

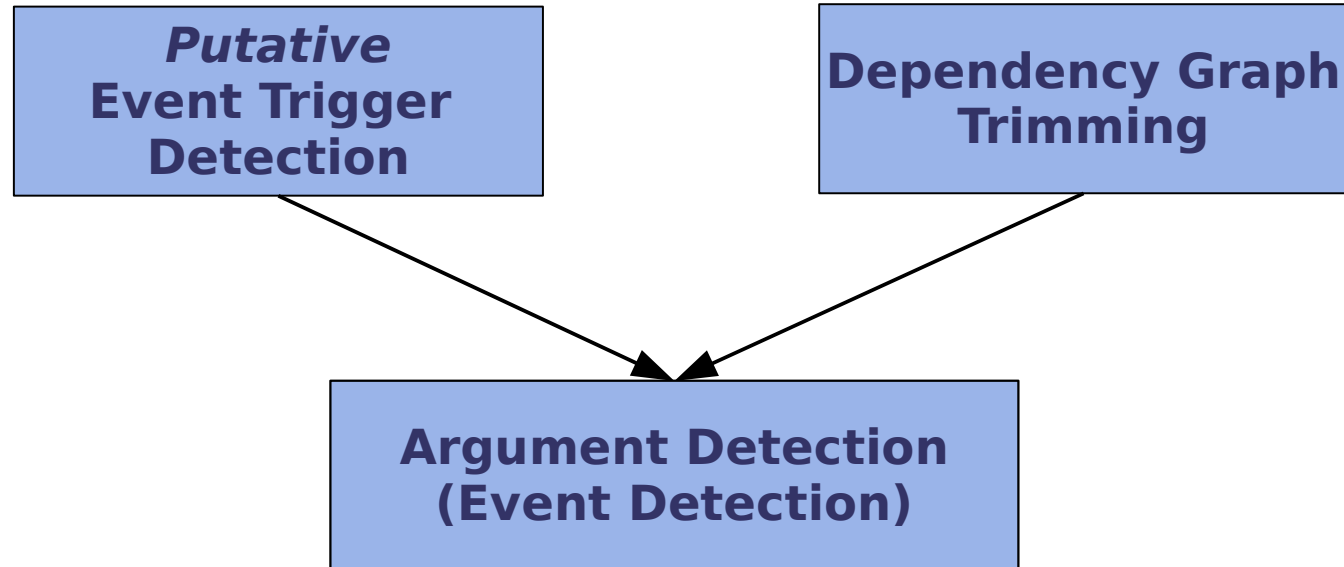
Event Extraction from Trimmed Dependency Graphs

Ekaterina Buyko, Erik Faessler,
Joachim Wermter and Udo Hahn

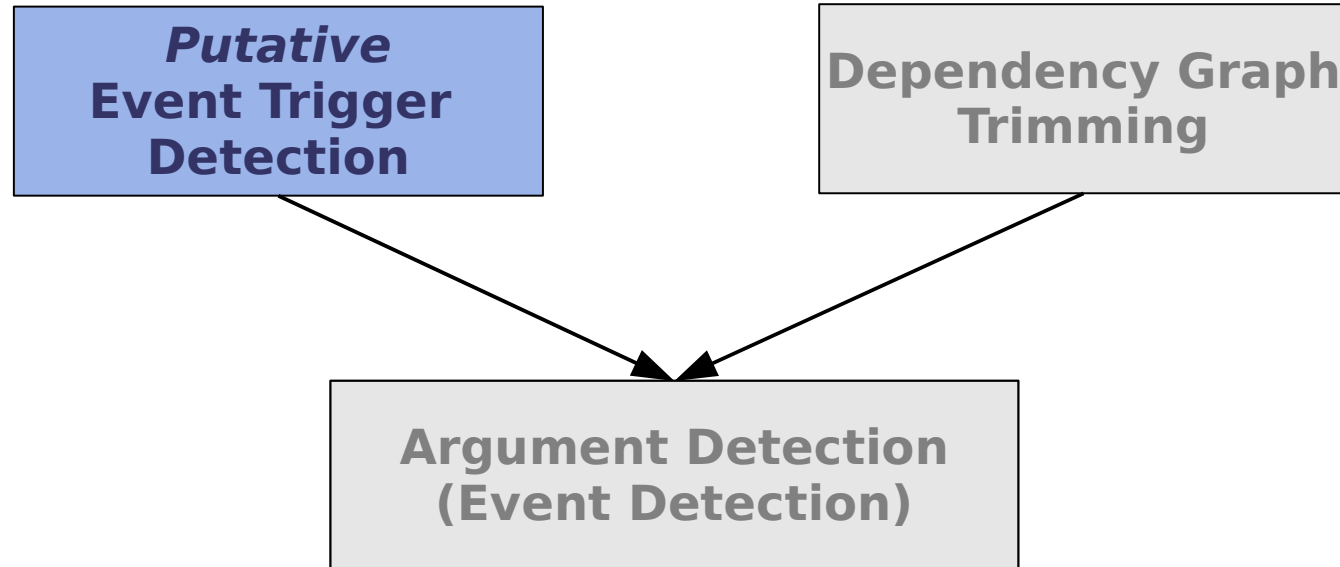
*Jena University
Language and Information Engineering (JULIE) Lab
Germany*



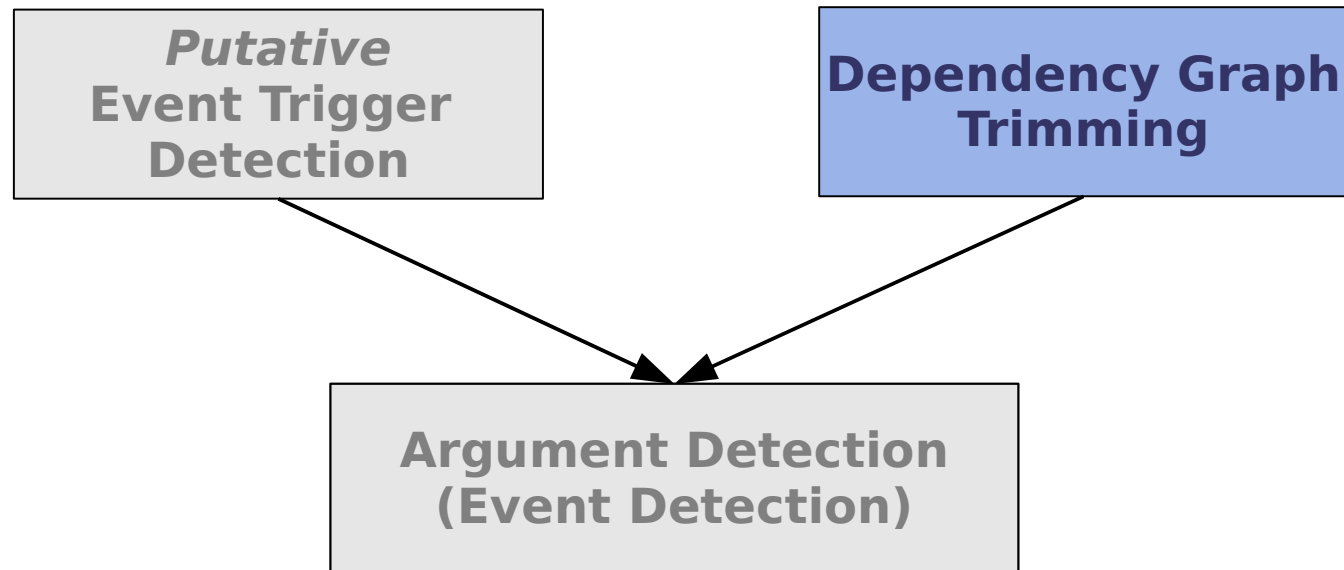
Event Extraction Solution



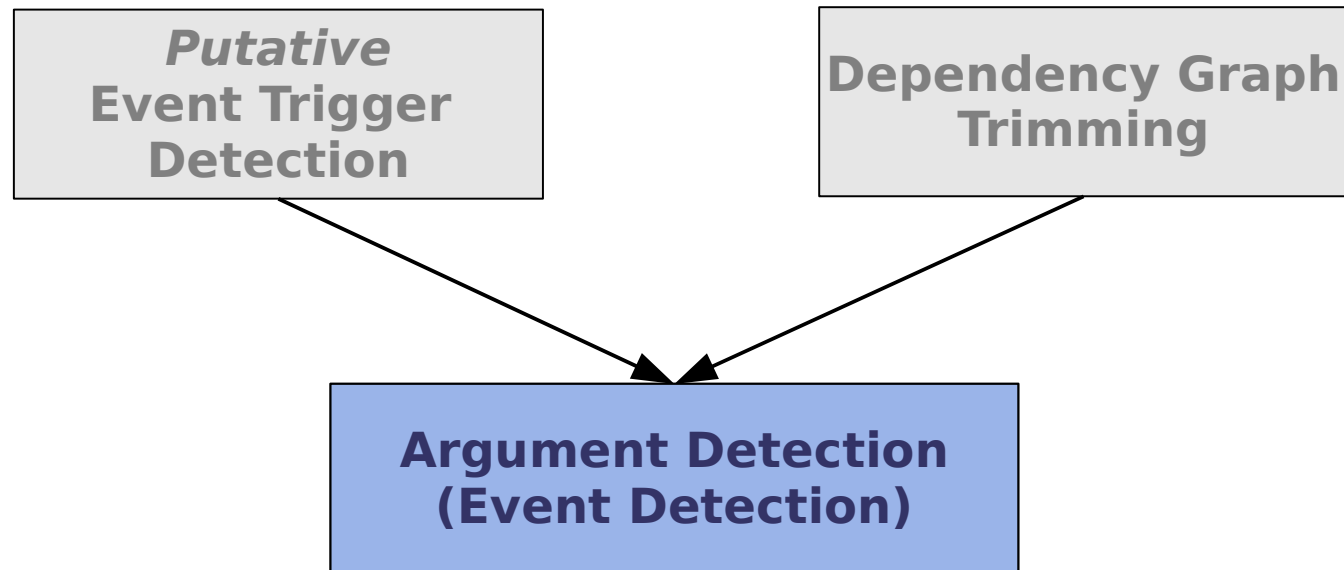
Event Extraction Solution



Event Extraction Solution



Event Extraction Solution



Event Trigger Detection (1/3)

To be or not to be an Event Trigger

Event Trigger Detection (1/3)

To be or not to be an Event Trigger

- wide variety
 - 3,730 alternative morphological forms for Positive regulation

Event Trigger Detection (1/3)

To be or not to be an Event Trigger

- wide variety
 - 3,730 alternative morphological forms for Positive regulation
- high ambiguity
 - 'induction' (Positive regulation, Transcription, Gene expression, not a trigger)

Event Trigger Detection (2/3)

Curation of Dictionaries

Event Trigger Detection (2/3)

Curation of Dictionaries

- important and discriminative
 - 'upregulate': Positive regulation

Event Trigger Detection (2/3)

Curation of Dictionaries

- important and discriminative
 - 'upregulate': Positive regulation
- important though not fully discriminative
 - 'cleavage': Protein catabolism

Event Trigger Detection (2/3)

Curation of Dictionaries

- important and discriminative
 - 'upregulate': Positive regulation
- important though not fully discriminative
 - 'cleavage': Protein catabolism
- by and large non-discriminative
 - 'presence': Localization, Gene expression

Event Trigger Detection (2/3)

Curation of Dictionaries

- important and discriminative
 - 'upregulate': Positive regulation
- important though not fully discriminative
 - 'cleavage': Protein catabolism
- by and large non-discriminative
 - 'presence': Localization, Gene expression
- absolutely non-discriminative
 - 'observe', 'demonstrate', 'function'

Event Trigger Detection (2/3)

Curation of Dictionaries

- important and discriminative
 - 'upregulate': Positive regulation
- important though not fully discriminative
 - 'cleavage': Protein catabolism
- by and large non-discriminative
 - 'presence': Localization, Gene expression
- absolutely non-discriminative
 - 'observe', 'demonstrate', 'function'

Event Trigger Detection (3/3)

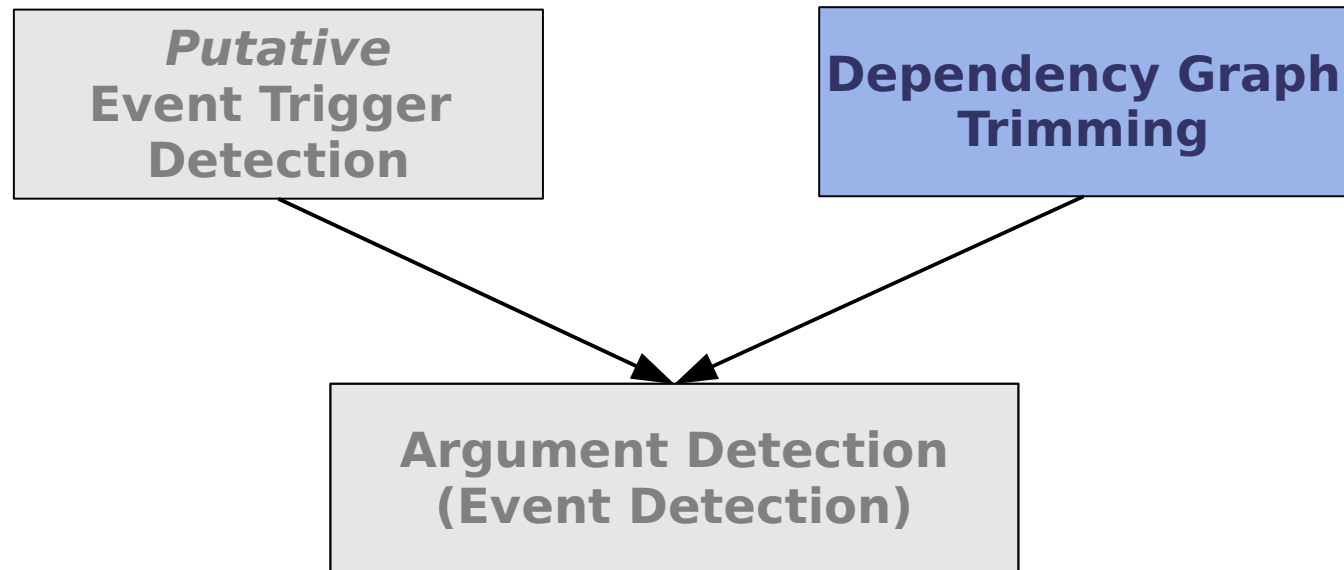
Disambiguation of Event Triggers

Importance of a trigger t_i for an event type e

$$Imp_e(t_i) := \frac{freq_e(t_i)}{\sum_j freq_e(t_j)}$$

$$type(t_i) = argmax(Imp_e(t_i))$$

Event Extraction Solution



Dependency Graph Trimming (1/3)

Introduction

Dependency Graph Trimming (1/3)

Introduction

- syntactic pruning
 - elimination of irrelevant nodes and relations
 - propagation of relations

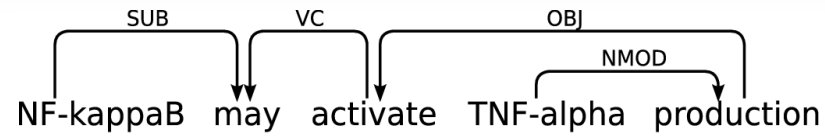
Dependency Graph Trimming (1/3)

Introduction

- syntactic pruning
 - elimination of irrelevant nodes and relations
 - propagation of relations
- conceptual overlaying
 - normalization of nodes
 - semantic decoration of nodes

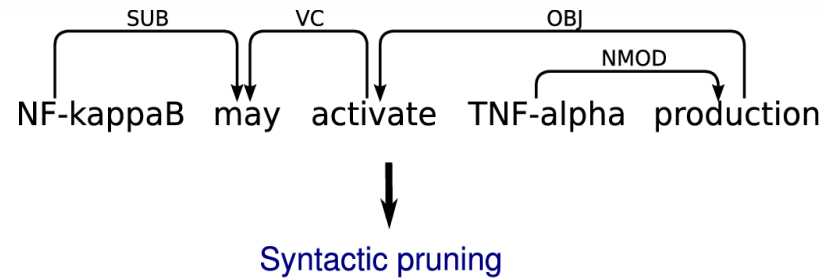
Dependency Graph Trimming (2/3)

Example



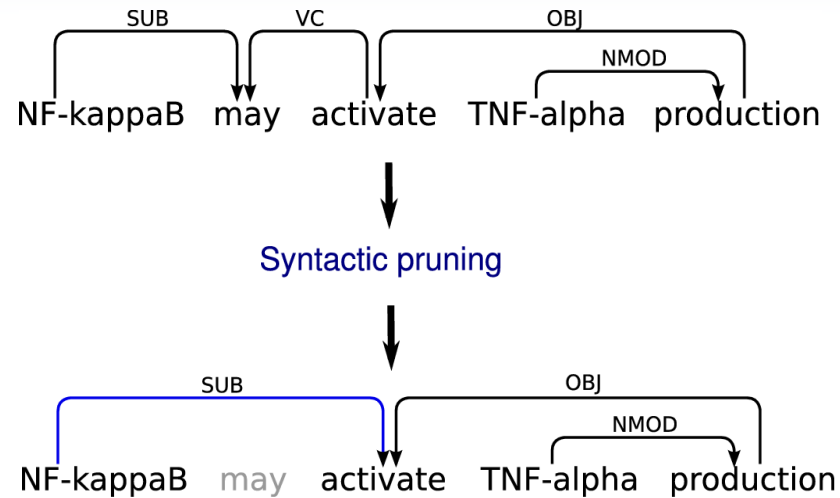
Dependency Graph Trimming (2/3)

Example



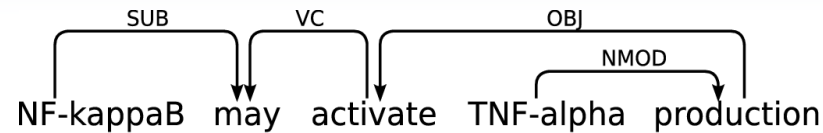
Dependency Graph Trimming (2/3)

Example

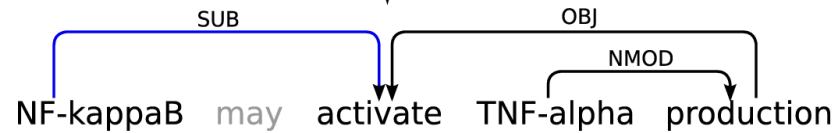


Dependency Graph Trimming (2/3)

Example



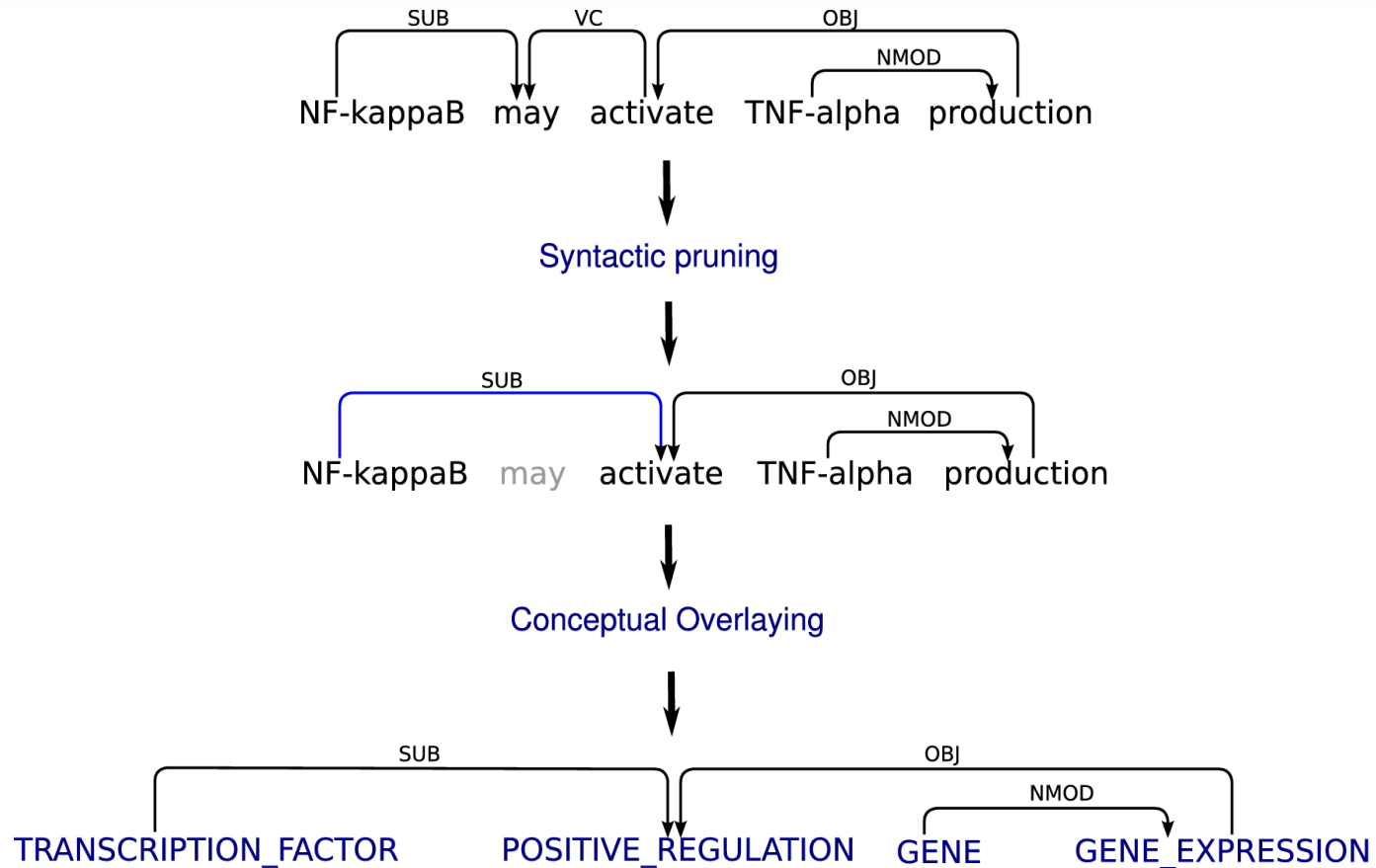
Syntactic pruning



Conceptual Overlaying

Dependency Graph Trimming (2/3)

Example

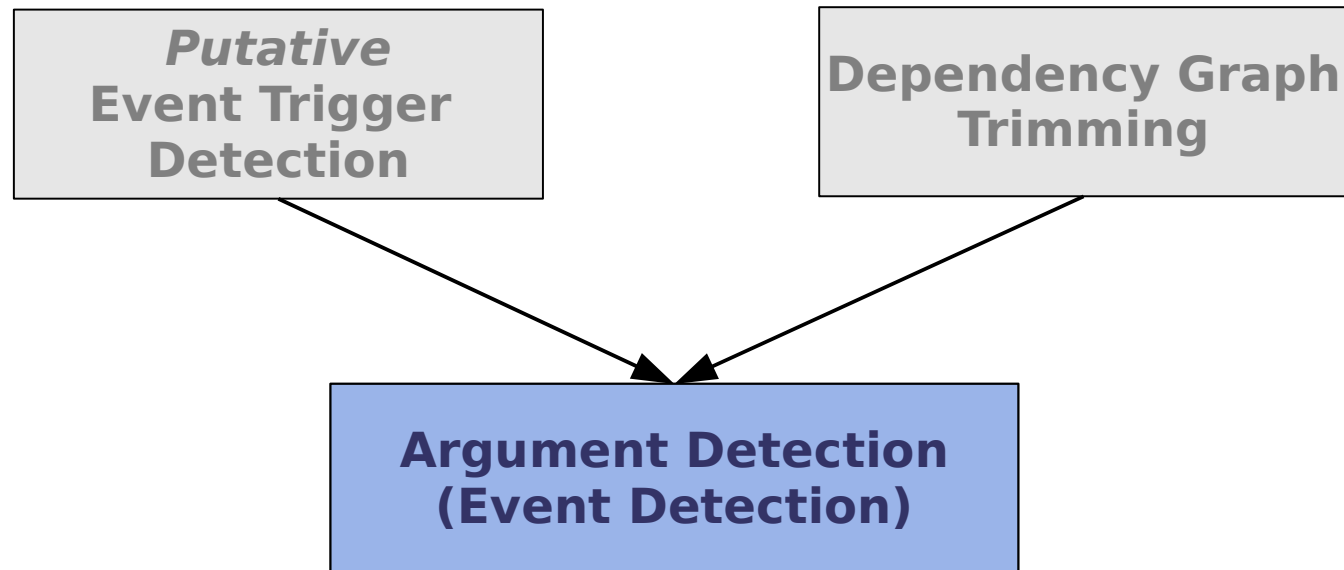


Dependency Graph Trimming (3/3)

Conceptual Overlaying

- specific semantic hierarchy
 - transcription factor, binding site, promoter
 - MeSH terms (except chemical and drugs brunch)
 - gene/protein
- Gene Ontology annotations (GOA)
 - event-specific categories of GO terms
 - mapping of genes via UniProt to GOA
- experimental methods (taken from GENIA)

Event Extraction Solution



Argument Identification (1/3)

Annotation of Shared Task Data

- addition of putative event triggers: training data (17,542 triggers (6,607 original triggers))

Argument Identification (1/3)

Annotation of Shared Task Data

- addition of putative event triggers: training data (17,542 triggers (6,607 original triggers))
- increase of negative event examples for learning

Argument Identification (1/3)

Annotation of Shared Task Data

- addition of putative event triggers: training data (17,542 triggers (6,607 original triggers))
→ increase of negative event examples for learning

Annotation Results for 7843230.xml in /mnt/data/data_bionlp_event_09/corpora/training-data/2-release/XML

Biphasic **control** of nuclear factor-kappa B **activation** by the T cell receptor **complex**: role of tumor necrosis factor alpha.

The regulation of nuclear factor (NF)-kappa B activation by the T cell receptor (TcR)/CD3 complex in primary human T cells has been studied at various times after activation. Only p50 NF-kappa B protein **bound** the kappa B element of interleukin-2 receptor (IL-2R) alpha chain promoter on resting T cells. However, immediately **after** TcR/CD3 **cross-linking** (after approximately 1 h; immediate) **binding** of p50.p65 **heterodimers** was observed. p50.c-rel **heterodimers** were also detected **bound** to this sequence at early time **points** (7-16 h; early), and both remained **active** at later time **points** (40 h; late) after **activation**. This **regulation** takes place mainly at the level of nuclear **translocation** of p65 and c-rel, at immediate and early time **points**. **Activation** also **induced** c-rel and p105/p50 mRNA **synthesis**, but not p65 mRNA whose **expression** was constitutive. Interestingly, all those early and late events, but not the immediate ones, were **inhibited** by a neutralizing **anti-tumor necrosis factor alpha** (TNF-alpha) monoclonal antibody. Similarly, cycloheximide **prevented** the p65

Click In Text to See Annotation Detail

- Annotations
 - EventTrigger
 - EventTrigger ("control")
 - begin = 9
 - end = 16
 - confidence = 1.0
 - componentId = de.julielab.
 - id = T1001
 - specificType = Regulation
 - ref = null
 - resourceEntryList = null
 - textualRepresentation = c

Legend

EventMenti... EventTrigger

Argument Identification (2/3)

Approaches

- feature-based
 - lexical features
 - shallow parsing features
 - dependency parsing features
- graph kernel-based (Airola et al., BioNLP 2008)
 - only shortest path information
 - modified representation of dependency relations

Argument Identification (3/3)

Event-specific classification

- graph kernel classifier
 - Phosphorylation
 - Localization
 - Protein catabolism
- feature based classifier
 - Regulation
 - Positive regulation, Negative regulation
- ensemble of feature and graph kernel classifiers
 - Binding
 - Gene expression & Transcription

Evaluation (1/3)

Baseline

shortest dependency path (SP) between putative argument and trigger

If

- no directional change
- no intervening event triggers

then

- argument is assigned the Theme role

Evaluation (2/3)

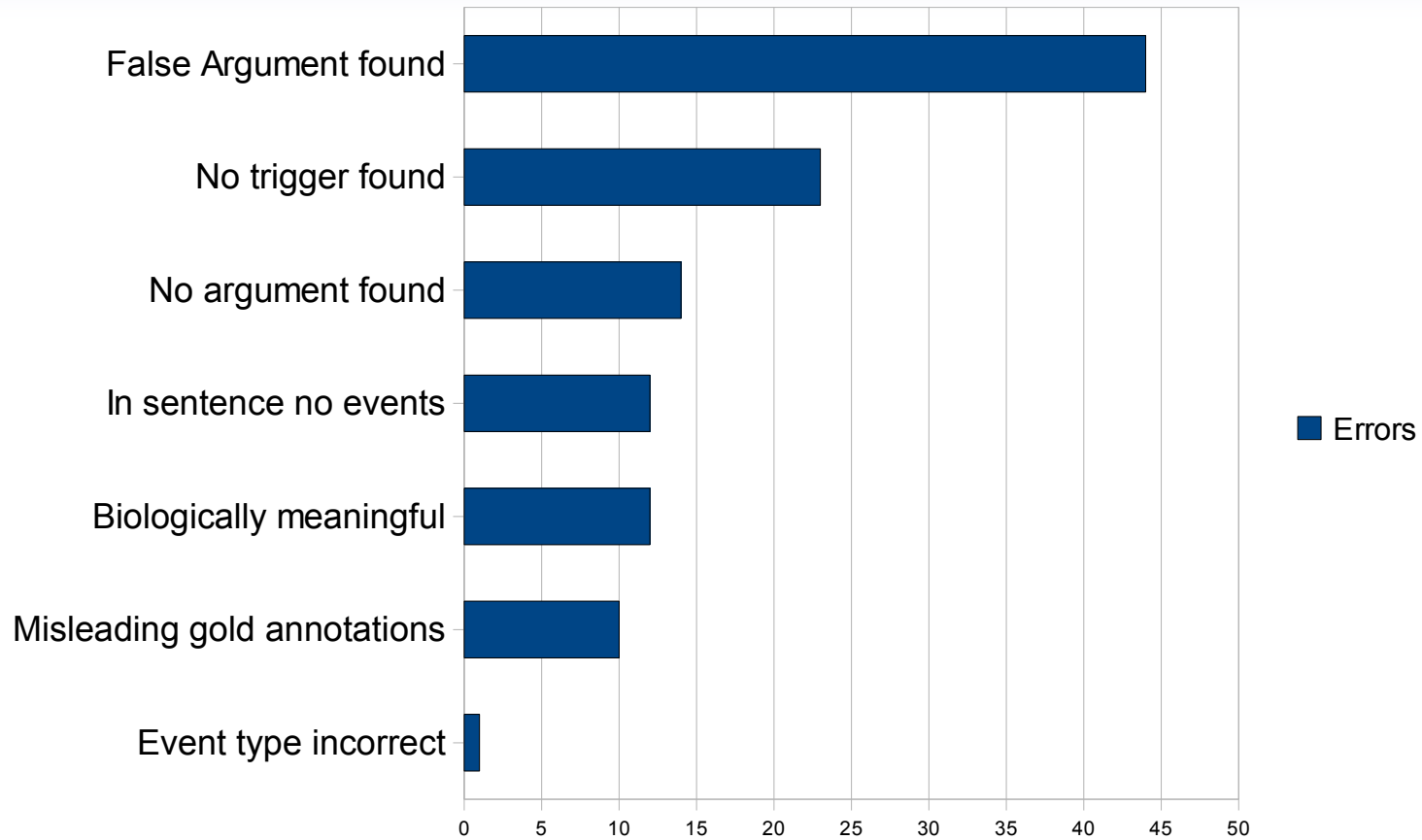
Results

		Baseline			Official run		
Event Class	gold	recall	prec.	F-score	recall	prec.	F-score
Localization	174	42.53	44.85	43.66	43.68	77.55	55.88
Binding	347	32.28	37.09	34.51	49.57	35.25	41.20
Gene expression	722	61.36	80.55	69.65	64.82	80.27	71.72
Transcription	137	39.42	35.06	37.11	35.77	62.03	45.37
Protein catabolism	14	71.43	66.67	68.97	78.57	84.62	81.48
Phosphorylation	135	65.93	90.82	76.39	76.30	91.15	83.06
EVT-TOTAL	1529	51.14	60.90	55.60	57.49	63.97	60.56
Regulation	291	9.62	11.72	10.57	31.27	30.13	30.69
Positive regulation	983	10.38	11.33	10.83	34.08	37.18	35.56
Negative regulation	379	14.25	19.22	16.36	40.37	31.16	35.17
REG-TOTAL	1653	11.13	12.96	11.98	35.03	34.18	34.60
ALL-TOTAL	3182	30.36	35.7	32.82	45.82	47.52	46.66

Table 1: Approximate Span Matching/Approximate Recursive Matching.

Evaluation (3/3)

Error Discussion



Summary

- 2nd rank in Shared Task with an f-score of 46.7%
- manual curation of dictionaries
- disambiguation of triggers
- trimming of dependency graphs
- enrichment of training data with negative examples
- event-specific configurations

Acknowledgements



**Boot
Strep**



<CA>LBC

SPONSORED BY THE



Federal Ministry
of Education
and Research

