« Semantic Role Labeling »

A challenging task in Natural Language Processing

Albert-Ludwigs-Universität Freiburg

Presented by	**Severin Gustorff**
Field of studies	Applied Computer Science
Seminar	Efficient Natural Language Processing
Date	14/12/11

UNI FREIBURG

Motivation

- What is a semantic role?
  - Semantic relationship that a participant has with the main verb in a clause or sentence.
  - *Example:*

John praised Mary.

agent	predicate	patient
Performs an action	The action to be done	Undergoes action and change ist state

  - There are other semantic roles:

Experiencer, Theme, Instrument, Force or Natural Cause, Direction or Goal, Recipient, Source or Origin, Time, Beneficiary, Manner, Purpose, Cause,…

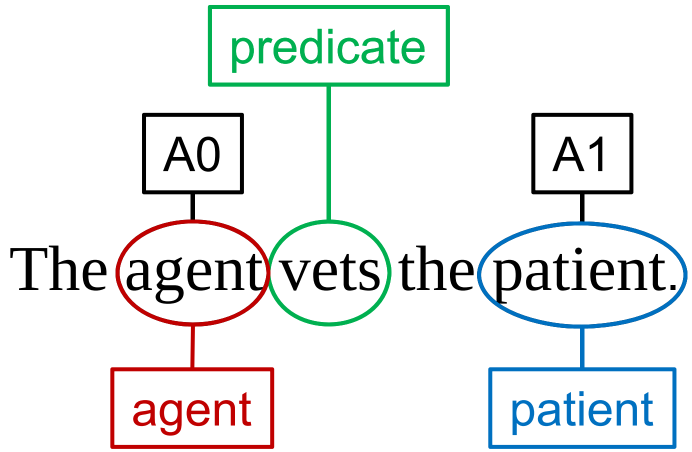
Motivation

- What is **Semantic Role Labeling** (SRL)?
  - task in natural language processing
  - the interpretation of a text requires the knowledge of the semantic roles of entities and events they participate in

identifying semantic arguments of the predicate

↓

classifying those arguments to their specific roles



Motivation

- Why do we need Semantic Role Labeling?
  - General: Finding semantic dependencies between words of certain classes
  - Applications:
    - **Question answering**

      Who **shot** Lee Harvey Oswald?

      agent — patient

    - **Gammar checking**

      When 900 years you reach, look as good, you will not.

      Comma doesn‘t belong here

    - **Translation**

English (SVO)	*Farsi (SOV)*	
[AGENT The little boy]	[AGENT pesar koocholo]	boy-little
[PRED kicked]	[THEME toop germezi]	ball-red
[THEME the red ball]	[ARGM-MNR moqtam]	hard-adverb
[ARGM-MNR hard]	[PRED zaad-e]	hit-past

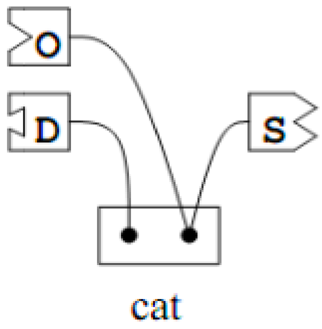
    - **Document Summarization** Predicates and Heads of Roles summarize content
    - **Information Extraction** (e.g. web mining, News tweets)

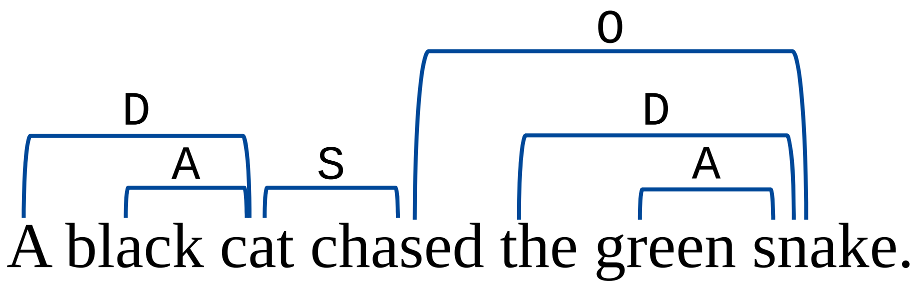
First Approaches

Link Parser

- Uses **Link Grammar**
- Roles as demands between the words
  - Example: The word 'cat'
    - can be a Subject (S)
    - can be an Object (O)
    - will have a Determiner (D)
    - Link Grammar description:

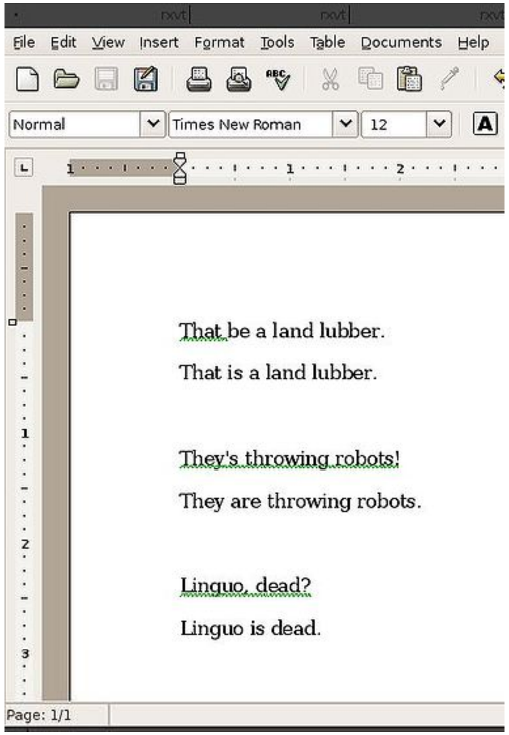
      `{A-}&D-{B+}&(O- or S+)`





Link Parser

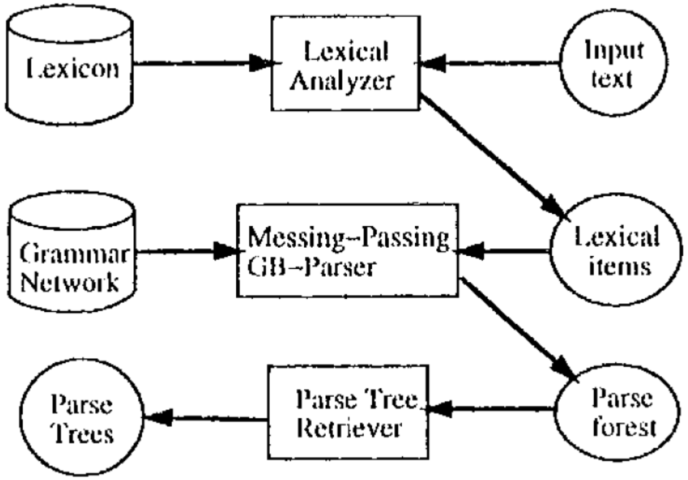
- Applications of Link Parsers
  - *AbiWord* grammar checking using the *RelEx semantic relationship extractor*
  - Information extraction of biomedical texts
  - Translation systems
  - Verification of natural language generation systems

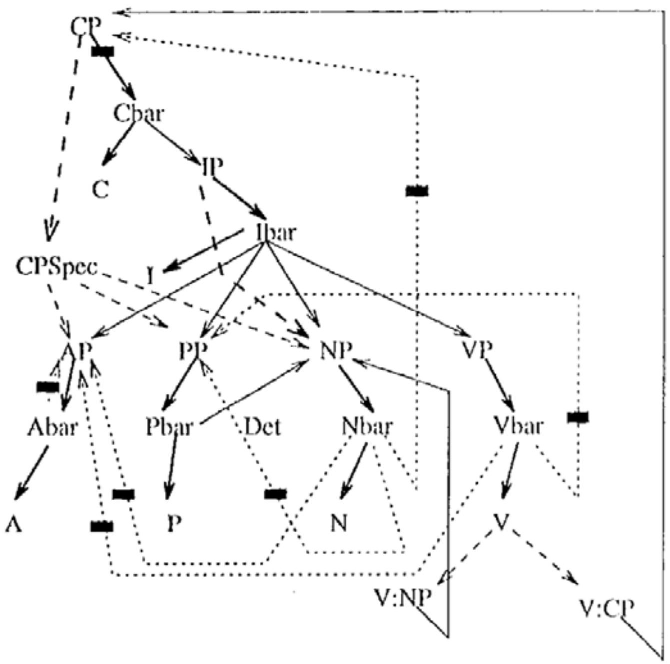


Syntactic Parser

PRINCIPAR (Principle-based English parser)

- Principles
  - any phrase can be moved anywhere (movement principle)
  - pronouns c ... ents (binding theory)
  - Other gram
- Grammar a
  - nodes = gra
  - links = type
  - Local const
  - Percolation constraints are attached to links



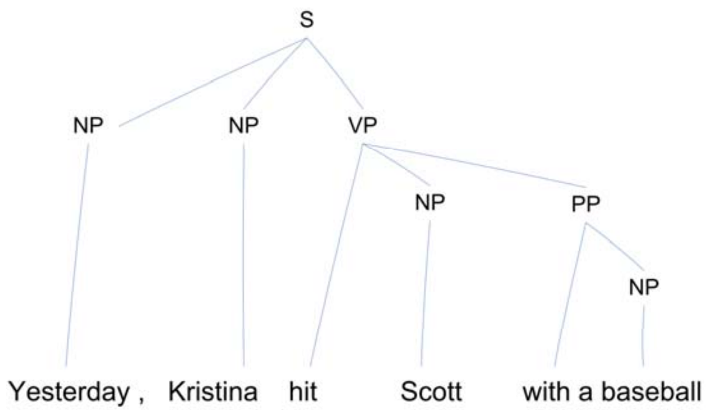


Syntactic Parser

Charniak‘s Parser

- Goal: Build / Expand a parse tree
- Maximum-entropy inspired (probabilistic) model
- Combine different conditioning events / features

  Lexical head of a word, pre-terminal, parent node, head of parent, grand parent node, left sibling
- Makes use of the PCFG (Probabilistic Context-Free Grammar)



Problems

- Only usefull for syntactic relationships
- But already well performing for grammar checking
- Certain basic level of semantics needed for grammar
- **No real semantic meaning!**

Where to get the meaning?

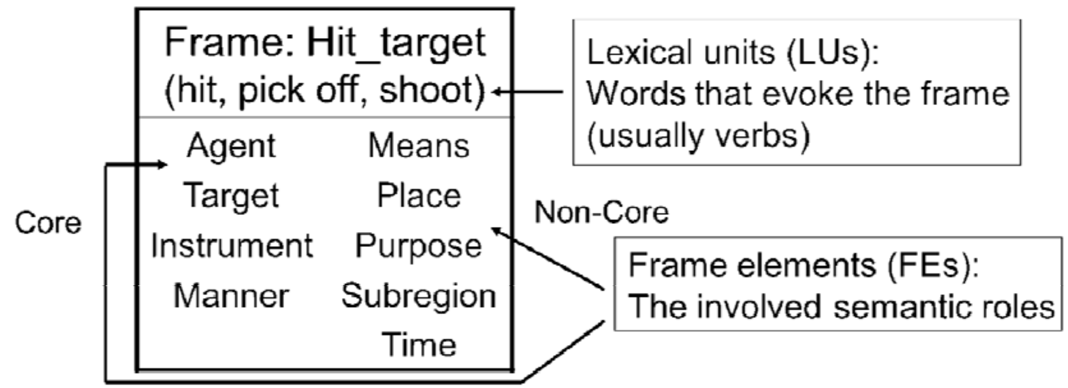
Hand-Labeled

Corpora

FrameNet [Fillmore et al. 2001] (Berkeley)

- Sentences from the British National Corpus (BNC)
- Annotated with frame-specific semantic roles

SIZE
>10,000 lexical units
>825 frames
>135,000 sentences



[$_{\text{Agent}}$ *Kristina*] ***hit*** [$_{\text{Target}}$ *Scott*] [$_{\text{Instrument}}$ *with a baseball*] [$_{\text{Time}}$ *yesterday*].

Corpora

PropBank (Proposition Bank) [Palmer et al. 05]

- Transfer sentences to verbal propositions
  - Kristina hit Scott → `hit(Kristina,Scott)`
- Based on *Penn TreeBank*
  - Add a semantic layer
  - Define a set of semantic roles for each verb
    - **A0** = Agent; **A1** = Patient or Theme; other arguments…
    - Adjunct-like arguments – **universal** to all verbs! (**AM-LOC**, **TMP**,…)
  - Uses Frame Files

`hit.01` "strike":

A0: agent, hitter; A1: thing hit;

A2: instrument, thing hit by or with

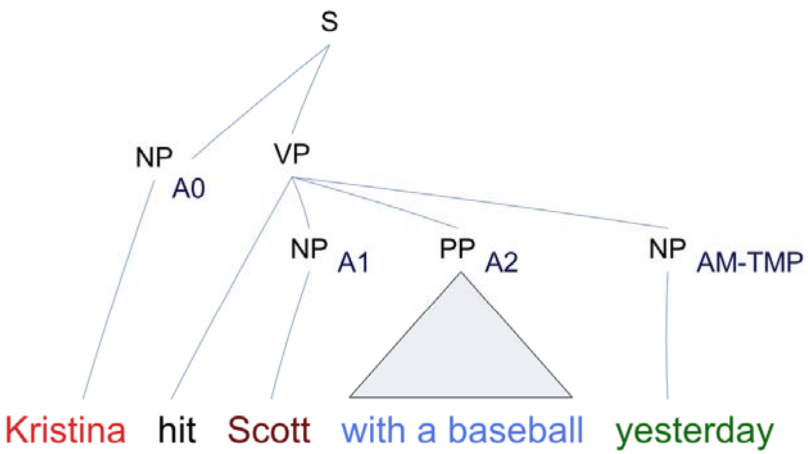
[$_{\mathbf{A0}}$ *Kristina*] ***hit*** [$_{\mathbf{A1}}$ *Scott*] [$_{\mathbf{A2}}$ *with a baseball*] *yesterday*.

SIZE
>3300 frame files
~113,000 propositions

Semantic roles
Syntactic annotations

Corpora

PropBank (continued)

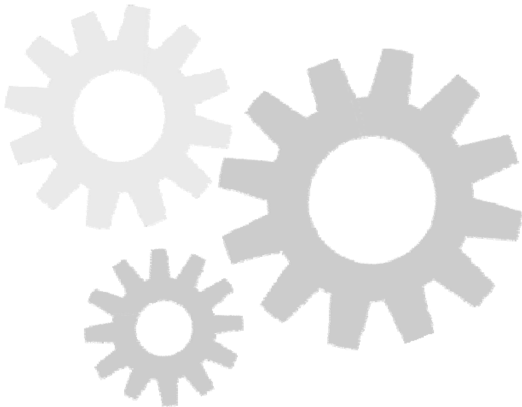


[$_{A0}$ *Kristina*] ***hit*** [$_{A1}$ *Scott*] [$_{A2}$ *with a baseball*] [$_{AM\text{-}TMP}$ *yesterday*].

A0	R-A0
A1	R-A1
A2	R-A2
A3	R-A3
A4	R-A4
A5	R-AA
AA	R-AM-ADV
AM	R-AM-CAU
AM-ADV	R-AM-DIR
AM-CAU	R-AM-EXT
AM-DIR	R-AM-LOC
AM-DIS	R-AM-MNR
AM-EXT	R-AM-PNC
AM-LOC	R-AM-TMP
AM-MNR	
AM-MOD	
AM-NEG	
AM-PNC	
AM-PRD	
AM-REC	
AM-TMP	

Other Corpora

- Chinese PropBank
- NomBank
- SemLink: Project to map between PropBank, VerbNet, FrameNet



Automatic SRL Systems

The Rise of Automatic SRL

- Gildea & Jurafsky 2002
  - First statistical model on FrameNet
- 7+ papers in major conferences in 2003
- 19+ papers in major conferences 2004, 2005
- 23+ papers in major conferences 2006, 2007
- 4 shared tasks
  - Senseval 3 (FrameNet) – 8 teams participated
  - CoNLL 04 (PropBank) – 10 teams participated
  - **CoNLL 05 (PropBank) – 19 teams participated**
  - SemEval 07 (FrameNet, NomBank, PropBank, Arabic SRL)

Function of A-SRL

Basic Notion

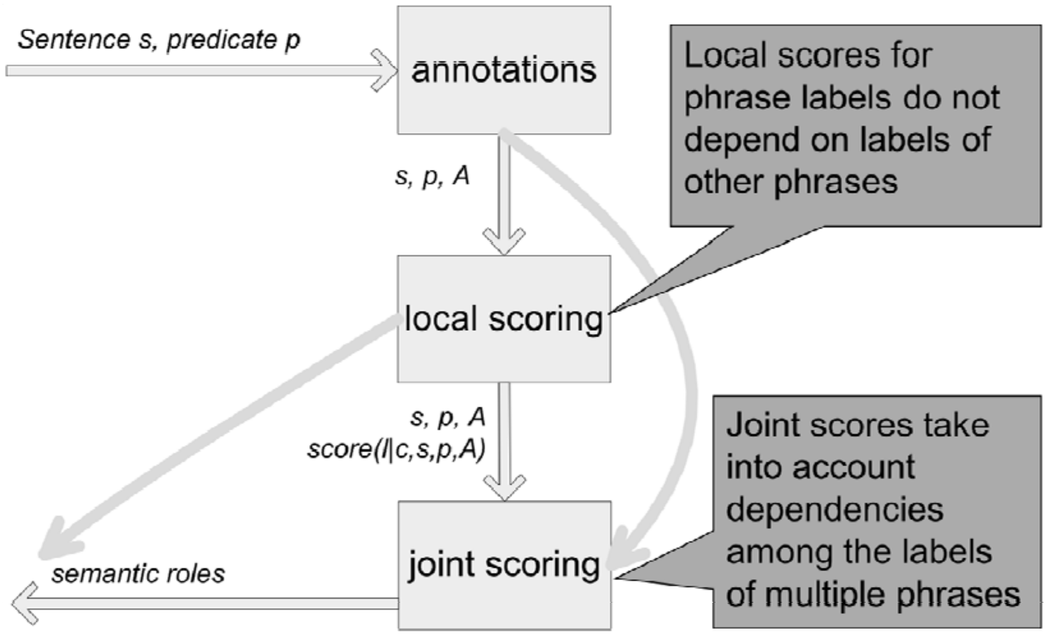
- *SRL is a mapping from the set of substrings of a string s to the label set L. L includes all argument labels and NONE.*

Subtasks

- Identification (arguments = A0 – A5; TMP, AM-LOC,…)
  - Separate the argument substrings from the rest in a sentence
  - Usually only 1 to 9 substrings are arguments and the rest have NONE for a predicate => **Hard task!**
- Classification
  - Given the set of arguments, decide the exact semantic label
  - Use features for classification

Function of A-SRL

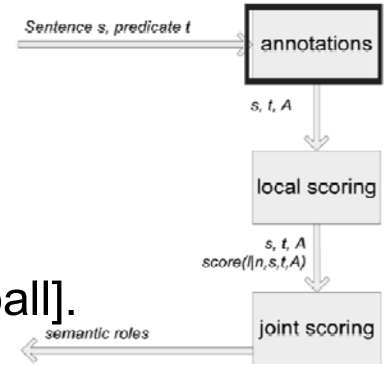
Basic Architecture SRL Systems



Subtasks of A-SRL

ANNOTATION – Syntactic Parsers

- Shallow parsing
- *Collins*‘ & *Charniak‘s* Parser



[$_{NP}$ Yesterday] , [$_{NP}$ Kristina] [$_{VP}$ hit] [$_{NP}$ Scott] [$_{PP}$ with] [$_{NP}$ a baseball].

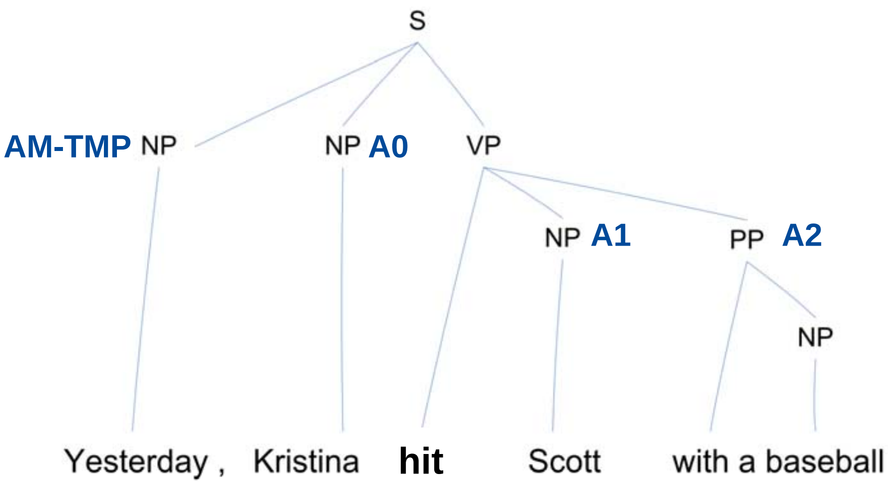
- Annotations from WordNet:

(v) **hit** (cause to move by striking)

⇒ WordNet hypernym

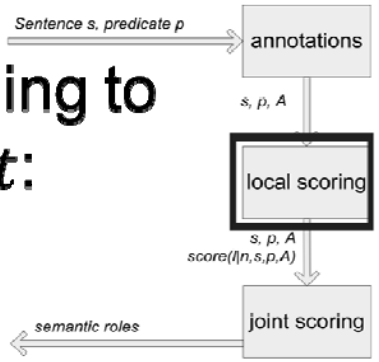
⇒ (cause to move forward with force)

- Also used: Pruning



Subtasks of A-SRL

LOCAL SCORING



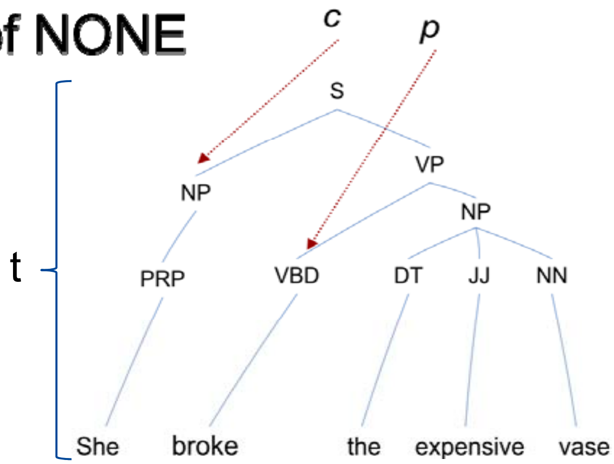
- Compute probabilities/scores for label l belonging to constituent c of predicate p given a parse tree t:

$$P(l \mid c,t,p)$$

  - Treat all words independently (local)
  - Filters out candidates with high probability of NONE
  - Candidates represented by features
  - Classification already incorporated

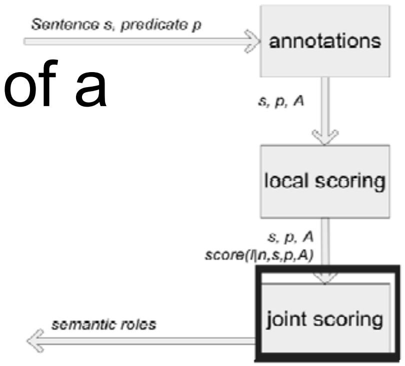
- Train classifier by Machine Learning:

  Back-off lattice-based relative frequency models,
  Decision trees, Support Vector Machines,
  Log-linear models, SNoW, AdaBoost, TBL, CRFs, IBL

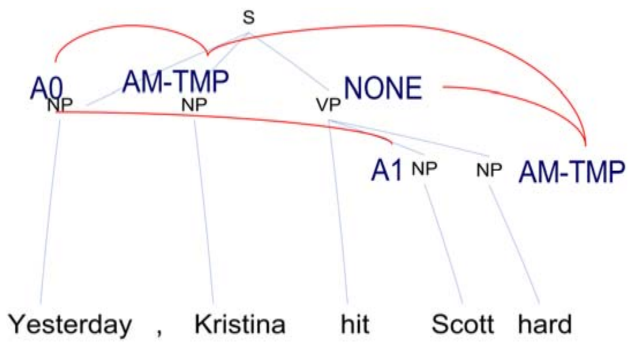


Subtasks of A-SRL

JOINT SCORING (a.k.a. global scoring)



- Use dependencies among several arguments of a predicate to ensure the assignments
- Can be done in different ways:
  - By constraints (e.g. arguments do not overlap)
  - Re-ranking of local scoring system; choose best assignment
  - Probabilistic models
    - Sequential Tagging
    - Conditional Random Fields
    - Generative models



Top Systems

Top Systems of the CoNLL´05

CoNLL-05 Shared Task on SRL

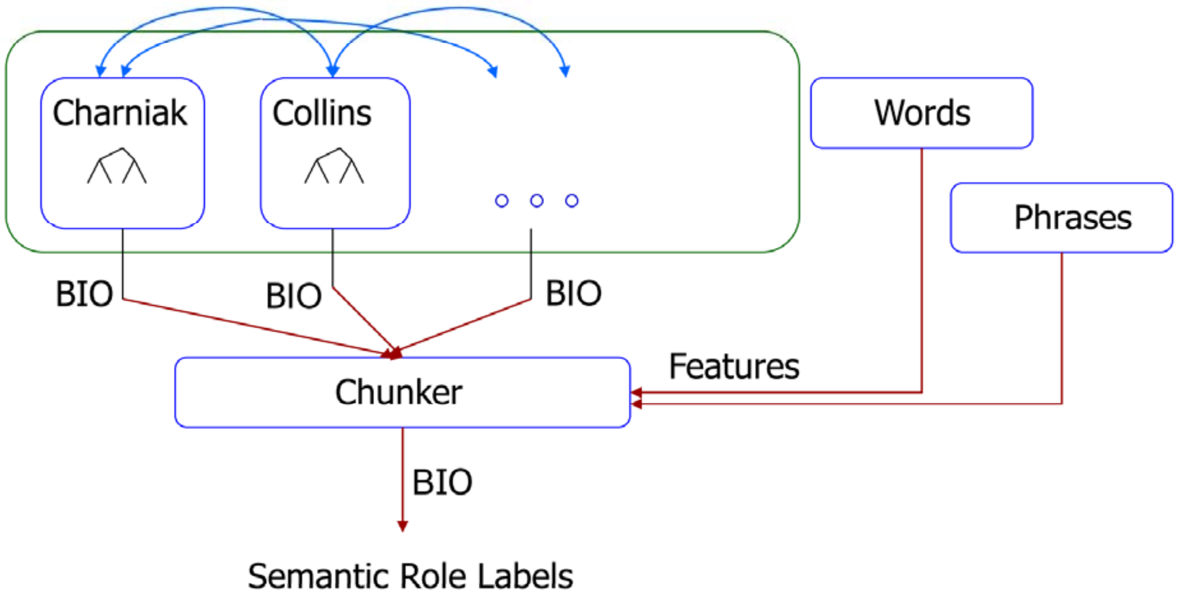
- Develop SRL systems using PropBank
- 19 teams participated
- 2 Testsets (Wall Street Journal, 2416 sent.; Brown corpus, 426 sent.)

Top Systems

#1 Punyakanok et al.	(University of Illinois)
#2 Haghighi et al.	(Stanford University)
#3 Màrquez et al.	(Technical University of Catalonia)
#4 Pradhan et al.	(University of Colorado)

#4 Pradhan et al.

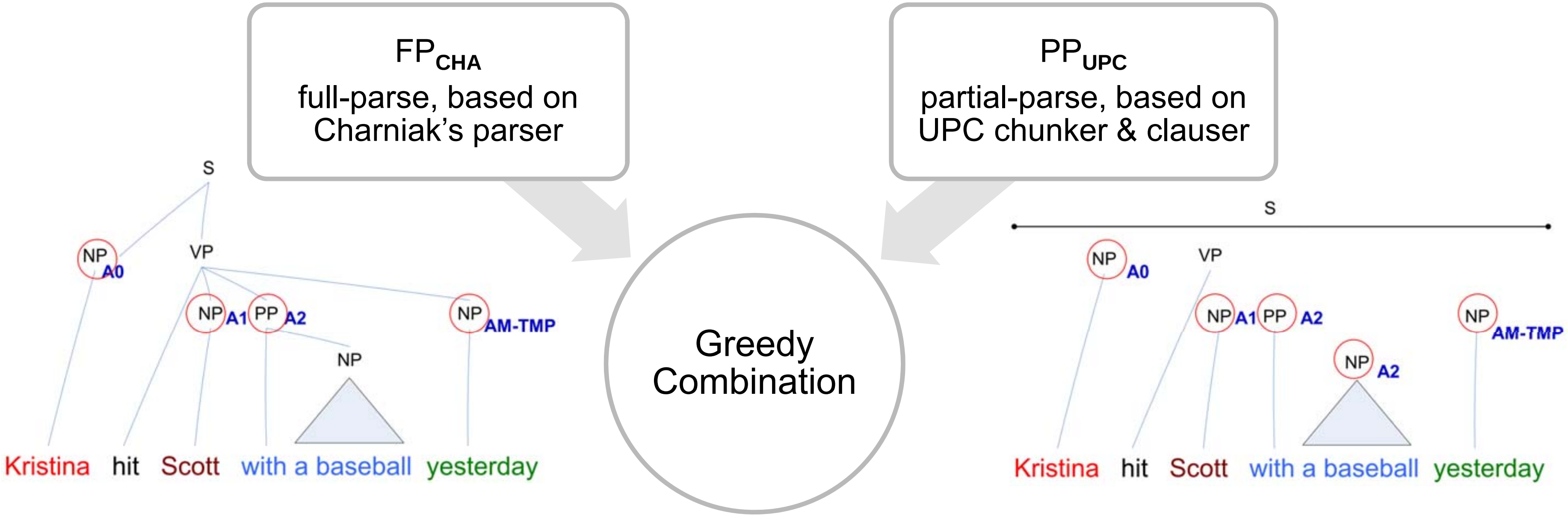
- Observation: the performance (F1) of an SRL system depends heavily on the syntactic view



- Classification by SVMs over 25 different features

#3 Màrquez et al.

- SRL is treated as a flat sequential labeling problem represented in the BIO format.



- Learning via AdaBoost

#1 Punyakanok et al.

- Output of the argument classifier often violates some constraints, especially when the sentence is long!
- Use **Integer Linear Programming**
  - **Input**: local scores (by the argument classifier), and structural and linguistic constraints
  - **Output**: the best legitimate global predictions
  - Formulated as an optimization problem
  - Allows incorporating expressive constraints on the argument types
  - This step is called Joint Inference

Results of the shared task

- Top performing systems (all combined systems):

	Development			Test WSJ			Test Brown			Test WSJ+Brown		
	P(%)	R(%)	F_1	P(%)	R(%)	F_1	P(%)	R(%)	F_1	P(%)	R(%)	F_1
punyakanok	80.05	74.83	77.35	82.28	76.78	79.44	73.38	62.93	67.75	81.18	74.92	77.92
haghighi	77.66	75.72	76.68	79.54	77.39	78.45	70.24	65.37	67.71	78.34	75.78	77.04
marquez	78.39	75.53	76.93	79.55	76.45	77.97	70.79	64.35	67.42	78.44	74.83	76.59
pradhan	80.90	75.38	78.04	81.97	73.27	77.37	73.73	61.51	67.07	80.93	71.69	76.03
baseline	50.00	28.98	36.70	51.13	29.16	37.14	62.66	33.07	43.30	52.58	29.69	37.95

- Syntactic parsers:

	Devel.			Test WSJ			Test Brown		
	P(%)	R(%)	F_1	P(%)	R(%)	F_1	P(%)	R(%)	F_1
UPC Chunker	94.66	93.17	93.91	95.26	94.52	94.89	92.64	90.85	91.73
UPC Clauser	90.38	84.73	87.46	90.93	85.94	88.36	84.21	74.32	78.95
Collins (1999)	85.02	83.55	84.28	85.63	85.20	85.41	82.68	81.33	82.00
Charniak (2000)	87.60	87.38	87.49	88.20	88.30	88.25	80.54	81.15	80.84

- Runtime
  - Punyakanok: Complete algorithm on both test sets: 1,7h [Collobert09]
  - *SENNA:* Complete algorithm on both test sets: 51s [Collobert09]

SENNA Sementaic/Syntactic Extraction using Neural Network Architecture

UNI FREIBURG

- NLP System for POS, Chunking, NER, and SRL
  - Unifies the different tasks by a multilayer neural network
    1. Extract word-features from Lookup-tables (single word)
    2. Extract word-features considering a surrounding window of words
    3. Combine feature vectors (convolution)
    4. Feed the NN with the features to get tagged output
  - Training (using *stochastic gradient ascent*)
    - Neural networks are trained with Word-Level- and Sentence-Level Log-Likelihoods
    - Entire English Wikipedia (!) + Reuters Corpus + 100.000 most common words from WSJ ($\Sigma = 850M$ words)
  - Program & Performance:
    - 2500 lines of C-Code
    - Computing all tags: < 0.001s/word
    - SRL $F_1 = 75.49$ % (Punyakanok et al.: $F_1 = 77.92$ %)

Benchmark machine: 3GHz Intel single core

Performance measures

- Test data:
  - 2433 sentences from WSJ (CoNLL2005 test set) ~59.000 words
- SENNA (only SRL)
  - All data processed in 58s = 0.001s/word = 0.024s/sentence
  - Main part of runtime consumed by convolution step
- Illinois Semantic Role Labeler [Punyakanok et al.]
  - All data processed in 87min = 0.1s/word = 2.15s/sentence
  - Main part of runtime consumed by Charniak‘s Full Parser!
- F1?

Conclusion

- SRL is an important problem in NLP
- Strong connections to applications requiring some degree of semantic interpretation
- Active topic of research, which has generated an important body of work in the last 8 years
- Latest works using enhanced learning methods show good results in speed and accuracy
- SRL still has to face some challenges before usage in real open-domain applications:
  - Widening the language domain (mosts systems only speak English)
  - More general corpora needed
  - Efficiency for massive text processing must be improved
  - Faster syntanctic parsers needed

References

[1] L.Carreras, X.Márquez, *Introduction to the CoNLL-2005 Shared Task: Semantic Role Labeling*, Technical University of Catalonia, 2005

[2] W.Yih & K.Toutanova, *Automatic Semantic Role Labeling*, Microsoft Research presentation, 2007

[3] L.Màrquez et.al., *Semantic Role Labeling as Sequential Tagging*, TU Catalonia 2005

[4] G.Hirst, *Semantic Interpretation and Ambiguity*, University of Toronto, 1988

[5] L.Carreras, X.Márquez, K.Litkowski, S.Stevenson, *Semantic Role Labeling: An Introduction to the Special Issue*, Universitat de Catalunya, MIT, University of Toronto, CL Research, 2008

[6] R.Gourab, *Semantic Role Labeling – M.Tech Seminar Report*, Indian Institute of Technology Bombay, 2010

[7] Illinois Semantic Role Labeler: http://cogcomp.cs.illinois.edu/page/software_view/12

[8] Link Grammar Parser: http://www.abisource.com/projects/link-grammar/

[9] R.Collobert, J.Weston et al., *Natural Language Processing (almost) from Scratch*, (SENNA project), NEC Labs America Princeton, Google New York, 2009

[10] E.Charniak, *A Maximum Entropy-Inspired Parser*, Brown University, 1999

[11] D.Lin, PRINCIPAR – An Efficient, Broad-coverage, Principle-based Parser, University of Manitoba, 1994

I	PRP	S-NP	O	-	S-A0
thank	VBP	S-VP	O	thank	S-V
you.	PRP	S-NP	O	-	S-A2

$\text{I}_{[A0]}$ $\text{thank}_{[V]}$ $\text{you}_{[A1]}$.