Corpus-based Learning of Formal Semantic Concepts:

Genericity and Presupposition

Anette Frank
Universität Heidelberg

Séminaire de Recherche en Linguistique
Université de Geneve
Département de Linguistique

9. April 2013

Corpus-based learning of formal semantic concepts

Distributional semantics: Novel research questions

- Can statistical semantics profit from formal semantics?
- Can formal semantics profit from statistical semantics?

Today's talk: Looking at Presupposition and Genericity

Hypotheses

- Formal semantics can deliver crucial insights to guide statistical models of semantics
- Statistical semantics can yield novel insights for formal semantics

Outline

Corpus-based learning of formal semantic concepts

- Statistical models of semantics
- **Presupposition:** Discriminative learning of fine-grained semantic relations between verbs
- **Genericity:** Classifying generic NPs and generic sentences
- Wrap-up

Acknowledgements

This talk reports research conducted by my doctoral students **Galina Tremper** and **Nils Reiter** under my (collaborative) supervision.

Statistical Models of Semantics

Distributional Hypothesis

- *"a word is characterized by the company it keeps"* (Firth, 1957)
- *"words which are similar in meaning occur in similar contexts"*

  (Rubenstein & Goodenough, 1965)
- *"words with similar meanings will occur with similar neighbors if enough text material is available"'* (Schütze & Pedersen, 1995)
- *"words that occur in the same contexts tend to have similar meanings"* (Pantel, 2005)

Statistical Models of Semantics

Lexical Distributional Semantics

- Word Senses and Sense Disambiguation
- Semantic Similarity and 'Semantic Relatedness'
- Meaning relations: Synonymy, Antonymy, Hyponymy, Meronymy, Causation

Approaches

- Pattern-based Acquisition (Hearst 1992, Pantel & Pennacciotti 2006)
- Contextual features: word-level, syntactic, semantic
- Vector Space Models (VSM) (Schütze 1998)
- Compositional VSMs
- Textual Entailment, 'Natural Logic'

Questions to Statistical Semantics

How do theoretical linguistic concepts align with corpus-based, statistical models of semantics?

- Can theoretical-linguistic concepts guide statistical models, to make them more effective?
- Can corpus-based, statistical models of semantics contribute novel insights for linguistic theory?

Presupposition: Discriminative learning of fine-grained semantic relations between verbs

(Tremper and Frank, DGfS 2011)

(Tremper and Frank, to appear, Discourse&Dialogue)

Drawing Inferences about Events

Lexical presupposition and entailment relations between verbs

(1) Spain *won* the finals of the 2010 World Cup.
$\vdash$ Spain *played* the finals of the 2010 World Cup.

(2) President John F. Kennedy was *assassinated*.
$\vdash$ President John F. Kennedy *died*.

$\rightarrow$ **Inferential relations between verbs are crucial for NLU**

Presupposition is preserved under Negation (Persistence)

(3) a. President John F. Kennedy was *not assassinated*.
$\nvdash$ President John F. Kennedy *died*.

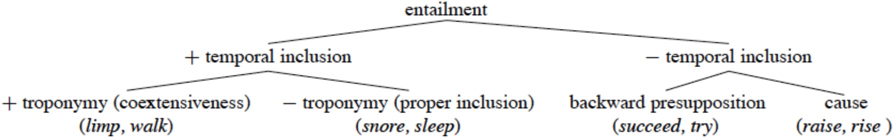
b. Spain *didn't win* the finals of the 2010 World Cup.
$\vdash$ Spain *played* the finals of the 2010 World Cup.

$\rightarrow$ **Presupposition and entailment need to be distinguished**

Acquisition of lexical semantic relations

WordNet

- Synonymy, antonymy, hypernymy (troponymy), meronymy
- Verb entailment relations

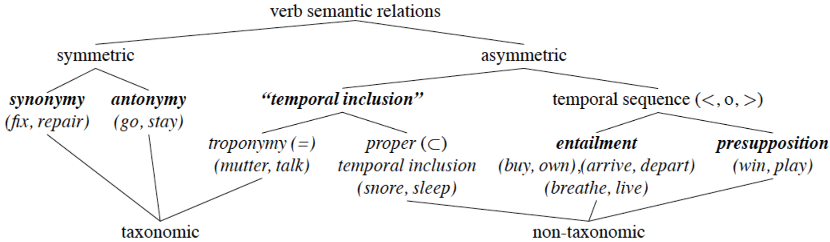


VerbOcean (Chklovski and Pantel, 2004)

- automatic acquisition of semantic relations between verbs: *similarity, strength, antonymy, enablement, happens-before*

Discriminative classification of semantic relations

Selected relation classes



Approach

How to address this task?

Difficult to distinguish such fine-grained inferential relations!

Prior work

- Pattern-based approaches (Chklovski and Pantel, 2004)
- Distributional methods for asymmetrical inference relations (Bhagat, Pantel and Hovy, 2007)

Here

- Exploit *linguistic properties* for *discriminative* classification
  - inferential behaviour under negation
  - temporal sequence properties
- using (minimally) supervised corpus-based approach

Inference Patterns under Negation

Relation	Temp.Rel(V_1,V_2)	Inference patterns (V_1,V_2) I_x : $p_{\pm v_1}$ *op* $p_{\pm v_2}$	Example
Entailment (*buy, own*)	temp.rel: V_1 ($<$,o,$>$) V_2	I_1: $+ \Box\!\!\rightarrow +$ I_2: $- \Diamond\!\!\rightarrow +^{\textbf{exception}}$ I_3: $\neg(+ \Diamond\!\!\rightarrow -)$ I_4: $- \Diamond\!\!\rightarrow -$	*I buy - I own* *I don't buy, but I (still) own* *I don't buy, so I (normally) don't own*
Presupposition (*win, play*) Temp. Incl. (*snore,sleep*)	$V_2 < V_1$ $V_1 \subset / = V_2$	I_1: $+ \Box\!\!\rightarrow +$ I_2: $- \Diamond\!\!\rightarrow +^{\textbf{persistence}}$ I_3: $\neg(+ \Diamond\!\!\rightarrow -)$ I_4: $- \Diamond\!\!\rightarrow -^{\textbf{cancell.}}$	*I win - I played* *I didn't win but/when I played* *I didn't win - because I didn't play*
Antonymy (*love,hate*)	no temp. seq.	I_1: $\neg(+ \Diamond\!\!\rightarrow +)$ I_2: $- \Box\!\!\rightarrow +^{\textbf{t.n.d.}}$ I_3: $+ \Box\!\!\rightarrow -^{\textbf{t.n.d.}}$ I_4: $\neg(- \Diamond\!\!\rightarrow -)^{\textbf{t.n.d.}}$	 *you don't love – you hate* *you love – you don't hate*
Synonymy (*fix,repair*)	no temp. seq.	I_1: $+ \Box\!\!\rightarrow +$ I_2: $\neg(- \Diamond\!\!\rightarrow +)$ I_3: $\neg(+ \Diamond\!\!\rightarrow -)$ I_4: $- \Box\!\!\rightarrow -$	*I fix - I repair* *I don't fix – I don't repair*

Inference Patterns under Negation

Relation	Temp.Rel(V_1,V_2)	Inference patterns (V_1,V_2) I_x : $p_{\pm v_1}$ *op* $p_{\pm v_2}$	Example
		I_1: $+ \Box\!\!\rightarrow +$	*I buy - I own*
Entailment	temp.rel:	I_2: $- \Diamond\!\!\rightarrow +^{\textbf{exception}}$	*I don't buy, but I (still) own*
(*buy, own*)	V_1 ($<$,o,$>$) V_2	I_3: $\neg(+ \Diamond\!\!\rightarrow -)$	
		I_4: $- \Diamond\!\!\rightarrow -$	*I don't buy, so I (normally) don't own*
Presupposition	$V_2 < V_1$	I_1: $+ \Box\!\!\rightarrow +$	*I win - I played*
(*win, play*)		I_2: $- \Diamond\!\!\rightarrow +^{\textbf{persistence}}$	*I didn't win but/when I played*
Temp. Incl.		I_3: $\neg(+ \Diamond\!\!\rightarrow -)$	
(*snore,sleep*)	$V_1 \subset/= V_2$	I_4: $- \Diamond\!\!\rightarrow -^{\textbf{cancell.}}$	*I didn't win - because I didn't play*
		I_1: $\neg(+ \Diamond\!\!\rightarrow +)$	
Antonymy	no temp. seq.	I_2: $- \Box\!\!\rightarrow +^{\textbf{t.n.d.}}$	*you don't love – you hate*
(love,hate)		I_3: $+ \Box\!\!\rightarrow -^{\textbf{t.n.d.}}$	*you love – you don't hate*
		I_4: $\neg(- \Diamond\!\!\rightarrow -)^{\textbf{t.n.d.}}$	
		I_1: $+ \Box\!\!\rightarrow +$	*I fix - I repair*
Synonymy	no temp. seq.	I_2: $\neg(- \Diamond\!\!\rightarrow +)$	
(fix,repair)		I_3: $\neg(+ \Diamond\!\!\rightarrow -)$	
		I_4: $- \Box\!\!\rightarrow -$	*I don't fix – I don't repair*

Inference Patterns under Negation

Relation	Temp.Rel(V_1,V_2)	Inference patterns (V_1,V_2) I_x : $p_{\pm v_1}$ *op* $p_{\pm v_2}$	Example
Entailment	temp.rel:	I_1: $+ \Box\!\!\rightarrow +$	*I buy - I own*
(*buy, own*)	V_1 ($<$,o,$>$) V_2	I_2: $- \Diamond\!\!\rightarrow +^{\textbf{exception}}$	*I don't buy, but I (still) own*
		I_3: $\neg(+ \Diamond\!\!\rightarrow -)$	
		I_4: $- \Diamond\!\!\rightarrow -$	*I don't buy, so I (normally) don't own*
Presupposition	$V_2 < V_1$	I_1: $+ \Box\!\!\rightarrow +$	*I win - I played*
(*win, play*)		I_2: $- \Diamond\!\!\rightarrow +^{\textbf{persistence}}$	*I didn't win but/when I played*
Temp. Incl.		I_3: $\neg(+ \Diamond\!\!\rightarrow -)$	
(*snore,sleep*)	$V_1 \subset / = V_2$	I_4: $- \Diamond\!\!\rightarrow -^{\textbf{cancell.}}$	*I didn't win - because I didn't play*
Antonymy	no temp. seq.	I_1: $\neg(+ \Diamond\!\!\rightarrow +)$	
(*love,hate*)		I_2: $- \Box\!\!\rightarrow +^{\textbf{t.n.d.}}$	*you don't love – you hate*
		I_3: $+ \Box\!\!\rightarrow -^{\textbf{t.n.d.}}$	*you love – you don't hate*
		I_4: $\neg(- \Diamond\!\!\rightarrow -)^{\textbf{t.n.d.}}$	
Synonymy	no temp. seq.	I_1: $+ \Box\!\!\rightarrow +$	*I fix - I repair*
(*fix,repair*)		I_2: $\neg(- \Diamond\!\!\rightarrow +)$	
		I_3: $\neg(+ \Diamond\!\!\rightarrow -)$	
		I_4: $- \Box\!\!\rightarrow -$	*I don't fix – I don't repair*

Discriminative Properties

Temporal Sequence and Behaviour under Negation

		Behaviour under Negation			
		$(+V_1, +V_2)$	$(-V_1, +V_2)$	$(+V_1, -V_2)$	$(-V_1, -V_2)$
Temp. Seq.	V_1 prec V_2	E	(E)[e]		E
	V_1 succ V_2	E	(E)[e]		E
		P	P		(P)[c]
	V_1 ovlp V_2	E	(E)[e]		E
No temp. sequence		T	T		(T)[c]
		{A}	A	A	{A}
		S			S

Discriminative Properties

Temporal Sequence and Behaviour under Negation

		Behaviour under Negation			
		$(+V_1,+V_2)$	$(-V_1,+V_2)$	$(+V_1,-V_2)$	$(-V_1,-V_2)$
Temp. Seq.	V_1 prec V_2	E	$(\text{E})^{\mathbf{e}}$		E
	V_1 succ V_2	E P	$(\text{E})^{\mathbf{e}}$ P		E $(\text{P})^{\mathbf{c}}$
	V_1 ovlp V_2	E	$(\text{E})^{\mathbf{e}}$		E
No temp. sequence		T {A} S	T A	 A	$(\text{T})^{\mathbf{c}}$ {A} S

Discriminative Properties

Temporal Sequence and Behaviour under Negation

		Behaviour under Negation			
		$(+V_1,+V_2)$	$(-V_1,+V_2)$	$(+V_1,-V_2)$	$(-V_1,-V_2)$
Temp. Seq.	V_1 prec V_2	E	$(\text{E})^{\mathbf{e}}$		E
	V_1 succ V_2	E P	$(\text{E})^{\mathbf{e}}$ P		E $(\text{P})^{\mathbf{c}}$
	V_1 ovlp V_2	E	$(\text{E})^{\mathbf{e}}$		E
No temp. sequence		T {A} S	T A	A	$(\text{T})^{\mathbf{c}}$ {A} S

Discriminative Properties

Temporal Sequence and Behaviour under Negation

		Behaviour under Negation			
		$(+V_1,+V_2)$	$(-V_1,+V_2)$	$(+V_1,-V_2)$	$(-V_1,-V_2)$
Temp. Seq.	V_1 prec V_2	E	(E)$^{\mathbf{e}}$		E
	V_1 succ V_2	E P	(E)$^{\mathbf{e}}$ P		E (P)$^{\mathbf{c}}$
	V_1 ovlp V_2	E	(E)$^{\mathbf{e}}$		E
No temp. sequence		T {A} S	T A	 A	(T)$^{\mathbf{c}}$ {A} S

Discriminative Classification

Using temporal sequence and negation properties for classification

- Corpus-based approach, using small set of training relation pairs
- Observe co-occurring verbs within syntagmatically related contexts
- Determine their (typical) temporal order and negation contexts as features for (type-based) classification

Challenges and Approach

Annotation of a gold standard data set

- **Type-based:** Labeling pairs of verbs:
  difficult to imagine (and agree on) all possible relevant readings and contexts
  $\mathcal{K} = 0.47$
- **Token-based:** Labeling verb pairs in context:
  contexts difficult to decide ($\pm$ related?), not all readings covered
  $\mathcal{K} = 0.44$
- Deciding complex inferential properties is difficult!
- Type-based annotation is less expensive
- Solution: **question-based annotation** using verb pairs with prototypical arguments
  $\mathcal{K} = 0.64$

Question-based Annotation

- “Decision Tree” breaks down complex decision into ’simple’ decision tasks
  (temporal sequence, negation, strength of inference)
- Prototypical arguments determine relevant readings based on selectional preference classes – Resnik(1996)

Computing selectional preference classes of arguments

Resnik(1996)

- *Selectional association score* between predicate p_i and semantic class c

$$A(p_i, c) = \frac{P(c|p_i) log \frac{P(c|p_i)}{P(c)}}{S(p_i)}$$

- *Selectional preference strength* $S(p_i)$:

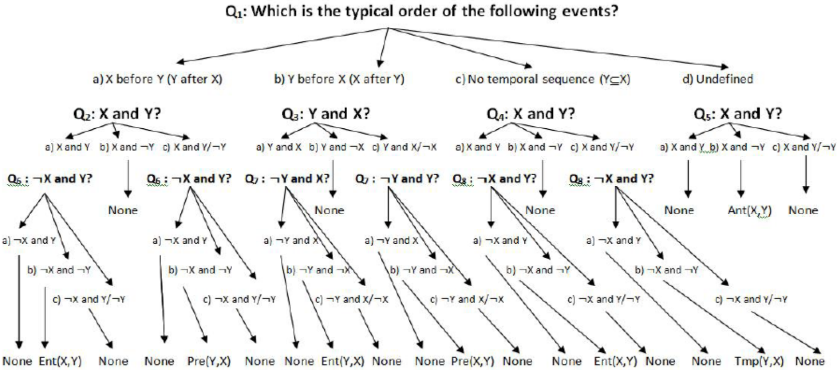
$$S(p_i) = \sum_c P(c|p_i) log \frac{P(c|p_i)}{P(c)}$$

- Modification to pairs of verbs:

$$A(p_i, p_j, c) = \frac{P(c|p_i,p_j) log \frac{P(c|p_i,p_j)}{P(c)}}{S(p_i,p_j)}$$

$$S(p_i, p_j) = \sum_c P(c|p_i, p_j) log \frac{P(c|p_i,p_j)}{P(c)}$$

Question-based Annotation: Decision Tree



Question-based Annotation: Example

Q_0:	**// Characterizing the interpretation of the events: //** Please give a translation for the verbs *learn* and *speak* in these readings: X: *John learns Spanish.* translation: ___ Y: *John speaks Spanish.* translation: ___
Q_1:	**// Determining the temporal order of events: //**
Q_2:	**// Determining negation properties: X and Y? //**
Q_6:	**// Determining negation properties: $\neg$X and Y? //**

Question-based Annotation: Example

Q_0:	**// Characterizing the interpretation of the events: //** X: *John learns Spanish.* translation: lernen Y: *John speaks Spanish.* translation: sprechen
Q_1:	**// Determining the temporal order of events: //** What is the typical order of the following events? a) *John learns Spanish and then he speaks Spanish.* X before Y b) *John speaks Spanish and then he learns Spanish.* X after Y c) *John learns Spanish and he speaks Spanish at the same time.* X during Y d) More than one order of events is possible. e) Not sure (difficult to define)
Q_2:	**// Determining negation properties: X and Y? //**
Q_6:	**// Determining negation properties: $\neg$X and Y? //**

Question-based Annotation: Example

Q_0:	**// Characterizing the interpretation of the events: //** X: *John learns Spanish.* translation: lernen Y: *John speaks Spanish.* translation: sprechen
Q_1:	**// Determining the temporal order of events: //** a) *John learns Spanish and then he speaks Spanish.* X before Y
Q_2:	**// Determining negation properties: X and Y? //** *John learns Spanish. Will he speak Spanish?* a) Yes (X and Y) b) No (X and ¬Y) c) **Maybe** (X and Y or ¬Y) - Persistence under Negation → presupposition
Q_6:	**// Determining negation properties: ¬X and Y? //**

Question-based Annotation: Example

Q_0:	**// Characterizing the interpretation of the events: //** Please give a translation for the verbs *learn* and *speak* in these readings: X: *John learns Spanish.* translation: lernen Y: *John speaks Spanish.* translation: sprechen
Q_1:	**// Determining the temporal order of events: //** What is the typical order of the following events? a) *John learns Spanish and then he speaks Spanish.* X before Y
Q_2:	**// Determining negation properties: X and Y? //** *John learns Spanish. Will he speak Spanish?* c) Maybe (X and Y or ¬ Y) – Persistence under Negation → presupposition
Q_6:	**// Determining negation properties: ¬X and Y? //** *John does not learn Spanish. Will he speak Spanish?* a) Yes (¬X and Y) → none b) **No** (¬X and ¬Y) - Cancellation → presupposition c) Maybe (¬X and ¬Y or Y) → none

Question-based Annotation: Example

Q_0:	**// Characterizing the interpretation of the events: //** Please give a translation for the verbs *learn* and *speak* in these readings: X: *John learns Spanish.* translation: lernen Y: *John speaks Spanish.* translation: sprechen
Q_1:	**// Determining the temporal order of events: //** What is the typical order of the following events? a) *John learns Spanish and then he speaks Spanish.* X before Y
Q_2:	**// Determining negation properties: X and Y? //** *John learns Spanish. Will he speak Spanish?* a) Maybe (X and Y or ¬ Y) Persistence under Negation → presupposition
Q_6:	**// Determining negation properties: ¬X and Y? //** *John does not learn Spanish. Will he speak Spanish?* b) **No (¬X and ¬Y)** - Cancellation → presupposition
Result: PRESUPPOSITION(SPEAK,LEARN)	

Classification Experiments

Classification Task

- Type-based classifier C: $\mathcal{X} \rightarrow \mathcal{Y}$ assigns classification instances $\mathcal{X}$ consisting of pairs of verb types (V_1,V_2) one label $\mathcal{R} \in \mathcal{Y}$.
- Two classification architectures:
  - **Flat:** Classify instances into 4 core relations plus 'Unrelated': $\mathcal{Y} = \{$ E, P, T, A, U $\}$
  - **Hierarchical:**
    Step 1: Binary classification: Related vs. Unrelated
    Step 2: Sub-classify instances of 'Related' class into 4 core relations: $\mathcal{Y} = \{$ E, P, T, A $\}$

Experiments

- Features
- Data sets and model building
- Evaluation and results on test set

Feature vectors for 4-/5-way classification

Feature sets

Feature type	Feature	Classification flat	Classification hier
typical temp. rel.	F$_0$: {before,during,after,undef}	✓	✓
polarity	F$_1$ – F$_4$: $P(< \pm V_1, \pm V_2 >\mid V_1, V_2)$	✓	✓
relatedness	F$_5$: avg. distance betw. V_1 and V_2 in tokens	✓	–
	F$_6$: $PMI(V_1, V_2)$	✓	–
	F$_7$ – F$_n$: conjunction c_i: $P(c_i \mid V_1, V_2)$	✓	✓

Feature Computation - Details

Temporal Sequence (Classifier)

- typical temporal order of events: *before, after, during*, *undefined*
- Performance (QA-annotation data): P: 71, R: 74, F_1: 73

Negation

- compute verb polarities: negative particles, adverbs, adjectives, verbs
- Performance: P: 84, R: 86, F_1: 85
- Conditional probabilities: $P(\langle \pm V_1, \pm V_2 \rangle \mid V_1, V_2)$

Data sets and Model Learning

Data

- Training: 48 verb pairs (equally distributed over relation types)
- Testing: 250 verb pairs (created by Q-based annotation)
- Corpus for feature extraction: ukWaC (Baroni et al. 2009): 30-500 sents with co-occurring verbs (per verb pair candidate)

Learning Algorithm

We use BayesNet for all experiments (Weka implementation), unless noted otherwise

Contiguity and Preprocessing

Preprocessing: Contiguity Filter

Selecting informative samples for feature extraction:
contiguously related verb pair contexts

Features used:

- length and form of relating grammatical path
- coreferring subj/obj: s-s, o-o, s-o, no coref
- distance in tokens and nb. of intervening verbs
- connectives

J48 classifier: classifies contexts as [$\pm$contiguous] with F_1: 0.793

Using contiguity for [$\pm$ related] classification (hier. class.)

classify verb pairs as [$\pm$ related]:

[$-$related]	if cnt([$+$cont]) $<$ cnt([$-$cont]) & temprel $=$ *undefined*
[$+$related]	otherwise.

Experiment I: Flat Classification

Semantic Relation	Precision	Recall	F_1-score	Baseline F_1-score
Presupposition	41%	45%	43%	25%
Entailment	47%	43%	44%	25%
Temporal Inclusion	38%	47%	42%	26%
Antonymy	68%	71%	70%	47%
Other/Unrelated	54%	53%	54%	12%
All	50%	51%	51%	27%

Table: Results for Flat Classification (BL: best feature: Conjunctions).

- classifier results clearly **outperform baseline**
- **balanced** recall and precision
- with 51% F_1-score: *modest performance*
- antonymy outperforms inferential relations (70 vs. low 40 F_1)

Experiment II: Hierarchical Classification

1st stage classification: [+/- related]

- ratio of contiguous/non-contiguous contexts (preprocessing)
- typical temporal relation

assign

[−related]	if cnt([+cont]) < cnt([−cont]) & temprel = *undefined*
[+related]	otherwise.

2nd stage classifier: 4-way flat classification

- input: verb pairs classified as [+related] by 1st stage classifier
- Feature set: all except contiguity features: F5 (distance) and F6 (PMI); yet keeping F7 (conjunctions)

Experiment II: Hierarchical Classification

Semantic	Baseline	Flat Classification			Hierarchical Classification		
Relation	F_1	P	R	F_1	P	R	F_1
Presupposition	25%	41%	45%	43%	**50%**	**46%**	**48%**
Entailment	25%	**47%**	43%	44%	44%	**46%**	**45%**
Temp. Incl.	26%	38%	47%	42%	**41%**	**47%**	**44%**
Antonymy	47%	68%	71%	70%	**72%**	**74%**	**73%**
Unrelated	12%	54%	53%	54%	**68%**	**63%**	**66%**
All	27%	50%	51%	51%	**55%**	**55%**	**55%**

Table: Hierarchical vs. Flat Classification (BL: best feature - Conjunctions)

- hierarchical classification outperforms flat classification
- strongest gains for presupposition (precision, w/ constant recall)
- balanced precision and recall
- antonymy scores highest
- 100% improvement over baseline

Impact of Features

Sem. Rel	Flat Classification				Hierarchical Classification			
	All	w/o Neg	w/o Tmp	w/o Conj	All	w/o Neg	w/o Tmp	w/o Conj
P	43%	37%	**24%**	35%	48%	41%	**22%**	34%
E	44%	41%	**14%**	28%	45%	43%	**14%**	25%
T.	42%	42%	**12%**	38%	44%	43%	**11%**	36%
A	70%	64%	64%	**15%**	73%	68%	59%	**14%**
U	54%	47%	45%	**35%**				
All	51%	46%	32%	**30%**	55%	52%	34%	**35%**

Table: Results using different feature sets. All figures are F_1-scores.

- *Conjunctions* is strongest feature for *antonymy* and *unrelated*
- *Temporal Relation* is strongest for the inferential relations
- *Negation* contributes most for *presupposition*

Conclusions

Contributions

- In-depth **analysis of semantic properties** of semantic relations between verbs
  → determined **discriminative properties** for classification: *negation* and *temporal sequence* properties
  → *question-based annotation* for improved consistency
- Corpus-based *type-based discriminative* classification for **four semantic relation types**
  - using identified feature types plus *'contiguity features'*
  - hierarchical classification outperforms flat classification
  - 100% improvement over baseline
  - weakly supervised: 10 instances/relation type
  - performance is competitive (but not strictly comparable to related work) (Tremper and Frank, to appear)

Intermediate questions?

Genericity:
Classifying generic NPs and generic sentences

(Reiter and Frank, ACL 2010)
(Reiter and Frank 2011, Tech. Report)

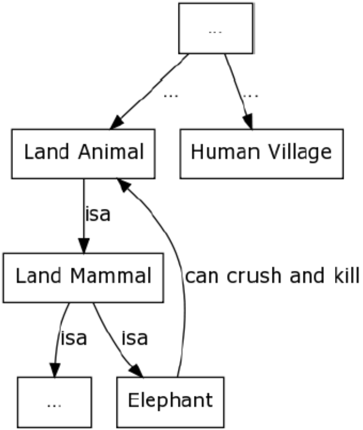
Elephants

[Elephants] can crush and kill any other land animal [...]
In Africa, groups of young teenage elephants attacked human villages after cullings done in the 1970s and 80s.

Wikipedia (2010)

Knowledge Acquisition

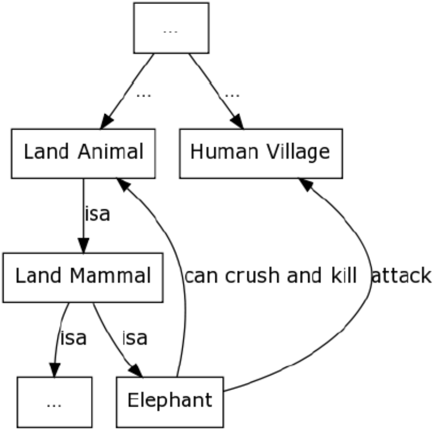
Elephants can crush and kill any other land animal.
Groups of teenage elephants attacked human villages.



Hearst (1992), Cimiano (2006), Bos (2009)

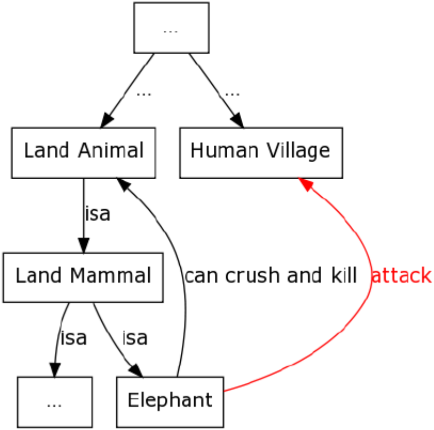
Knowledge Acquisition

Elephants can crush and kill any other land animal.
Groups of teenage elephants attacked human villages.



Knowledge Acquisition

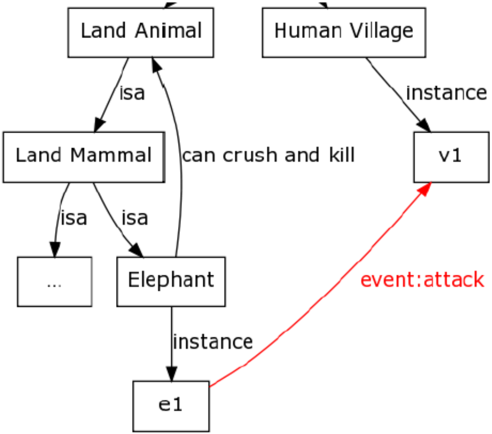
Elephants can crush and kill any other land animal.
Groups of teenage elephants attacked human villages.



This is not a property of the class Elephant!

Knowledge Acquisition

Elephants can crush and kill any other land animal.
Groups of teenage elephants attacked human villages.



It is a property of an instance of the class Elephant!

Relevance of Recognizing Genericity in K-Acquisition

1. Generic expressions express (rule-like) world knowledge
   - Generic noun phrases
     Horses are able to sleep while standing. Wikipedia
   - Generic (habitual) sentences
     Chimpanzees make tools and use them to acquire foods and for social displays. Wikipedia
     After 1971 [he = Paul Erd"os] also took amphetamines. Wikipedia

Relevance of Recognizing Genericity in K-Acquisition

2. Need to distinguish *classes* and *instances*

otherwise

- *Instance-level information* is generalized to the *class*, or
- *Class-level knowledge* is attached to *instances*

Relevance of Recognizing Genericity in K-Acquisition

3. Challenges: Quantifier interpretation and inferential properties

- *Rock ballads are popular with exactly one fan.*
  i. Rock ballads as a kind have only one fan.
  ii. The is only one fan who likes rock ballads.
- *The lion was the most wide-spread mammal.* Wikipedia
- *Birds fly.*
- *The Black Robin [...] is an endangered bird from the Chatham Islands. [...] It was first described by Walter Buller in 1872.*

Wikipedia

Aim

$\Rightarrow$ Automatically identify and distinguish
generic (vs. non-generic)
noun phrases and sentences.

Generic Noun Phrases

- Refer to a kind or class of individuals

Examples

- *The lion was the most widespread animal.*
- *Lions eat up to 30 kg in one sitting.*

Krifka et al. (1995)

Generic Sentences

- Express rule-like knowledge about habitual actions
- Do not express a particular event

Examples

- *After 1971 [he] also took amphetamines.*
- *Lions eat up to 30 kg in one sitting.*

Krifka et al. (1995)

Co-Occurrence

Both phenomena can (but don't have to) co-occur in a single sentence

	S[+gen]	S[−gen]
NP[+gen]	*Lions eat up to 30 kg in one sitting.*	*The lion was the most widespread mammal.*
NP[−gen]	*After 1971 [Paul Erd˝os] also took amphetamines.*	*Paul Erd˝os was born [...] on March 26, 1913.*

Interpretations of Generic Noun Phrases

Quantification

- Quantification over individuals
- Exact determination of the quantifier restriction is difficult
- Quantification over "relevant" or "normal" individuals

Dahl (1975), Declerck (1991), Cohen (1999)

Kind-Referring

- A generic NP refers to a kind
- Kinds are individuals that have properties on their own

Carlson (1977)

Interpretation of Generic Sentences

$$Q[x_1, ..., x_i](\underbrace{[x_1, ..., x_i]}_{\textit{Restrictor}}; \underbrace{\exists y_1, ..., y_i[x_1, .., x_i, y_1, ..., y_i]}_{\textit{Matrix}})$$

- Dyadic operator Q relates restrictor and matrix
- Generic operator quantifies over situations and events
- Exact determination of the quantifier restriction is extremely difficult

Heim (1982), Krifka et al. (1995)

- Classification of generic sentences

Mathew and Katz (2009)

Interpretation of Generic Sentences

$$Q[x_1, ..., x_i](\underbrace{[x_1, ..., x_i]}_{\textit{Restrictor}}; \underbrace{\exists y_1, ..., y_i[x_1, .., x_i, y_1, ..., y_i]}_{\textit{Matrix}})$$

- Dyadic operator Q relates restrictor and matrix
- Generic operator quantifies over situations and events
- Exact determination of the quantifier restriction is extremely difficult

Heim (1982), Krifka et al. (1995)

- Classification of generic sentences

Mathew and Katz (2009)

Characteristics

- No specific linguistic marking of generic expressions

Examples (Noun Phrases)

- <u>The lion</u> was the most widespread mammal.
- <u>A lioness</u> is weaker [...] than a male.
- <u>Elephants</u> can crush and kill any other land animal.

Examples (Sentences)

- John walks to work.
- John walked to work (when he lived in California).
- John will walk to work (when he moves to California).

Aim 1: Classifying generic (vs. non-generic) NPs

Most of the tests and criteria for genericity given in the literature can't be directly operationalised for corpus-based analysis

- some predicates only allow kind-readings (*be extinct, invent*)
- reference to established kinds allows creation of kind-readings
  The Coke bottle has a narrow neck.
- meaning changes when inserting *usually, typically*
- generic sentences express 'essential' (vs. accidental) properties
  A madrigal is ?? popular / polyphonic.
  A football hero is popular.

Krifka et al. (1995)

Phenomena are context-sensitive

⇒ Corpus-based approach to identify generic noun phrases

Aim 1: Classifying generic (vs. non-generic) NPs

Most of the tests and criteria for genericity given in the literature can't be directly operationalised for corpus-based analysis

- some predicates only allow kind-readings (*be extinct*, *invent*)
- reference to established kinds allows creation of kind-readings
  The Coke bottle has a narrow neck.
- meaning changes when inserting *usually, typically*
- generic sentences express 'essential' (vs. accidental) properties
  A madrigal is ?? popular / polyphonic.
  A football hero is popular.

Krifka et al. (1995)

Phenomena are context-sensitive

⇒ Corpus-based approach to identify generic noun phrases

Features

	Syntactic	Semantic
NP-level	Number, Person, PoS, DeterminerType, BarePlural	Countability, Granularity, Sense[0-3, Top]
S-level	Clause.{PoS, Passive, NbModifiers}, DependencyRelation[0-4], Clause.Adjunct.{VerbType, AdverbType}, XLE.Quality	Clause.{Tense, Progressive, Perfective, Mood, Pred, HasTempModifier}, Clause.Adjunct.{Time, Pred}, EmbeddingPredicate.Pred

Table: Feature Classes

Feature Selection

Feature Combinations

- Each triple, pair and single feature tested in isolation

Ablation Testing

1. A single feature in turn is removed from the feature set
2. The feature whose omission causes the biggest drop in f-score is considered a strong feature
3. Remove strong feature and start over

   In the end, we have a list of features sorted by their impact

Experiment: Corpus and Algorithm

Corpus

- ACE-2 corpus Mitchell et al. (2003)
- Newspaper texts
- 40,106 annotated entities
- 5,303 (13.2 %) marked as generic
- Balancing training data: $\sim$ 10,000 entities for each class
  - Over-sampling generic entities
  - Under-sampling non-generic entities

Bayesian Network

- Weka implementation of a Bayesian net Witten and Frank (2002)
- A Bayesian network represents dependencies between random variables as graph edges

Experiment: Corpus and Algorithm

Corpus

- ACE-2 corpus — Mitchell et al. (2003)
- Newspaper texts
- 40,106 annotated entities
- 5,303 (13.2 %) marked as generic
- Balancing training data: $\sim$ 10,000 entities for each class
  - Over-sampling generic entities
  - Under-sampling non-generic entities

Bayesian Network

- Weka implementation of a Bayesian net — Witten and Frank (2002)
- A Bayesian network represents dependencies between random variables as graph edges

Results of Feature Selection – Ablation

	Syntactic	Semantic
NP-level	Number, Person, Pos, DeterminerType, BarePlural	Countability, Granularity, Sense[0], Sense[1-3, Top]
S-level	Clause.PoS, Clause.{Passive, NbModifiers}, DependencyRelation[2], DependencyRelation[0-1,3-4], Clause.Adjunct.{VerbType, AdverbType}, XLE.Quality	Clause.{Tense, Pred}, Clause.{Progressive, Perfective, Mood, HasTempModifier}, Clause.Adjunct.{Time, Pred}, Embedding Predicate.Pred

Table: Feature Classes, selected features highlighted (ablation, Set5)

Baselines

Majority	Each entity is non-generic
Person	Use the feature Person
Suh	Results of a pattern-based approach on detection of generic NPs Suh (2006)

	Generic			Overall		
	P	R	F	P	R	F
Majority	0	0	0	75.3	86.8	80.6
Person	60.5	10.2	17.5	84.3	87.2	85.7
Suh (2006)	28.9					

Table: Baseline results

Classification Results – Feature Selection

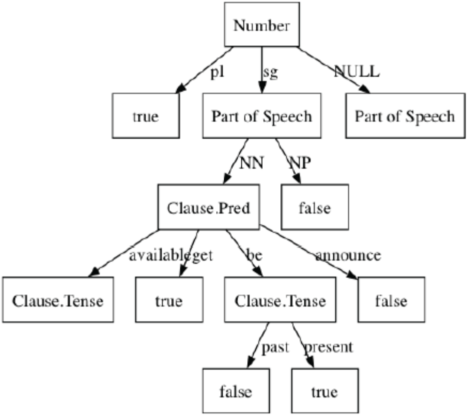
	Feature Set	Generic			Overall		
		P	R	F	P	R	F
Baseline	Majority	0	0	0	75.3	86.8	80.6
	Person	60.5	10.2	17.5	84.3	87.2	85.7
	Suh (2006)	28.9					
Unbal.	5 best single features	49.5	37.4	42.6	85.3	86.7	86.0
	Feature groups	42.7	69.6	52.9	88.0	83.6	85.7
	Ablation set	45.7	64.8	53.6	87.9	85.2	86.5
Bal.	5 best single features	29.7	71.1	41.9	85.9	73.9	79.5
	Feature groups	35.9	83.1	50.1	88.7	78.2	83.1
	Ablation set	37.0	81.9	51.0	88.8	79.2	83.7

Table: Results of the classification for Feature Selection

- Ablation testing yields the feature set that outperforms every other feature set

Conclusion 1: Classifying generic NPs

- Corpus-based classification is feasible
- Features from all levels in combination perform best
  (Sentence vs. NP,
  Syntax vs. Semantics)
- Contextual factors with impact on the phenomenon can be uncovered
  $\rightarrow$ allow deeper investigations of 'factors' for generic interpretation



Generic Sentences

What about generic sentences?
(How) do noun phrase and sentence genericity interact?

Classifying Generic Noun Phrases and Sentences

Cross-classifying generic NPs and sentences

- Sentence-level features are relevant for classifying (non-)generic NPs (Reiter & Frank 2010)
- Definiteness of the noun phrases is relevant for classifying (non-)generic sentences (Mathew & Katz 2009)

Questions and hypotheses

- Both types of genericity are characterized by properties at the NP and S levels, but in different ways.
- Do the two types of genericity interact, and in which ways?
- Can any/one of the two classifiers 'help' the other? ($\rightarrow$ joint classification)

Aim 2: Cross-classifying Generic NPs and Sentences

Exp I: Investigation of feature sets

- Learn base classifiers C_{genS} and C_{genNP}
- What type of features discriminate the two types?
  - Human interpretation: mostly semantic
    (tense & aspect, specific object reference, temporal modifiers)

Aim 2: Cross-classifying Generic NPs and Sentences

Cross-classification by *Stacked Classification*

Exp II:

- *base classifier* C_S: pre-classify sentences: **S.Gen**
- *target class classifier* C_{NP}: assign target class **NP.Gen** using prediction/learning from base classifier
- ACE data (ground truth for NP.gen)

Exp III:

- *base classifier* C_{NP}: pre-classify noun phrases: **Subj/Obj.Gen**
- *target class classifier* C_S: assign target class **S.Gen** using prediction/learning from base classifier
- PTB data (ground truth for S.gen)

Aim 2: Cross-classifying Generic NPs and Sentences

	Exp IIa (ACE): target class: NP.Gen	**Exp IIIa** (PTB): target class: S.Gen
target class classifier	$\mathbf{C}_{NP}$((NP-Level/Sel_{NP}) $\cup$ S.Gen)	$\mathbf{C}_{S}$((S-Level/Sel_{S}) $\cup$ Subj/Obj.Gen)
	⬆	⬆
base classifier	$\mathbf{C}_{S}$(S-Level/Sel_{S}) → S.Gen	$\mathbf{C}_{NP}$(NP-Level/Sel_{NP}) → Subj/Obj.Gen

Exp I: Calibrating base classifiers and feature sets

Generic NPs

	Feature Set	P	R	F
Generic	R&F: NP-Level	30.1	71.0	42.2
	R&F: Set 5	37.0	81.9	**51.0**
	S-Level	21.7	69.6	33.1
	NP-Level	33.1	72.5	45.4
	Sel.$_{np}$	**37.2**	73.0	49.2
	RF	36.2	**82.8**	50.4

Table: Results for NP genericity: generic class only, balanced data, 10CV; Feature sets: S-Level/NP-Level only; Sel(ected); RF = R&F reconstructed

- replicated feature set RF comparable to R&F results
- complementary class features (S-level) clearly lag behind
- mixed feature sets clearly outperform NP-level features

Exp I: Calibrating base classifiers and feature sets

Generic Sentences

	Feature Set	P	R	F
Habitual	NP-Level	36.0	52.4	42.7
	MK	56.1	63.0	59.4
	S-Level	65.9	73.2	69.4
	Sel.$_s$	**66.6**	**74.8**	**70.5**
Episodic	NP-Level	86.1	76.0	80.7
	MK	90.1	87.3	88.7
	S-Level	92.9	90.2	91.5
	Sel.$_s$	93.3	90.3	91.8

Table: Results for sentence genericity (Exp Ib)

- Replicated feature set underperforms M&K results (unbal. data)
- Complementary feature set (NP-level) lags behind
- S-level clearly outperforms (mixed) MK feature set; almost reaches best (mixed) selected features

Analysis of Feature Sets: Best feature sets

Level	NP genericity: *Sel.*$_{np}$	
	Syntactic	Semantic
NP	*BarePlural, Definiteness, Determiner, Number, Person*, MWE	**Granularity, Sense[0,1,2]**
S	*PP[at,on]*, *Rel. S. Position*, **Conditional**, DepRel[0,2], **Modal**	*Aspect, Sense[root], Temporal*, **Modifiers**

Level	Sentence genericity: *Sel.*$_{s}$	
	Syntactic	Semantic
NP	*Subj&Obj: BarePlural, Definiteness, Determiner, Number, Person*, ∃**Object, Obj: PoS**	
S	*PP[at,in,on]*, *Rel. S. Position*, PoS	*Aspect, Sense[root], Temporal*, **Tense**

Table: Best feature sets: NP and sentence genericity

Analysis of Feature Sets

Observations

- large overlap in feature sets
- both types make use of NP- and S-level features
- sentence genericity: S-level features rival best (selected) features
- sentence genericity: no semantic NP-features
- both types: semantic sentence-level features

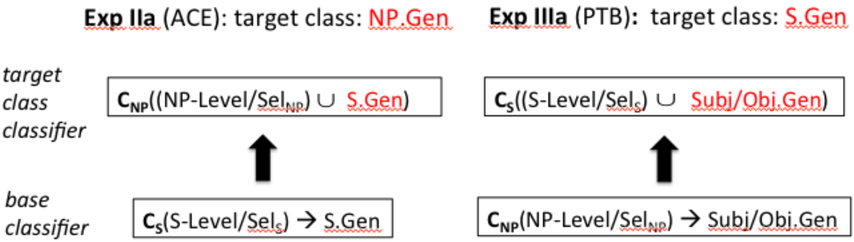
This suggests a *dependence of NP genericity on S-level features*, but not the other way round

Differentiating features

NP genericity: Semantic NP class, S modifiers; Conditional, Modal
S genericity: Presence and form of object; in-PP; tense

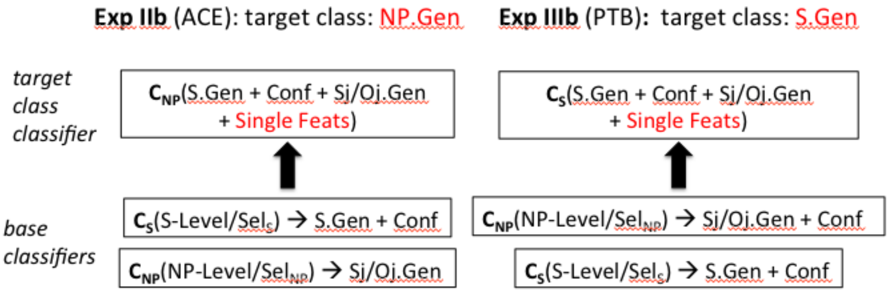
Aim 2: Cross-classifying Generic NPs and Sentences

Exp II/IIIa: using base classifier prediction as additional feature in target classifier



Aim 2: Cross-classifying Generic NPs and Sentences

Exp II/IIIb: Meta Learning: Target classifier uses predictions and confidences of both classifiers (opt: plus some strong features)



Exp II: Generic NPs with Stacked Classification

		Feature Set	P	R	F
Exp IIa	Generic	NP-Level	33.2	73.6	45.8
		NP-Level+S.Gen	33.7	**78.3**	47.1
		Sel.$_{np}$	37.1	72.9	49.2
		Sel.$_{np}$+S.Gen	**37.5**	75.8	**50.2**
Exp IIb	Generic	Person	32.1	75.8	45.1
		Tense	32.1	79.2	45.7
		Subj/Obj	31.9	80.4	45.7
		All	**32.5**	78.5	45.9
		Meta	**32.5**	**82.6**	**46.7**

Table: Classification results for Exp II

- Exp IIa: S.Gen prediction yields recall gains, at comparable precision
- Injection of S.Gen as a feature outperforms meta learning (Exp IIb)
- Dependencies: P(NP[gen+] | S[hab+]) > P(NP[gen−] | S[hab+])

Exp III: Generic Sentences with Stacked Classification

		Feature Set	P	R	F
Exp IIIa	Habitual	S-Level	65.3	73.2	69.0
		S-Level+NP.Gen	64.9	73.7	69.0
		Sel.	**66.4**	73.6	**69.8**
		Sel.+NP.Gen	60.8	**76.0**	67.5
Exp IIIb	Habitual	Meta	50.7	**78.6**	61.7
		Person	56.4	74.4	64.2
		POS	63.9	73.4	68.3
		Tense	64.0	73.4	68.4
		All	**67.1**	72.3	**69.6**

Table: Classification results for Exp III

- Exp IIIa: Injection of NP.Gen prediction harms Sel.$_S$ results
- Exp IIIb: Small improvements in precision
- In general, comparable performance to base classifier
- Dependencies: P(S[gen+] | Subj[gen+]) > P(S[gen+] | Subj[gen-])

Conclusions (I + II)

Feature analysis

- **NP.Gen:** features distributed over all feature groups
  S.Gen: S-level features are sufficient; no semantic NP features
  → Asymmetric dependence of NP on sentence genericity
- Many overlapping syntactic NP features
  → insights to be gained from semantic and S-level features

Interaction analysis using Cross-classification

- Significant effects on NP.Gen classifier using S.Gen predictions (compared to both NP-level and Sel.$_{np}$ base classifiers)
- Inspection of the models reveals insights about interactions: probabilistic dependencies in line with linguistic intuitions
- Meta learning: able to correct misclassifications of base classifiers → cast genericity as joint classification problem

Wrap-up

Distributional semantics: Novel research questions

- Can statistical semantics profit from formal semantics?
- Can formal semantics profit from statistical semantics?

I hope to have shown that we can ...

→ ... account for a difficult (and relevant) classification problem (*presupposition vs. entailment*) using insights of formal semantics

→ ... gain insights into factors that determine *genericity* by investigating corpus-based features and classification dependencies

Thanks for your attention!

References

- Nils Reiter and Anette Frank (2010): Identifying generic noun phrases. In: *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL 2010)*, Uppsala, Sweden, pp. 40-49.
- Nils Reiter and Anette Frank (2011): Cross-classifying Generic Noun Phrases and Sentences. Technical Report, Heidelberg University.
- Galina Tremper and Anette Frank (to appear): A Discriminative Analysis of Fine-Grained Semantic Relations including Presupposition: Annotation and Classification. To appear in: *Dialogue and Discourse, Special issue on: Beyond semantics: The challenges of annotating pragmatic and discourse phenomena*, edited by Stefanie Dipper, Heike Zinsmeister and Bonnie Webber.
- Galina Tremper and Anette Frank (2011): Extending Fine-Grained Semantic Relation Classification to Presupposition Relations between Verbs. In: Stefanie Dipper and Heike Zinsmeister (ed.): *Beyond Semantics: Corpus-based Investigations of Pragmatic and Discourse Phenomena. Proceedings of the DGfS Workshop Göttingen, February 23-25.* Bochumer Linguistische Arbeitsberichte 3 (3), pp. 129-144.

Semantic Relation	Precision	Recall	F_1-score	Baseline F_1-score
Presupposition	62%	50%	56%	30%
Entailment	53%	49%	51%	33%
Temporal Inclusion	44%	62%	52%	25%
Antonymy	76%	80%	78%	63%
All	59%	60%	59%	38%

Table: Results for Flat Classification (BL: best feature: Conjunctions).

Exp IV: Performance of NP.Gen classification for habitual sentences

Exploiting dependencies

- NP genericity seems dependent on sentence genericity
- → Evaluate NP genericity classification for habitual sentences (co-occurrence class c: [+NP.Gen, +S.Gen])

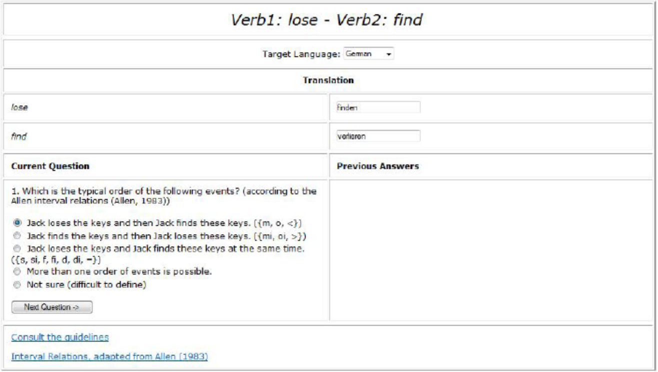
		Feature Set	P	R	F
All	Generic	NP-Level	32.5	82.6	46.7
		Sel.	35.4	79.4	48.9
		RF	**35.9**	**83.5**	**50.2**
Hab.	Generic	NP-Level	40.0	82.9	54.0
		Sel.	**42.1**	80.2	55.3
		RF	41.7	**84.5**	**55.9**

Table: Classification results for Exp IV

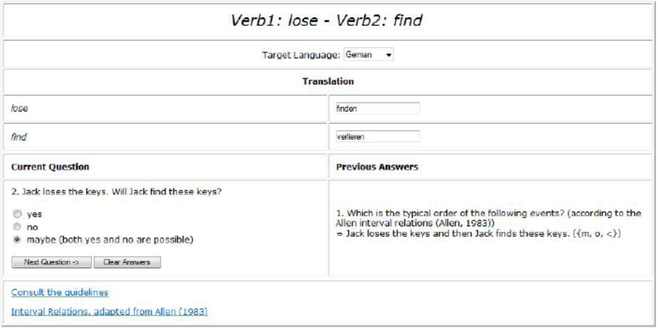
Examples of verb relations

Relation	Example	*Inference pattern*
Presupposition	*win - play*	*winning* presupposes *playing* *not winning* presupposes *playing*
Entailment	*kill - die*	*killing* implies *dying* *not killing* doesn't imply *dying*
Temporal *Inclusion*	*snore - sleep* *mutter - talk*	*snoring* happens during *sleeping* *muttering* is a special form of *talking*
Antonymy	*go - stay*	either *going* or *staying* *going* is the opposite of *staying*
Other/unrel.	*jump - sing*	none of the above

Web-based Annotation Interface



Web-based Annotation Interface



Web-based Annotation Interface