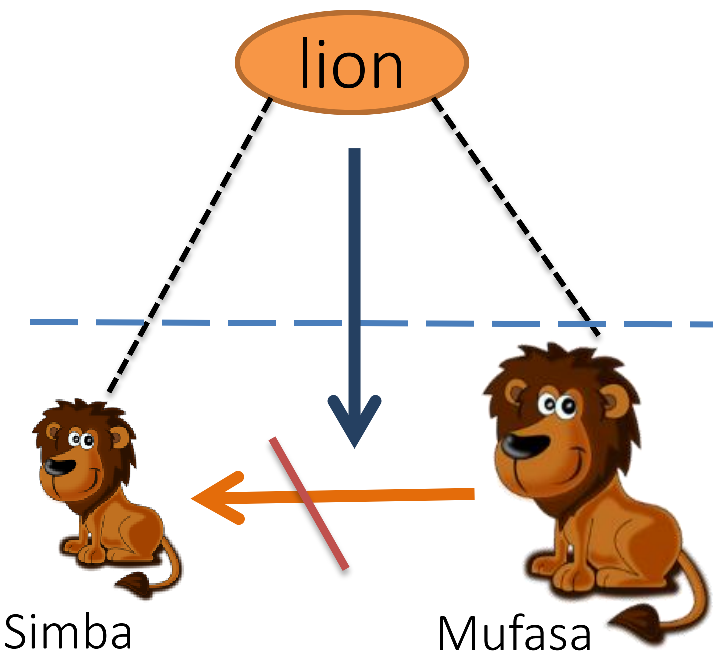
<u>Mufasa</u> is dangerous.
<u>Simba</u> is dangerous.

non-generic

Genericity



Krifka, Manfred, et al.
Introduction to genericity.
In *The Generic Book* (1995).



different
entailment properties

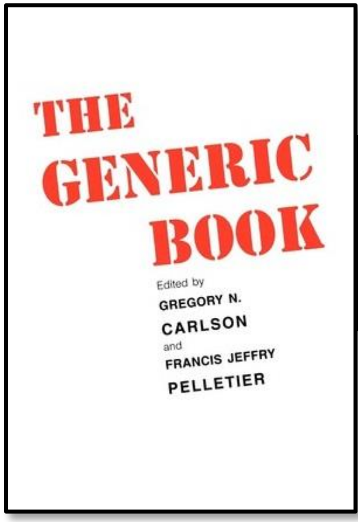
Lions are dangerous.

kind-referring
generic

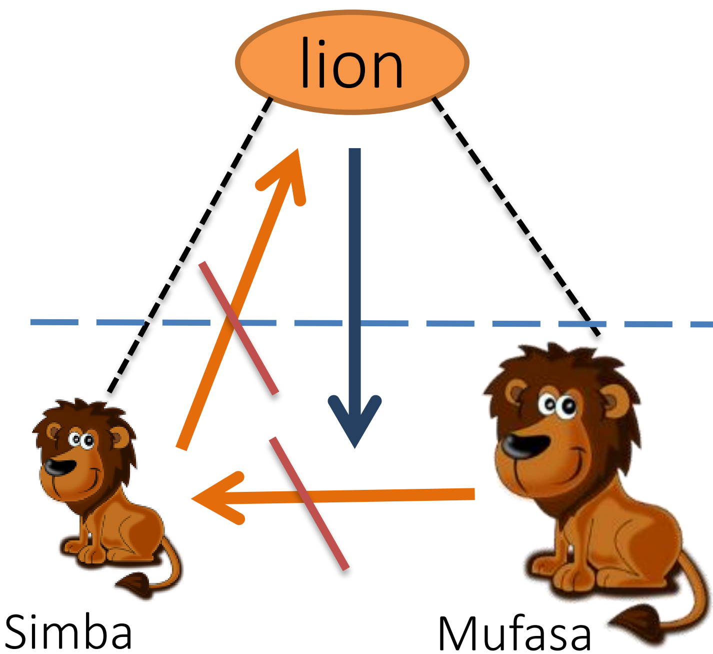
Mufasa is dangerous.
Simba is dangerous.

non-generic

Genericity



Krifka, Manfred, et al.
Introduction to genericity.
In *The Generic Book* (1995).



different
entailment properties

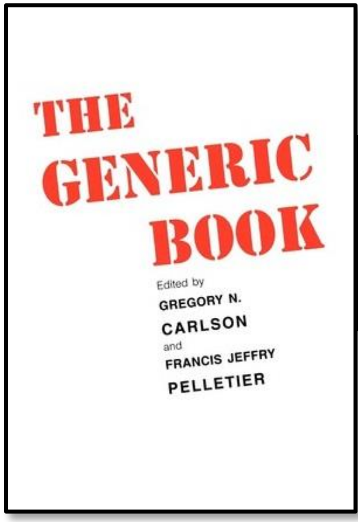
kind-referring
generic

Lions are dangerous.

Mufasa is dangerous.
Simba is dangerous.

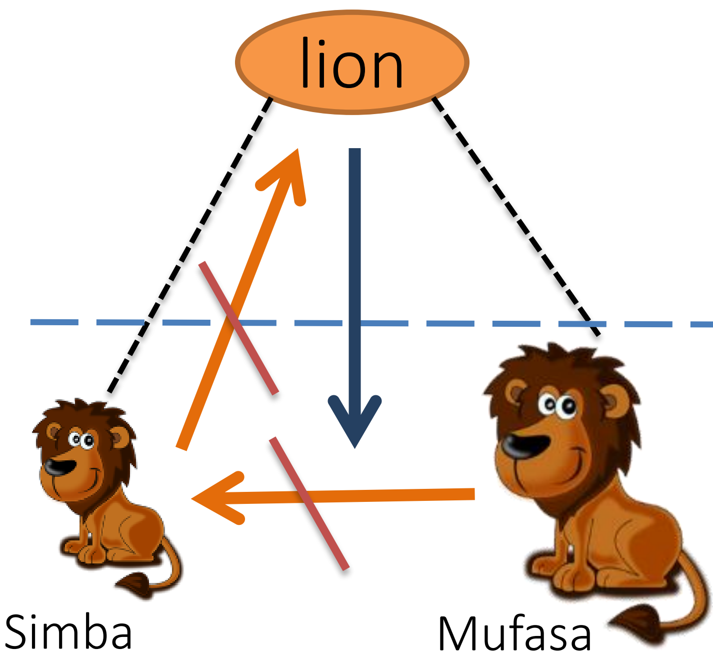
non-generic

Genericity



Krifka, Manfred, et al.
Introduction to genericity.
In *The Generic Book* (1995).

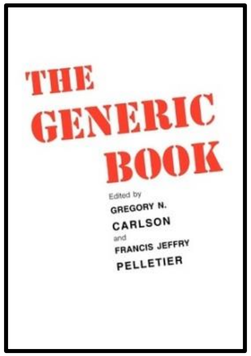
- ✓ information / event extraction
- ✓ knowledge acquistion from text



different
entailment properties

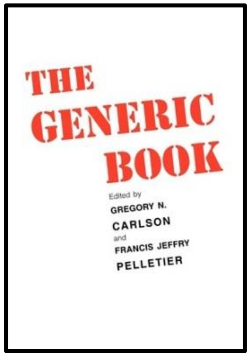
Lions are dangerous. — kind-referring generic

Mufasa is dangerous.
Simba is dangerous. — non-generic



Reference to kinds

	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.

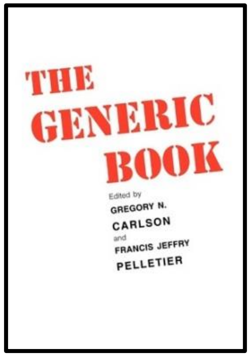


Reference to kinds

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.

Reference to kinds

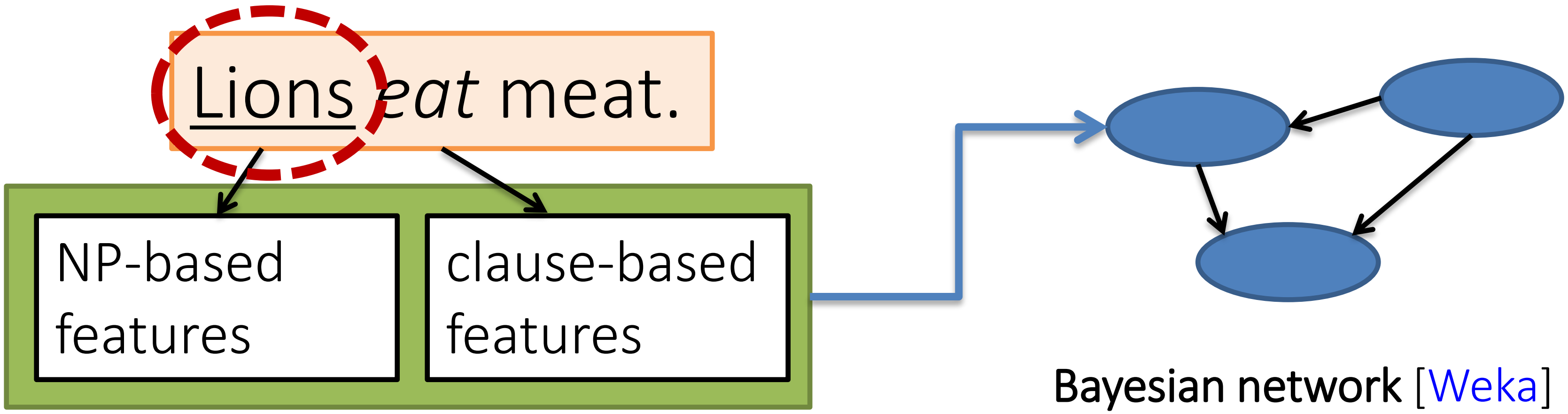


	kind-referring	non-kind-referring
definite NPs	The lion is a predatory cat.	The cat chased the mouse.
indefinite NPs	Lions eat meat.	Dogs were barking outside.
quantified NPs	Some (type of) dinosaur is extinct.	Some dogs were barking outside.
proper names	Panthera leo persica was first described by the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

form of NP not sufficient

clause / context matters

Baseline: identifying generic noun phrases



Bayesian network [Weka]

Nils Reiter and Anette Frank. **Identifying generic noun phrases.** ACL 2010.

Data: ACE-2 & ACE-2005

→ largest corpora annotated with NP-level genericity to date, ~40k NPs

- SPC = specific / non-generic
- GEN = generic
- USP = underspecified

Syntactic-semantic features

→ reimplementation of R&F using freely available resources
→ extracted from dependency parses (Stanford parser)
https://github.com/annefried/sitent

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, false
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

Discourse-sensitive approach

WIKIPEDIA
The Free Encyclopedia

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

Annemarie Friedrich and Manfred Pinkal. **Discourse-sensitive automatic identification of generic expressions.** ACL 2015.

Discourse-sensitive approach

WIKIPEDIA
The Free Encyclopedia

[**The recent year's growth twigs**]
are green and turn dark brown.

Annemarie Friedrich and Manfred Pinkal. **Discourse-sensitive automatic identification of generic expressions.** ACL 2015.

Discourse-sensitive approach

WIKIPEDIA
The Free Encyclopedia

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

Annemarie Friedrich and Manfred Pinkal. **Discourse-sensitive automatic identification of generic expressions.** ACL 2015.

Discourse-sensitive approach

WIKIPEDIA
The Free Encyclopedia

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

Annemarie Friedrich and Manfred Pinkal. **Discourse-sensitive automatic identification of generic expressions.** ACL 2015.

Discourse-sensitive approach

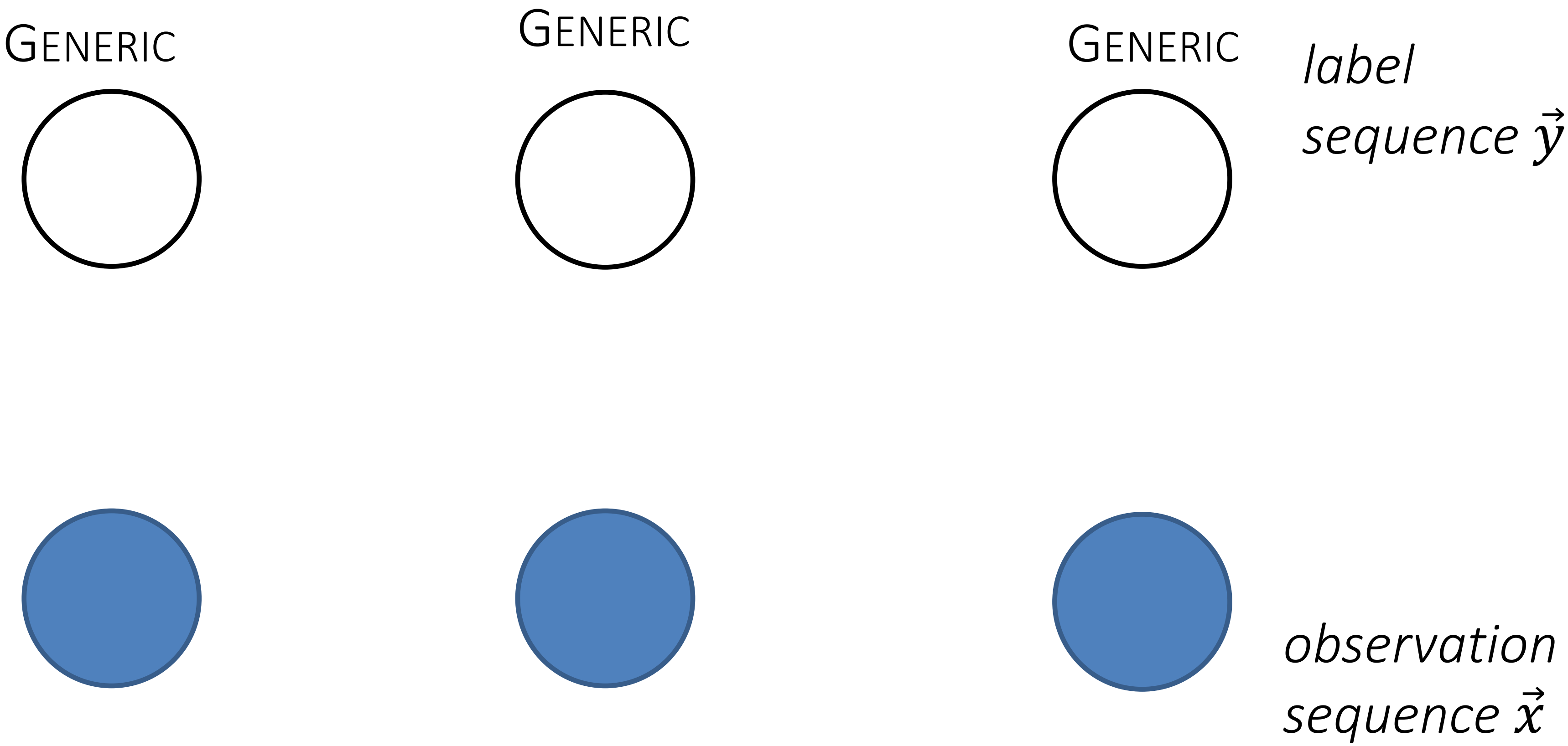
Wikipedia
The Free Encyclopedia

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

genericity labeling of noun phrases in entire texts
→ sequence labeling task

Annemarie Friedrich and Manfred Pinkal. **Discourse-sensitive automatic identification of generic expressions.** ACL 2015.

Conditional random field (CRF)

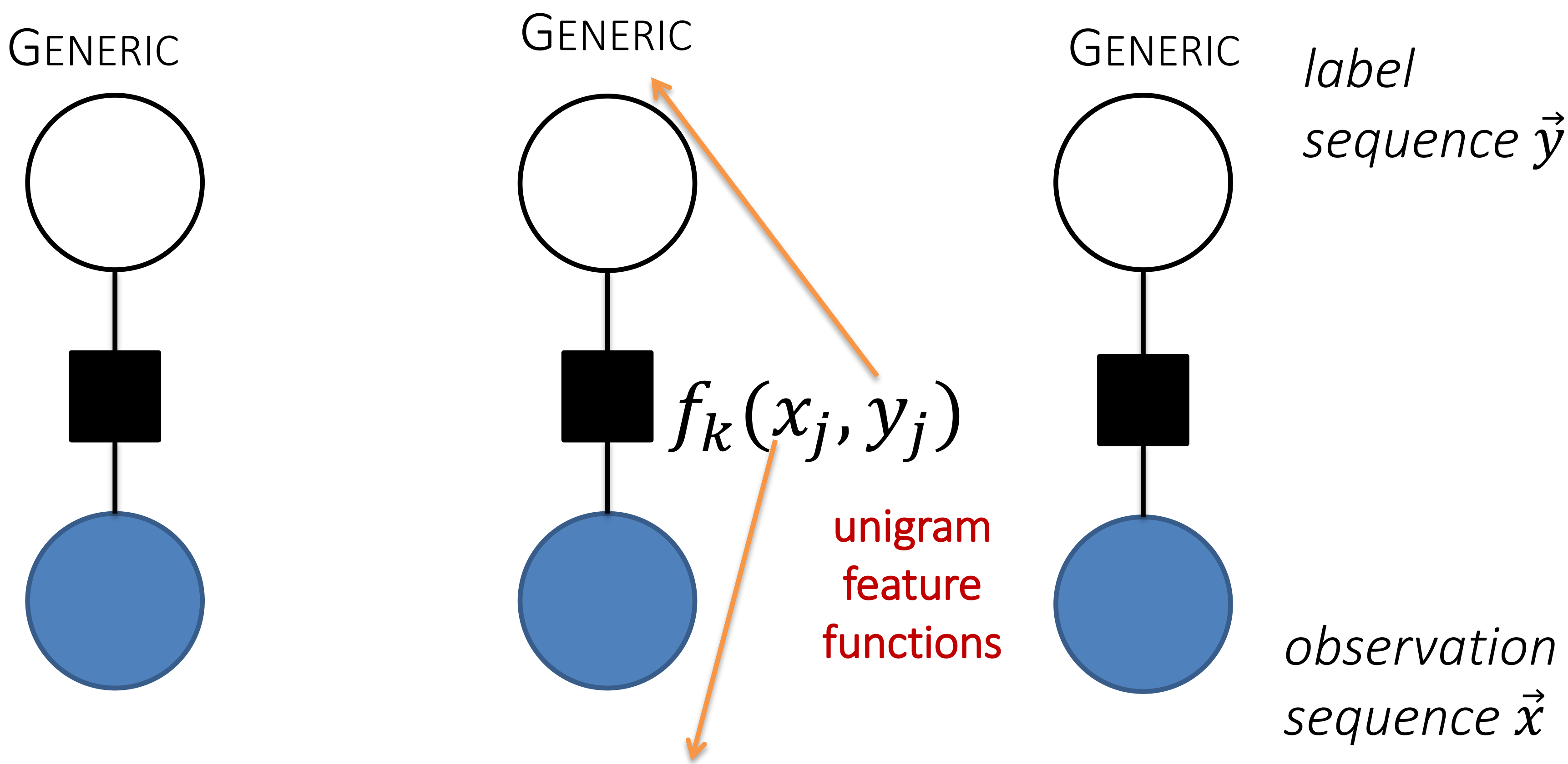


Acer saccharum is a deciduous tree.

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

The recent year's growth twigs are green.

Conditional random field (CRF)

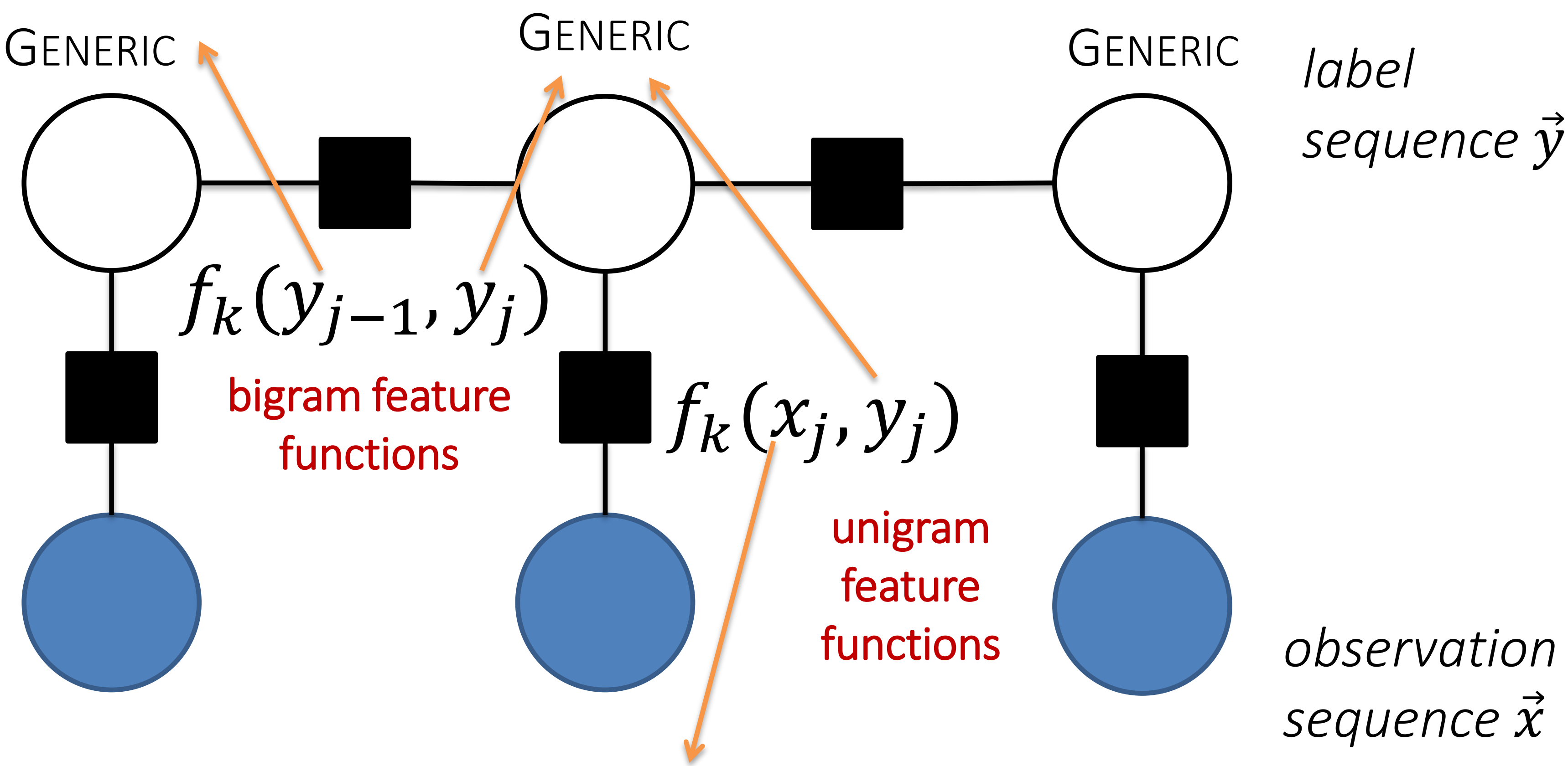


Acer saccharum is a deciduous tree.

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

The recent year's growth twigs are green.

Conditional random field (CRF)

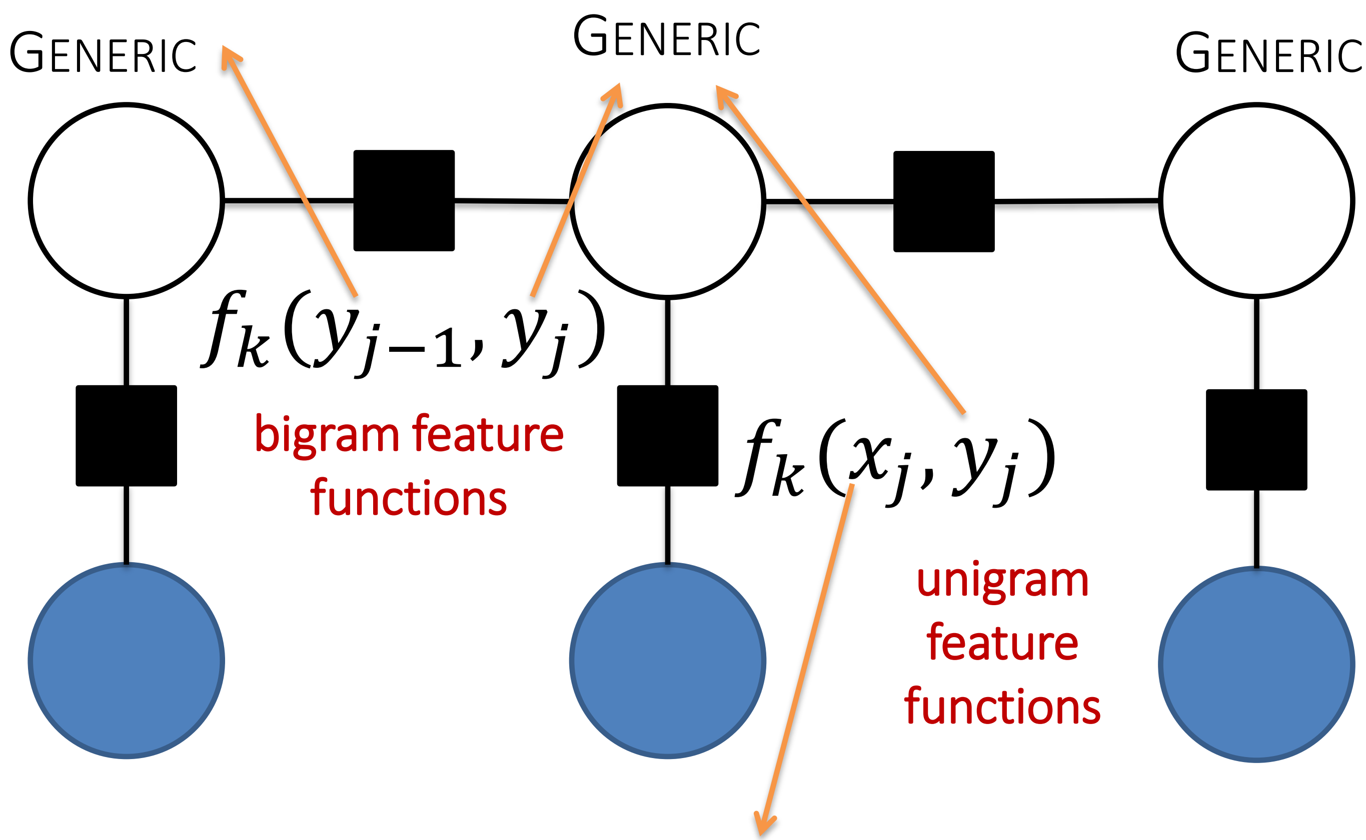


<u>Acer saccharum</u> is a deciduous tree.

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

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

Conditional random field (CRF)



label sequence $\vec{y}$

$$P(\vec{y}|\vec{x}) \sim \sum_k \lambda_k f_k$$

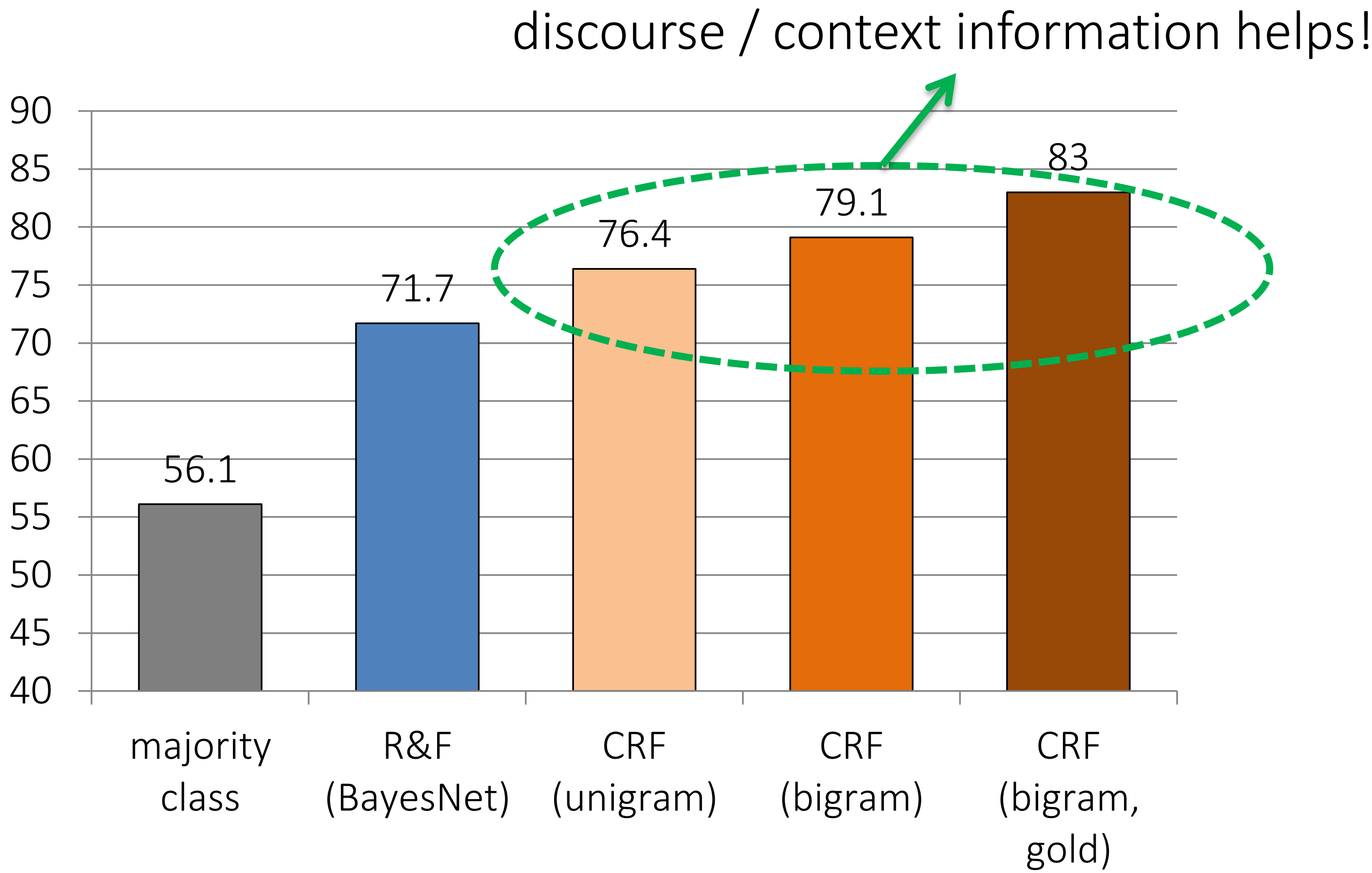
observation sequence $\vec{x}$

<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.

Accuracy: Wikipedia data (main referent)



all differences statistically significant

Computational modeling of situation entity types

[ACL 2015, LAW 2015]

is the main referent generic?

[ACL 2014]

lexical aspectual class

[EMNLP 2015]

recognize habituality

entire documents, segmented into clauses

automatic classification of situation entity types

[ongoing work]

Computational modeling of situation entity types

[ACL 2015, LAW 2015]

is the main referent generic?

[ACL 2014]

lexical aspectual class

[EMNLP 2015]

recognize habituality

entire documents, segmented into clauses

automatic classification of situation entity types

[ongoing work]

Lexical aspectual class

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

Lexical aspectual class

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

Juice **fills** the glass.
stative

Lexical aspectual class

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

Juice **fills** the glass.
stative

The glass **was filled** with juice.
both interpretations possible

Lexical aspectual class

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

Juice **fills** the glass.
stative

The glass **was filled** with juice.
both interpretations possible

Vendler [1957]: time schemata of verbs
lexical aspect / aktionsart

<table>
<tr><td>states</td><td>love, own</td><td>stative</td></tr>
<tr><td>activities</td><td>run</td><td rowspan="3">dynamic</td></tr>
<tr><td>accomplishments</td><td>write a letter</td></tr>
<tr><td>achievements</td><td>realize</td></tr>
</table>

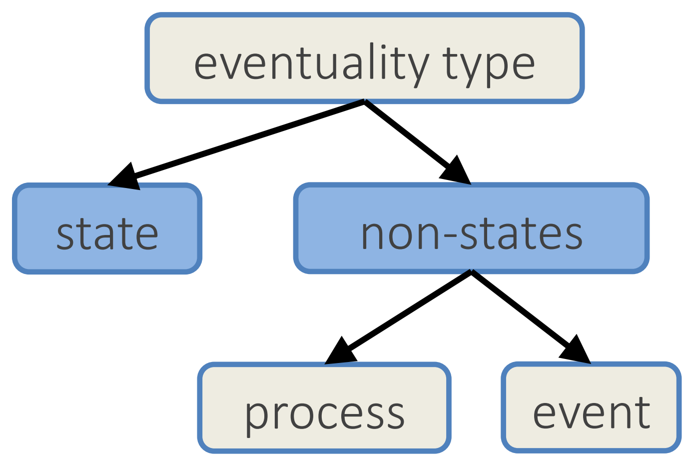
Lexical aspectual class

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

Juice **fills** the glass.
stative

The glass **was filled** with juice.
both interpretations possible

Bach [1986]: time schemata of <u>sentences</u>



Vendler [1957]: time schemata of <u>verbs</u>
lexical aspect / aktionsart

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

Predicting fundamental lexical aspectual class

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

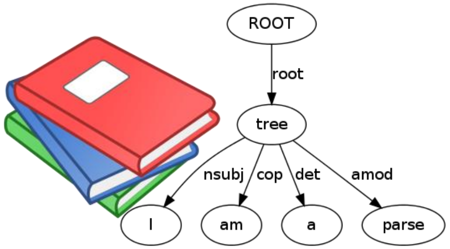
Eric Siegel and Kathleen McKeown. **Learning methods to combine linguistic indicators.** *Computational Linguistics*, 2000.

Predicting fundamental lexical aspectual class

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

Linguistic indicators

large parsed text corpus
(Gigaword)



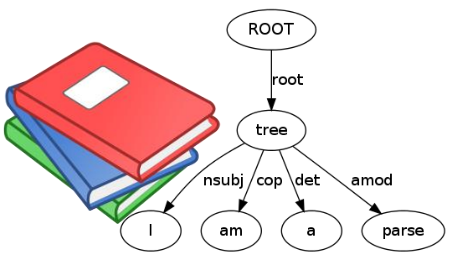
Eric Siegel and Kathleen McKeown. **Learning methods to combine linguistic indicators.** *Computational Linguistics*, 2000.

Predicting fundamental lexical aspectual class

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

Linguistic indicators

large parsed text corpus (Gigaword)



frequency	negated	no subject
present	perfect	evaluation adverb
past	progressive	continuous adverb
future	for-PP	manner adverb
particle	in-PP	temporal adverb

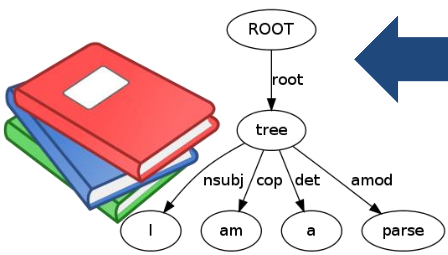
Eric Siegel and Kathleen McKeown. **Learning methods to combine linguistic indicators.** *Computational Linguistics*, 2000.

Predicting fundamental lexical aspectual class

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

Linguistic indicators

large parsed text corpus (Gigaword)



counts for each **verb type**

frequency	negated	no subject
present	perfect	evaluation adverb
past	progressive	continuous adverb
future	for-PP	manner adverb
particle	in-PP	temporal adverb

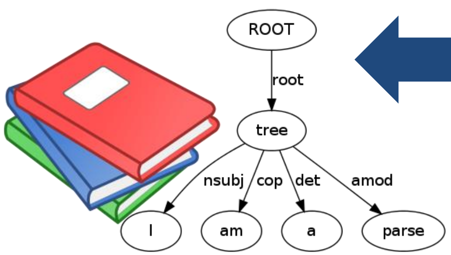
Eric Siegel and Kathleen McKeown. **Learning methods to combine linguistic indicators.** *Computational Linguistics*, 2000.

Predicting fundamental lexical aspectual class

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

Linguistic indicators

large parsed text corpus (Gigaword)



counts for each **verb type**

frequency	negated	no subject
present	perfect	evaluation adverb
past	progressive	continuous adverb
future	for-PP	manner adverb
particle	in-PP	temporal adverb

verb type: drink -- ling_ind_past = 0.0927
→ 9.27% of all instances of *drink* in corpus are in past tense

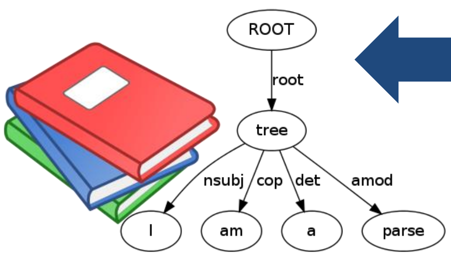
Eric Siegel and Kathleen McKeown. **Learning methods to combine linguistic indicators.** *Computational Linguistics*, 2000.

Predicting fundamental lexical aspectual class

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

Linguistic indicators

large parsed text corpus (Gigaword)



counts for each **verb type**

frequency	negated	no subject
present	perfect	evaluation adverb
past	progressive	continuous adverb
future	for-PP	manner adverb
particle	in-PP	temporal adverb

verb type: drink -- ling_ind_past = 0.0927
→ 9.27% of all instances of *drink* in corpus are in past tense

→15 features for each **verb type**

Eric Siegel and Kathleen McKeown. **Learning methods to combine linguistic indicators.** *Computational Linguistics*, 2000.

Fundamental lexical aspectual class

Eric Siegel and Kathleen McKeown, 2000.

She **filled** the glass with juice.

linguistic indicator features for **fill**:

present	0.0927
negation	0.00024
…	…

training: labeled data

She **likes** flowers. stative

Mary **bought** a cat. dynamic

Random Forest classifier

dynamic

Fundamental lexical aspectual class

Eric Siegel and Kathleen McKeown, 2000.

The glass is **filled** with juice.

She **filled** the glass with juice.

linguistic indicator features for **fill**:

present	0.0927
negation	0.00024
…	…

training: labeled data

She **likes** flowers. stative

Mary **bought** a cat. dynamic

Random Forest classifier

dynamic

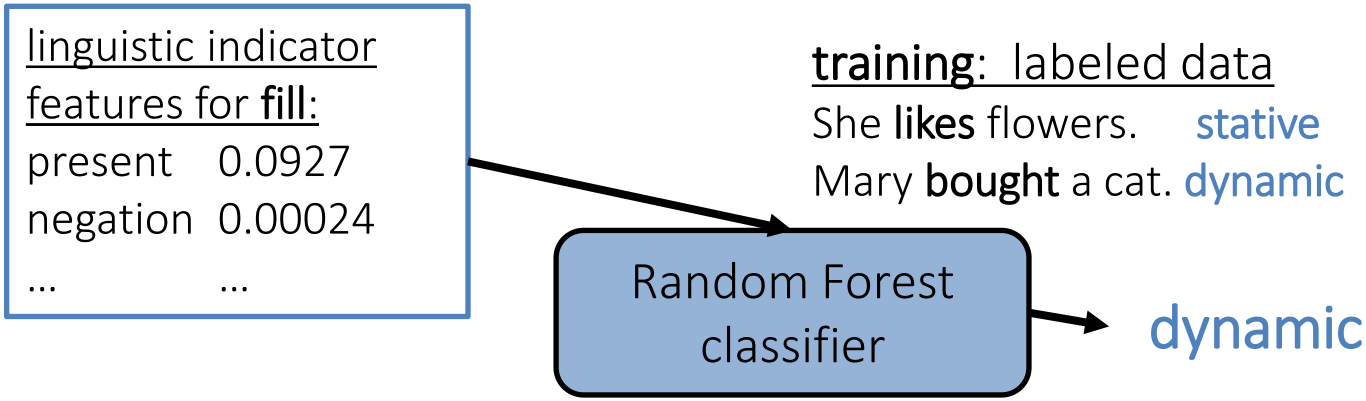
Fundamental lexical aspectual class

Eric Siegel and Kathleen McKeown, 2000.

Classification always results in majority class of verb type. Dataset not available.

The glass is **filled** with juice.

She **filled** the glass with juice.



Fundamental lexical aspectual class

Eric Siegel and Kathleen McKeown, 2000.

Classification always results in majority class of verb type. Dataset not available.

The glass is **filled** with juice.

She **filled** the glass with juice.

linguistic indicator features for **fill**:

present	0.0927
negation	0.00024
…	…

training: labeled data

She **likes** flowers. stative

Mary **bought** a cat. dynamic

~6000 clauses from MASC, complete texts, κ=0.7

Random Forest classifier

dynamic

Annemarie Friedrich and Alexis Palmer. **Automatic prediction of aspectual class of verbs in context.** ACL 2014.

Fundamental lexical aspectual class

Eric Siegel and Kathleen McKeown, 2000.

Classification always results in majority class of verb type. Dataset not available.

The glass is **filled** with juice.

She **filled** the glass with juice.

linguistic indicator features for **fill**:

present	0.0927
negation	0.00024
…	…

Random Forest classifier → dynamic

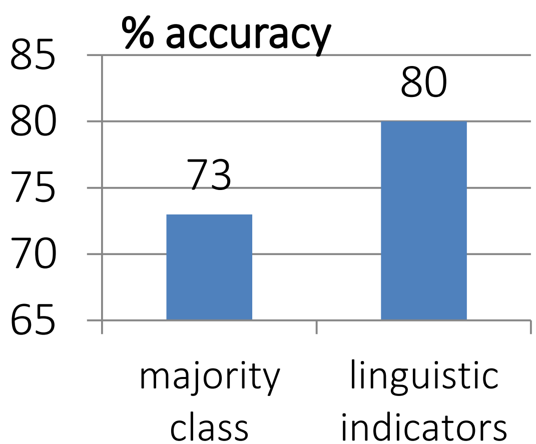
training: labeled data

She **likes** flowers. stative

Mary **bought** a cat. dynamic

~6000 clauses from MASC, complete texts, κ=0.7

10-fold cross validation: UNSEEN VERBS



linguistic indicators generalize across verb types

Annemarie Friedrich and Alexis Palmer. **Automatic prediction of aspectual class of verbs in context.** ACL 2014.

Fundamental lexical aspectual class

linguistic indicator features for **fill**:

present	0.0927
negation	0.00024
…	…

instance-based features for clause:

tense	past
subject	noun.person
voice	active
…	…

training/test: labeled data

She **likes** flowers. stative

Mary **bought** a cat. dynamic

Random Forest classifier

dynamic

Annemarie Friedrich and Alexis Palmer. **Automatic prediction of aspectual class of verbs in context.** ACL 2014.

Fundamental lexical aspectual class

linguistic indicator features for **fill**:

present	0.0927
negation	0.00024
…	…

instance-based features for clause:

tense	past
subject	noun.person
voice	active
…	…

training/test: labeled data

She **likes** flowers. stative

Mary **bought** a cat. dynamic

2667 sentences from Brown corpus for 20 frequent ambiguous verbs

2 annotators, κ = 0.6

Leave-One-Out CV

Random Forest classifier → dynamic

Annemarie Friedrich and Alexis Palmer. **Automatic prediction of aspectual class of verbs in context.** ACL 2014.

Fundamental lexical aspectual class

linguistic indicator features for **fill**:

present	0.0927
negation	0.00024
…	…

instance-based features for clause:

tense	past
subject	noun.person
voice	active
…	…

training/test: labeled data

She **likes** flowers. stative

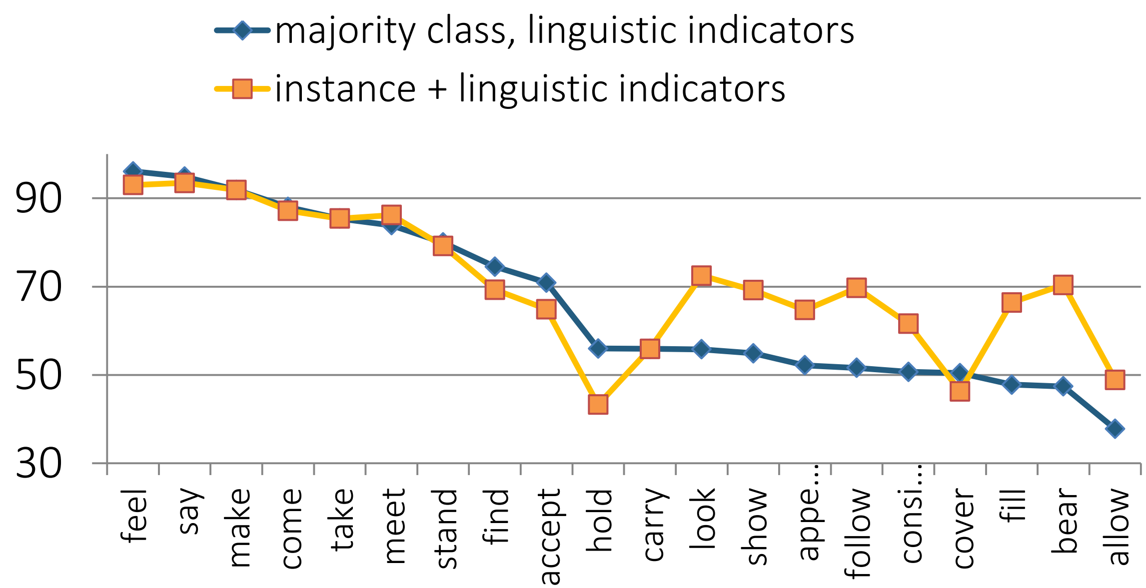
Mary **bought** a cat. dynamic

2667 sentences from Brown corpus for 20 frequent ambiguous verbs

2 annotators, κ = 0.6

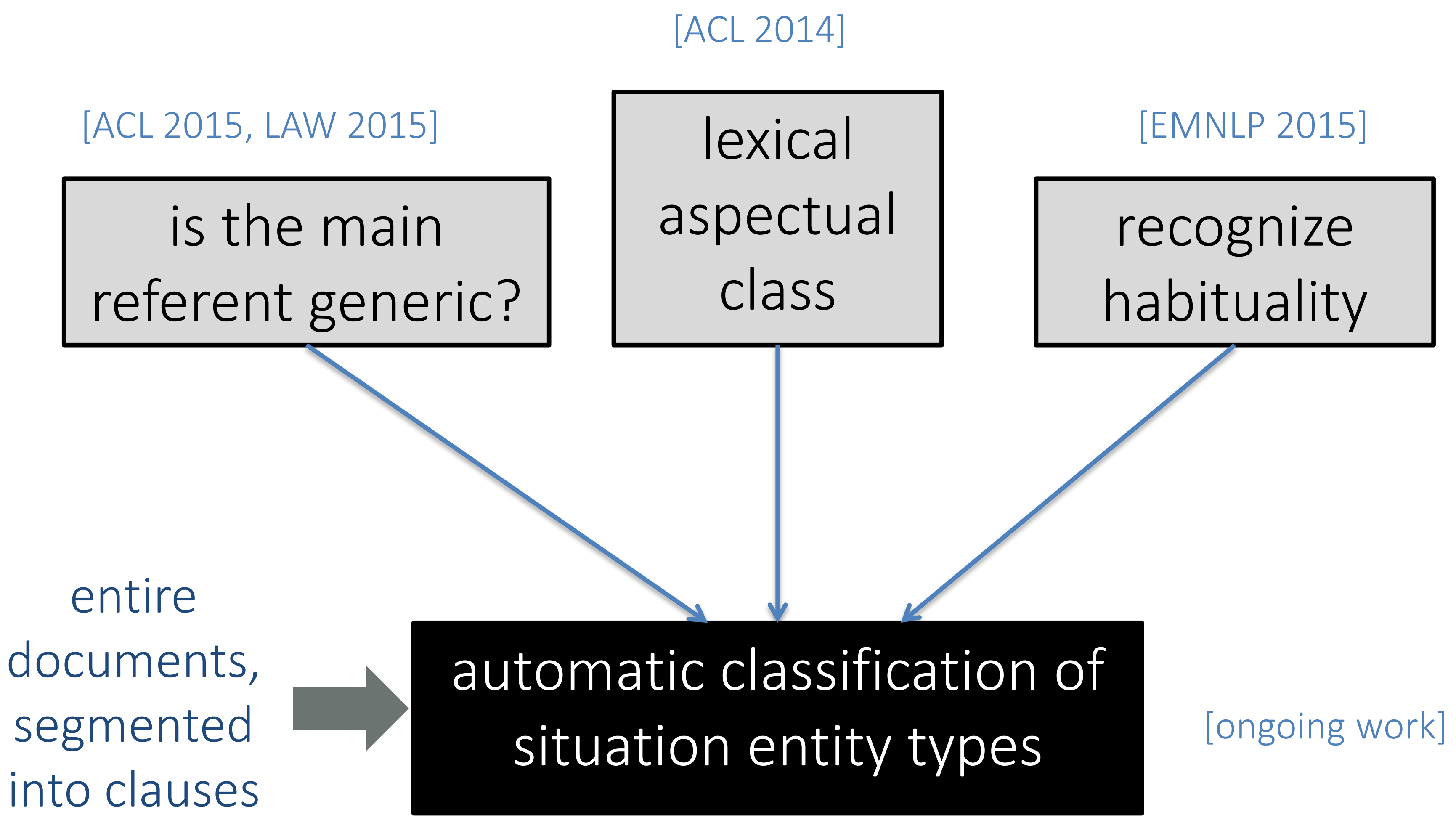
Leave-One-Out CV

Random Forest classifier → dynamic

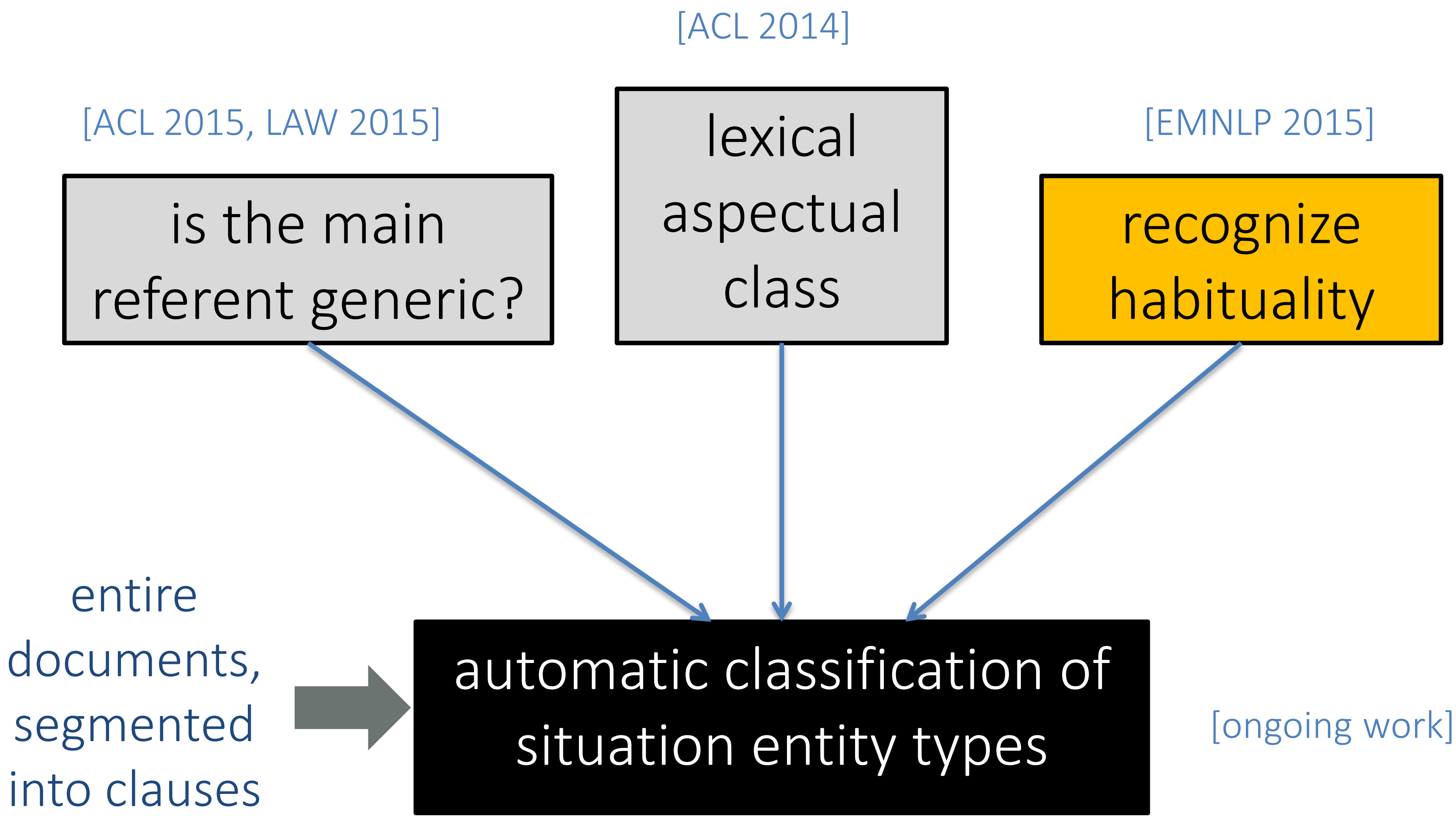


Annemarie Friedrich and Alexis Palmer. **Automatic prediction of aspectual class of verbs in context.** ACL 2014.

Computational modeling of situation entity types



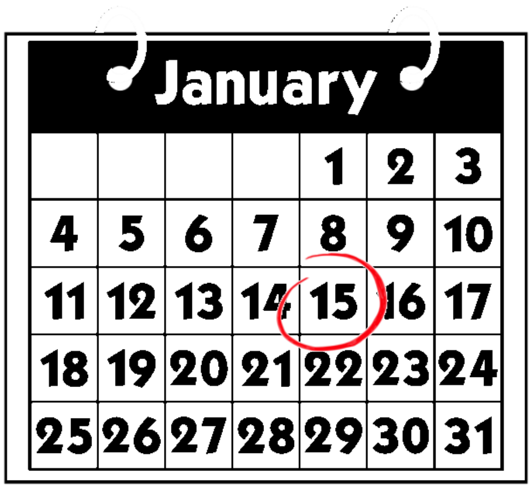
Computational modeling of situation entity types



Habituality

episodic

a particular event

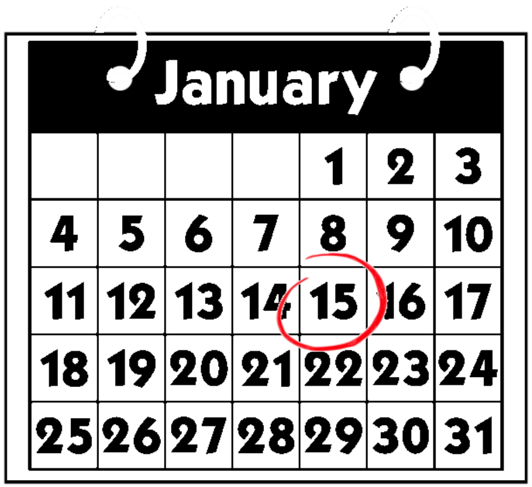


John went swimming yesterday!

Habituality

episodic

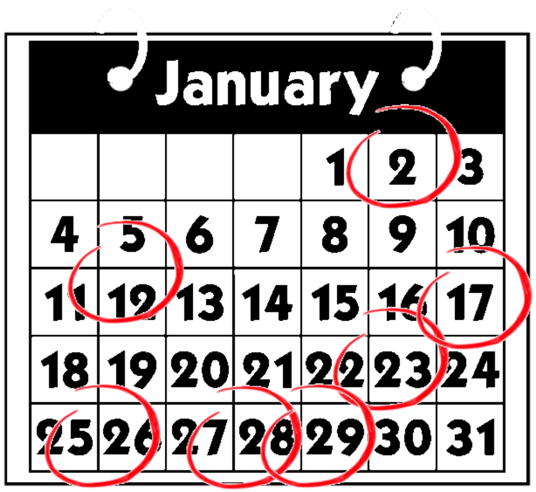
a particular event



John went swimming yesterday!

habitual

generalization over situations, exceptions are tolerated

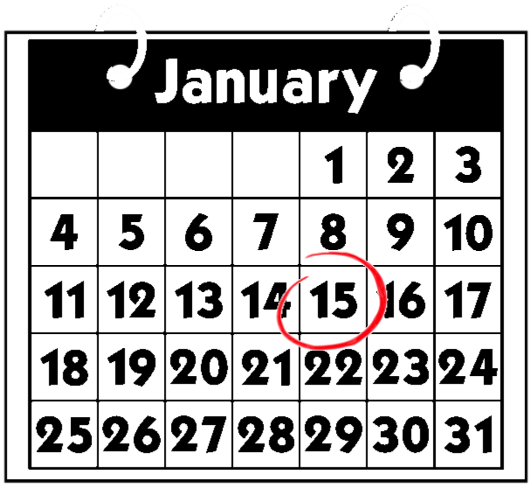


Bill often goes swimming.

Habituality

episodic

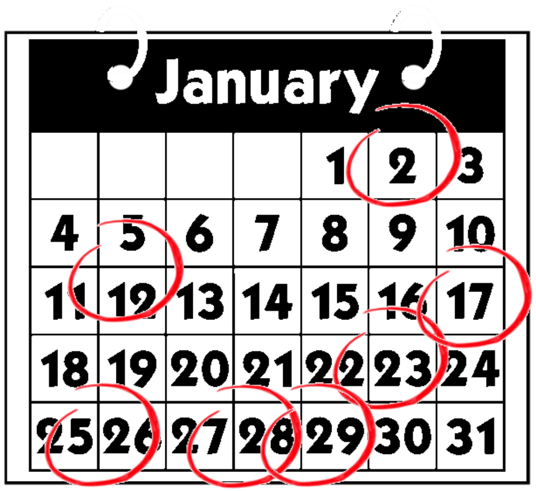
a particular event



John went swimming yesterday!

habitual

generalization over situations, exceptions are tolerated



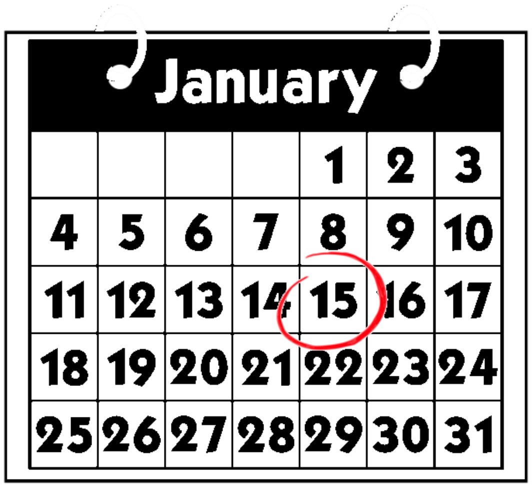
Bill often goes swimming.

Thomas Mathew and Graham Katz. **Supervised categorization of habitual and episodic sentences.** *Sixth Midwest Computational Linguistics Colloquium, Bloomington, Indiana*. 2009.

Habituality

episodic

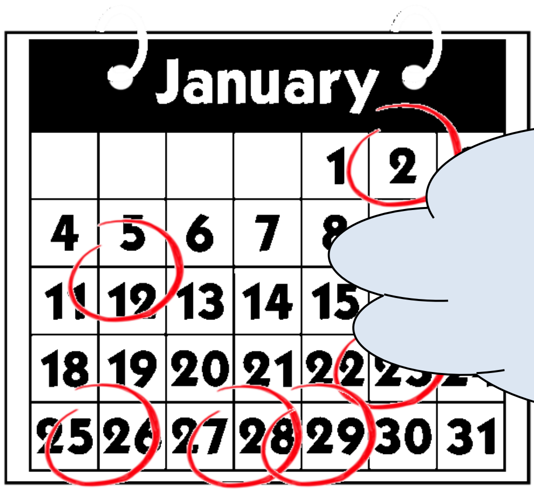
a particular event



John went swimming yesterday!

habitual

generalization over situations, exceptions are tolerated



Bill often goes swimming.

Bill **likes** coffee.
Bill **didn‘t** go swimming.
Bill **can** swim.

Thomas Mathew and Graham Katz. **Supervised categorization of habitual and episodic sentences.** *Sixth Midwest Computational Linguistics Colloquium, Bloomington, Indiana*. 2009.

A three-way classification of clausal aspect

clausal aspect		lexical aspect
episodic	Bill **drank** a coffee after lunch.	*dynamic*

Annemarie Friedrich and Manfred Pinkal. **Recognising habituals: a three-way classification of clausal aspect.** EMNLP 2015.

A three-way classification of clausal aspect

clausal aspect		lexical aspect
episodic	Bill **drank** a coffee after lunch.	*dynamic*
habitual	Bill *usually* **drinks** coffee after lunch.	*dynamic*
	Italians **drink** coffee after lunch.	*dynamic*
	Sloths *sometimes* **sit** on top of branches.	*stative*
	John *never* **drinks** coffee.	*dynamic*

Annemarie Friedrich and Manfred Pinkal. **Recognising habituals: a three-way classification of clausal aspect.** EMNLP 2015.

A three-way classification of clausal aspect

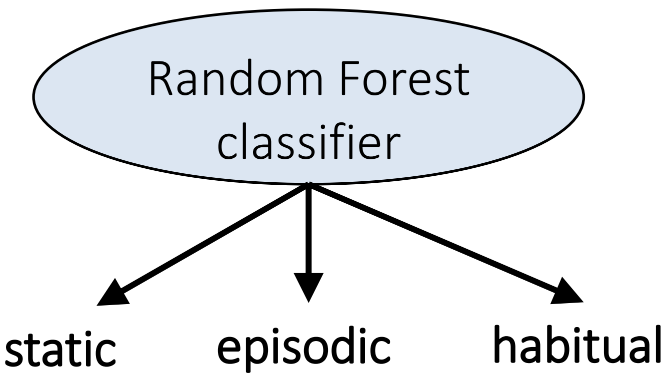
clausal aspect		lexical aspect
episodic	Bill **drank** a coffee after lunch.	*dynamic*
habitual	Bill *usually* **drinks** coffee after lunch.	*dynamic*
	Italians **drink** coffee after lunch.	*dynamic*
	Sloths *sometimes* **sit** on top of branches.	*stative*
	John *never* **drinks** coffee.	*dynamic*
static	Bill **likes** coffee.	*stative*
	Bill *can* **swim**.	*dynamic*
	Bill *didn't* **drink** coffee yesterday.	*dynamic*
	Mary *has* **made** a cake.	*dynamic*

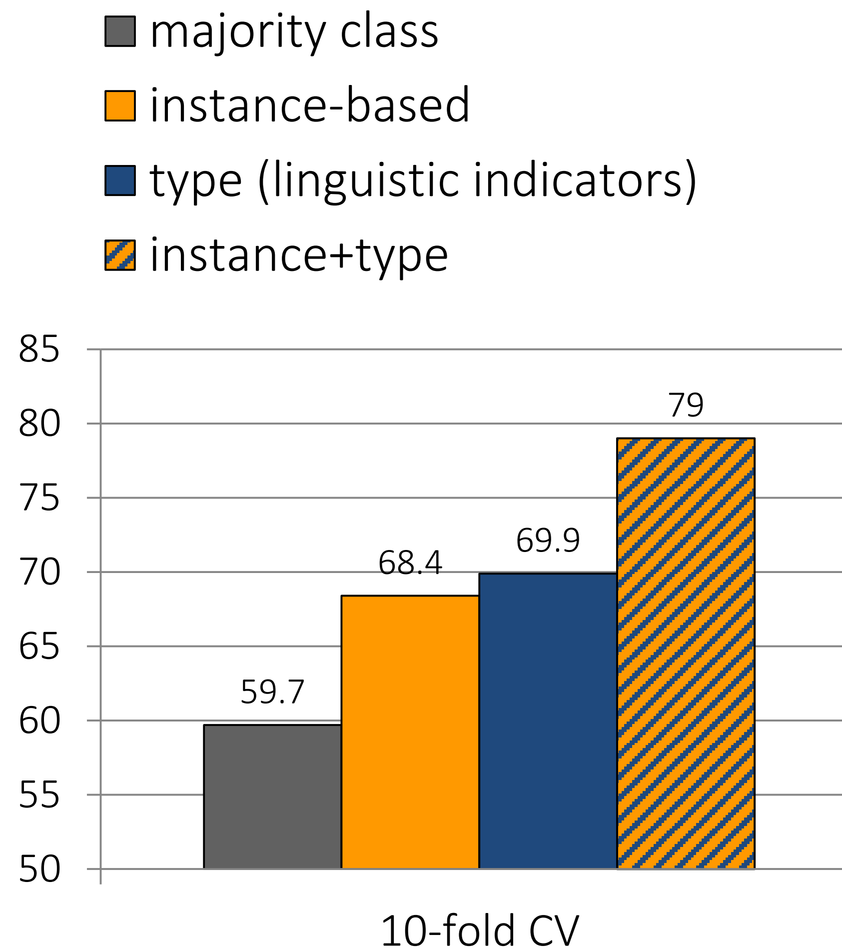
Annemarie Friedrich and Manfred Pinkal. **Recognising habituals: a three-way classification of clausal aspect.** EMNLP 2015.

Automatic classification of clausal aspect

102 texts, 10355 clauses
3 annotators, κ=0.61

60% **static**
20% **episodic**
20% **habitual**





Annemarie Friedrich and Manfred Pinkal. **Recognising habituals: a three-way classification of clausal aspect.** EMNLP 2015.

Automatic classification of clausal aspect

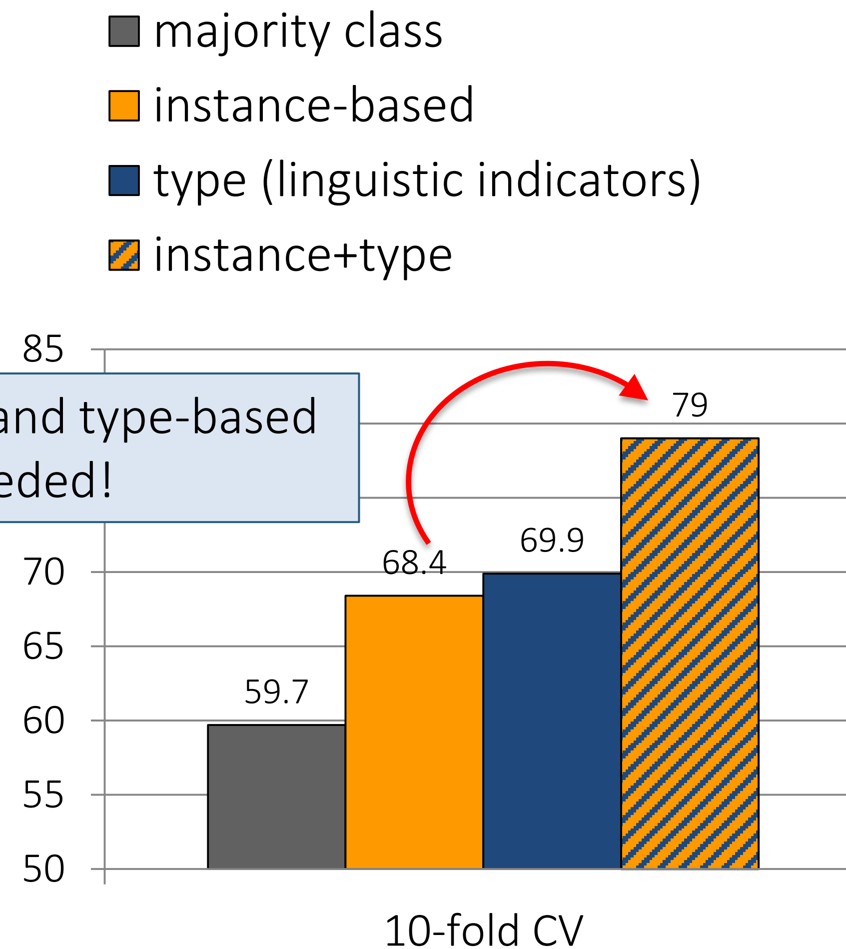
102 texts, 10355 clauses
3 annotators, κ=0.61

60% **static**
20% **episodic**
20% **habitual**

Random Forest classifier

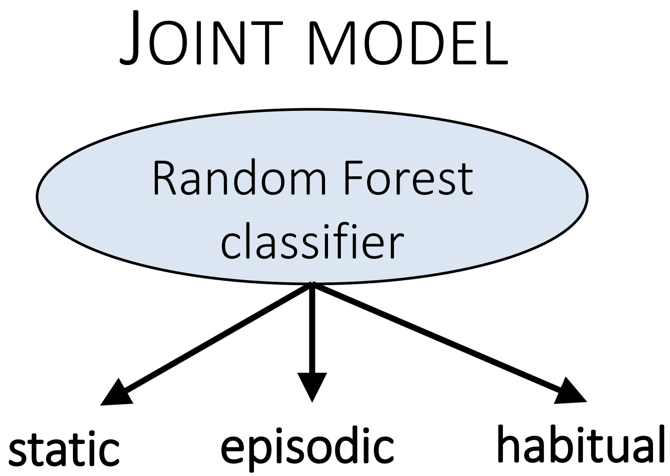
static **episodic** **habitual**

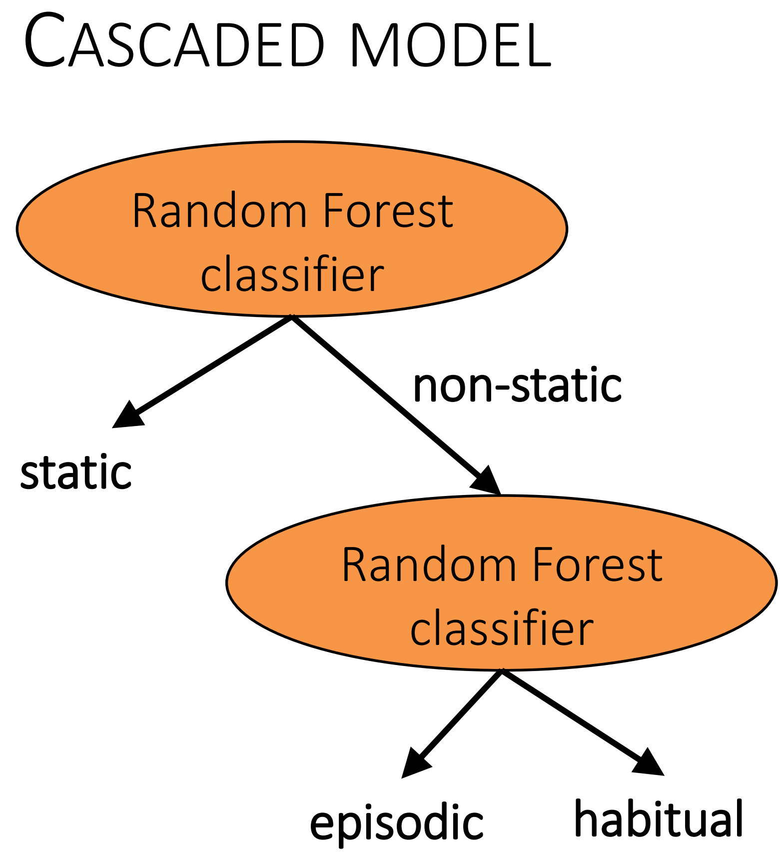
Both instance- and type-based features are needed!



Annemarie Friedrich and Manfred Pinkal. **Recognising habituals: a three-way classification of clausal aspect.** EMNLP 2015.

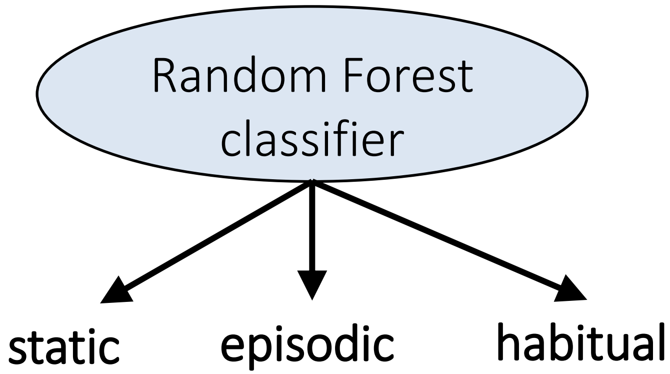
Automatic classification of clausal aspect



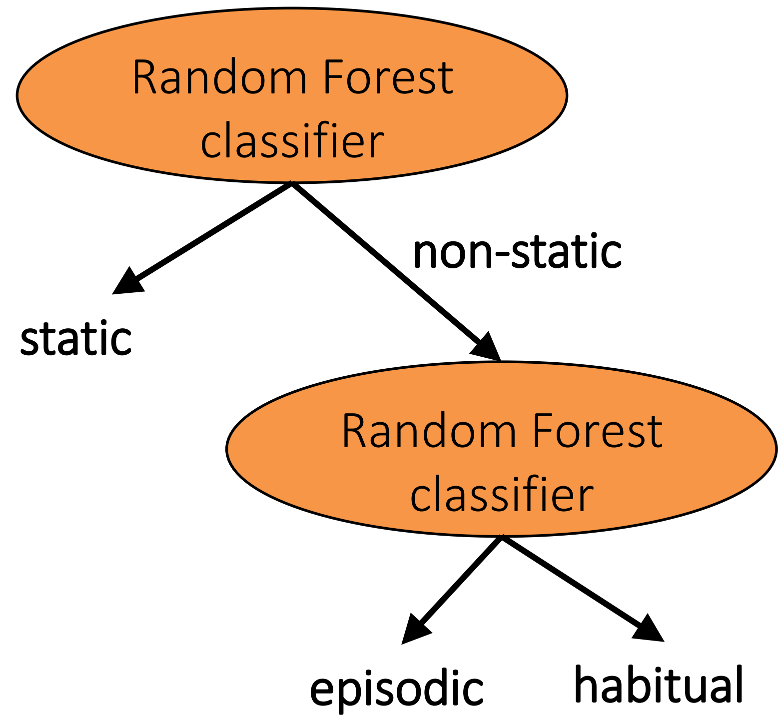


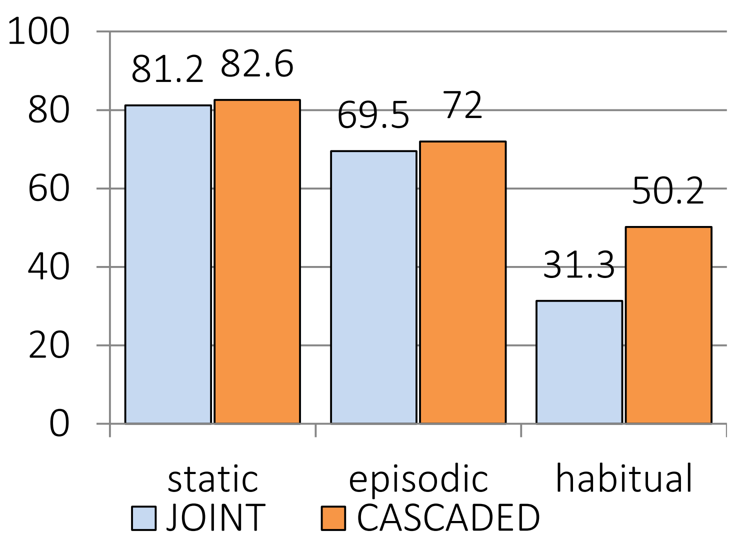
Automatic classification of clausal aspect

JOINT MODEL



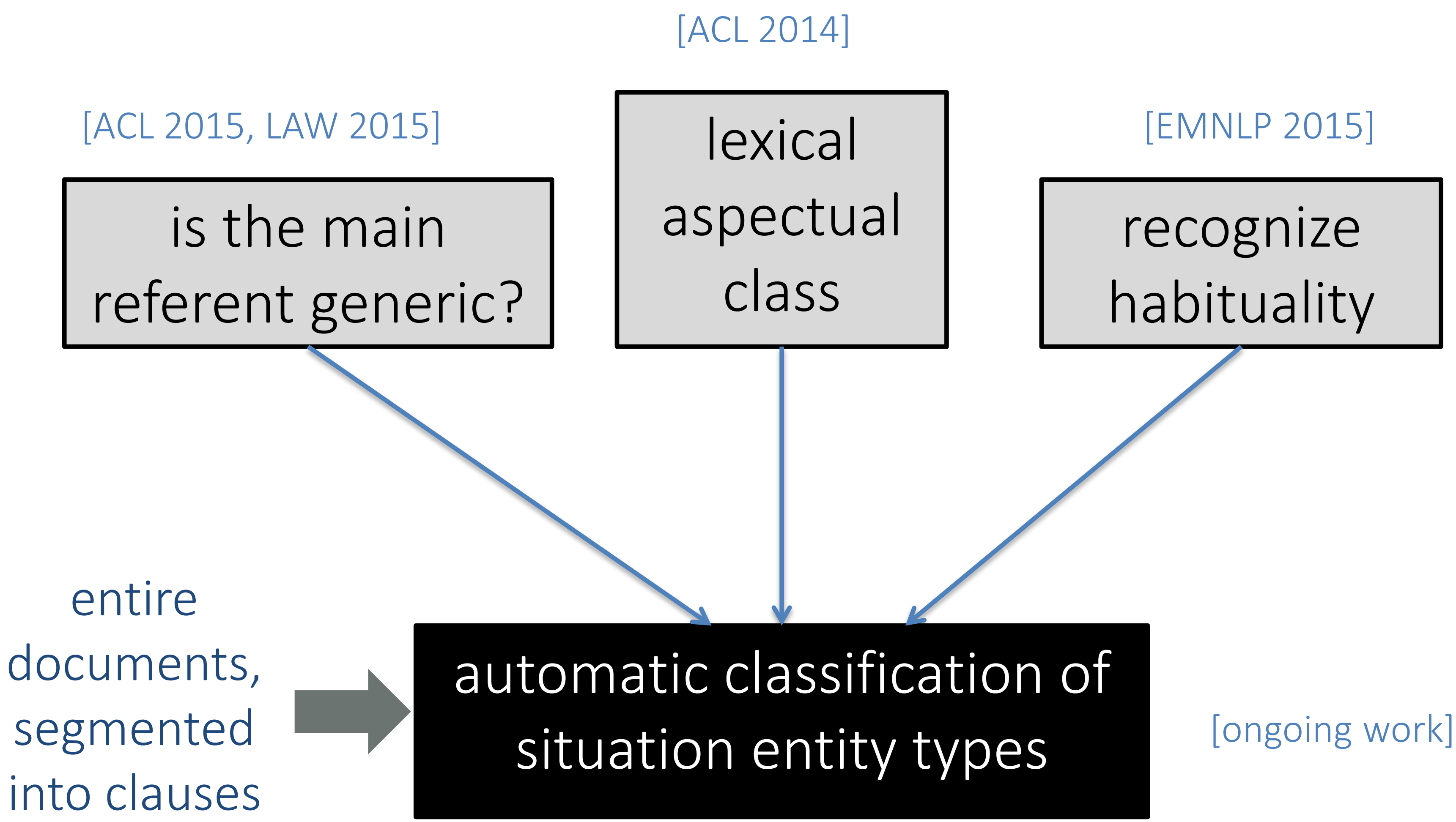
CASCADED MODEL



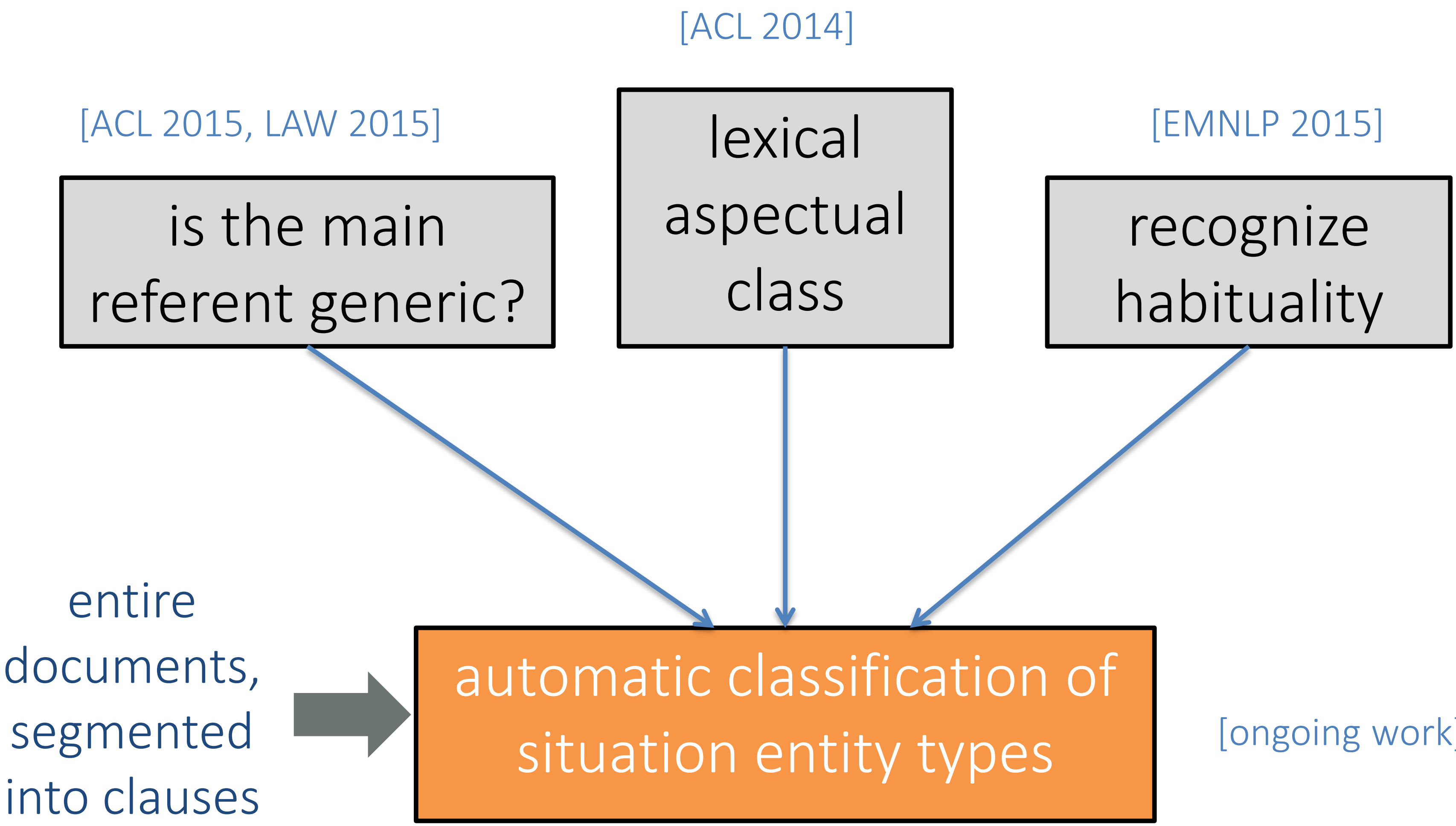


Cascaded model improves identification of habituals in free text.

Computational modeling of situation entity types



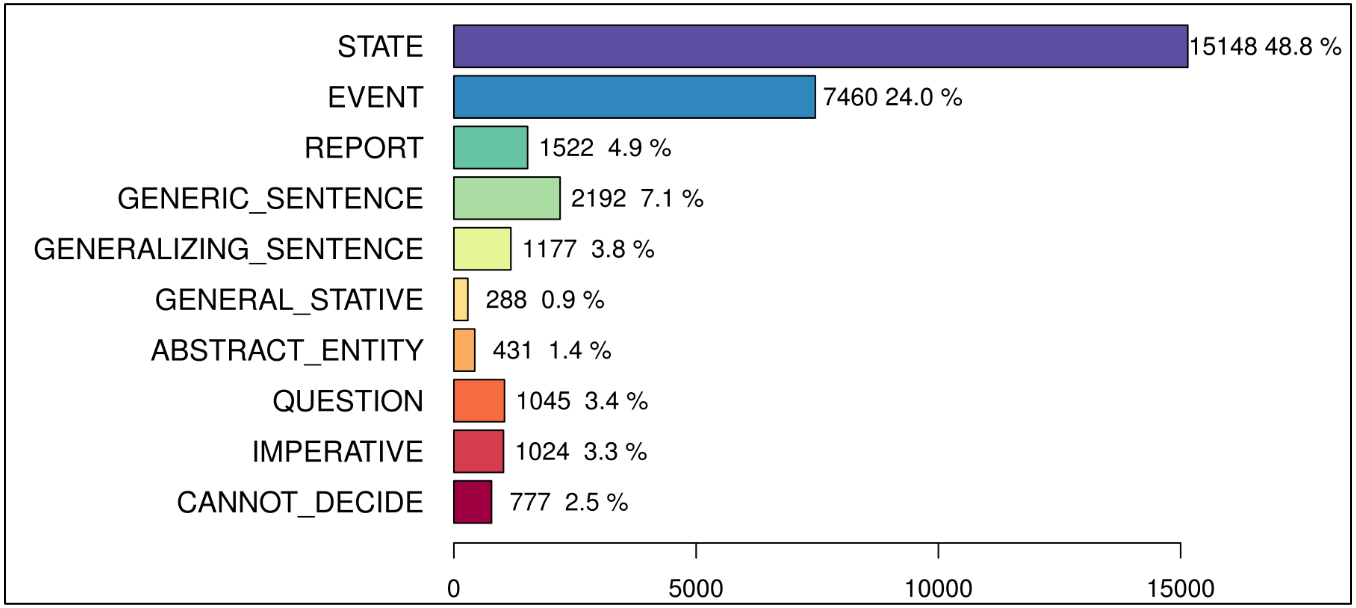
Computational modeling of situation entity types



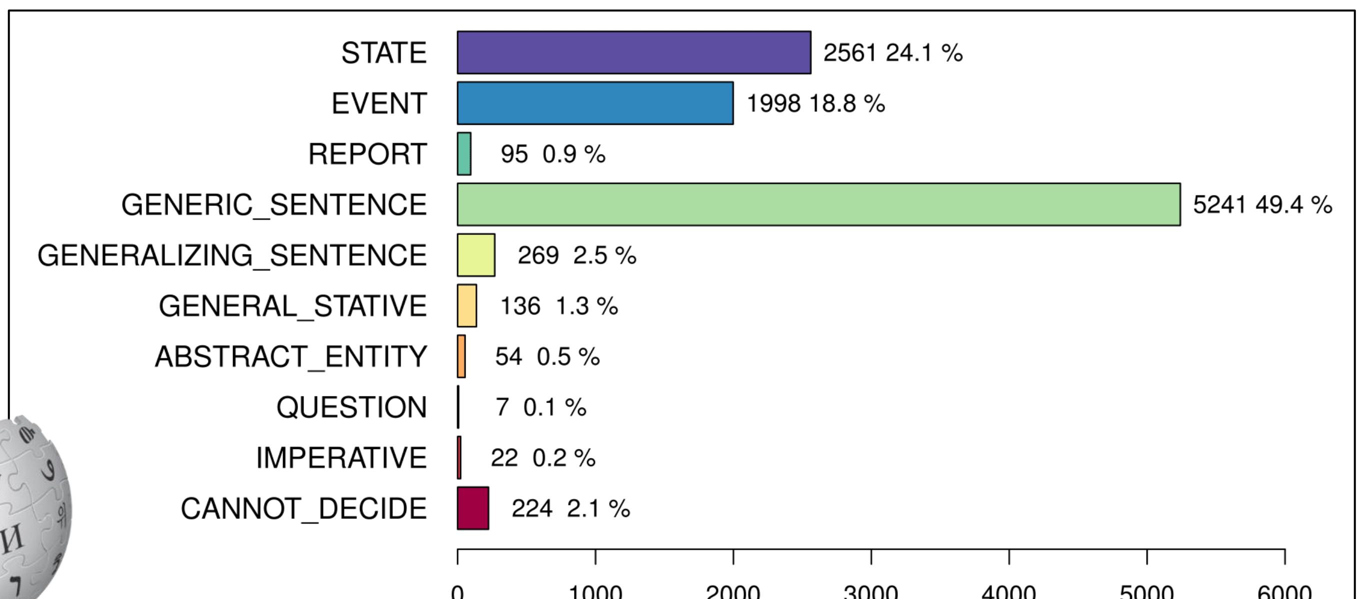
Situation entity type distributions

MASC

- blog
- email
- essays
- ficlets
- fiction
- govt-docs
- jokes
- journal
- letters
- news
- technical
- travel

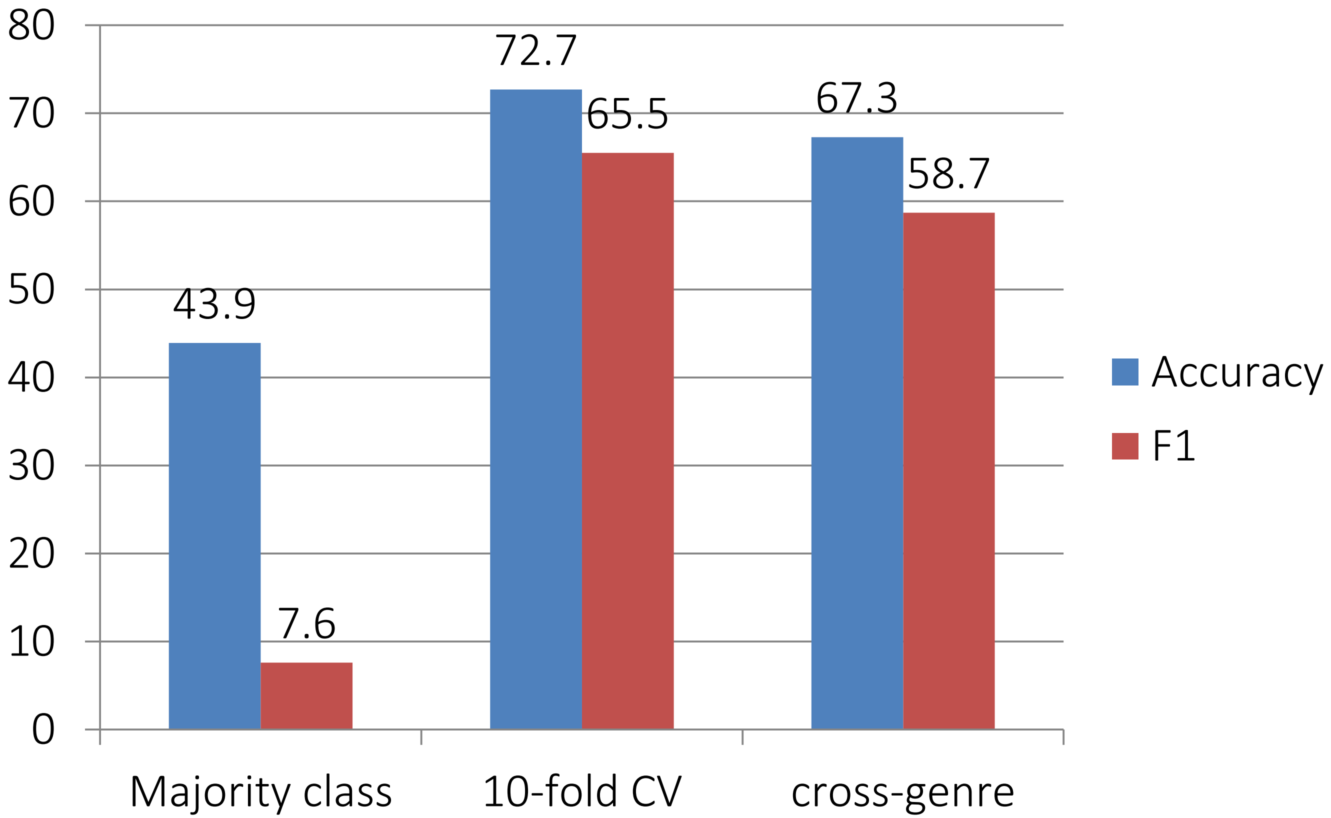


Wikipedia



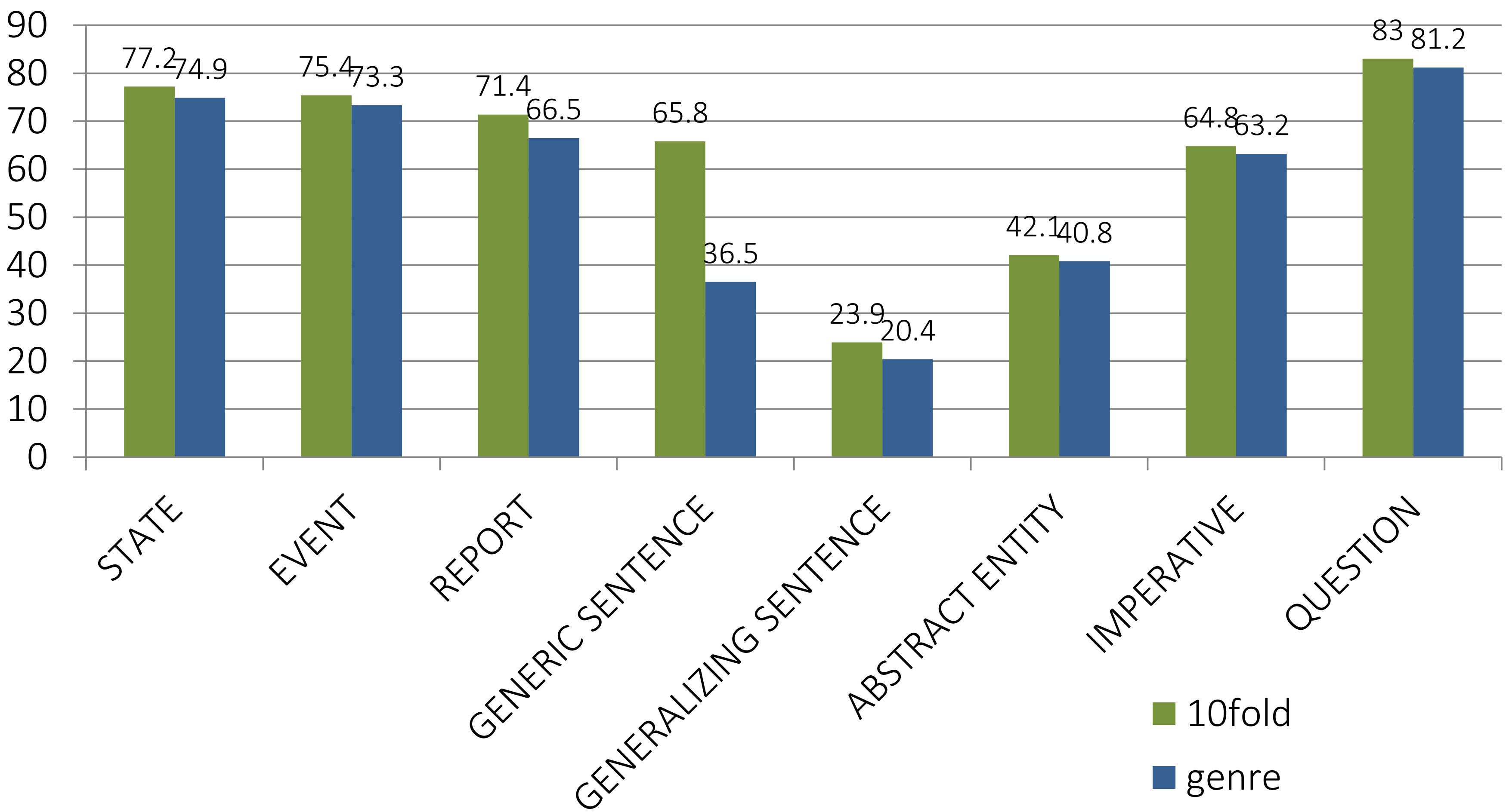
Situation entity types (intermediate results)

development set, ~32550 clauses from MASC+Wiki
8-way classification task
Conditional Random Field, selection of syntactic-semantic features



Situation entity types (intermediate results)

development set, ~32550 clauses
8-way classification task
Conditional Random Field



On-going / future work

- improving classification of **situation entity types**

On-going / future work

- improving classification of **situation entity types**
- investigate interaction of prediction of **features** (main referent, clausal aspect) and **situation entity types**

On-going / future work

- improving classification of **situation entity types**
- investigate interaction of prediction of **features** (main referent, clausal aspect) and **situation entity types**
- investigate impact of different **genres / domains**

On-going / future work

- improving classification of **situation entity types**
- investigate interaction of prediction of **features** (main referent, clausal aspect) and **situation entity types**
- investigate impact of different **genres / domains**
- create models for labeling situation entity types and **discourse modes**

On-going / future work

- improving classification of **situation entity types**
- investigate interaction of prediction of **features** (main referent, clausal aspect) and **situation entity types**
- investigate impact of different **genres / domains**
- create models for labeling situation entity types and **discourse modes**
- integrate situation entity type information in computational models of discourse, e.g., identification of **coherence relations** or **temporal processing**

On-going / future work

- improving classification of **situation entity types**
- investigate interaction of prediction of **features** (main referent, clausal aspect) and **situation entity types**
- investigate impact of different **genres / domains**
- create models for labeling situation entity types and **discourse modes**
- integrate situation entity type information in computational models of discourse, e.g., identification of **coherence relations** or **temporal processing**
- other languages (extend work of Mavridou et al. 2015)

Summary

- Groundwork for computational models of a novel approach to discourse analysis: complementary to existing approaches such as RST, Penn / Prague DTB, SDRT.

Summary

- Groundwork for computational models of a novel approach to discourse analysis: complementary to existing approaches such as RST, Penn / Prague DTB, SDRT.
- Computational modeling of various aspectual distinctions (habituality, lexical aspectual class): useful for text understanding tasks such as temporal processing

Summary

- Groundwork for computational models of a novel approach to discourse analysis: complementary to existing approaches such as RST, Penn / Prague DTB, SDRT.
- Computational modeling of various aspectual distinctions (habituality, lexical aspectual class): useful for text understanding tasks such as temporal processing
- Recognition of genericity: knowledge acqusition from text

Summary

- Groundwork for computational models of a novel approach to discourse analysis: complementary to existing approaches such as RST, Penn / Prague DTB, SDRT.
- Computational modeling of various aspectual distinctions (habituality, lexical aspectual class): useful for text understanding tasks such as temporal processing
- Recognition of genericity: knowledge acqusition from text

different types of clauses contribute differently to structure of discourse

Summary

- Groundwork for computational models of a novel approach to discourse analysis: complementary to existing approaches such as RST, Penn / Prague DTB, SDRT.
- Computational modeling of various aspectual distinctions (habituality, lexical aspectual class): useful for text understanding tasks such as temporal processing
- Recognition of genericity: knowledge acqusition from text

different types of clauses contribute differently to structure of discourse

Summary

- Groundwork for computational models of a novel approach to discourse analysis: complementary to existing approaches such as RST, Penn / Prague DTB, SDRT.
- Computational modeling of various aspectual distinctions (habituality, lexical aspectual class): useful for text understanding tasks such as temporal processing
- Recognition of genericity: knowledge acqusition from text

different types of clauses contribute differently to structure of discourse

Thank you!

References

- Jean-Michel Adam. *Les textes: types et prototypes: récit, description, argumentation, explication et dialogue*. Armand Colin, 2011.
- Alexis Palmer, Elias Ponvert, Jason Baldridge, and Carlota Smith. A sequencing model for situation entity classification. ACL 2007.
- Kleio-Isidora Mavridou, Annemarie Friedrich, Melissa Peate Sørensen, Alexis Palmer and Manfred Pinkal: Linking discourse modes and situation entity types in a cross-linguistic corpus study. September 2015. LSDSem. Lisbon, Portugal.
- Carlota S Smith. *Modes of discourse: The local structure of texts*, volume 103. Cambridge University Press, 2003.
- Carlota S Smith. Aspectual entities and tense in discourse. In *Aspectual inquiries*, pages 223–237. Springer, 2005.
- Radu Soricut and Daniel Marcu. Sentence Level Discourse Parsing using Syntactic and Lexical Information. NAACL 2003.
- Egon Werlich. *Typologie der Texte*. UTB für Wissenschaft, 1989.

Computational model for genericity

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

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

GENERIC

The recent year's growth twigs are green.

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

GENERIC

CRF

sequence of clauses (entire document)

features: indicator functions

sequence of labels

Linear-chain Conditional Random Field

sum over observations in $\vec{x}$

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

sum over feature functions

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

weights for feature functions

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

Discriminative training
(maximum likelihood, CRF++ toolkit)

Accuracy: ACE-2 and ACE-2005