Function Tagging

Don Blaheta

Thesis Proposal

Committee:

Eugene Charniak (Advisor)

Mark Johnson (Cog. & Ling. Sciences)

Michael Collins (AT&T)

23 August 2002

Function Tagging

The Penn treebank

- Much prior work using treebank
- Virtually all concentrated on *parses*—either quality or speed

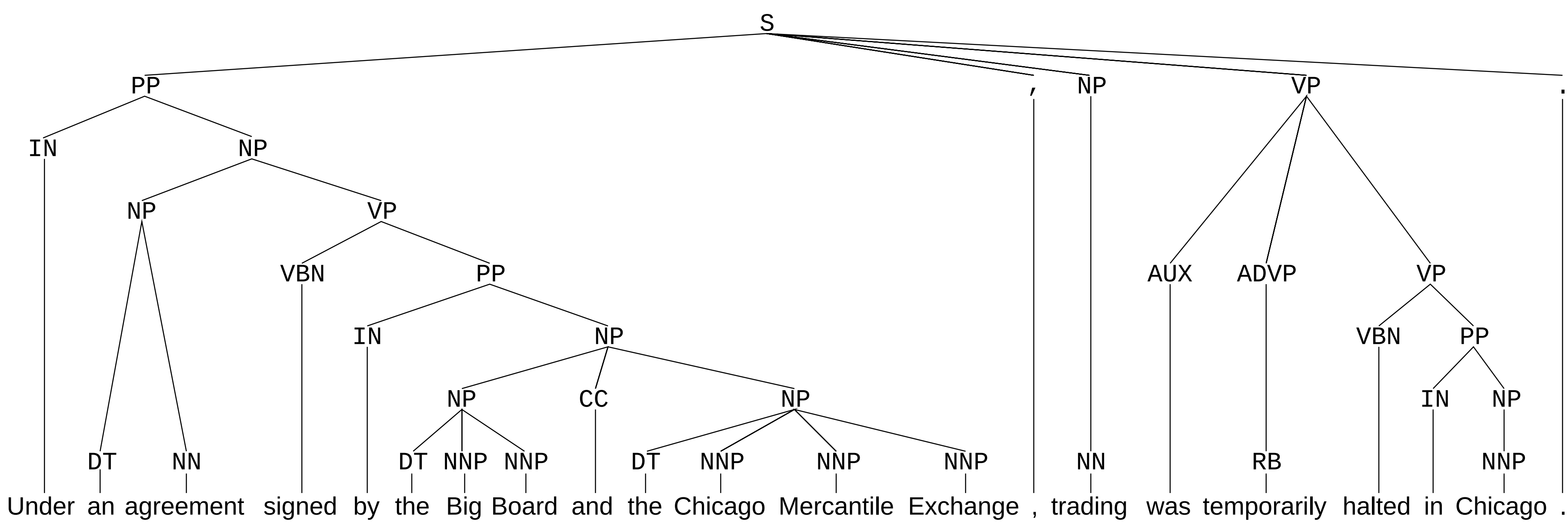


Figure 1: A sample treebank parse

Function Tagging

The Penn treebank

- Other data of interest encoded in treebank
- E.g. *function tags*

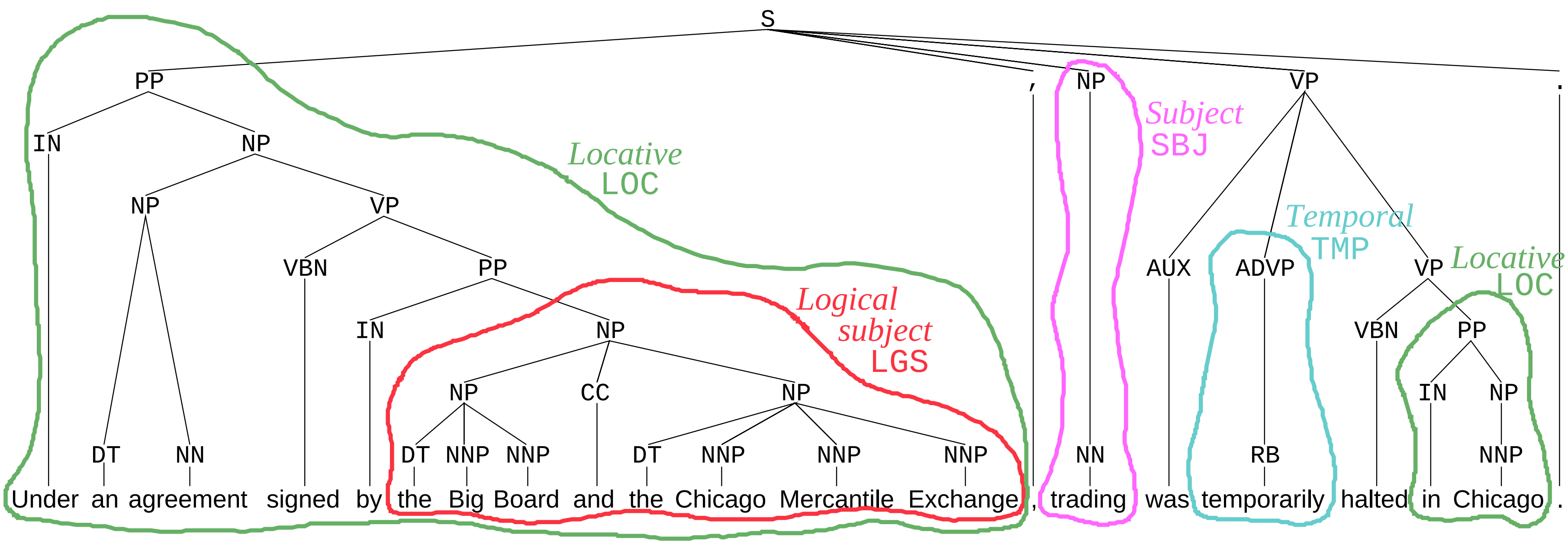


Figure 2: A sample treebank parse including function tags

Function tags

Function tags indicate additional information about the syntactic or semantic role of a constituent.

Subject (SBJ) The subject of an S (possibly embedded).

- *Futures traders* say *the S&P* was signaling that *the Dow* could fall 200 points.

Logical subject (LGS) The logical subject of a passive sentence.

- The largest distributor was already owned by *Quantum*.

Function tags continued

Temporal (TMP) Denotes time and duration.

- The CEO said *yesterday* in a statement that he has not *yet* seen the bid.

Locative (LOC) Denotes location (real or metaphorical).

- Martinair Holland is based *in Amsterdam*.
- *In the U.S. polyethylene market*, Quantum has claimed the largest share.

Function Tagging

The tags

		Within Category	*All Constits*
Grammatical			**11.2%**
DTV	Dative	.5%	.1%
LGS	Logical subject	3.0%	.3%
PRD	Predicate	17.8%	2.0%
PUT	'Put' object	.3%	.0%
SBJ	Subject	78.5%	8.8%
VOC	Vocative	.0%	.0%
Miscellaneous			**.12%**
CLF	'It'-cleft	5.4%	.01%
HLN	Headline	42.8%	.05%
TTL	Title	51.8%	.06%

		Within Category	*All Constits*
Form/Function			**7.8%**
ADV	Adverbial	11.5%	.9%
BNF	Benefactive	.0%	.0%
DIR	Direction	8.2%	.6%
EXT	Extent	3.2%	.3%
LOC	Locative	25.3%	2.0%
MNR	Manner	6.2%	.5%
NOM	Nominal	6.8%	.5%
PRP	Purpose	5.3%	.4%
TMP	Temporal	33.4%	2.6%
Topicalisation			**.5%**
TPC	Topicalised	100.0%	.5%

Features

- *Features* are simply boolean-valued functions.
  - "Is this node's label NP?"
  - "Is this node's head's part-of-speech NN?"
  - "Is this node's form/function tag TMP?"
- Informally, 'feature' also can refer to a group of related features.
  - "What is this node's parent's label?"
  - "What is this node's grammatical function tag?"
- Features can be used to guess other features...

Feature lookup

- When we guess a feature f based on some set of other features, we essentially look up f's probability given that set.

$$f \qquad a - b - c - d$$

Figure 3: f and the set of features used to guess it

- For sparse data, ignore the last few features.

Feature chains

- Alternate method:
  - start with a prior value $P(f)$
  - *refine:* multiply in a new term consisting of the likelihood of f, given the previous features in the chain.

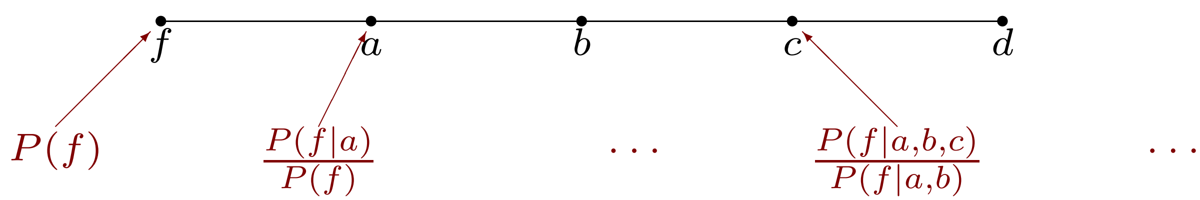


Figure 4: Refining the probability of f

  - repeat until sparse data problems arise.

- Equivalent to table lookup method.

Function Tagging

Feature trees

- In the 'chain', each 'link' expresses a dependency relationship. What if some terms are independent?

- Each independence assumption causes a fork in the chain, yielding a feature *tree*.

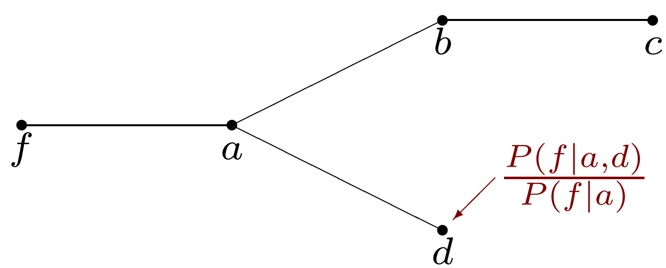
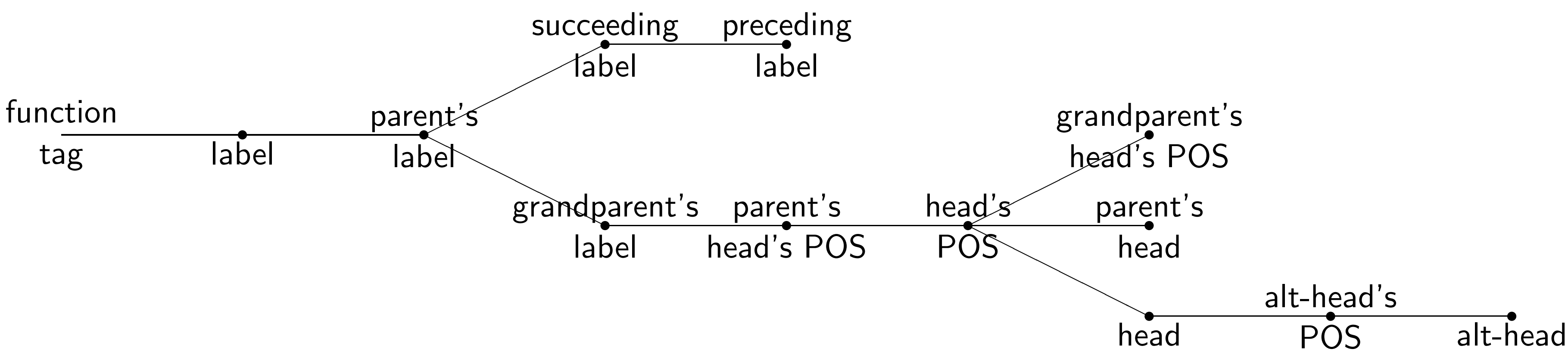


Figure 5: A feature tree: d is independent of b and c

Function Tagging

A feature tree



Optimising the feature tree

- Start with zero-node tree
- Perturb
  - Add feature at leaf
  - Swap two features
  - Move leaf to different (possibly new) branch
- Train, test on development corpus
- Keep best trees
- Repeat

Experiment

Training

- Gather statistics on sections 2-21 of treebank.
- Record conditioning environments of function tags.

Testing

- Begin with output of parser on section 23.
- For each constituent,
  - calculate likelihood of each function tag, and
  - for each category, assign most likely tag (possibly null).

Function Tagging

Performance by category

	All nodes	——Only tagged nodes——		
	Accuracy	Precision	Recall	F-measure
Grammatical	99.0%	96.5%	95.3%	95.9%
Form/Function	97.6%	86.7%	80.3%	83.4%
Topicalisation	99.9%	95.4%	95.4%	95.4%
Miscellaneous	99.8%	63.9%	21.9%	32.6%
Overall	99.1%	92.7%	88.7%	90.7%

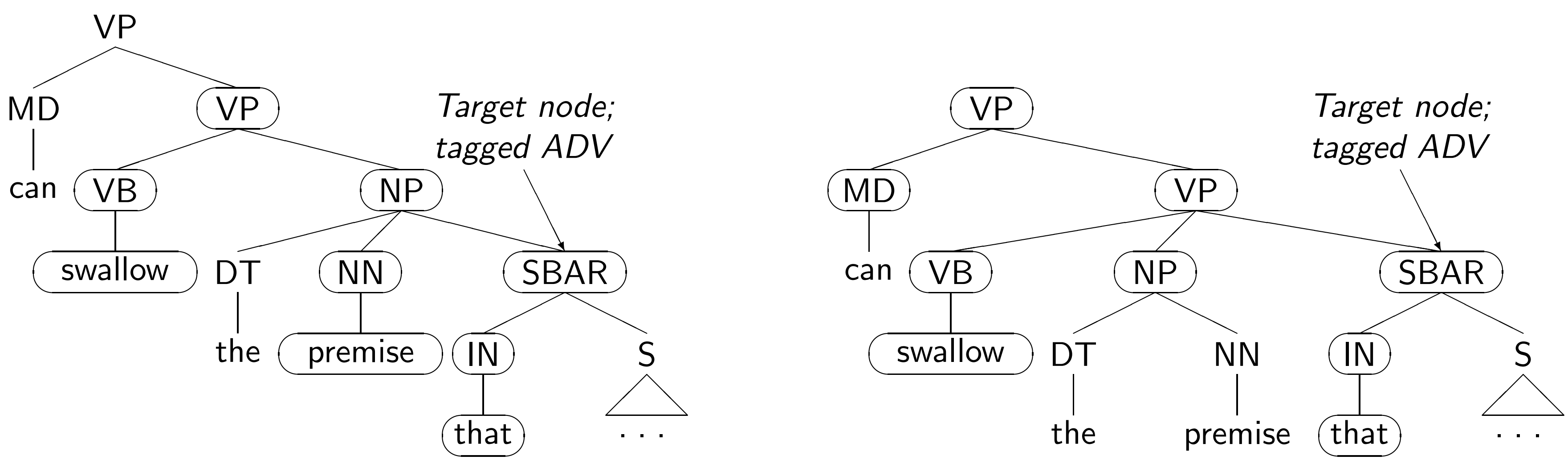
Function Tagging

Error analysis

Parser error	20%
Type A, B error	18%
Type C error	13%
Dubious	6%
Algorithm error	44%

Function Tagging

Outside sources of error I: Parser error



Outside sources of error II: Treebank error

- Type A: Detectable
  - LGS "attaches to the NP object of *by* and not to the PP node."
  - "President Bush has been weakened *by the Panama fiasco*."
- Type B: Fixable
  - LOC can be metaphorical, but not idiomatic
  - "think *about national service*" shouldn't be LOC
- Type C: Inconsistent
  - MNR indicates the manner in which an action is performed
  - "*impatiently*", "*suddenly*", "*significantly*", "*clearly*"

Related work: Collins 1997

- Parsing can be improved with complement/adjunct knowledge
- Function tags are used to indicate this
  - e.g. SBJ is complement, TMP is adjunct
- Results reported only on parser quality

Related work: Gildea and Jurafsky 2000

- FrameNet corpus project
- Composed primarily of “frames” of discourse, e.g. conversation
- Phrases tagged as “frame elements”, e.g. TOPIC, MEDIUM
- Every frame has different frame elements
- Both harder and easier; difficult to compare

Related work: Brants, Skut, and Krenn 1997

- German-language treebank from POS-tagged newspaper text
- *Every* item has "function label" e.g. SB, HD
- Order-2 Markov model, one per parent label type

Brants, Skut, and Krenn	
PP children	97.9%
S children	89.1%
Overall accuracy	94.2%

Blaheta	
No-null precision	96.5%
No-null recall	95.3%
No-null F-measure	95.9%
With-null accuracy	99.0%

System improvements I: decision trees

- Ask a series of questions to determine prediction
- Later questions depend on earlier answers
- Several systems for automatically learning them
- Handles large number of features
- Variant: boosted decision trees

Preliminary decision tree results

	Precision	Recall
Grammatical		
Feature trees	98.2%	97.4%
Decision trees	98.8%	97.6%
Form/function		
Feature trees	80.6%	76.9%
Decision trees	81.1%	71.1%

System improvements II: neural nets

- Neural nets have very simple algorithm
  - Encode guesses as feature vectors
  - Dot-product with weight vector
  - If best guess is wrong, subtract from weights and add in correct
- Can handle huge numbers of features
- Shown useful for parsing, POS tagging (Collins 2002)
- Related: support vector machines

Additional features I: object heads

- Already have 'alt' heads: second child of PP
- Doesn't work if premodifiers are present
  - only *in* Boston
  - an hour *before* the concert
- Define true object head for PPs
- . . . and other constituents?

Additional features II: conjunctions

- Function tag should appear on conjunction node
- Already use CCNP, CCVP—CCPP may be sufficient
- . . . or, more general CC feature

Additional features III: multiple siblings

- Currently, one sibling to either side
- Not helpful in rules like S $\rightarrow$ NP , NP VP
- Features for two or three siblings in each direction

Additional features IV: other function tags

- Tags not independent
- Different categories
  - SBJ isn't also TMP, LOC, etc.
- Different constituents
  - TMP only appears on object of PP if PP is marked DIR

Additional features V: lexical backoff

- Lexical items make for sparse data
- Possible backoffs:
  - word stem (Porter, CELEX)
  - semantic cluster (automatic, hand-built, wordnet)
  - POS tag (already use this!)
  - POS 'supertag'

Additional features VI: window contexts

- "Does word X occur within 5 words of this constituent?"
- Straightforward as binary feature
- Infeasible in some systems?

Function Tagging

Summary

- Function tags annotate variety of syntactic, semantic info
- Heretofore underexploited part of the Penn treebank
- Created successful feature-based system to perform task
- Propose to try improving performance
  - Other feature-based systems
  - New features

I'd like to thank

- Eugene Charniak
- Mark Johnson
- Michael Collins
- Members of BLLIP
- The Academy
- You

Any questions?

Function Tagging

Feature chains, technical

If a feature f can be guessed from features $f_1, \ldots, f_n$, we usually estimate its probability as

$$P(f|f_1, f_2, \ldots, f_n) \approx \hat{P}(f|f_1, f_2, \ldots, f_j), \quad j \leq n\ .$$

This is equivalent to

$$P(f|f_1, f_2, \ldots, f_n) \approx \hat{P}(f) \frac{\hat{P}(f|f_1)}{\hat{P}(f)} \frac{\hat{P}(f|f_1, f_2)}{\hat{P}(f|f_1)} \cdots \frac{\hat{P}(f|f_1, f_2, \ldots, f_j)}{\hat{P}(f|f_1, f_2, \ldots, f_{j-1})}$$

or

$$P(f|f_1, f_2, \ldots, f_n) \approx \prod_{i=0}^{j} \frac{\hat{P}(f|f_1, \ldots, f_{i-1}, f_i)}{\hat{P}(f|f_1, \ldots, f_{i-1})}\ .$$

Function Tagging

Feature trees, technical

À propos Figure 5, if d were still dependent on c (and we had complete data), the probability estimate would be

$$P(f|a,b,c,d) \approx \hat{P}(f)\,\frac{\hat{P}(f|a)}{\hat{P}(f)}\,\frac{\hat{P}(f|a,b)}{\hat{P}(f|a)}\,\frac{\hat{P}(f|a,b,c)}{\hat{P}(f|a,b)}\,\frac{\hat{P}(f|a,b,c,d)}{\hat{P}(f|a,b,c)}\;.$$

Noting d's independence from b and c, this becomes

$$P(f|a,b,c,d) \approx \hat{P}(f)\,\frac{\hat{P}(f|a)}{\hat{P}(f)}\,\frac{\hat{P}(f|a,b)}{\hat{P}(f|a)}\,\frac{\hat{P}(f|a,b,c)}{\hat{P}(f|a,b)}\,\frac{\hat{P}(f|a,d)}{\hat{P}(f|a)}\;,$$

which cancels to

$$P(f|a,b,c,d) \approx \frac{P(f|a,b,c)P(f|a,d)}{P(f|a)}\;.$$

Baseline performance

	Baseline 1 (never tag)	Baseline 2 (choose most likely tag)			
		Tag	Precision	Recall	F-measure
Grammatical	87.4%	SBJ	10.2%	80.9%	18.1%
Form/Function	91.8%	TMP	3.0%	36.1%	5.5%
Topicalisation	99.4%	TPC	0.6%	100.0%	1.2%
Miscellaneous	99.7%	TTL	0.1%	44.8%	0.2%
Overall	94.6%	—	3.5%	64.0%	6.6%

Function Tagging

Machine parse vs Treebank parse

	All nodes	——Only tagged nodes——		
	Accuracy	Precision	Recall	F-measure
Machine parsed	99.0%	92.7%	88.7%	90.7%
Treebank parse	99.2%	94.0%	89.8%	91.8%

Related work: Gildea and Jurafsky 2000

- FrameNet corpus
  - Domain: communication (cognition, motion)
  - Frame: conversation (statement, judgement)
  - Words: argue, debate, discussion, tiff
  - Frame elements: PROTAGONIST, TOPIC, MEDIUM
- Probabilistic, with lattice backoff model
- Given a sentence with marked frame elements, label them: 81.2%
- Given a sentence, mark frame elements: 66% (+ 15% partial)

Function Tagging

The tags

ADV	Non-specific adverbial	HLN	Headline	PUT	Locative complement of 'put'
BNF	Benefactive	LGS	Logical subject	SBJ	Subject
CLF	It-cleft	LOC	Location	TMP	Temporal
CLR	'Closely related'	MNR	Manner	TPC	Topic
DIR	Direction	NOM	Nominal	TTL	Title
DTV	Dative	PRD	Predicate	VOC	Vocative
EXT	Extent	PRP	Purpose		

Grammatical		**11.24%**
DTV	0.48%	0.05%
LGS	2.97%	0.33%
PRD	17.80%	2.01%
PUT	0.26%	0.03%
SBJ	78.48%	8.81%
VOC	0.03%	0.00%

Form/Function		**7.80%**
ADV	11.47%	0.89%
BNF	0.07%	0.01%
DIR	8.28%	0.64%
EXT	3.23%	0.25%
LOC	25.29%	1.98%
MNR	6.21%	0.48%
NOM	6.79%	0.53%
PRP	5.27%	0.41%
TMP	33.39%	2.61%

Topicalisation		**0.47%**
TPC	100.00%	0.47%

Miscellaneous		**0.12%**
CLF	5.39%	0.01%
HLN	42.75%	0.05%
TTL	51.84%	0.06%