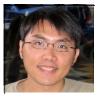


Non-local Dependencies and Semantic Role Labeling

Slides from

Scott Wen-tau Yih Kristina Toutanova
Microsoft Research (formerly UIUC, Stanford)
Roger Levy
UCSD (formerly Stanford)

1

Syntactic Variations versus Semantic Roles

Yesterday, *Kristina* hit *Scott* *with a baseball*

Scott was hit by *Kristina* *yesterday* *with a baseball*

Yesterday, *Scott* was hit *with a baseball* by *Kristina*

With a baseball, *Kristina* hit *Scott* *yesterday*

Yesterday *Scott* was hit by *Kristina* *with a baseball*

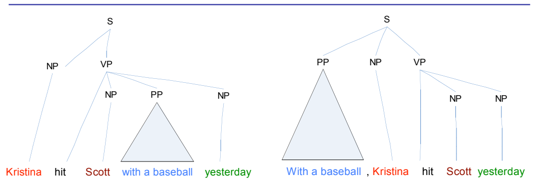
The baseball with which *Kristina* hit *Scott* *yesterday* was hard

Kristina hit *Scott* *with a baseball* *yesterday*

Agent, hitter | Thing hit | Instrument | Temporal adjunct

2

Syntactic Variations (as trees)



3

Semantic Role Labeling – *Giving Semantic Labels to Phrases*

- [$_{\text{AGENT}}$ John] **broke** [$_{\text{THEME}}$ the window]
- [$_{\text{THEME}}$ The window] **broke**
- [$_{\text{AGENT}}$Sotheby's] .. **offered** [$_{\text{RECIPIENT}}$ the Dorrance heirs] [$_{\text{THEME}}$ a money-back guarantee]
- [$_{\text{AGENT}}$ Sotheby's] **offered** [$_{\text{THEME}}$ a money-back guarantee] to [$_{\text{RECIPIENT}}$ the Dorrance heirs]
- [$_{\text{THEME}}$ a money-back guarantee] **offered** by [$_{\text{AGENT}}$ Sotheby's]
- [$_{\text{RECIPIENT}}$ the Dorrance heirs] will [$_{\text{ARM-NEG}}$ not] be **offered** [$_{\text{THEME}}$ a money-back guarantee]

4

Recovering non-local dependencies

- We want to interpret non-local dependencies:

  Who did Mary want to *see* _? (WH question)
  What did Paul expect _ *to arrive* ?

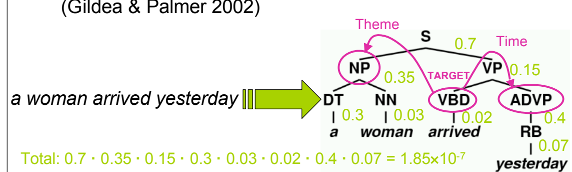
- We know how to disambiguate (in parsing) with probabilistic context-free grammars
- But non-local dependencies aren't *transparently* represented by context-free grammars

A woman arrived *who I knew*. (RC extraposition)
We discussed *plans* yesterday *to redecorate the house*.(VP extraposition)
The story was about *the children* that the witch wanted to *eat*.
These are *the children* that Mary wanted to *compete*. (relativization)

5

What is the state of the art in robust computation of linguistic meaning?

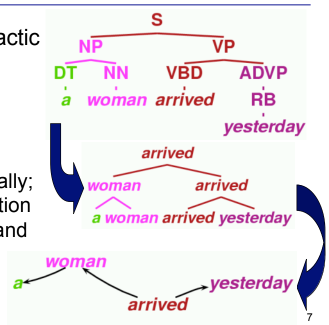
- *Probabilistic context-free grammars* trained on syntactically annotated corpora (***Treebanks***) yield robust, high-quality syntactic parse trees
- Nodes of these parse trees are often reliable indicators of *phrases* corresponding to semantic units (Gildea & Palmer 2002)



6

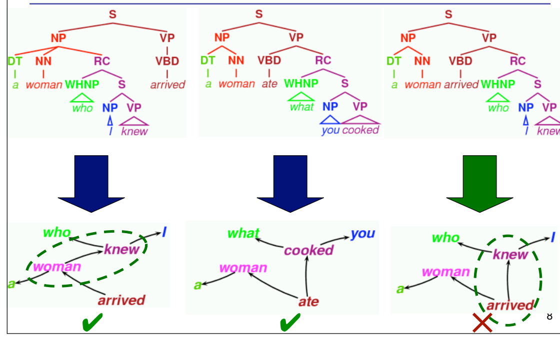
Dependency trees from CF trees

- Alternatively, syntactic parse trees can directly induce *dependency trees*
- Can be interpreted pseudo-propositionally; high utility for Question Answering (Pasca and Harabagiu 2001)



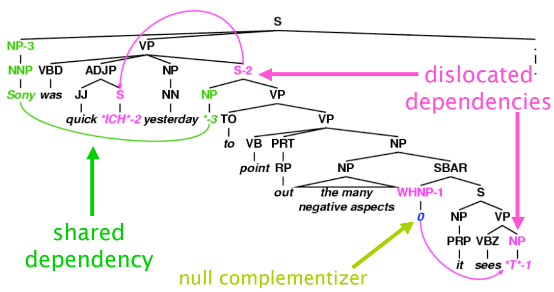
7

Parses to dependencies: limits



8

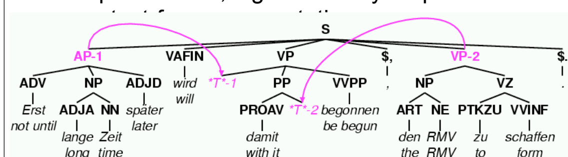
Nonlocal annotation in Penn Treebank



9

Nonlocal annotation in NEGRA (German)

- Intuitively, much more nonlocal dependency in German
- NEGRA directly annotates crossing dependencies, algorithmically maps to a

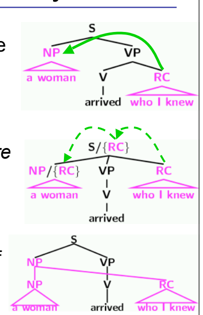


"The RMV will not begin to be formed until much later."

10

Three methods of non-local dependency recovery

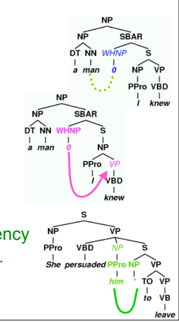
- Approximate dependency recovery with a context-free parser; correct the output post-hoc (Johnson 2002; present work; also akin to traditional LFG parsing)
- Incorporate non-local dependency information into the *category structure* of chart parser entries (Collins 1999; Dienes 2003; also akin to traditional G/HPSG, CCG parsing)
- Incorporate non-local dependency information into the *edge structure* of chart parser entries (Plaehn 2000; TAG)



11

Tree reshaping via cascaded classification

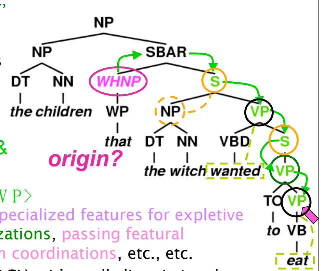
1. Null complementizers (mediate relativization)
   - A. Identify sites for null node insertion
   - B. Find best daughter position and insert.
2. Dislocated dependencies
   - A. Identify dislocated nodes
   - B. relocate to original/"deep" mother node
   - C. Find best daughter position and insert
3. Shared dependencies
   - A. Identify sites of nonlocal shared dependency
   - B. Identify best daughter position and insert.
   - C. Find controller for each control locus.



Use sequence of maxent classifiers. Feature types:

- Syntactic category, parent, grandparent (subj vs obj extraction; VP finiteness
- Head words (*wanted* vs *to* vs. *eat*)
- Presence of daughters (NP under S)
- *Syntactic path* (Gildea & Jurafsky 2002):

⟨↑SBAR,↓S,↓VP,↓S,↓VP⟩

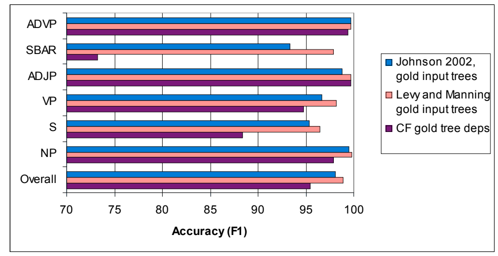


Plus: feature conjunctions, specialized features for expletive subject dislocations, passivizations, passing featural information properly through coordinations, etc., etc.

cf. Campbell (2004, ACL) – it's really linguistic rules

13

Evaluation on dependency metric: gold-standard input trees



14

Why is SRL Important?
Applications as a simple meaning rep'n

- Question Answering
  - Q: When was Napoleon defeated?
  - Look for: [PATIENT Napoleon] [PRED **defeat-synset**] [ARGM-TMP *ANS*]
- Machine 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
  - SRL can be used to construct useful rules for IE

15

Application: Semantically precise search

Query: *afghans destroying opium poppies*

16

Some History

- Fillmore 1968: The case for case
  - Recognize underlying semantic cases = semantic roles
- Minsky 1974: *frames* describe events or situations
  - Multiple participants, "props", and "conceptual roles"
- Levin 1993: verb class defined by sets of frames (meaning-preserving alternations) a verb appears in
  - *{break,shatter,..}: Glass X's easily; John Xed the glass, …*
  - *Cut* is different: *The window broke; *The window cut.*
- FrameNet, late '90s: based on Fillmore's work: large corpus of sentences annotated with *frames*
- PropBank: addresses tragic flaw in FrameNet corpus, but with use of much more impoverished representation

17

Alternations for verbs of contact:

conative:

Jean moved the table.
*Jean moved at the table.

body-part possessor ascension:

Janet broke Bill's finger.
*Janet broke Bill on the finger.

middle construction:

Bread cuts easily.
*Cats touch easily.

Verb Class

Alternation	Touch	Hit	Cut	Break
conative	N	Y	Y	N
body-part possessor ascension	Y	Y	Y	N
middle	N	N	Y	Y

Examples of verbs for each class:
Touch: kiss, sting, tickle
Hit: bash, hammer, tap
Cut: chip, hack, scratch
Break: hack, split, tear

Underlying hypothesis: *verbal meaning determines syntactic realizations*

Beth Levin analyzed thousands of verbs and defined hundreds of classes.

18

Frames in FrameNet

frame(TRANSPORTATION)
frame_elements(MOVER(S), MEANS, PATH)
scene(MOVER(S) move along PATH by MEANS)

frame(DRIVING)
inherit(TRANSPORTATION)
frame_elements(DRIVER (=MOVER), VEHICLE (=MEANS), RIDER(S) (=MOVER(S)), CARGO (=MOVER(S)))
scenes(DRIVER starts VEHICLE, DRIVER controls VEHICLE, DRIVER stops VEHICLE)

frame(RIDING_1)
inherit(TRANSPORTATION)
frame_elements(RIDER(S) (=MOVER(S)), VEHICLE (=MEANS))
scenes(RIDER enters VEHICLE,
VEHICLE carries RIDER along PATH,
RIDER leaves VEHICLE)

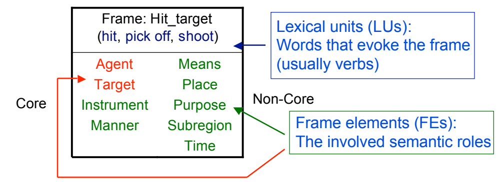
Figure 1: A subframe can inherit elements an[d] semantics from its parent

[Baker, Fillmore, Lowe, 1998]

FEG	Annotated Example from BNC
D	[$_D$ Kate] **drove** [$_P$ home] in a stupor.
V, D	A pregnant woman lost her baby after she fainted as she waited for a bus and fell into the path of [$_V$ a lorry] **driven** [$_D$ by her uncle].
D, P	And that was why [$_D$ I] **drove** [$_P$ eastwards along Lake Geneva].
D, R, P	Now [$_D$ Van Cheele] was **driving** [$_R$ his guest] [$_P$ back to the station].
D, V, P	[$_D$ Cumming] had a fascination with most forms of transport, **driving** [$_V$ his Rolls] at high speed [$_P$ around the streets of London].
D+R, P	[$_D$ We] **drive** [$_P$ home along miles of empty freeway].
V, P	Over the next 4 days, [$_V$ the Rolls Royces] will **drive** [$_P$ down to Plymouth], following the route of the railway.

Figure 2: Examples of Frame Element Groups and Annotated Sentences

FrameNet [Fillmore et al. 01]



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

Methodology for FrameNet

While (remaining funding > 0) do

1. Define a frame (eg DRIVING)
2. Find some sentences for that frame
3. Annotate them

- Corpora
  - FrameNet I – British National Corpus only
  - FrameNet II – LDC North American Newswire corpora
- Size
  - >8,900 lexical units, >625 frames, >135,000 sentences

http://framenet.icsi.berkeley.edu

Annotations in PropBank

- Based on Penn TreeBank
- Goal is to annotate *every tree* systematically
  - so statistics in the corpus are meaningful
- Based on Levin's verb classes (via VerbNet)
  - But annotated with lowest common denominator ARG0, ARG1 roles.
- Generally more data-driven & bottom up
  - No level of abstraction beyond verb senses
  - Annotate every verb you see, whether or not it seems to be part of a frame

Some verb senses and "framesets" for propbank

Frameset: **decline.01** "go down incrementally"
  Arg1: entity going down
  Arg2: amount gone down by, EXT
  Arg3: start point
  Arg4: end point
Ex: ...[$_{\text{Arg1}}$ its net income] *declining* [$_{\text{Arg2-EXT}}$ 42%] [$_{\text{Arg4}}$ to \$121 million] [$_{\text{ArgM-TMP}}$ in the first 9 months of 1989]. (wsj_0067)

Frameset: **decline.02** "demure, reject"
  Arg0: agent
  Arg1: rejected thing
Ex: [$_{\text{Arg0}}$ A spokesman$_i$] *declined* [$_{\text{Arg1}}$ *trace*$_i$ to elaborate] (wsj_0038)

FrameNet vs PropBank -1

FRAMENET ANNOTATION:

[$_{\text{Buyer}}$ Chuck] *bought* [$_{\text{Goods}}$ a car] [$_{\text{Seller}}$ from Jerry] [$_{\text{Payment}}$ for \$1000].

[$_{\text{Seller}}$ Jerry] *sold* [$_{\text{Goods}}$ a car] [$_{\text{Buyer}}$ to Chuck] [$_{\text{Payment}}$ for \$1000].

PROPBANK ANNOTATION:

[$_{\text{Arg0}}$ Chuck] *bought* [$_{\text{Arg1}}$ a car] [$_{\text{Arg2}}$ from Jerry] [$_{\text{Arg3}}$ for \$1000].

[$_{\text{Arg0}}$ Jerry] *sold* [$_{\text{Arg1}}$ a car] [$_{\text{Arg2}}$ to Chuck] [$_{\text{Arg3}}$ for \$1000].

FrameNet vs PropBank -2

FRAMENET ANNOTATION:

[$_{Goods}$ A car] was *bought* [$_{Buyer}$ by Chuck].

[$_{Goods}$ A car] was *sold* [$_{Buyer}$ to Chuck] [$_{Seller}$ by Jerry].

[$_{Buyer}$ Chuck] was *sold* [$_{Goods}$ a car] [$_{Seller}$ by Jerry].

PROPBANK ANNOTATION:

[$_{Arg1}$ A car] was *bought* [$_{Arg0}$ by Chuck].

[$_{Arg1}$ A car] was *sold* [$_{Arg2}$ to Chuck] [$_{Arg0}$ by Jerry].

[$_{Arg2}$ Chuck] was *sold* [$_{Arg1}$ a car] [$_{Arg0}$ by Jerry].

25

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

- Transfer sentences to propositions
  - Kristina hit Scott → hit(Kristina,Scott)
- Penn TreeBank → PropBank
  - Add a semantic layer on Penn TreeBank
  - Define a set of semantic roles for each verb
  - Each verb's roles are numbered

…[**A0** the company] to … *offer* [**A1** a 15% to 20% stake] [**A2** to the public]
…[**A0** Sotheby's] … *offered* [**A2** the Dorrance heirs] [**A1** a money-back guarantee]
…[**A1** an amendment] *offered* [**A0** by Rep. Peter DeFazio] …
…[**A2** Subcontractors] will be *offered* [**A1** a settlement] …

26

Proposition Bank (PropBank) Define the Set of Semantic Roles

- It's difficult to define a general set of semantic roles for all types of predicates (verbs).
- PropBank defines semantic roles for each verb and sense in the frame files.
- The (core) arguments are labeled by numbers.
  - A0 – Agent; A1 – Patient or Theme
  - Other arguments – no consistent generalizations
- Adjunct-like arguments – ***universal*** to all verbs
  - AM-LOC, TMP, EXT, CAU, DIR, PNC, ADV, MNR, NEG, MOD, DIS

27

Proposition Bank (PropBank) Frame Files

- hit.01 "strike"
  - A0: agent, hitter; A1: thing hit; A2: instrument, thing hit by or with

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

AM-TMP
Time

- look.02 "seeming"
  - A0: seemer; A1: seemed like; A2: seemed to

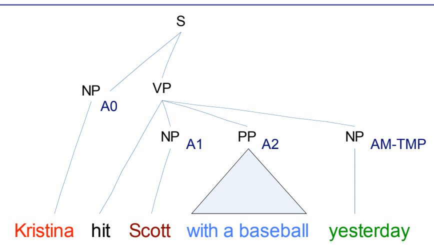
  [$_{A0}$ *It*] ***looked*** [$_{A2}$ *to her*] *like* [$_{A1}$ *he deserved this*].
- deserve.01 "deserve"
  - A0: deserving entity; A1: thing deserved; A2: in-exchange-for

  It looked to her like [$_{A0}$ *he*] ***deserved*** [$_{A1}$ *this*].

Proposition:
A sentence and a target verb

28

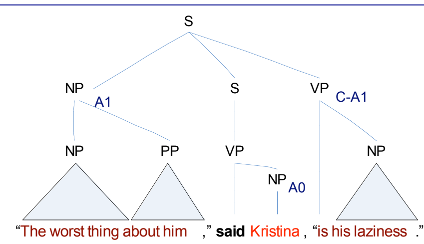
Proposition Bank (PropBank) Add a Semantic Layer



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

29

Proposition Bank (PropBank) Add a Semantic Layer – Continued



[$_{A1}$ *The worst thing about him*] ***said*** [$_{A0}$ *Kristina*] [$_{C-A1}$ *is his laziness*].

30

Proposition Bank (PropBank) Final Notes

- Current release (Mar 4, 2005): Proposition Bank I
  - Verb Lexicon: 3,324 frame files
  - Annotation: ~113,000 propositions
    http://www.cis.upenn.edu/~mpalmer/project_pages/ACE.htm
- Alternative format: CoNLL-04,05 shared task
  - Represented in table format
  - Has been used as standard data set for the shared tasks on semantic role labeling
    http://www.lsi.upc.es/~srlconll/soft.html

31

---

1. lie(“he”,…)
2. leak(“he”, “information obtained from … he supervised”)
3. obtain(X, “information”, “from a wiretap he supervised”)
4. supervise(“he”, “a wiretap”)

```
He            -              (A0*)     (A0*)      *       *
is            -                 *         *       *       *
also          -                 *         *       *       *
accused       -                 *         *       *       *
of            -                 *         *       *       *
lying         lie             (V*)        *       *       *
under         -           (AM-LOC*        *       *       *
oath          -                *)         *       *       *
and           -                 *         *       *       *
of            -                 *         *       *       *
leaking       leak              *      (V*)       *       *
information   -                 *      (A1*    (A1*)      *
obtained      obtain            *         *     (V*)      *
from          -                 *         *       *       *
a             -                 *         *    (A2*    (A1*
wiretap       -                 *         *       *      *)
he            -                 *         *       *    (A0*)
supervised    supervise         *        *)      *)     (V*)
.             -                 *         *       *       *
```

32

Information Extraction versus Semantic Role Labeling

Characteristic	IE	SRL
Coverage	narrow	broad
Depth of semantics	shallow	shallow
Directly connected to application	sometimes	no

33

Overview of SRL Systems

- Definition of the SRL task
  - Evaluation measures
- General system architectures
- Machine learning models
  - Features & models
  - Performance gains from different techniques

34

Subtasks

- Identification: $2^{\{1,2,\ldots,m\}} \longmapsto \{NONE, ARG\}$
  - Very hard task: to separate the argument substrings from the rest in this exponentially sized set
  - Usually only 1 to 9 (avg. **2.7**) substrings have labels ARG and the rest have NONE for a predicate
- Classification: $2^{\{1,2,\ldots,m\}} \longmapsto L \setminus \{NONE\}$
  - Given the set of substrings that have an *ARG* label, decide the exact semantic label
- Core argument semantic role labeling: (easier)
  - Label phrases with core argument labels only. The modifier arguments are assumed to have label NONE.

35

Evaluation Measures

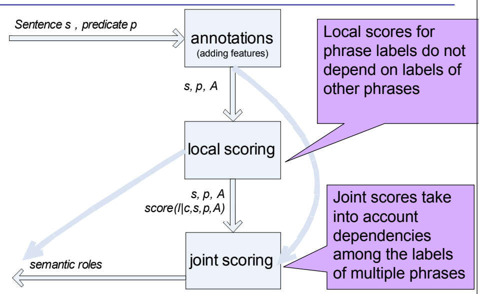
Correct: [$_{A0}$ The queen] **broke** [$_{A1}$ the window] [$_{AM\text{-}TMP}$ yesterday]
Guess: [$_{A0}$ The queen] broke the [$_{A1}$ window] [$_{AM\text{-}LOC}$ yesterday]

Correct	Guess
{The queen} →A0 {the window} →A1 {yesterday} ->AM-TMP all other → NONE	{The queen} →A0 {window} →A1 {yesterday} ->AM-LOC all other → NONE

- Precision ,Recall, F-Measure {*tp*=1,*fp*=2,*fn*=2} *p*=*r*=*f*=1/3
- Measures for subtasks
  - Identification (Precision, Recall, F-measure) {tp=2,fp=1,fn=1} *p*=*r*=*f*=2/3
  - Classification (Accuracy) acc = .5 (labeling of correctly identified phrases)
  - Core arguments (Precision, Recall, F-measure) {tp=1,fp=1,fn=1} *p*=*r*=*f*=1/2

36

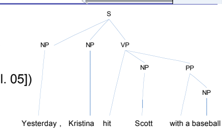
Basic Architecture of a Generic SRL System



37

Annotations Used

- Syntactic Parsers
  - Collins', Charniak's (most systems)
    CCG parses
    ([Gildea & Hockenmaier 03],[Pradhan et al. 05])
    TAG parses ([Chen & Rambow 03])



- Shallow parsers
  [$_{NP}$Yesterday] , [$_{NP}$Kristina] [$_{VP}$hit] [$_{NP}$Scott] [$_{PP}$with] [$_{NP}$a baseball].
- Semantic ontologies (WordNet, automatically derived), and named entity classes
  (v) **hit** (cause to move by striking)
  WordNet hypernym
  propel, impel *(cause to move forward with force)*

38

Annotations Used - Continued

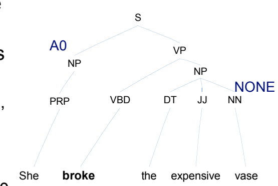
Most commonly, substrings that have argument labels correspond to syntactic constituents

- In Propbank, an argument phrase corresponds to exactly one parse tree constituent in the correct parse tree for **95.7**% of the arguments;
  - when more than one constituent correspond to a single argument (**4.3**%), simple rules can join constituents together (in 80% of these cases, [Toutanova 05]);
- In Propbank, an argument phrase corresponds to exactly one parse tree constituent in Charniak's automatic parse tree for approx **90.0%** of the arguments.
  - Some cases (about 30% of the mismatches) are easily recoverable with simple rules that join constituents ([Toutanova 05])
- In FrameNet, an argument phrase corresponds to exactly one parse tree constituent in Collins' automatic parse tree for **87%** of the arguments.

39

Labeling Parse Tree Nodes

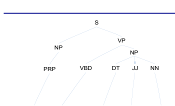
- Given a parse tree *t*, label the nodes (phrases) in the tree with semantic labels
- To deal with discontiguous arguments
  - In a post-processing step, join some phrases using simple rules
  - Use a more powerful labeling scheme, i.e. C-A0 for continuation of A0



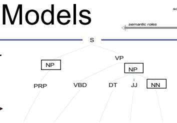
Another approach: labeling chunked sentences. Will not describe in this section.

40

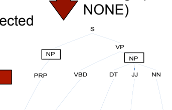
Combining Identification and Classification Models



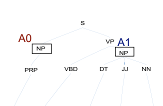
Step 1. *Pruning.* Using a hand-specified filter.



Step 2. *Identification.* Identification model (filters out candidates with high probability of NONE)



Step 3. *Classification.* Classification model assigns one of the argument labels to selected nodes (or sometimes possibly NONE)



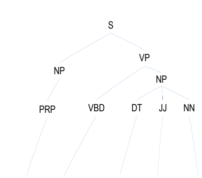
41

Combining Identification and Classification Models – Continued

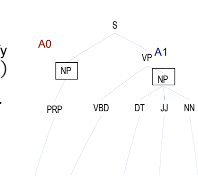
$$P(l|c,t,p) = P_{ID}(Id(l)|\Phi(c,t,p)) * P_{CLS}(l|Id(l),\Phi(c,t,p))$$

or

$$P(l|c,t,p) = P(l|\Phi(c,t,p))$$

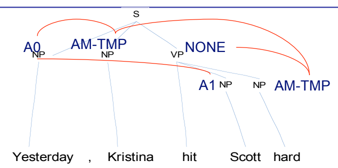


One Step. Simultaneously identify and classify using $P(l|c,t,p)$



42

Joint Scoring Models



- These models have scores for a whole labeling of a tree (not just individual labels)
  - Encode some dependencies among the labels of different nodes

$$P_{JOINT}(l_1,\ldots,l_n|n,t,p)! = \prod_i P(l_i|n_i,t,p)$$

43

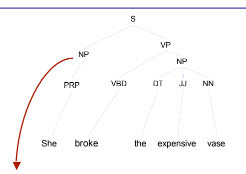
Combining Local and Joint Scoring Models

- Tight integration of local and joint scoring in a **single probabilistic model** and exact search [Cohn&Blunsom 05] [Màrquez et al. 05],[Thompson et al. 03]
  - When the joint model makes strong independence assumptions
- **Re-ranking** or approximate search to find the labeling which maximizes a combination of local and a joint score [Gildea&Jurafsky 02] [Pradhan et al. 04] [Toutanova et al. 05]
  - Usually exponential search required to find the exact maximizer
- Exact search for **best assignment by local model satisfying hard joint constraints**
  - Using Integer Linear Programming [Punyakanok et al 04,05] (worst case NP-hard)
- More details later

44

Gildea & Jurafsky (2002) Features

- Key early work
  - Future systems use these features as a baseline
- Constituent Independent
  - Target predicate (lemma)
  - Voice
  - Subcategorization
- Constituent Specific
  - Path
  - Position (*left, right*)
  - Phrase Type
  - Governing Category (*S* or *VP*)
  - Head Word



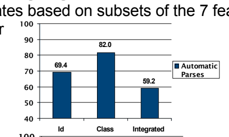
Target	*broke*
Voice	*active*
Subcategorization	$VP \rightarrow VBD\ NP$
Path	$VBD\uparrow VP\uparrow S\downarrow NP$
Position	*left*
Phrase Type	*NP*
Gov Cat	*S*
Head Word	*She*

45

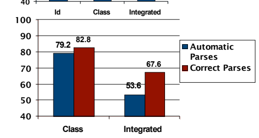
Performance with Baseline Features using the G&J Model

- **Machine learning algorithm:** interpolation of relative frequency estimates based on subsets of the 7 features introduced earlier

FrameNet Results



Propbank Results



46

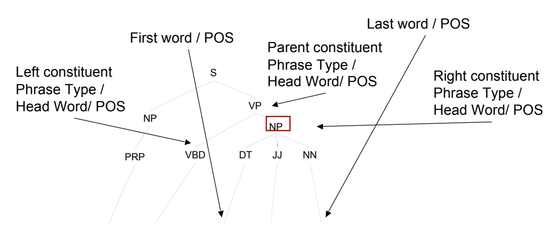
Performance with Baseline Features using the G&J Model

- Better ML: 67.6 → **80.8** using SVMs [Pradhan et al. 04]).
- Content Word (different from head word)
- Head Word and Content Word POS tags
- **NE labels (Organization, Location, etc.)**
- Structural/lexical context (phrase/words around parse tree)
- Head of PP Parent
  - If the parent of a constituent is a PP, the identity of the preposition

47

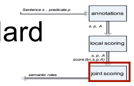
Pradhan et al. (2004) Features

- More (**31%** error reduction from baseline due to these + Surdeanu et al. features)



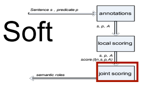
48

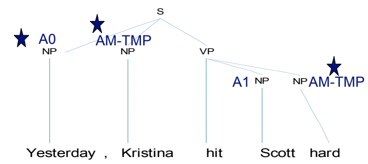
Joint Scoring: Enforcing Hard Constraints



- Constraint 1: Argument phrases do not overlap
  *By [$_{A1}$ working [$_{A1}$ hard] , he] **said** , you can achieve a lot.*
  - Pradhan et al. (04) – greedy search for a best set of non-overlapping arguments
  - Toutanova et al. (05) – exact search for the best set of non-overlapping arguments (dynamic programming, linear in the size of the tree)
  - Punyakanok et al. (05) – exact search for best non-overlapping arguments using integer linear programming
- Other constraints ([Punyakanok et al. 04, 05])
  - no repeated core arguments (good heuristic)
  - phrases do not overlap the predicate

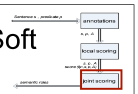
Joint Scoring: Integrating Soft Preferences





- There are many statistical tendencies for the sequence of roles and their syntactic realizations
  - When both are before the verb, AM-TMP is usually before A0
  - Usually, there aren't multiple temporal modifiers
  - Many others which can be learned automatically

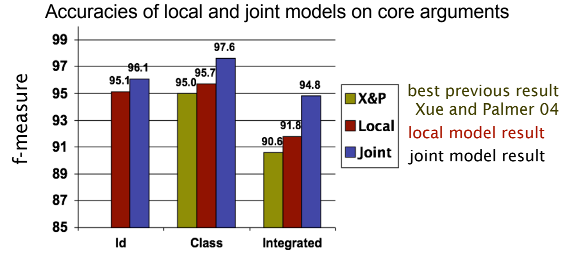
Joint Scoring: Integrating Soft Preferences



- Gildea and Jurafsky (02) – a smoothed relative frequency estimate of the probability of frame element multi-sets:
  $$P(\{A0, AM_{TMP}, A1, AM_{TMP}\}|hit)$$
  - Gains relative to local model 59.2 → 62.9 FrameNet automatic parses
- Pradhan et al. (04) – a language model on argument label sequences (with the predicate included)
  $$P(A0, AM_{TMP}, hit, A1, AM_{TMP})$$
  - Small gains relative to local model for a baseline system 88.0 → 88.9 on core arguments PropBank correct parses
- Toutanova et al. (05) – a joint model based on CRFs with a rich set of joint features of the sequence of labeled arguments
  - Gains relative to local model on PropBank correct parses 88.4 → 91.2 (24% error reduction); gains on automatic parses 78.2 → 80.0
- Also tree CRFs [Cohn & Brunson] have been used

Semantic roles: joint models boost results [Toutanova et al. 2005]

Accuracies of local and joint models on core arguments



Error reduction from best published result:

***44.6%** on Integrated **52%** on Classification*

CoNLL 2005 Shared Task Results on WSJ and Brown Tests

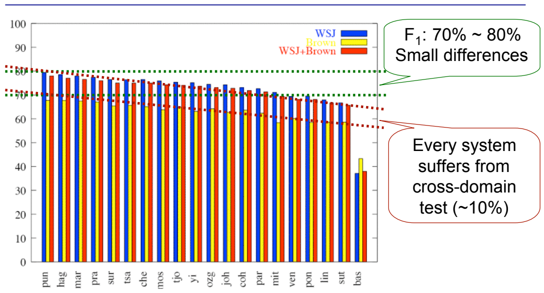


Figure from Carreras&Màrquez's slide (CoNLL 2005)

System Properties

- Learning Methods
  - SNoW, MaxEnt, AdaBoost, SVM, CRFs, etc.
  - ➢ *The choice of learning algorithms is less important.*
- Features
  - All teams implement more or less the standard features with some variations.
  - *A must-do for building a good system!*
  - *A clear feature study and more feature engineering will be helpful.*

System Properties – Continued

- Syntactic Information
  - Charniak's parser, Collins' parser, clauser, chunker, etc.
  - Top systems use Charniak's parser or some mixture
  - *Quality of syntactic information is very important!*
- System/Information Combination
  - 8 teams implement some level of combination
  - Greedy, Re-ranking, Stacking, ILP inference
  - *Combination of systems or syntactic information is a good strategy to reduce the influence of incorrect syntactic information!*

55

Per Argument Performance

CoNLL-05 Results on WSJ-Test

- Core Arguments (Freq. ~70%)

	Best F_1	Freq.
A0	88.31	25.58%
A1	79.91	35.36%
A2	70.26	8.26%
A3	65.26	1.39%
A4	77.25	1.09%

- Adjuncts (Freq. ~30%)

	Best F_1	Freq.
TMP	78.21	6.86%
ADV	59.73	3.46%
DIS	80.45	2.05%
MNR	59.22	2.67%
LOC	60.99	2.48%
MOD	98.47	3.83%
CAU	64.62	0.50%
NEG	98.91	1.36%

Arguments that need to be improved

Data from Carreras&Màrquez's slides (CoNLL 2005)

56