

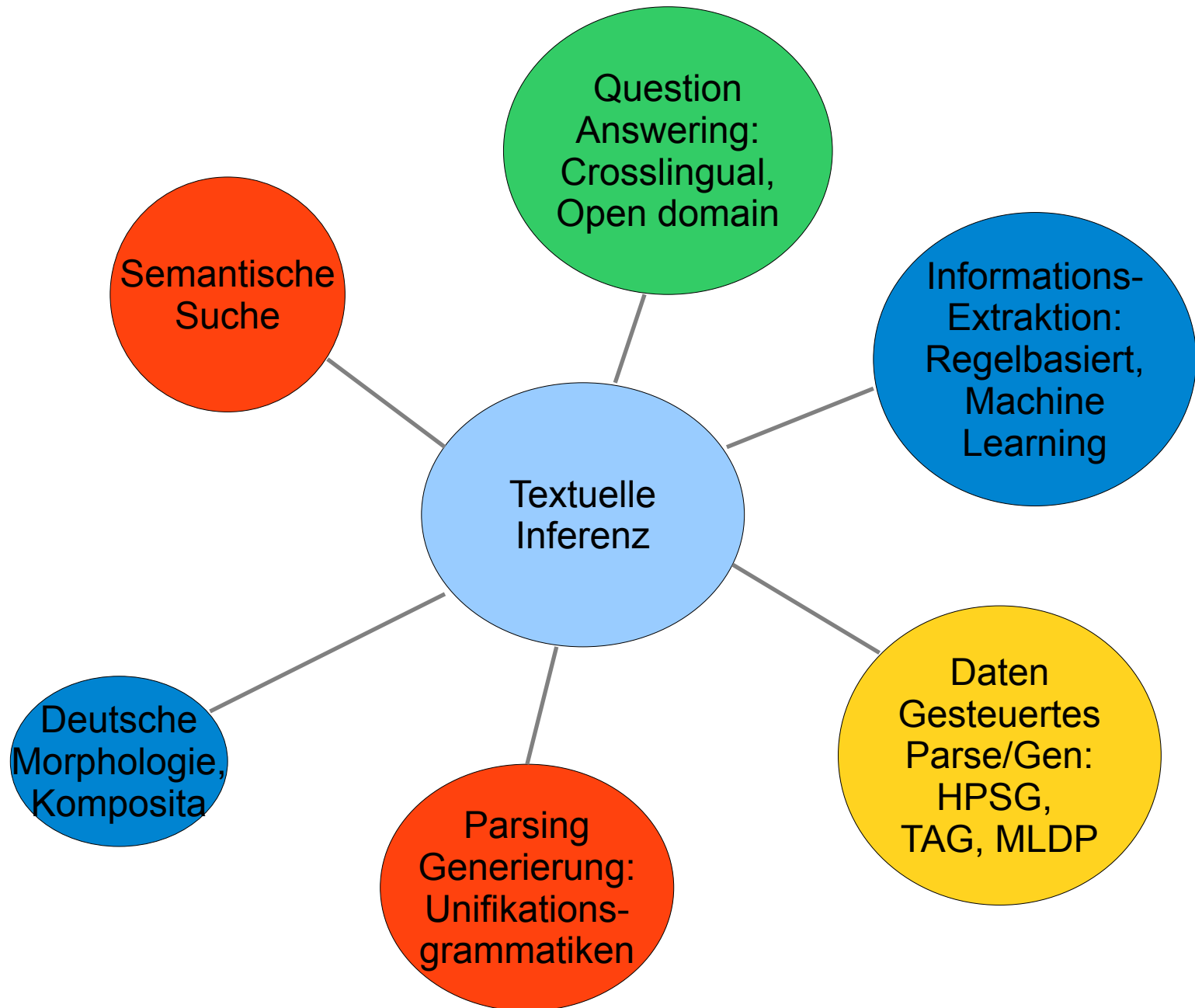
Textual Entailment - Methoden und Anwendungen

PD Dr. Günter Neumann

Forschungsbereich Sprachtechnologie
Deutsches Forschungszentrum für
Künstliche Intelligenz (DFKI)
&
Fachbereich Computerlinguistik
Universität des Saarlandes

Würzburg, 25th January 2010

Überblick: Forschungsarbeiten



Motivation

- Textbasierte Anwendungen benötigen robuste semantische Inferenzmaschinen

Q: Who is John Lennon's widow?

A: Yoko Ono unveiled a bronze statue of her late husband, John Lennon, to complete the official renaming of England's Liverpool Airport as Liverpool John Lennon Airport.

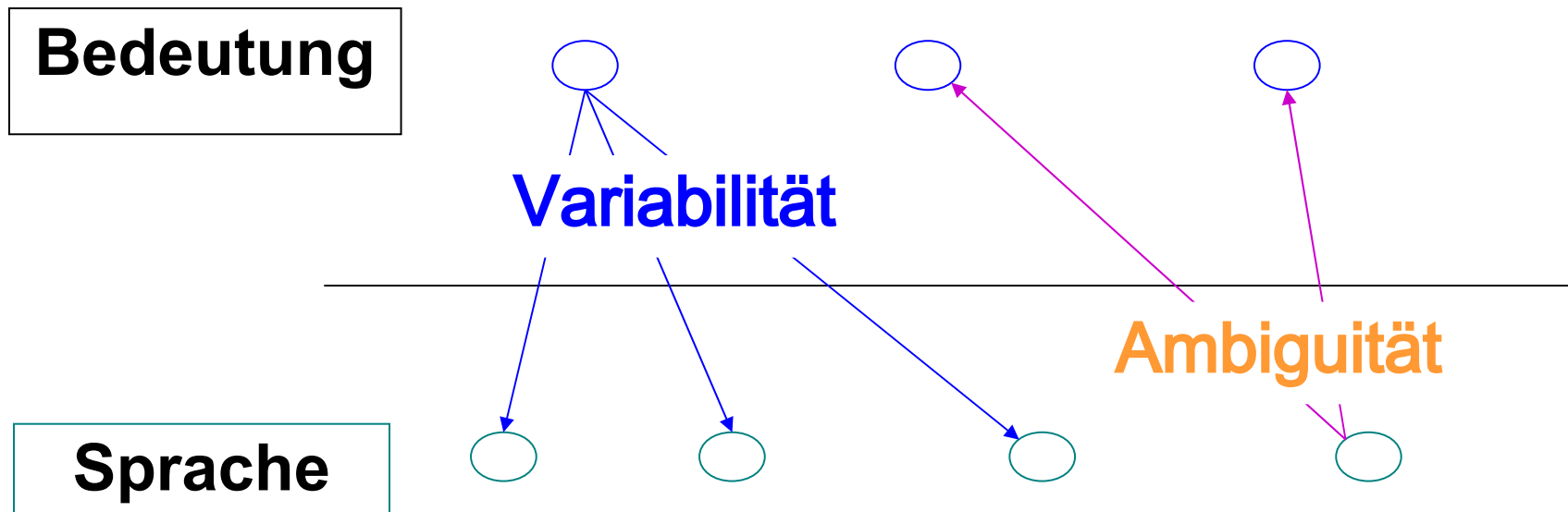
Motivation

- Textbasierte Anwendungen benötigen robuste semantische Inferenzmaschinen

Q: Who is John Lennon's widow?

A: Yoko Ono unveiled a bronze statue of her late husband, John Lennon, to complete the official renaming of England's Liverpool Airport as Liverpool John Lennon Airport.

Natürliche Sprache und Bedeutung



Variabilität von semantischen Ausdrücken

All major stock markets surged

Dow gains 255 points

Dow ends up

Dow climbs 255



Stock market hits a record high

The Dow Jones Industrial Average closed up 255

Erkennung von Variabilität – Wesentliche Inferenz in Anwendungen

Question Answering (QA)

Information Extraction (IE)

Information Retrieval (IR)

Multi Document Summarization (MDS)

Typische Anwendungsinferenz

Frage

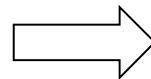
Wer übernahm Overture?

>>

Antwortmuster

X übernahm Overture

Overtures Übernahme
durch Yahoo



Yahoo übernahm Overture

Text

Hypothetische Antwort

- Ähnlich für IE: *X übernehmen Y*
- Ähnlich für “semantisches” IR: *t: Overture wurde von ... übernommen.*
- Textzusammenfassung (Mehrfachquellen) – bestimme redundante Informationen

Angewandtes textuelles Entailment: Beziehung zwischen Texten bezüglich semantischem Enthaltensein

Hypothesis (h): *John Wayne was born in Iowa*

- QA: “Where was John Wayne Born?”
 - Answer: Iowa

inference

Text (t): The birthplace of John Wayne is in Iowa

Generisches Entailment als Aufgabe

Hypothesis (h): *John Wayne was born in Iowa*

- Sei text t gegeben, können wir inferieren, dass h (sehr wahrscheinlich) wahr ist ?

inference

Text (t): The birthplace of John Wayne is in Iowa

Klassisches Entailment

- Chierchia & McConnell-Ginet (2001):
Ein Text t enthält eine Hypothese h , wenn h in allen Umständen (möglichen Welten) wahr ist, in denen t wahr ist.
- Sehr strikt – betrachtet nicht Unsicherheiten, die in Anwendungen erlaubt sind.

“Fast genaues” Entailment

t: The technological triumph known as GPS ... was incubated in the mind of Ivan Getting.


h: Ivan Getting invented the GPS.

t: According to the Encyclopedia Britannica, Indonesia is the largest archipelagic nation in the world, consisting of 13,670 islands.

h: 13,670 islands make up Indonesia.

Textual Entailment \approx

Menschliches Leseverständnis

- Aus einem Lernbuch für Schüler (Sela and Greenberg):
- **Referenztext:** “...*The Bermuda Triangle lies in the Atlantic Ocean, off the coast of Florida.* ...”
...

- **Hypothese (Wahr/Falsch?):** *The Bermuda Triangle is near the United States*

Maschinelles Leseverständnis

By Canadian Broadcasting Corporation

T: The school has turned its one-time metal shop – lost to budget cuts almost two years ago - into a money-making professional fitness club.

Q: When did the metal shop close?

A: Almost two years ago

Zwei mögliche Ansätze:

- a. System beantwortet Fragen, die von außen gestellt werden (QA)
- b. System erzeugt selber Fragen, die von außen beantwortet werden (E-Learning)

Wissenschaftliche Motivation

- Allein der Einsatz traditioneller, diskreter Verarbeitungskomponenten nicht ausreichend, z.B. POS-Tagging, Dependenzanalysen, WSD, etc.
- Sondern: sprachverstehende Anwendungen brauchen die Fähigkeit
 - Zu bestimmen, ob zwei Strings in einem bestimmten Kontext „dasselbe meinen“ unabhängig ihrer oberflächigen Ähnlichkeit
 - Zu bestimmen, ob ein String die Bedeutung eines anderen Strings „enthält“
 - Strings in einer bedeutungserhaltenden Weise zu reformulieren
- Daher: empirische Modelle über semantische Überlappung
 - gemeinsames Framework für eine angewandte Semantik, die skalierbare, robuste, effiziente semantische Inferenzen ermöglicht

Recognizing Textual Entailment (RTE) Challenge – Wissenschaftlicher Wettbewerb

- Internationaler wissenschaftlicher Wettbewerb seit 2005, cf. Ido Dagan et al.
- Zentrale Motivation:
 - Wissenschaftler aus der ganzen Welt zusammenbringen, um gemeinsam das Forschungsgebiet „angewandte Semantik“ voranzutreiben („open collaboration“)
- RTE im Rahmen der Text Analysis Conference (TAC)
 - <http://www.nist.gov/tac/2009/workshop/> (National Institute of Standards and Technology Gaithersburg, Maryland USA)
 - RTE-5 Nov/2009: 26 Teams weltweit
 - DFKI-Team: 2. bestes Ergebnis

Datenformate

<pair id="1" entailment="YES" task="IE" length="short" >

<t>The sale was made to pay Yukos' US\$ 27.5 billion tax bill, Yuganskneftegaz was originally sold for US\$ 9.4 billion to a little known company Baikalfinansgroup which was later bought by the Russian state-owned oil company Rosneft . </t>

<h>Baikalfinansgroup was sold to Rosneft. </h>

</pair>

<pair id="2" entailment="NO" task="IE" length="short" >

<t>The sale was made to pay Yukos' US\$ 27.5 billion tax bill, Yuganskneftegaz was originally sold for US\$9.4 billion to a little known company Baikalfinansgroup which was later bought by the Russian state-owned oil company Rosneft . </t>

<h>Yuganskneftegaz cost US\$ 27.5 billion. </h>

</pair>

<pair id="3" entailment="NO" task="IE" length="long" >

<t>Loraine besides participating in Broadway's Dreamgirls, also participated in the Off-Broadway production of "Does A Tiger Have A Necktie". In 1999, Loraine went to London, United Kingdom. There she participated in the production of "RENT" where she was cast as "Mimi" the understudy. </t>

<h>"Does A Tiger Have A Necktie" was produced in London. </h>

</pair>

<pair id="4" entailment="YES" task="IE" length="long" >

<t>"The Extra Girl" (1923) is a story of a small-town girl, Sue Graham (played by Mabel Normand) who comes to Hollywood to be in the pictures.

This Mabel Normand vehicle, produced by Mack Sennett, followed earlier films about the film industry and also paved the way for later films about Hollywood, such as King Vidor's "Show People" (1928). </t>

<h>"The Extra Girl" was produced by Sennett. </h>

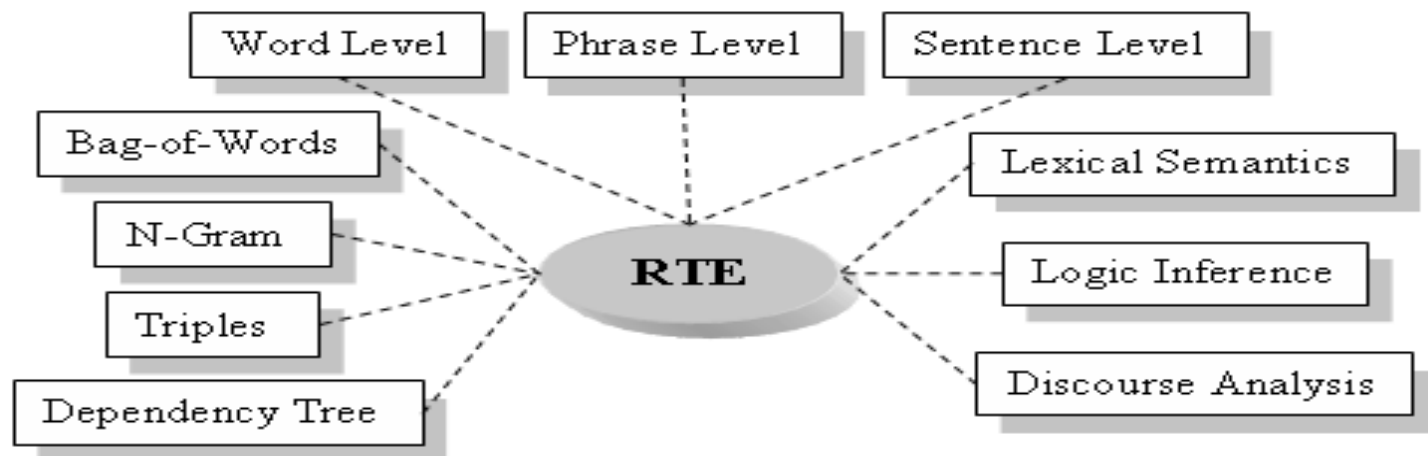
</pair>

Robuste Textinferenz

Verarbeitung freier Texte

- Semantische Unterspezifikation
 - Unpräzise Formulierung semantischer Beziehungen
 - Wagheit, Mehrdeutigkeit
- Fehlertolerante Methoden nötig
 - Verrauschte Eingabedaten
 - Verrauschte Ausgabe von Teilkomponenten

Unterschiedliche Ansätze betrachten/integrieren Merkmale aus verschiedenen linguistischen Beschreibungsebenen



Aktuelle Anätze & Methoden

- Konventionelle Methoden
 - Annahme einer Unabhängigkeit zwischen Worten (Bag of Words) (*Corley and Mihalcea, 2005*)
Wir benutzen diese als Baseline.
 - Messen von Abständen von syntaktischen Bäumen, um Entailmentbeziehungen zu entdecken (*Kouylekov and Magnini, 2006*)
Wir benutzen nur Teile von syntaktischen Bäumen, anstatt die kompletten Syntaxbäume.

Aktuelle Anätze & Methoden

- Logikbasierte Regeln
 - Logic rules (*Bos and Markert, 2005*)
 - Sequenzen von erlaubten Transformationsregeln (*de Salvo Braz et al., 2005*)
 - Wissensrepräsentationbasierte Modelle auf Basis von logischen Beweissystemen (*Tatu et al., 2006*)
- Normalerweise müssen diese Regeln manuell spezifiziert und implementiert werden, was einen enormen Aufwand bedeutet und die Verfügbarkeit von Computerlinguisten, die in vielen Anwendungsgebieten geschult sind.

Aktuelle Anätze & Methoden

- Vergrößerung von Trainingsmaterial
 - Automatisch bestimmten zusätzliches Trainingsmaterial (*Hickl et al., 2006*) (1st in RTE-2)
Nicht einfach, solche Daten zu bestimmen und eine Übereinstimmung bei den Annotatoren zu erzielen (Annotator Agreement).
- Kernel Methoden
 - Maschinelle Lernverfahren auf Basis von Baum-Kernels(*Zanzotto and Moschitti, 2006*) (3rd in RTE-2)
Wir verwenden Kernels über Teilsequenzen anstatt Kernels über den vollständigen Syntaxbäumen zu spezifizieren.

Frage: Wie weit kommt man mit Syntax alleine?

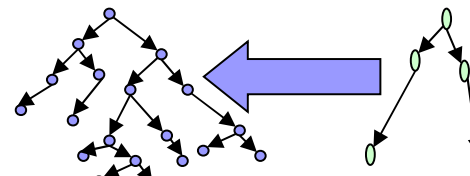
Ziel: Möglichst maximale syntaktische Baseline erzielen

Es werden neben der syntaktischen Dependenzanalyse keine weiteren linguistischen noch wissensbasierten Ressourcen eingesetzt.

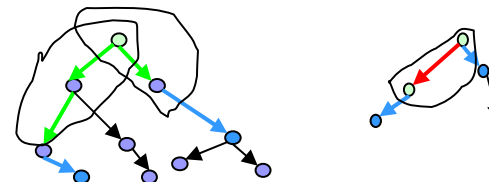
- Teilbäumen zwischen Baum H und relevanten Teilbäumen in T
- Baumkompression (Reduktion von Redundanz)
 - Reduziere Rauschen des Parsers
 - Liefert komprimierte Pfad-Wurzel-Pfad-Sequenzen
- Merkmalsextraktion für Teilsequenzen
 - Betrachte alle möglichen Teilfolgen der Pfad (Spine)-Differenzen
 - Support Vector Machine (SVM) zum Erlernen eines Klassifikators

<T, H>

Dependency parsing



Subtree alignment



Spine Difference

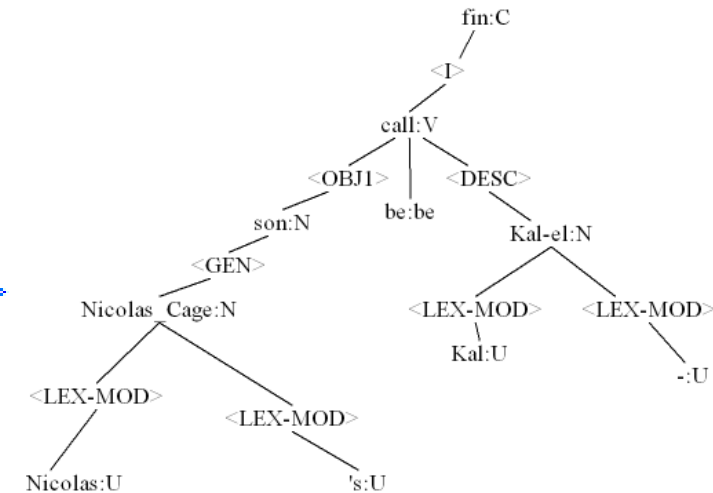
Finde: matchende Args/NP's mit kompatiblen Event/Roots

Predicate	Left spine diff.	Right spine diff.
<u>Text</u>	←—●—→	→
<u>Hypothese</u>	←	

Sentence representation

- A sentence is represented as a set of triples of general form <head relation modifier>
 - Ex: Nicolas Cage's son is called Kal'el

```
<triple left="E0" right="6">fin:C i call:V </triple>
<triple left="2" right="1">Nicolas_Cage:N lex-mod Nicolas:U </triple>
<triple left="2" right="3">Nicolas_Cage:N poss 's:U </triple>
<triple left="4" right="2">son:N gen Nicolas_Cage:N </triple>
<triple left="6" right="4">call:V s son:N </triple>
.....
```



- Dependency Structure
 - A DAG where nodes represent words and edges represent directed grammatical functions
 - We consider this as a “shallow semantic representation”
 - We use Minipar (Lin, 1998) and StanfordParser (Klein and Manning, 2003) for English and SMES (Neumann et al., 1997) for German.

A Natural Language Example

- Pair: id="61" entailment="YES" task="IE"
source="RTE"

- Text:

*Although they were born on different planets, Oscar-winning actor **Nicolas Cage's** new **son** and Superman have something in common, both were named **Kal-el**.*

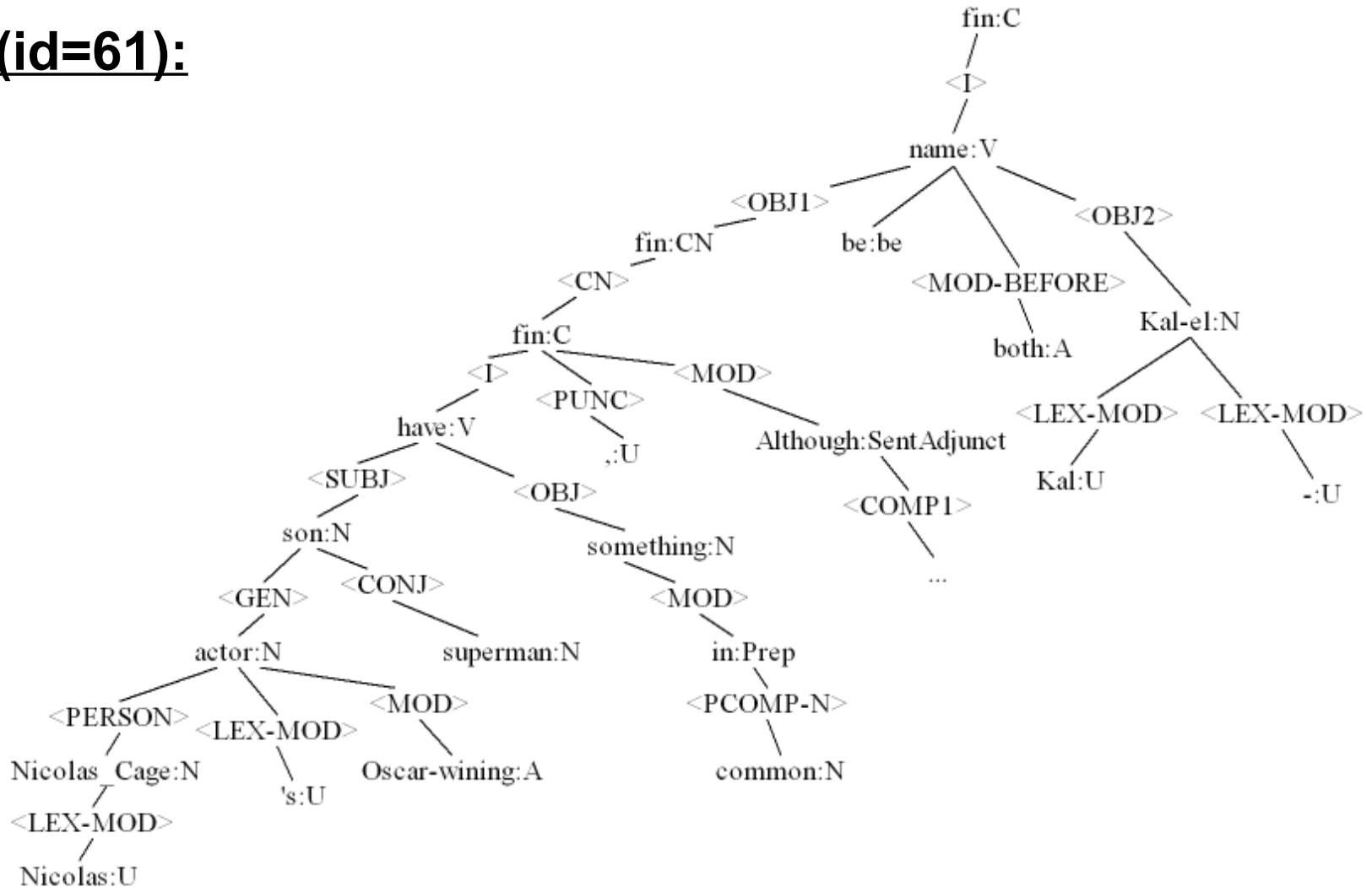
- Hypothesis:

Nicolas Cage's son is called Kal-el.

Dependency Graph

Dependency Tree of T

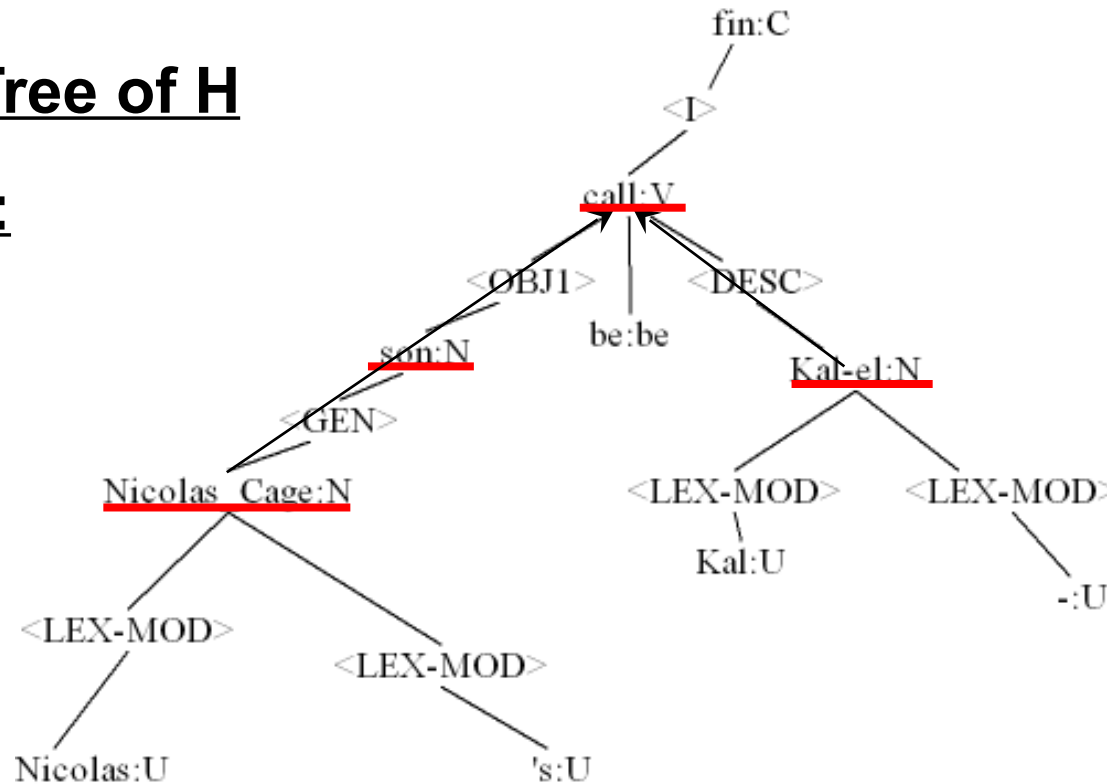
of pair (id=61):



Dependency Graph (cont.)

Dependency Tree of H

of pair (id=61):



Nicolas Cage's son is called Kal-el.

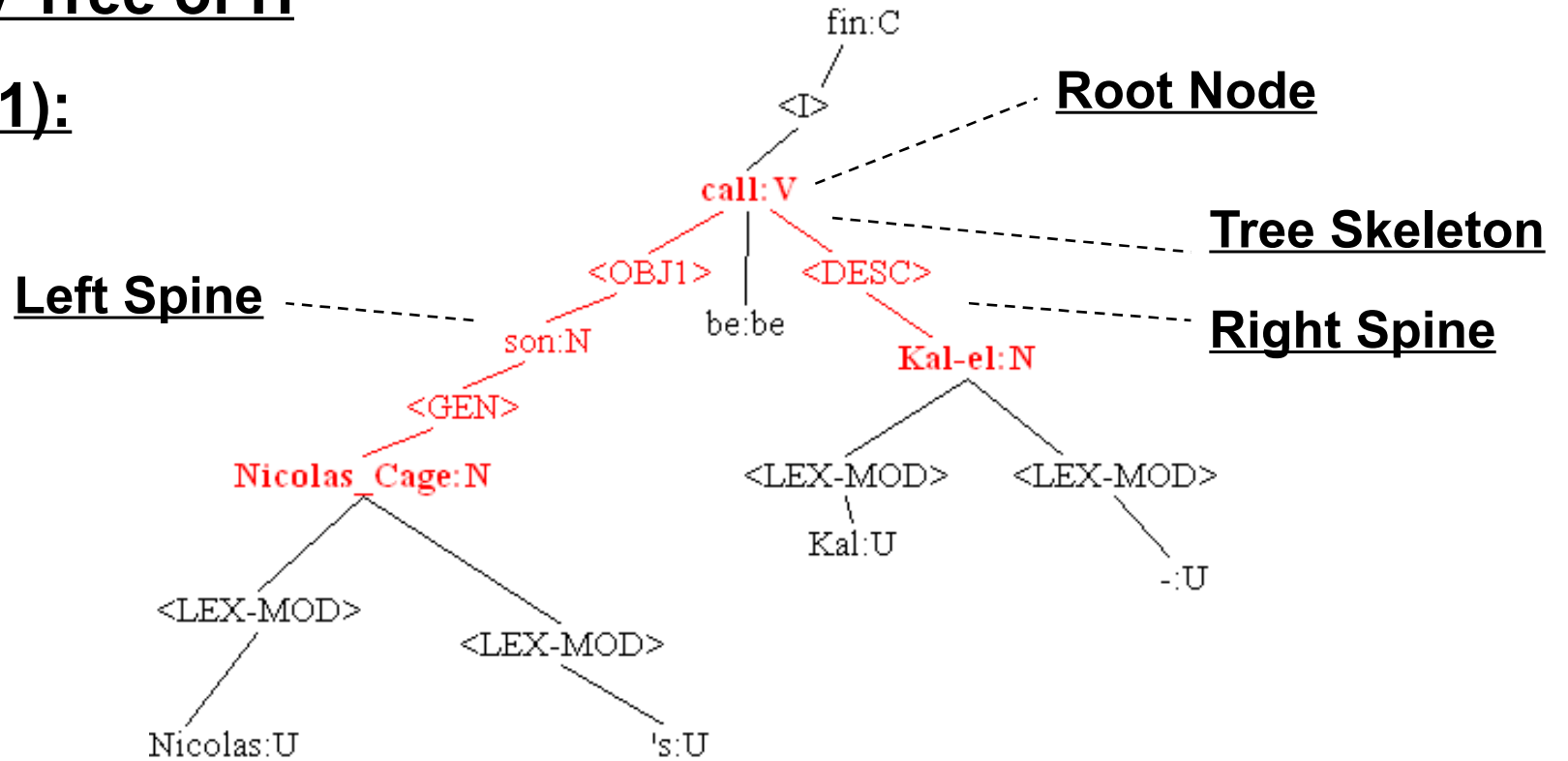
• Observations

- H is simpler than T
- H can help us to identify the relevant parts in T

Tree Skeleton

Dependency Tree of H

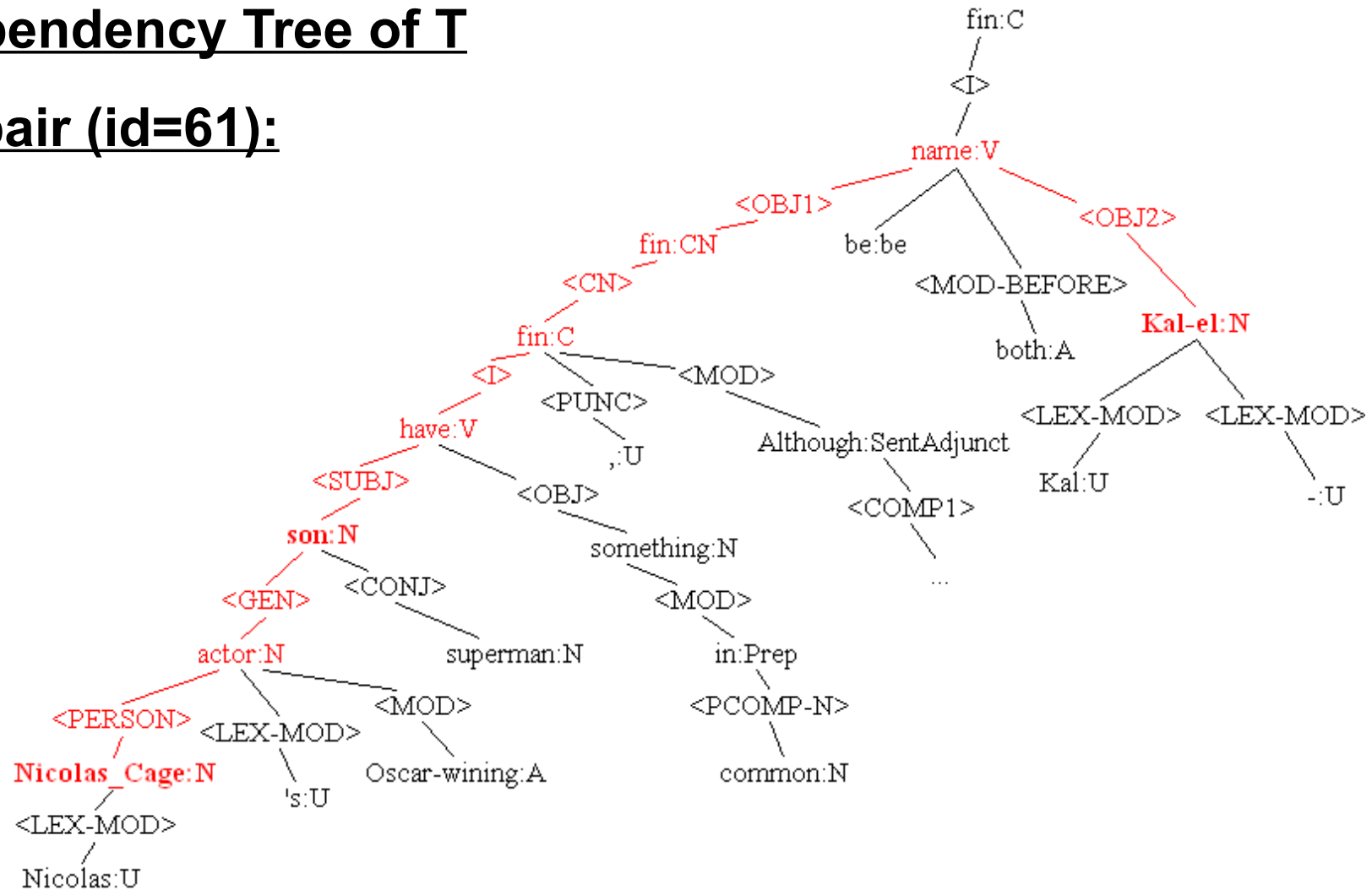
of pair (id=61):



Nicolas Cage's son is called Kal-el.

Tree Skeleton (cont.)

Dependency Tree of T of pair (id=61):



Generalization

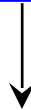
- Left Spine #Root Node# Right Spine

- Text

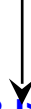
Nicolas_Cage:N <PERSON> actor:N <GEN> son:N <SUBJ> have:V <I> fin:C <CN> fin:CN <OBJ1>

#Name:V#

<OBJ2> Kal-el:N



Nicolas_Cage:N & N <GEN> son:N <SUBJ> V <I> C <CN> CN <OBJ1> #Name:V# <OBJ2> Kal-el:N



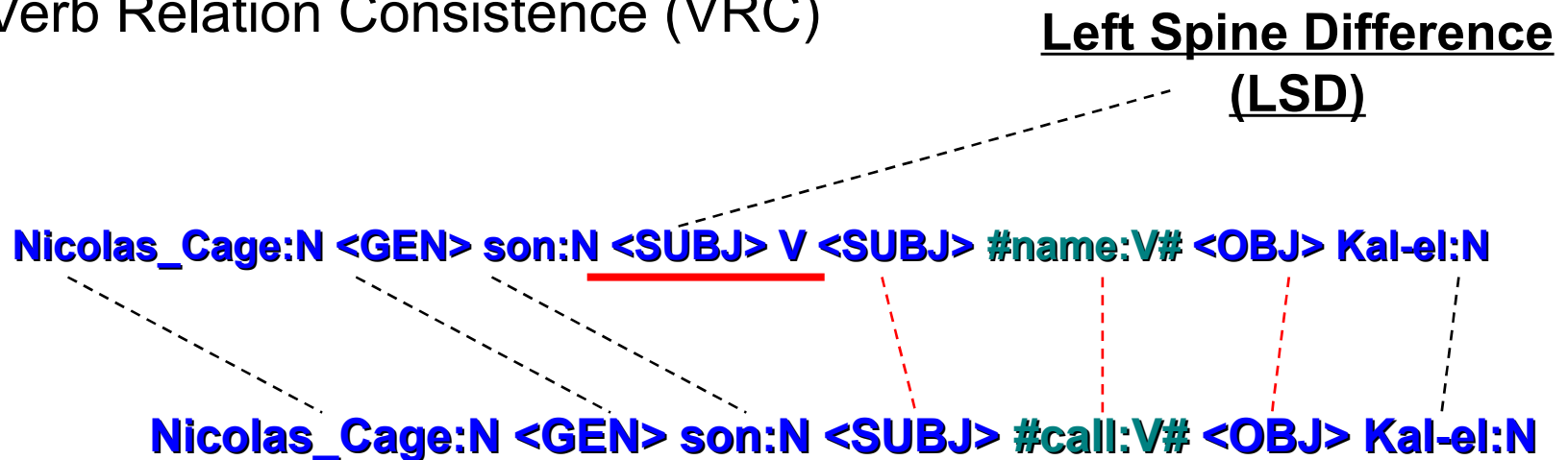
Nicolas_Cage:N <GEN> son:N <SUBJ> V <SUBJ> #name:V# <OBJ> Kal-el:N

- Hypothesis

Nicolas_Cage:N <GEN> son:N <SUBJ> #call:V# <OBJ> Kal-el:N

Spine Merging

- Merging
 - Left Spines: exclude Longest Common Prefixes
 - Right Spines: exclude Longest Common Suffixes
- RootNode Comparison
 - Verb Consistence (VC)
 - Verb Relation Consistence (VRC)



Pattern: Elementary predicate

- Pattern Format
 - $\langle \text{LSD}, \text{RSD}, \text{VC}, \text{VRC} \rangle \rightarrow \text{Predication}$
 - Example: $\langle \text{"SUBJ V"}, \text{""}, 1, 1 \rangle \rightarrow \text{YES}$
- Closed-Class Symbol (CCS)

Types	Symbols
Dependency Relation Tags	<i>SUBJ, OBJ, GEN, ...</i>
POS Tags	<i>N, V, Prep, ...</i>

- LSD and RSD are either *NULL* or CCS sequences

Testing Phase

- Pair: id="247" entailment="YES" task="IE"
source="BinRel"

- Text:

Author Jim Moore was invited to argue his viewpoint that Oswald , acting alone , killed Kennedy.

- Hypothesis:

Oswald killed Kennedy.

Testing Phase (cont.)

Text

.....
 <triple left="17" right="E0">kill:V mod-before vpsc:C</triple>
 <triple left="17" right="16">kill:V punc ,:U</triple>
 <triple left="17" right="E8">kill:V subj Oswald:N</triple>
 <triple left="17" right="18">kill:V obj Kennedy:N</triple>

Hypothesis

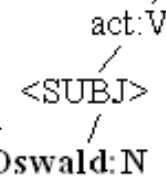
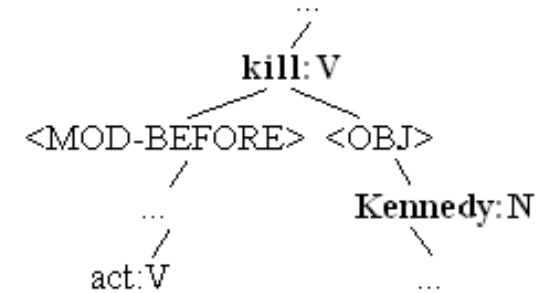
.....
 <triple left="E0" right="2">fin:C i kill:V</triple>
 <triple left="2" right="1">kill:V s Oswald:N</triple>
 <triple left="2" right="E2">kill:V subj Oswald:N</triple>

Oswald:N <SUBJ> V <SUBJ> #kill:V# <OBJ> Kennedy:N

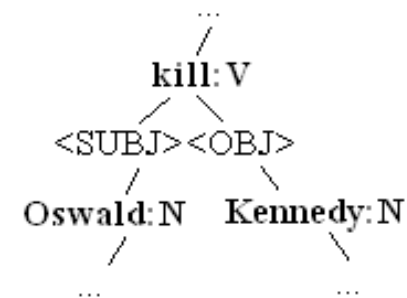
Oswald:N <SUBJ> #kill:V# <OBJ> Kennedy:N

<"SUBJ V", "", 1, 1> → YES

Text



Hypothesis



Experiments: System

- **Entailment methods:**

- Bag-of-Words (BoW)
- Triple Set Matcher (TSM)
- Minipar + Sequence Kernel + Backup Strategies (Mi+SK+BS)
- StanfordParser + Sequence Kernel + Backup Strategies (SP+SK+BS)

- **Classifier:**

- SVM (SMO) classifier from the WEKA ML toolkit

Experiments: Data

- **From RTE challenges:**

- RTE-2 Dev Set (800 T-H pairs) + Test Set (800 T-H pairs)
- RTE-3 Dev Set (800 T-H pairs) + Test Set (800 T-H pairs)

- **Additional data for IE and QA tasks:**

- Automatically collected from MUC6, BinRel (*Roth and Yih, 2004*), TREC-2003
- Manually classified into yes/no concerning entailment relation

Results on RTE-2 Data

Systems\Tasks	IE	IR	QA	SUM	ALL
Exp A1: 10-Fold Cross-Validation on Dev+Test Set					
BoW	50%*	58.8%	58.8%	74%	60.4%
TSM	50.8%	57%	62%	70.8%	60.2%
Mi+SK+BS	61.2%	58.8%	63.8%	74%	64.5%
Exp A2: Train: Dev Set (50%); Test: Test Set (50%)					
BoW	50%	56%	60%	66.5%	58.1%
TSM	50%	53%	64.5%	65%	58.1%
Mi+SK+BS	62%	61.5%	64.5%	66.5%	63.6%

* The accuracy is actually 47.6%. Since random guess will achieve 50%, we take this for comparison.

Results on RTE-3 Data

Systems\Tasks	IE	IR	QA	SUM	All
Exp B1: 10-fold Cross Validation on RTE-3 Dev Data					
BoW	54.5%	70%	76.5%	68.5%	67.4%
TSM	53.5%	60%	68%	62.5%	61.0%
Mi+SK+BS	63%	74%	79%	68.5%	71.1%
SP+SK+BS	60.5%	70%	81.5%	68.5%	70.1%
Exp B2: Train: Dev Data; Test: Test Data					
Mi+SP+SK+BS	58.5%	70.5%	79.5%	59%	66.9%*

* THE 5TH PLACE OF RTE-3 AMONG 26 TEAMS IN 2007 USING THE MOST

PURSTIC APPROACH

Components of the 5th best systems

Systems	Acc. %	Lx*	Ng	Sy	Se	LI	C	ML	B
Hickl et al.	80,00	X	X	X	X		X	X	X
Tatu et al.	72,25	X				X			X
Iftene	69,13	X		X					X
Adams	67,00	X	X				X	X	
DFKI	66,87			X				X	

* Following the notation in (*Giampiccolo et al., 2007*):

Lx: Lexical Relation DB;

Ng: N-Gram / Subsequence overlap;

Sy: Syntactic Matching / Alignment;

Se: Semantic Role Labeling;

LI: Logical Inference;

C: Corpus/Web;

ML: ML Classification;

B: Entailment corpora/Background Knowledge;

Comments

- Puristic approach:
 - We do not exploit any additional knowledge source beside the dependency trees nor have we extended the RTE training data
- Relational method:
 - For the **IE** task, SK method gives highest improvements
 - Kernel method seem to be more appropriate if the underlying task reveals a more “relational nature”
- Fallback strategies:
 - The “shallow” methods realized through BoW and TSM seem to work better for IR and SUM.

Extra Data Collection

- **IE**: MUC6, BinRel Corpus

- **T**: relevant sentence(s)
- **H**: NE + Relation + NE

Dole had hoped to pull out a win in North Carolina, the home state of his wife, Elizabeth.



Elizabeth is born in North Carolina.

- **QA**: TREC2003 QA

- **T**: (ir)relevant sentence(s)
- **H**: question + answer

Vice-President Albert Gore described the book "critically important" and compared it with "Silent Spring," Rachel Carson's 1962 book that set off a movement to ban DDT and other pesticides.



What book did Rachel Carson write in 1962?
Silent Spring

Results for SK method

Only SK method on Extra data (460 out of 750)

Methods\ tasks	IE (MUC,BinRel)	QA (TREC2003)	Overall
BoW	62.9%	61.4%	62.3%
TSM	64.9%	62.3%	63.8%
SK	76.3%	65.7%	74.5%

Only SK method on RTE-2 data

Exps\Tasks	IE	IR	QA	SUM	ALL
ExpA1: coverage	63.3%	18.3%	36.3%	16.3%	536
ExpA1: acc. of matches	64%	67.1%	66.2%	73.9%	66.2%
ExpA2: coverage	63.5%	23.5%	44%	17%	296
ExpA2: acc. of matches	66.9%	70.2%	58.0%	64.7%	64.5%

Zwischenfazit

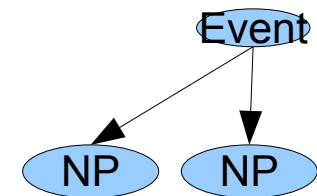
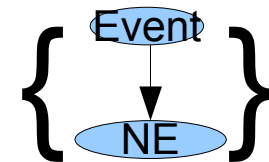
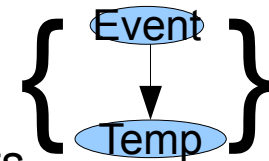
- Unsere rein syntaxbasierte Methode erreicht eine hohe syntaktische Baseline
 - Durch Konstruktion struktureller Merkmale über abstrakten Baumbeschreibungen von Dependenzanalysen
 - Diese Merkmale sind Eingabe für ein maschinelles Lernverfahren (Sequenz-Kernel)
 - Wichtig: Der Klassifikator wird nur dann angewendet, wenn für ein Textpaar $\langle T, H \rangle$ H sich strukturell mit T abgleichen läßt (Subtree Alignment) → erhöhte Akkuratheit!
- Zur Weiterentwicklung dieser Vorgehensweise
 - Feinkörnigere Kontrolle von Teilbaumbeziehungen

Divide-and-Conquer Methode

- Beobachtung: Wenn ein Subtree-Alignment möglich ist, dann kann mit hoher Akkuratheit Klassifizier Entailmentbeziehung erkennen.
- Bisher: einfacher exakter Stringvergleich von NP's und deren korrespondierenden Events/Roots
- Jetzt: fein-körnigerer Vergleich auf Basis von IE-Methoden und Spezifikation entsprechender Module.

Divide-and-Conquer Methode

- Kernidee:
 - Eine Menge von spezifischen RTE Methoden werden über einen Voting-Mechanismus kombiniert
- Jeder RTE-Spezialist betrachtet nur ein spezifisches RTE-Problem, so dass die zu erwartende Akkuratheit maximiert wird
 - T-M = Temporal Cases
 - Bestimme temporale Ausdrücke & korrespondierende Events
 - Spezifiziere Entailmentregeln für Time-Event Paare
 - NE-M = Named Entity Cases
 - Extrahiere Named Entities & korrespondierende Events
 - Spezifiziere Entailmentregeln für NE-Event Paare
 - TS-M = die bisherige NP-basierte Methode
 - Extrahiere NP's & korrespondierende Events
 - Bisheriger Sequenz-Kernel anwenden



Details zur T-M Methode

- Ziel: Zeige, das Entailment zwischen $\langle T, H \rangle$ NICHT gilt, ansonsten delegiere Problem
- Methode:
 - Bestimme alle temporalen Ausdrücke TMP im D-Baum
 - Bestimme für jeden TMP den regierenden Knoten (Ev) und bilde Paare $\langle \text{Ev}, \text{TMP} \rangle$
 - Setze: $\langle T, H \rangle ::= \{ \langle \langle \text{Ev}, \text{TMP} \rangle_T, \langle \text{Ev}, \text{TMP} \rangle_H \rangle \}$
 - Wenn alle Paare inkonsistent sind, dann antworte NO, sonst UNKNOWN
- Vergleichsmaße
 - WordNet, VerbOcean zw. Ev's
 - Manuell definierte Entailment-Regeln über temporalen Ausdrücken zw. TMP's

Semantisches Netz für Verben
mit hoher Abdeckung, Bsp-Eintrag:
transform [similar] integrate :: 10.728199

- Beispiel $\langle T, H \rangle$
 - **T: Released in 1995**, Tyson returned to boxing, **winning** the World Boxing Council title in **1996**. The same year, however, he lost to Evander Holyfield, and in a **1997 rematch bit** Holyfield's ear, for which he was temporarily banned from boxing.
 - **H: In 1996** Mike Tyson **bit** Holyfield's ear.
- Dependenzverhältnisse
 - **T:** 1995: released (verb); 1996: winning; 1997: rematch (noun), bit (verb)
 - **H:** 1996: bit (verb)
- Paarweiser Vergleich
 - $\langle \text{release}, 1995 \rangle, \langle \text{bit}, 1996 \rangle \rightarrow \text{NO}$
 - $\langle \text{win}, 1996 \rangle, \langle \text{bit}, 1996 \rangle \rightarrow \text{NO}$
 - $\langle \text{rematch}, 1997 \rangle, \langle \text{bit}, 1996 \rangle \rightarrow \text{NO}$
 - $\langle \text{bit}, 1997 \rangle, \langle \text{bit}, 1996 \rangle \rightarrow \text{NO}$
- $\langle T, H \rangle \rightarrow \text{NO}$

Evaluation im Rahmen der TAC-2008

BEST RANKING					
THREE-WAY TASK				TWO-WAY TASK	
3-W		2-W			
UAIC20081	0.685	UAIC20081	0.72	lcc1	0.746
OAQA1	0.616	OAQA1	0.688	UAIC20081	0.721
DFKI1	0.614	DFKI1	0.687	DFKI3	0.706
QUANTA1	0.588	QUANTA1	0.664	QUANTA1	0.659
UMD1	0.556	UMD1	0.619	DLSIUAES1	0.608

Teilnehmende Team:
26 (USA: 9, EU: 13, ASIEN: 4)
Teilnehmer/Task:
8 an 3-way, 13 an 2-way, 5 an beiden

2-Wege-Aufgabe (Details)

Tasks	T-M	TS-M	NE-M	BoW-BM	Tri-BM	Run1	Run2	Run3
IR(300)	75.0%/4	76.5%/85	61.0%/164	63.3%	54.3%	66.0%	72.3%	71.7%
QA(200)	90.0%/10	73.2%/82	54.8%/93	49.0%	53.5%	73.0%	72.0%	74.0%
SUM(200)	83.3%/6	74.5%/51	55.2%/67	63.5%	54.0%	64.0%	69.5%	71.5%
IE(300)	72.7%/11	74.2%/128	46.7%/152	50.0%	50.0%	66.7%	66.3%	66.7%
All(1000)	80.6%/31	74.6%/346	54.3%/477	56.5%	52.8%	67.2%	69.9%	70.6%

Run1: T-M, TS-M, and Tri-BM
Run2: T-M, TS-M, and BoW-BM
Run3: T-M, TS-M, NE-M, and Tri-BM, BoW-BM

Anwendungspotential

- RTE erfolgreich eingesetzt bereits für Antwortvalidierung
 - Beispiel
 - Q: „In which country was Edouard Balladur born?“, A: “France”
 - T: „*Paris, Wednesday CONSERVATIVE Prime Minister Edouard Balladur, defeated in France's presidential election, resigned today clearing the way for President-elect Jacques Chirac to form his own new government...*”
 - Entailed(Q+A, T) → YES/NO ?
 - Clef 2007/2008 → DFKI beste Ergebnisse für Englisch & Deutsch
 - Analoges Verfahren auch für Relationsextraktion, cf. Wang & Neumann, 2008.
- Neu: RTE für semantische Suche
 - Ist Frage X enthalten in (bereits beantworteten) Frage Y ?

Motivation

- Hybride Frageantwort (QA)-Systeme
 - Integration von strukturierten (DB) und unstrukturierten Antwortquellen
- Ontologiebasierte Informationsextraktion
 - Modellierung von Domänenwissen
- Textuelle Inferenz
 - Für einen kontrollierten semantischen Zugriff
 - Zur Bestimmung von impliziten Antworten
- Forschungsprojekte: QALL-ME (EU), ConQA (Saarland)

Der QA Flaschenhals

- Hybride QA:
 - zunehmende semantische Strukturierung (Semantic Web, Web 2.0) → Verschmelzung von Ontologie-basierten Datenbanken und Informationsextraktion aus Texten
 - Dynamik und Offenheit des Webs bedingt zusätzliche **neue** Komplexität in der NL Schnittstelle

"Wer schrieb das Drehbuch zu Saw III?"

Komplexe
linguistische &
wissensbasierte
Inferenzen

=

```
SELECT DISTINCT ?writerName WHERE  
{ ?movie name "Saw III"^^string . ?movie  
hasWriter ?writer . ?writer name ?writerName . }
```

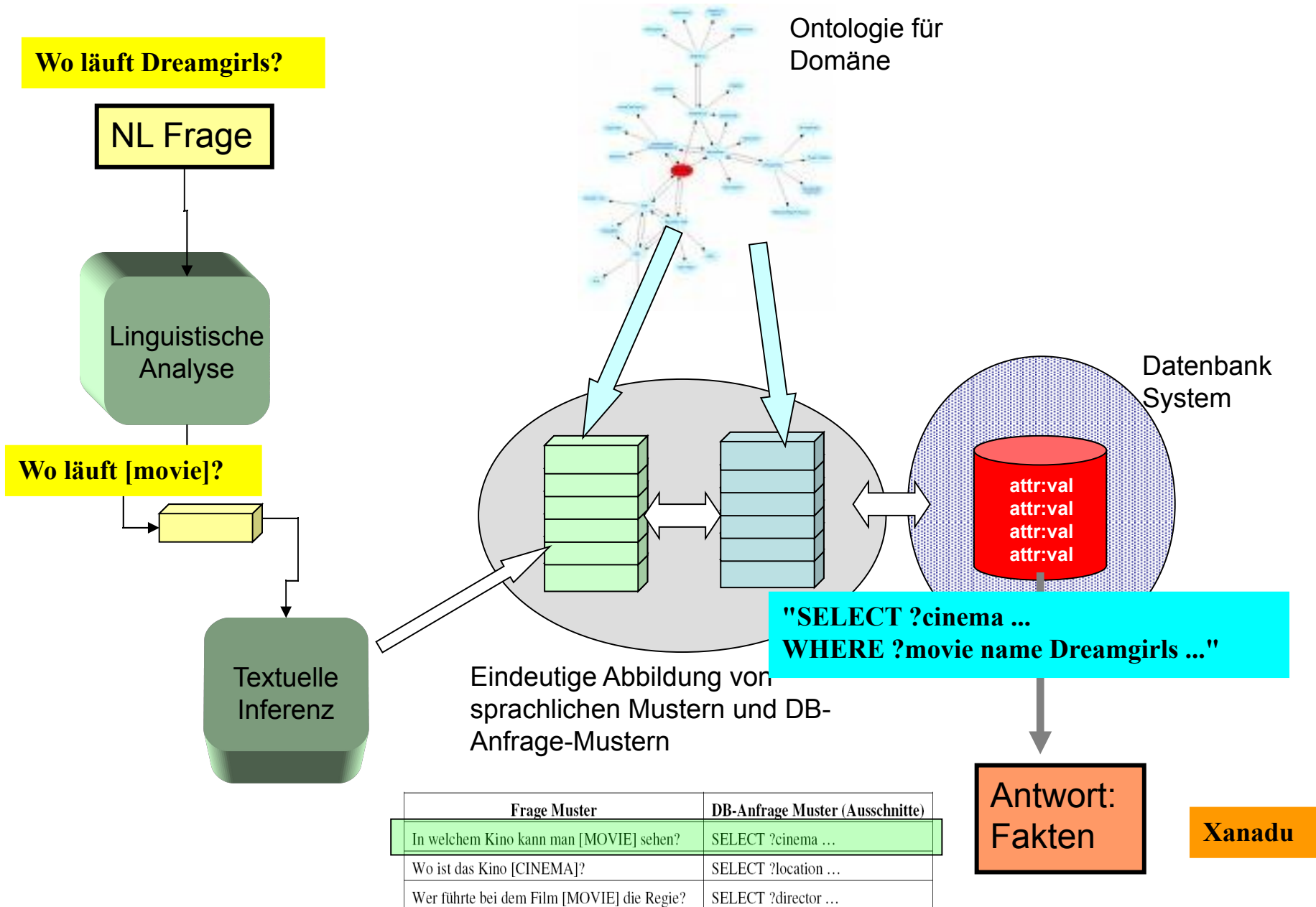
"Wer war Autor des Drehbuchs zum Film Saw III?"

Lösungsansätze

- Volle Berechnung (tiefe Inferenz)
 - AI complete; insbesondere, wenn auch unvollständige/falsche Anfragen erlaubt sind
- Kontrollierte Subsprache
 - Der Benutzer darf nur Fragen in einer bestimmten Form und mit eindeutiger Semantik formulieren
 - Nicht akzeptable kognitive Mehrbelastung
- Kontrollierte Abbildung
 - Eineindeutige Beziehung zwischen sprachlichen Mustern und DB-Anfrage Mustern
 - Sprachliche Freiheiten durch die Methode der Textuellen Inferenz ermöglichen

Entailment-based QA

Realisiert im EU-Projekt QALL-ME; open-source framework for entailment-based QA



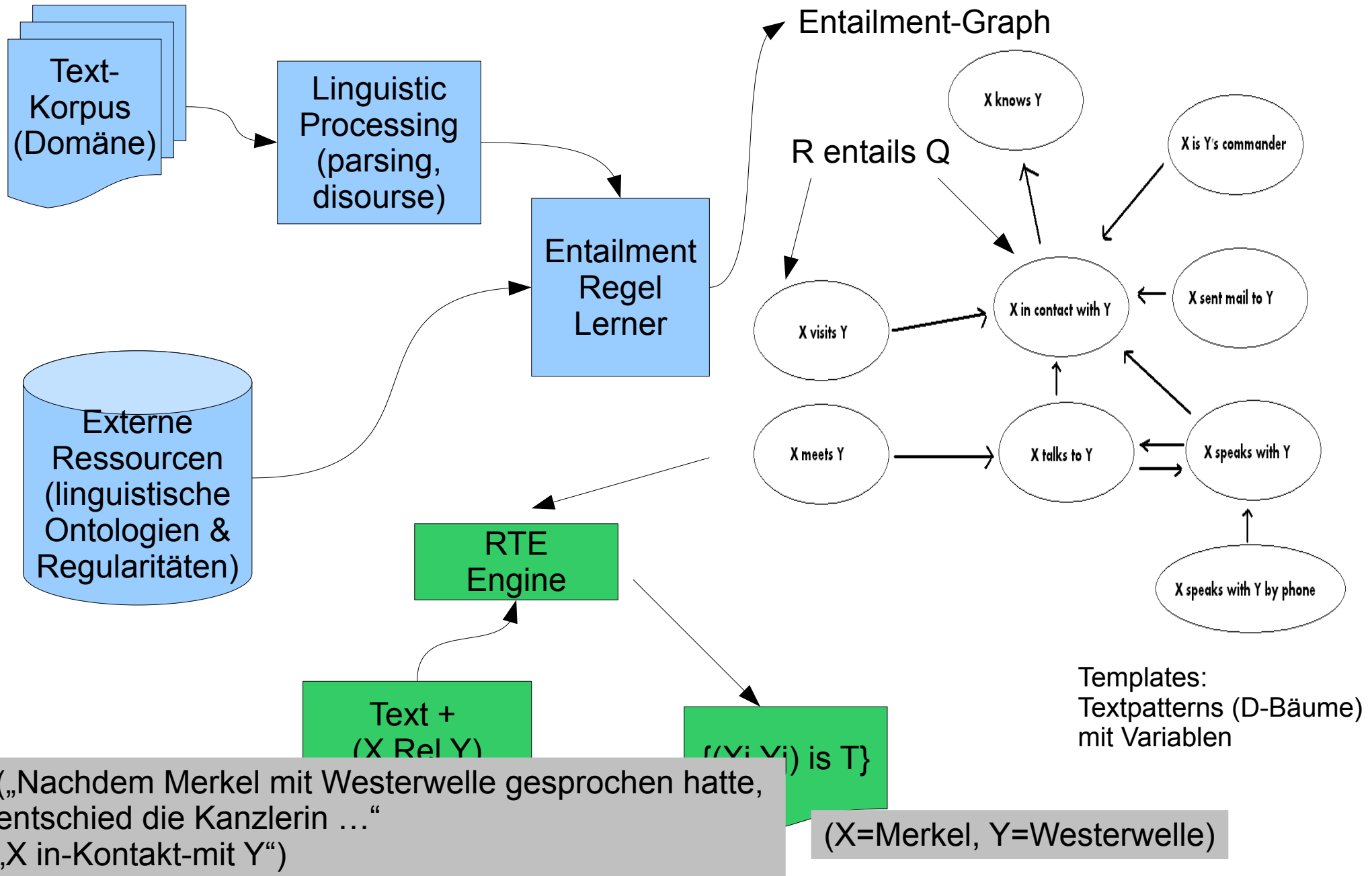
Vorteile

- Inferenzen verbleiben auf der sprachlichen Ebene
- RTE Methoden sind per se robust → unterstützen Verarbeitung von unvollständigen/fehlerhaften Anfragen
- Sehr gut geeignet auch für Zusammenspiel mit Informationsextraktion (IE)
 - Ontologiebasierte Informationsextraktion

RTE und Relationsextraktion

- Relationen zwischen Entitäten, z. B. „X ist in Kontakt mit Y“, „X spricht mit Y“, „X hat besucht Y“, „X ist angekommen in Y“
- Ziel: Relationen zwischen Entitäten automatisch extrahieren
- Zwei Schritte:
 - Das Lernen von Taxonomien für Relationen, und
 - Das Inferieren des Auftretens solcher Relationen in unstrukturierten Texten.

Ziel: hochskalierbare Methoden



Einige Herausforderungen

- Erkennung von Benennungen (mentioning) von Entitäten und Relationen („PERS reiste nach LOC“, „PERS erreicht LOC“)
- Bestimmung von Diskursinformation
- Integration/Erwerb lexikalischer, syntaktischer Entailment-Eigenschaften (Ontologien, lexiko-syntaktische Patterns, ...)
- Kontextuelle Präferenzen, z. B. es sollte eine bestehende Regel „X hit Y \rightarrow X attack Y“ nicht für einen Text feuern „I hit the ball over the net.“ \rightarrow zusätzliche Constraints, die Subtree-Alignment steuern, cf. Szpektor et al. 2008.
- Multilingualität: wie Relationen aus unterschiedlichen Sprachen integrieren?

Textuelle Inferenz: Fazit

- Erreicht wurde:
 - Hohe syntaktische Baseline
 - Weitere Verbesserung durch Integration IE-basierter Komponenten
 - Einbettung in größere Systemzusammenhänge → Anwendungspotential
- Aktuelle Schritte: (DFKI RTE-5 Erweiterungen)
 - Semantisches Annotieren von Abhängigkeitsrelationen
 - Referenzauflösung für satzübergreifende Analysen
 - 2bestes Ergebnis in Rahmen der TAC-2009
- Nächste Schritte:
 - Entailmentregeln → Entailment-Taxonomien (cf. Natural Logic) & maschinelle Lernverfahren
 - Multilingualität

Zusammenfassung & Zukunft

- Aktuell: Textuelle Inferenz
 - Als Kern für semantische Anwendungen (IE, QA, semantische Suche)
- Syntaktisch-Semantische Ansätze
 - Syntaktisch kontrollierte Verarbeitung (Beweissysteme)
 - Automatische Verfahren zur semantischen Role-Labeling und Typisierung von syntaktischen Einheiten (NG und VG)
- Hochskalierbare Methoden
 - Unüberwachtes semantisches Parsing (vgl. Pon&Domingos, 2009)
 - Effiziente Repräsentation und probabilistische Inferenzmaschinen, z. B. Markov Logic (vgl. Schoenmakers et al., 2008)