Lexicalisation of Long-Distance Dependencies in a Treebank-Based, Wide Coverage, Statistical LFG Grammar

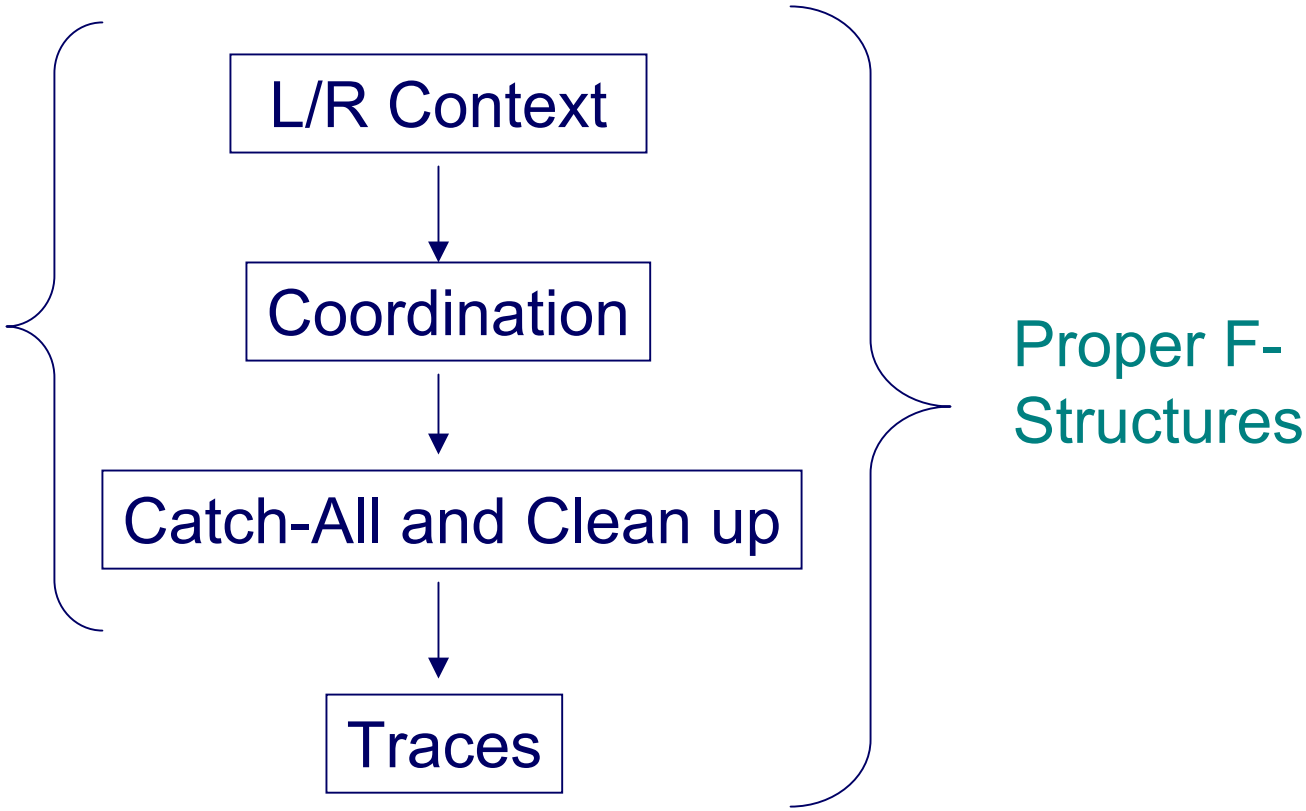
Aoife Cahill, Mairéad McCarthy, Ruth O' Donovan, Josef van Genabith, Andy Way
{acahill, mccarthy, rodonovan, josef, away} @computing.dcu.ie

School of Computer Applications
National Centre for Language Technology
Dublin City University
Dublin 9
Ireland

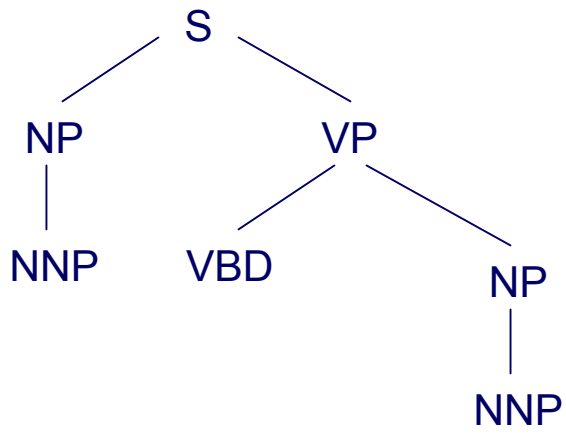
Outline

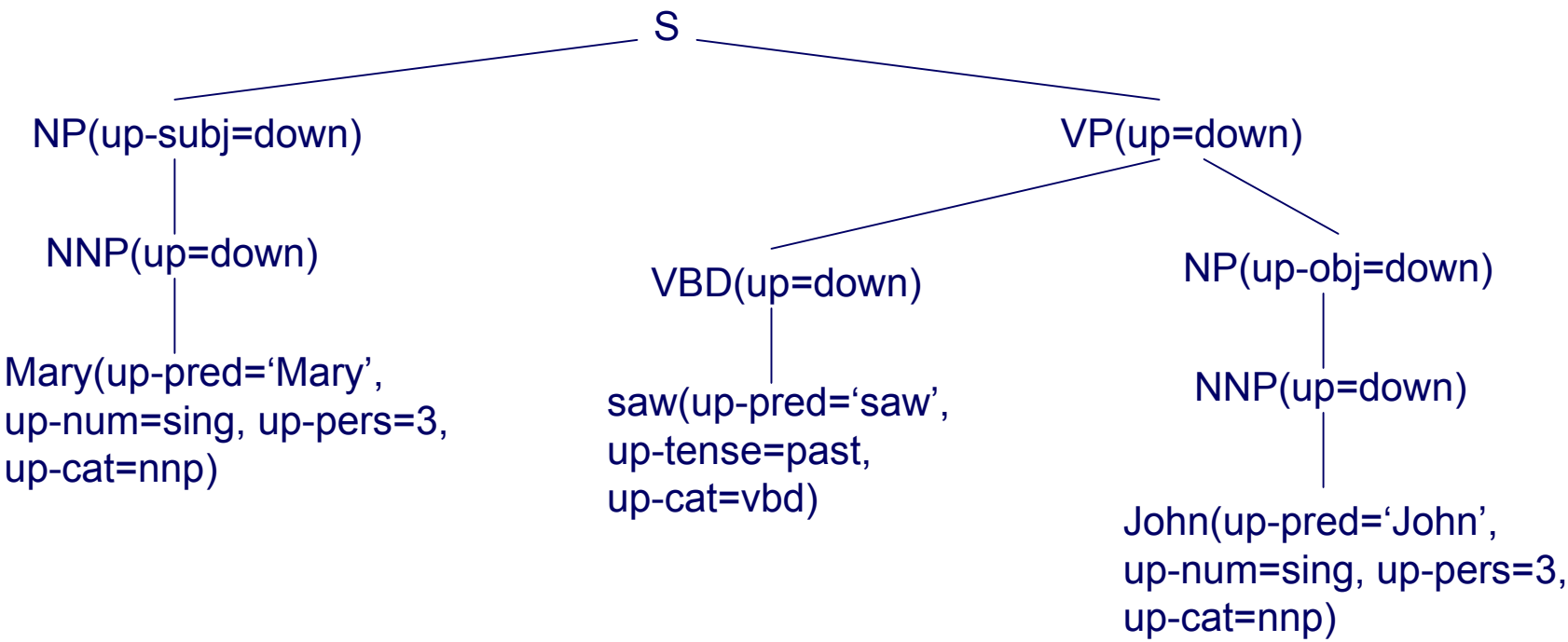
- Introduction
- Annotation Algorithm for Penn-II
- Parsing
  - 2 parsing architectures, Proto F-Structures
- Extraction of Semantic Forms and Functional Uncertainty Equations
- Proper F-Structures
- Results
- Conclusions

Penn-II F-Structure Annotation Algorithm

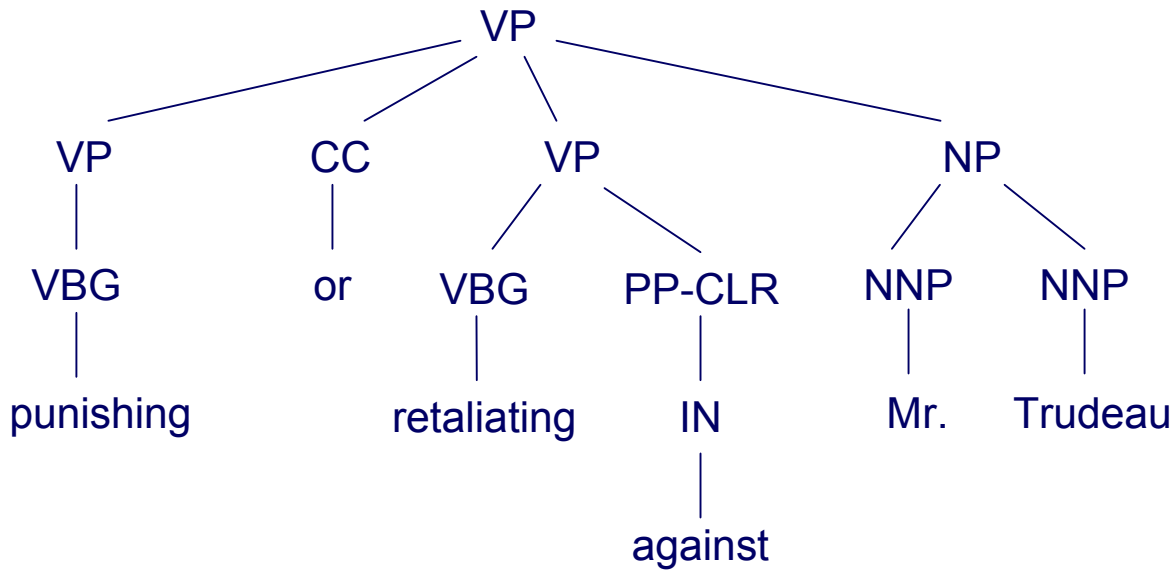


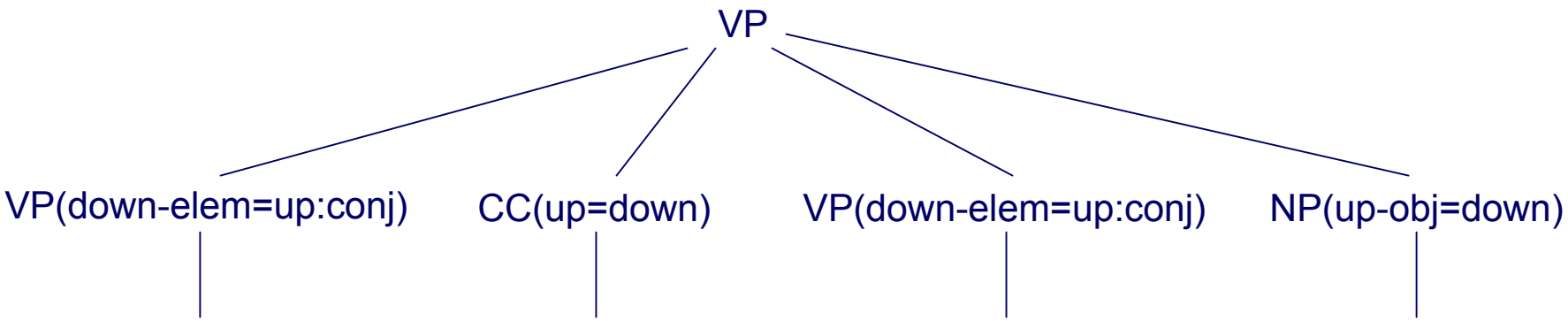
Example – L/R Context





Example - Coordination

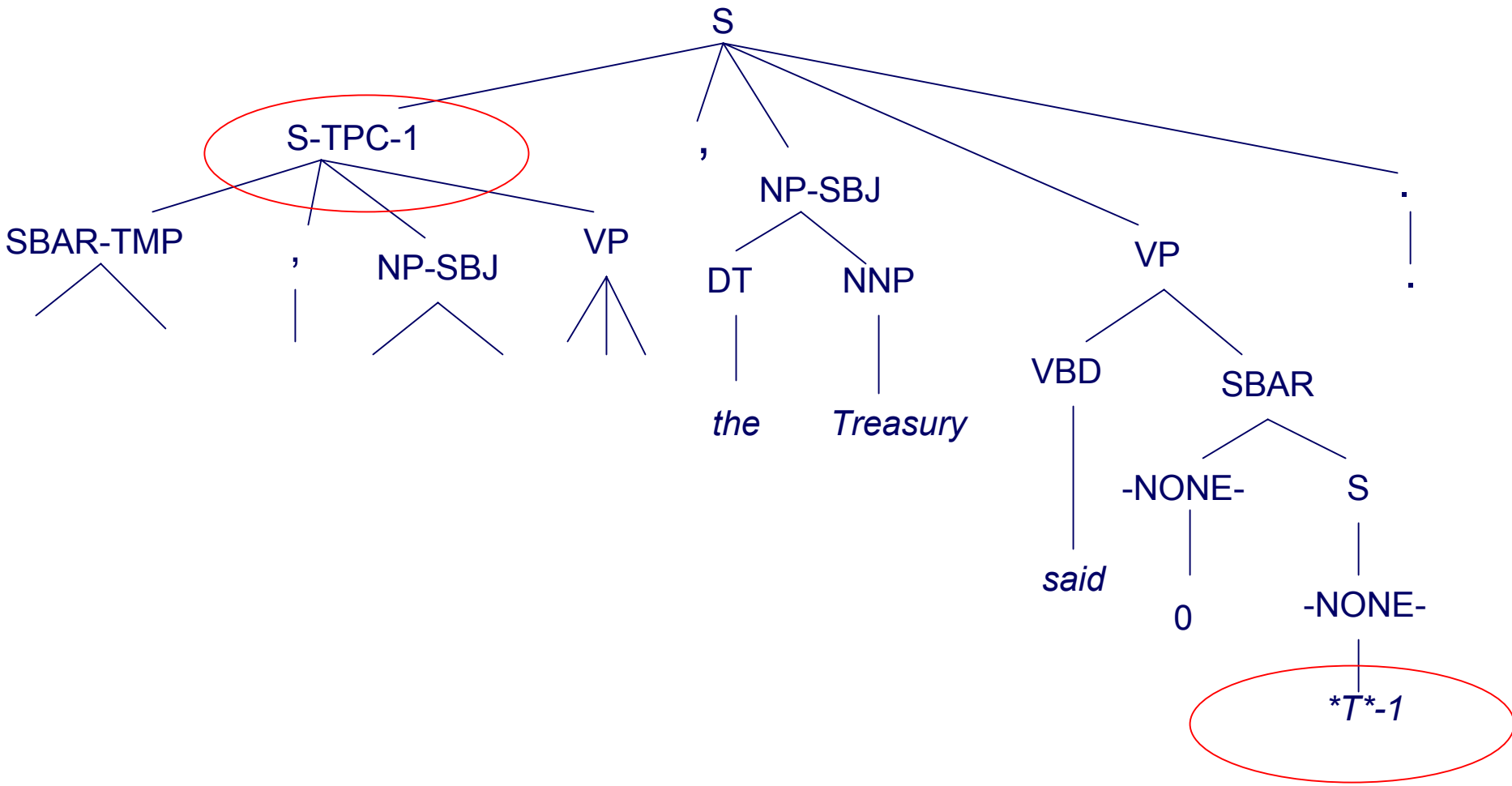




Example – Catch-all and Clean-up

- Catch-all
  - Use Functional Tags in the Penn-II Annotation Scheme
  - -SBJ ↑SUBJ=↓
- Clean-up
  - Resolves clashing annotations
  - ↑OBJ2=↓

Example – Traces



"Until Congress acts , the government hasn't any authority to issue new debt obligations of any kind , the Treasury said ."

Example – Traces

```
topic : index : 1
        subj : spec : det : pred : the
               num : sing
               pred : government
               pers : 3
         …
         …
        pred : have
        tense : pres
subj : spec : det : pred : the
       pers : 3
       pred : treasury
       num : sing
comp : index : 1
       subj : spec : det : pred : the
              num : sing
              pred : government
              pers : 3
         …
         …
       pred : have
       tense : pres
pred : say
tense : past
```

Precision and Recall

$$\text{precision} = \frac{\text{number of correct constituents in proposed parse}}{\text{number of constituents in proposed parse}}$$

$$\text{recall} = \frac{\text{number of correct constituents in proposed parse}}{\text{number of constituents in treebank parse}}$$

Evaluation - Qualitative

- Evalb – annotated trees

All sentences	
Bracketing Recall:	94.18
Bracketing Precision:	94.18
Complete Match:	35.24

Sentences length <=40	
Bracketing Recall:	94.02
Bracketing Precision:	94.02
Complete Match:	36.36

- Triples – F-Structures

(Crouch et. al, 2000)

– relation(predicate, argument)

All Grammatical Functions	
Precision:	0.94
Recall:	0.95
f-score:	0.94

Preds Only	
Precision:	0.91
Recall:	0.92
f-score:	0.91

Evaluation - Quantitative

- Fragmentation
  - % sentences receiving 1 f-structure

0 F-structures	295	0.609202048570957%
1 F-structures	48125	99.382537584668%
2 F-structures	4	0.00826036676028416%

- Coverage
  - % daughter nodes receiving an annotation

ADJP	1641	1639	99.87
ADVP	605	603	99.66
NP	30735	30726	99.97
PP	1073	1071	99.81
S	14817	14815	99.98
SBAR	409	409	100.0
SQ	650	648	99.69
VP	40824	40822	99.99

2 Parsing Architectures

- Pipeline
  - text $\Rightarrow$ PCFG $\Rightarrow$ LFG $\Rightarrow$ annotated trees
- Integrated
  - text $\Rightarrow$ PCFG$\cup$LFG $\Rightarrow$ annotated trees

Parsing into Proto F-Structures

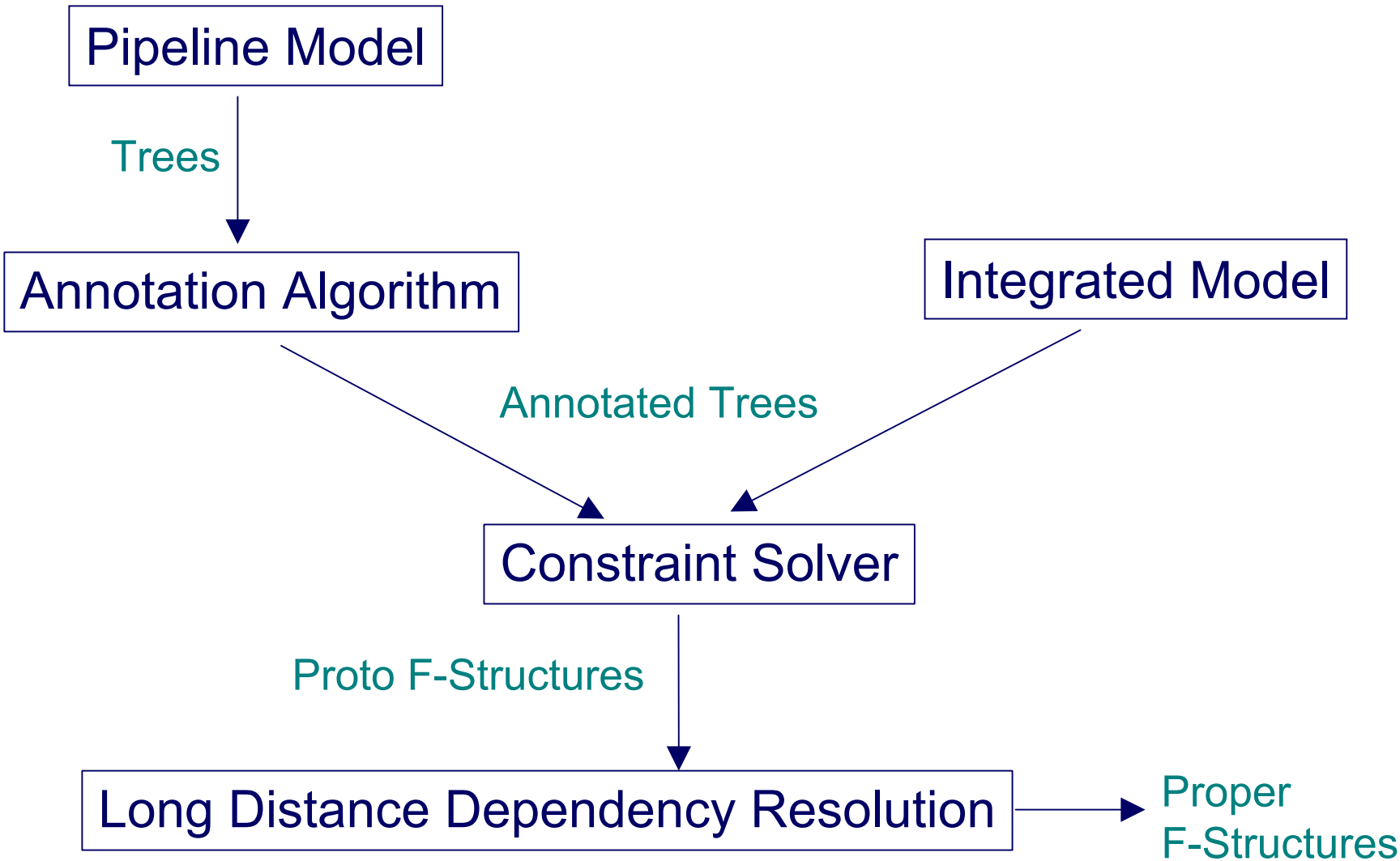
- PCFG consists of CFG rules with associated probabilities
- A-PCFG treats strings consisting of CFG categories followed by 1 or more functional annotation(s) as monadic categories (e.g. NP[up-obj=down])
- In both cases new text is parsed into proto f-structures.

Parsing into Proto F-Structures

Results

	Pipeline (full)	Pipeline (compacted 2)	Integrated (full)	Integrated (compacted 2)
General evaluation				
# Rules	22760	6106	34336	8818
# Parses	2190	2150	2180	2101
Time (hrs)	63.97	4.46	137.05	3.16
Parse Tree evaluation				
Labelled F-score	76.79	77.65	81.23	79.48
Unlabelled F-score	79.71	80.09	83.48	81.74
F-Structure evaluation				
Fragmentation (%)	94.61	97.25	99.63	99.33
Preds Only (f-score)	65.11	63.25	68.16	65.94
All grammatical functions (f-score)	72.63	71.80	74.92	73.64
Evalb – annotated trees (f-score)	64.23	62.72	57.01	54.30

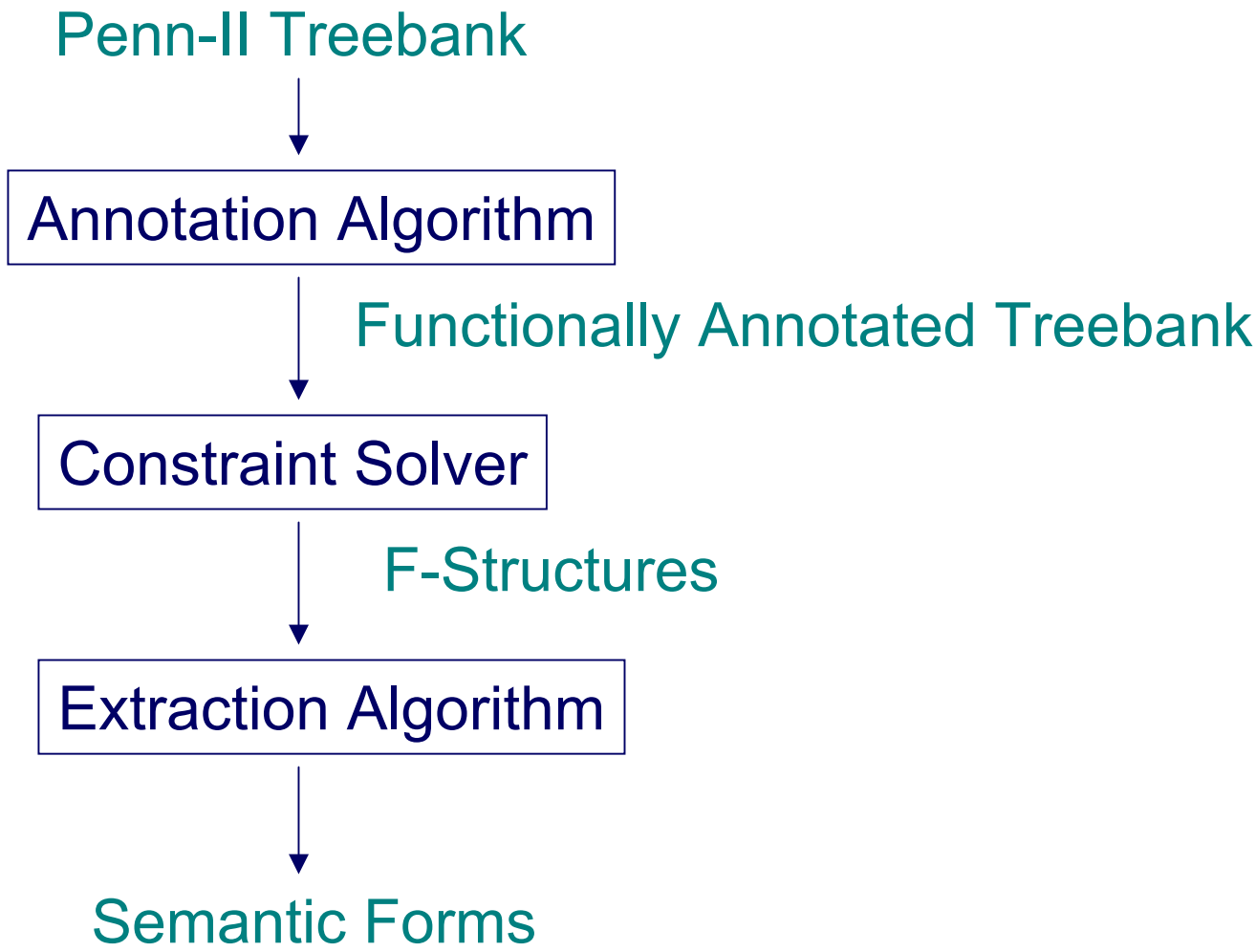
From Proto to Proper F-Structures



Parsing into proper F-Structures

- In LFG, LDDs are treated in f-structures in terms of:
  - **functional uncertainty equations**
  - **semantic forms**
- Can we generate these automatically given our resources?

Extraction of Lexical Resources



Extraction of Lexical Resources

For each f-structure, for each level of embedding the local predicate is determined and all subcategorisable grammatical functions present at that level of embedding are collected for that predicate.

(van Genabith, Sadler, Way, 1999)

Extraction of Lexical Resources

```
subj : num : sing
       pers : 3
       cat : nnp
       pred : 'John'
obl : obj : spec : det : pred : the
            pred : system
            num : sing
            pers : 3
            cat : nn
      pred : in
tense : present
cat : vbd
pred : believe
```

believe([subj,obl:in])
in([obj])
john([])
system([])

“John believes in the system”

Extraction of Lexical Resources

- We extract over 16,000 unique non-empty semantic forms.
- Associated with conditional probability *P(s|l)* where *l* is a lemma and *s* is a semantic form.
- Example (including indication of passive use):

Semantic Form	Frequency	Probability
`blame([obj,subj,obl:for])`	18	0.231
`blame([obj,subj,obl:on])`	18	0.231
`blame([obj,subj])`	13	0.167
`blame([subj,obl:for],p)`	13	0.167
`blame([subj,obl:on],p)`	4	0.051
`blame([subj],p)`	4	0.051
`Others`	8	0.103

Extraction of Lexical Resources

	#Sem. Forms (with just a general OBL category)	# Sem. Forms (with distinct OBLs)
# Distinct Sem. Forms	35	349
#Singletons	2	95
#Twice Occurring	3	42
#Occurring max. 5	9	191
#Occurring > 5	26	158

Most frequent semantic forms extracted

be ([subj,xcomp])	22297	do ([subj,xcomp])	2088
have ([subj,xcomp])	8132	could ([subj,xcomp])	1453
say ([comp,subj])	7038	can ([subj,xcomp])	1263
be ([obj,subj])	6026	be ([subj])	1180
will ([subj,xcomp])	4064	may ([subj,xcomp])	1060
would ([subj,xcomp])	2865	make ([obj,subj])	901
have ([obj,subj])	2626	expect ([subj,xcomp])	849
say ([subj,xcomp])	2319	take ([obj,subj])	713

Semantic Form Evaluation

- The extracted semantic forms were evaluated against the COMLEX Dictionary.
- Used different thresholds to examine the effect on precision and recall.
- Varying levels of prepositional detail:
  - Without PP
  - With PP
  - With Preposition
  - Directional Prepositions (P-Dir)

Semantic Form Evaluation

	Precision	Recall	F-Score
Without PP	75.2%	69.1%	72.0%
With PP	65.5%	63.1%	64.3%
With Preposition	71.8%	16.8%	27.3%

Threshold of 1%

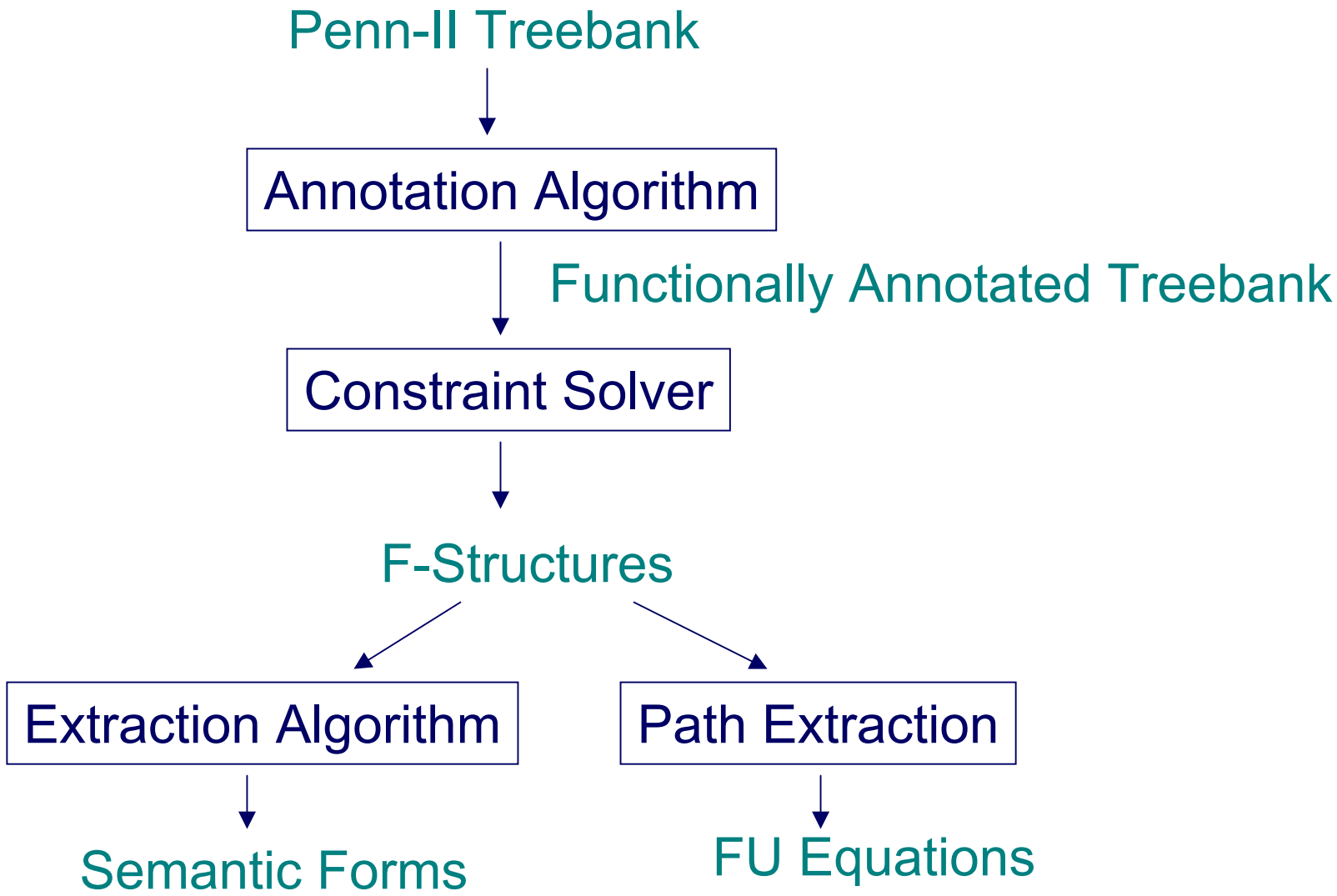
	Precision	Recall	F-Score
Without PP	80.2%	63.6%	70.9%
With PP	69.6%	56.9%	62.7%
With Preposition	76.7%	13.9%	23.5%

Threshold of 5%

	Precision	Recall	F-Score
With Preposition	86.9%	45.1%	59.4%

Using P-Dir List and Threshold of 1%

Functional Uncertainty Equations



Functional Uncertainty Equations

- Regular expressions over f-structure paths (e.g. ↑topic = (↑comp)* comp)
- Extract the shortest path in the f-structures between where linguistic material is encountered and where it should be interpreted semantically
- Approximate extension of FU equations

Functional Uncertainty Equations

```
focus : index : 1
        spec : det : pred : which
        pers : 3
        pred : car
        num : pl
subj : num : pl
       pers : 3
       pred : american
xcomp : subj : num : pl
                pers : 3
                pred : american
        obj : index : 1
              spec : det : pred : which
              pers : 3
              pred : car
              num : pl
        pred : favor
tense : present
pred : do
```

up-focus = up-xcomp:obj

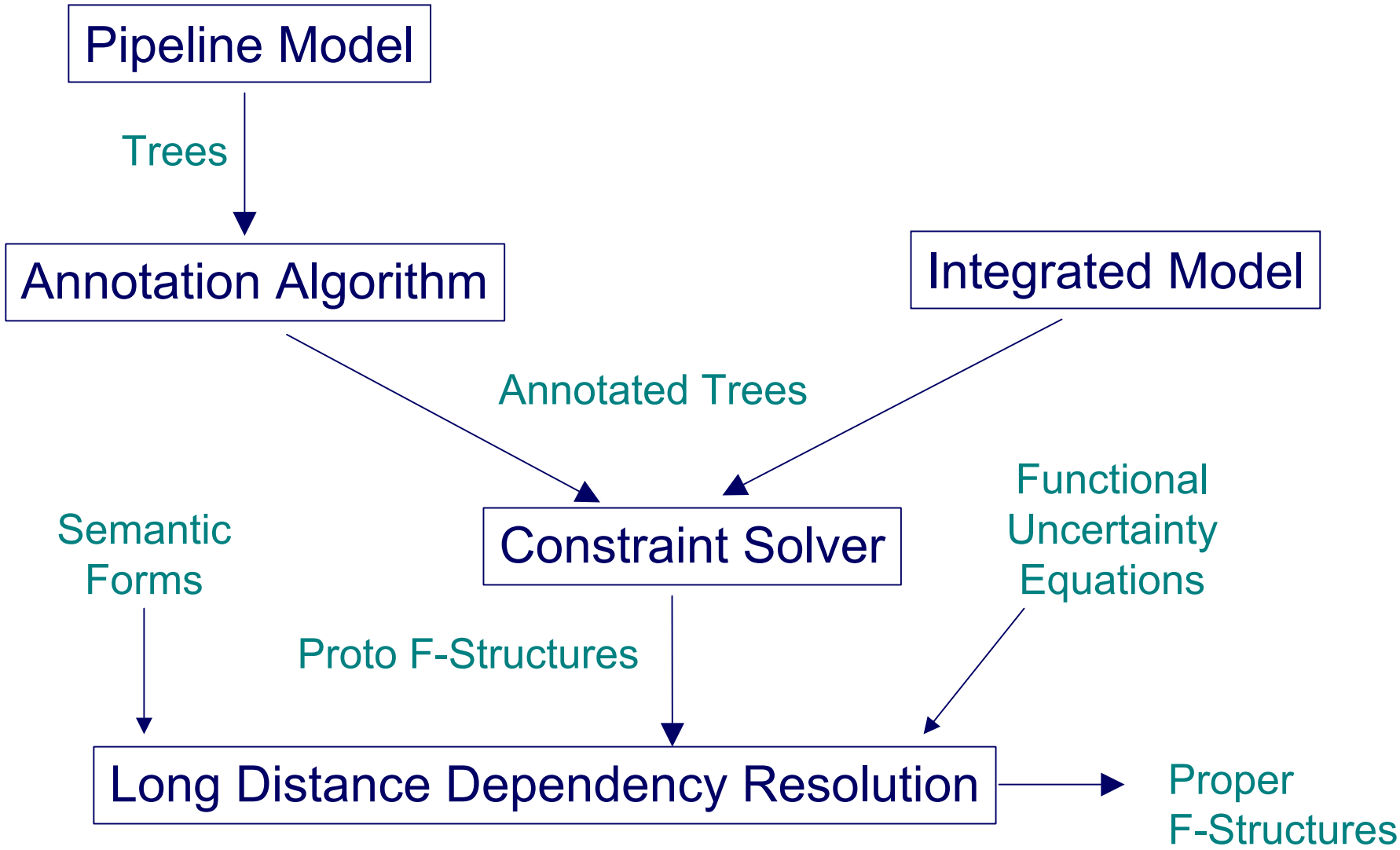
Functional Uncertainty Equations

- Associated with conditional probability *P(p|d)* where *p* is a path and *d* is either TOPIC or FOCUS
  - 23 TOPIC paths
  - 54 FOCUS paths

Sample FOCUS paths with frequencies

up-subj	7894	up-xcomp:xcomp	161
up-obj	1167	up-xcomp:xcomp:obj	135
up-xcomp	956	up-comp:subj	119
up-xcomp:obj	793	up-xcomp:subj	92

Resolving Long Distance Dependencies



Resolving Long Distance Dependencies

- Resolution Algorithm

  (recursively traverses an f-structure)

  (also applies to FOCUS)

  – find TOPIC:T attribute-value pair

  – retrieve TOPIC paths

  – for each path p of the form GF_1:…: GF_n:GF, traverse the f-structure along the path GF_1:…: GF_n to local sub f-structure h

    - retrieve local PRED:P_h
    - add GF:T to h iff
      - GF is not present at h
      - h together with GF is locally complete and coherent with respect to a semantic form s for P_h
    - multiply path and semantic form probabilities involved to rank resolution

Resolving Long Distance Dependencies

"John met Mary, the paper said."

Before Resolution:

```
topic : pred : meet
        subj : pred : mary
        obj : pred : john
pred : say
subj : spec : the
       pred : paper
```

After Resolution:

```
topic : pred : meet
        subj : pred : mary
        obj : pred : john
pred : say
subj : spec : the
       pred : paper
comp : pred : meet
       subj : pred : mary
       obj : pred : john
```

LDD Experiments

- 2 Parsers
  - A-PCFG (Integrated parsing model)
  - Collins model 2 parser
- Parsers trained on sections 02-21
- Results ranked by path prob.* sf prob, sf prob only, path prob only
- Evaluation against 105 f-structures (manually constructed gold standard)
- Evaluation against 2415 f-structures for Penn-II (automatically annotated)

LDD Results

	Before Resolution			After Resolution		
	P	R	F	P	R	F
A-PCFG	69.82	52.57	59.98	69.16	57.92	63.04
Collins	72.68	58.5	64.82	71.37	64.43	68.27

Triples evaluation: 105 gold standard sentences

	Before Resolution			After Resolution		
	P	R	F	P	R	F
A-PCFG	73.4	63.79	68.26	71.69	67.56	69.56
Collins	72.52	69.81	71.15	70.6	74.9	72.69

Triples evaluation: 2415 Automatically annotated sentences

LDD Resolution

- Varying probabilities for ranking

	SF * Path	SF	Path
A-PCFG	63.04	62.96	63.12
Collins	68.27	68.15	68.23

Conclusions and Future Work

- From Proto to Proper F-Structure
  - extraction of semantic forms with conditional probability
  - extraction of functional uncertainty equations with conditional probability
  - resolution of Long Distance Dependencies
- F-Structure Evaluation
  - Briscoe and Carroll, GRs
  - PARC 700
- Migration of annotation and parsing techniques to other treebanks and languages (Cahill et al., 2003)