Temporal Information Extraction
&
Biographical Summarization

Inderjeet Mani
MITRE
imani@mitre.org

Topics & Players

- Time Annotation Guidelines
  - *Beth Sundheim, SPAWARS*
  - *George Wilson, MITRE*
  - *Lisa Ferro, MITRE*
  - *Inderjeet Mani, MITRE*
- Temporal Information Extraction
  - *George Wilson, MITRE*
  - *Inderjeet Mani, MITRE*
- Biographical Summarization
  - *Barry Schiffman (Columbia)*
  - *Kris Concepcion (MITRE)*
  - *Inderjeet Mani (MITRE)*

Temporal Information Extraction

- Motivation
  - Question Answering
    - e.g., answering “when” questions, cf. TREC 1999 Q&A task
  - Single-Document Summaries of multiple-incident documents
  - Multi-Document Summarization (esp. merging and ordering information)
  - Event Tracking over time
- Developed
  - Initial Time Annotation Guidelines v1.0
  - TEMPEX v1.03 Time Tagger

Time Annotation Guidelines v1.0

Extends MUC-7 spec to flag more expressions, assign more values:

Indexicality <TIMEX2 VAL="2000-10-27">tomorrow</TIMEX2>

Duration <TIMEX2 VAL="PT30M">half an hour long</TIMEX2>;

<TIMEX2 VAL="1990" MOD="BEFORE">more than a decade ago</TIMEX2>

Granularity <TIMEX2 VAL="XXXX-WXX-2" SET="YES" PERIODICITY="F1W" GRANULARITY="G1D">every Tuesday</TIMEX2>

Fuzziness <TIMEX2 VAL="1990-SU">Summer of 1990 </TIMEX2>

Non-specificity <TIMEX2 VAL="XXXX-04" NON_SPECIFIC="YES">April</TIMEX2> is usually wet.

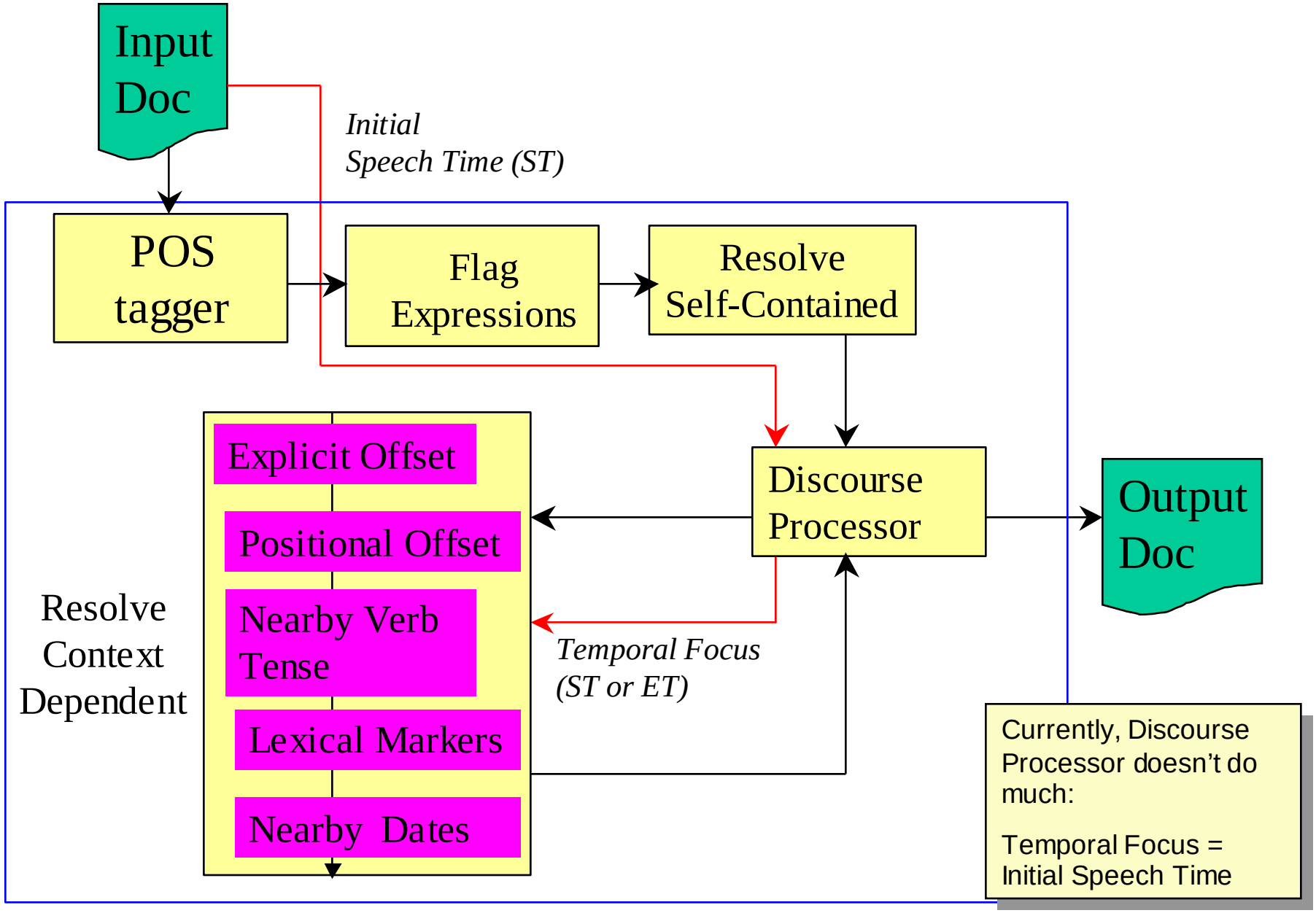
Time Annotation Plans (FY01)

- Obtain feedback from you on TIDES Time Annotation Guidelines v1.0 by December 25, 2000 (award scheme)
- Update existing small internal reference annotation, for distribution
  - NYT (currently 22 articles, 35,555 words) and TDT2 (221 articles, 78,171 words)
- Conduct small inter-annotator agreement study
- Launch a community-wide effort to annotate a larger corpus (at HLT meeting)
  - we provide: guidelines, instructions, small reference annotated corpus, annotation software (including time tagger), and scoring software

TEMPEX Time Tagger

- TEMPEX v1.03 Time Tagger developed
  - used in Qanda question-answerer (for TREC)
  - available for research (nonexclusive, royalty-free license)
- TEMPEX tagging of time values in news achieved 83.2% accuracy (v1.03)
- Major sources of error repaired through machine learning
  - 72% accuracy on “today” WSD
  - 97% accuracy on season word WSD
- First small steps towards event chronologies
  - initial versions of event ordering, simple visualization
- For more details, see ACL’2000 paper

TEMPEX Architecture



TEMPEX Accuracy (v1.01)

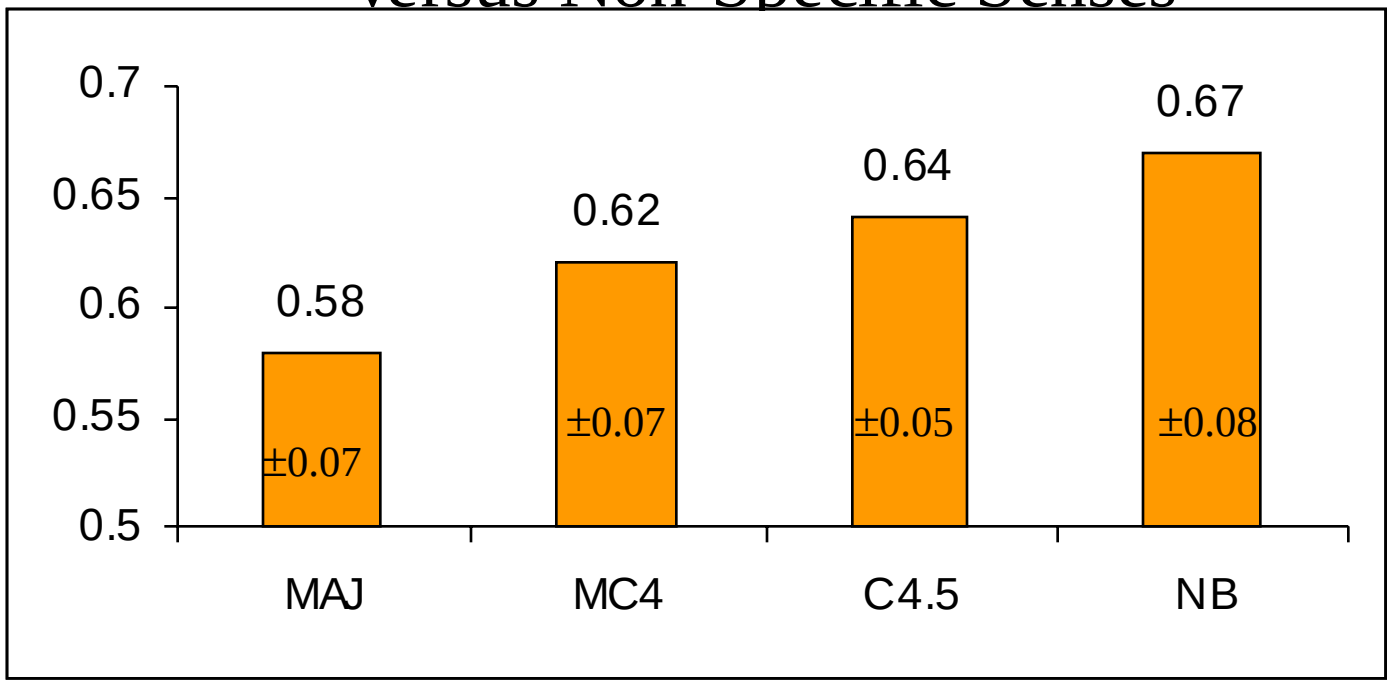
Source articles number of words	Type	Human Found (Correct)	System Found	System Correct	Precision	Recall	F-measure
NYT 22 35,555	TIMEX	302	302	296	98.0	98.0	98.0
	Values	302	302	249 (129)	82.5 (42.7)	82.5 (42.7)	82.5 (42.7)
Broadcast 199 42,616	TIMEX	426	417	400	95.9	93.9	94.9
	Values	426	417	353 (105)	84.7 (25.1)	82.9 (24.6)	83.8 (24.8)
Overall 221 78,171	TIMEX	728	719	696	96.8	95.6	96.2
	Values	728	719	602 (234)	83.7 (32.5)	82.7 (32.1)	83.2 (32.3)

Numbers in parentheses are the baseline scores from just tagging values of absolute, fully specified TIMEXs (e.g., "January 31st, 1999").

Error Analysis (v1.01)

- Extra values: 15/25 *Today's meeting/youth*
- Bad Extent: 11/16 *Friday the thirteenth*
- Wrong Value: 10/30 *This/next year* (confusion on New Year's Eve)
- Errors in source: 2 *40:00 hrs GMT* (broadcast news transcription errors)

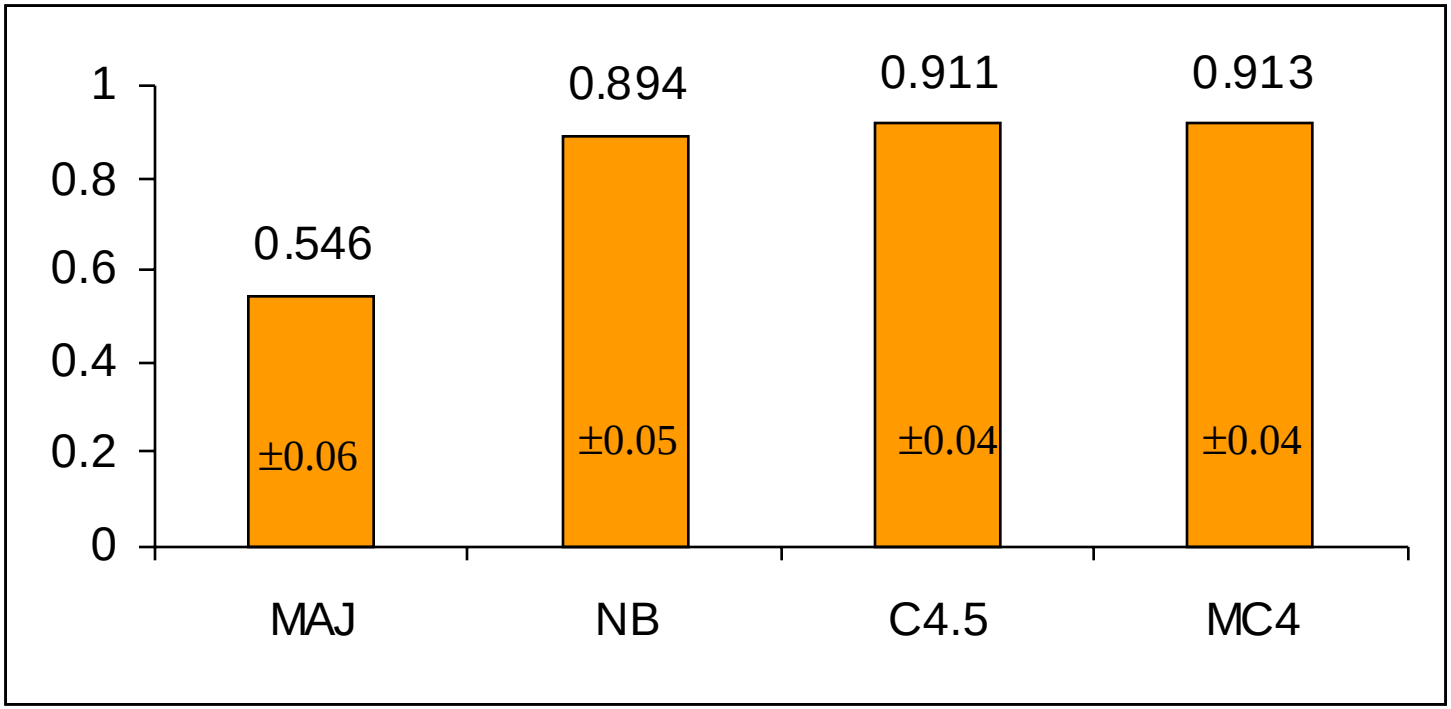
Machine Learning of Specific versus Non-Specific Senses



"today"
10-fold cv
511 training
57 test
C4.5 > MAJ*

- ***If "today" is immediately followed by a determiner, and is inside a quotation, and sentence has no day of the week, then"today" is Non-Specific (72%)***
- ***If sentence has "most", then "today" is Non-specific (68% accuracy)***

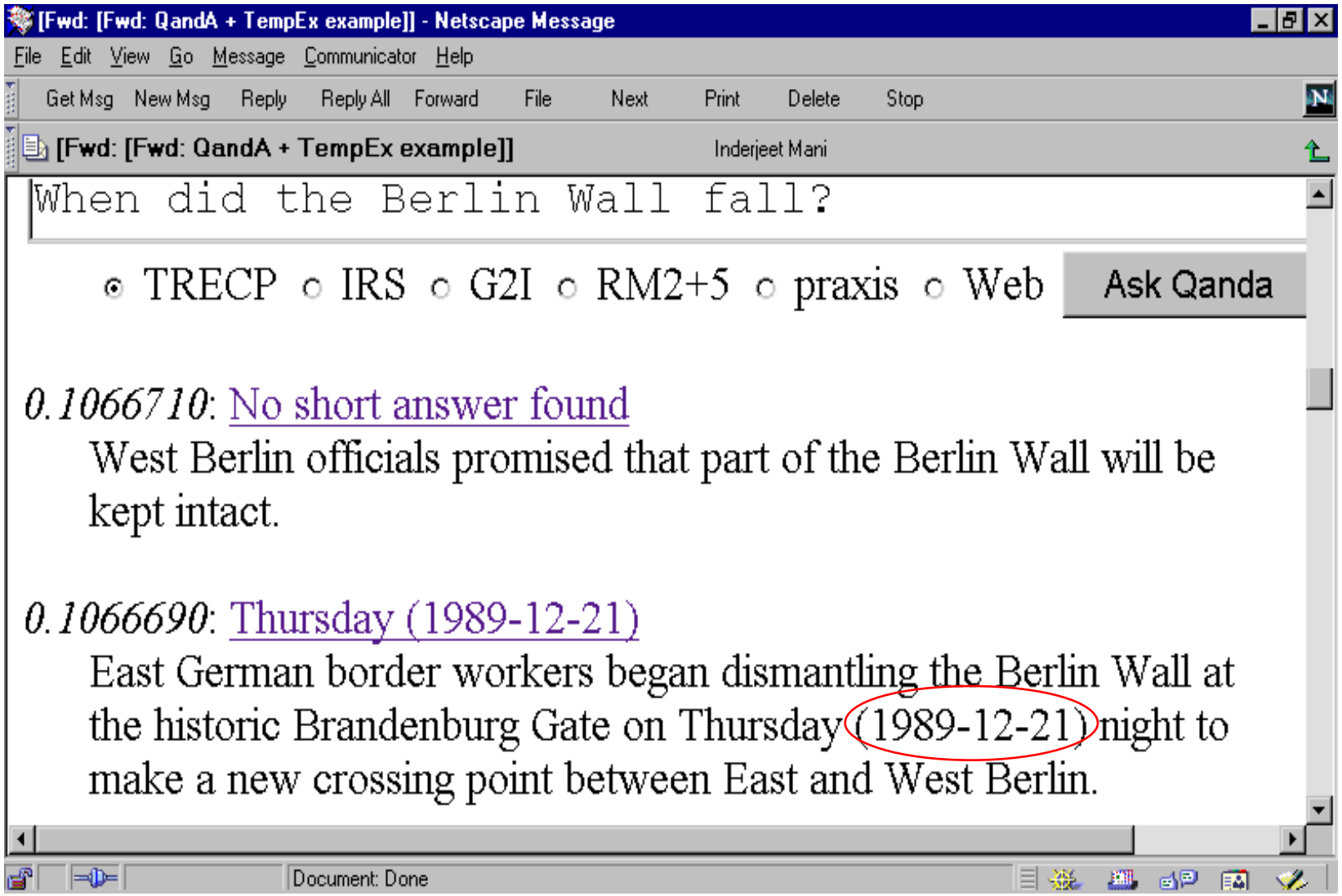
Machine Learning of Seasonal References



"fall"
10-fold cv
563 training
63 test
NB > MAJ*

- ***If "fall" is immediately preceded by "last |next" ["this"], then"fall" is seasonal 97.3% [94.5%]***
- ***If "fall" is followed 2 words after by a year expression, then "fall" is seasonal (86.3%)***
- ***If "fall" isn't immediately preceded by "this" or "last | next" and is immediately followed by a particle (e.g., "off", "down"), then "fall" is non-seasonal (89.5%)***

TEMPEX in Qanda



TEMPEX Plans (FY01)

- Extend coverage to phenomena in latest guidelines
- Improve Discourse Processor
  - Use event-ordering heuristics to adjust Temporal Focus
- Apply time tags to order extracts for Multi-Document Summarization

Biographical Summarizer (BioSumm)

- Multi-document, topic-focused biographical summarizer
- Extracts descriptions of people across documents, merging them to remove redundant descriptions
  - *lawyer for the defendant + an attorney for Paula Jones = a lawyer for Paula Jones (synonymy, PNs, length.)*
- Outputs text or template (slot fills are underlined here)

bio-summary.csh "Vernon Jordan"
-target-len 10 clinton-corpus

V*ernon Jordan is a presidential friend and a Clinton adviser. He is 63 years old. He helped Ms . Lewinsky find a job. He has numerous acquaintances, including Susan Collins, Betty Currie, Pete Domenici, Bob Graham, James Jeffords and Linda Tripp.*

H*enry Hyde is a Republican chairman of House Judiciary Committee and a prosecutor in Senate impeachment trial. He will lead the Judiciary Committee's impeachment review. He spoke of soldiers who died defending the United States.*

V*ictor Polay is the Tupac Amaru rebels' top leader, founder and the organization's commander-and-chief. His associates include Alberto Fujimori, Tupac Amaru Revolutionary, and Nestor Cerpa.*

Knowledge Sources for BioSumm

- Syntax (Uses CASS parser and pattern-matching rules)
  - to get valid person descriptions, and to merge descriptions
- Thesaurus concepts under person, from WordNet
  - for filtering person descriptions and merging descriptions
- Corpus statistics
  - person associates are extracted using mutual information
  - modifiers with higher corpus frequency are preferred (e.g., exclude ‘beleagured’ in ‘a beleagured President’)
  - merged descriptions sorted by corpus frequency of description head

BioSumm Component Evaluation Overview

- ***Person-noun tagger***
  - <person> *string*</person> if threshold % of senses of *string* in WordNet under Person, e.g., *accessory* - non-person; *bum* - person
  - scored against Brown Corpus SEMCOR semantic concordance of 6000 sentences with 42,000 nouns
- ***Relative clause tagger*** scored on 36 instances over 22 docs (scores for coref id and extent)
- ***MDS Description merger*** scored in automatic comparison of system merge of system-generated descriptions against human merge of them (1,300 articles, 226 multi-description people)
  - System Descriptions: *Washington lawyer, Washington white-collar defense lawyer, former federal prosecutor*
  - System Merge: *Washington white-collar defense lawyer*
  - Human Merge: *a Washington lawyer and former federal prosecutor*
  - Baseline carried out over just 15 people

Component	Accuracy
Person Noun Tagger	.97
Relative Clause Tagger	.82
Description Merger	.42
Baseline string match Description Merger	.21

Bio SummPlans (FY01)

- Add serious **cross-document coreference** (major problem)
- **Fine-tune hierarchy** using corpus
- Figure out what **biographical features** are salient for different types of people (cf. lack of appositives for famous people)
  - find verbs strongly associated with agent head nouns in the corpus
    - *politician - vote, elect; executive - resign, appoint; police - arrest, shoot*
    - Include **essential** modifiers: *former; the late; president in 1866*
- Tie in **temporal tagging**, to be able to track biographical changes
- Evaluate, in document context, whether description is an **adequate** one (very human-time-intensive)

DONE!

TURN PAGE FOR BACKUPS

Small Steps towards Event Chronologies

- Simple visualization developed
- Ordered events in time based on "temporal coherence of narrative"
  - Ordered verb occurrences based on rules for coherent tense sequences (based on Song&Cohen AAAI'91)
- Plan to associate "occurrence" events with times, based on local syntactic analysis:
  - *crashed on Wednesday*, *the 1999 crash*, etc.
  - Baseline: just propagating times to neighboring verb occurrences without analysis gives ~59.4% *correct*
    - *Correctness*: percentage of correct time fills for correctly recognized taggable verbs
    - Excludes auxiliaries, modals, verbs following *to*, *not*, or specific modals

Personifier Evaluation

- **Evaluation tested how accurately tagger could identify whether a description was appropriate as a person description**
- Personifier treats a candidate noun as a person if at least 3/4 of the WordNet senses of the noun have person as ancestor.
- Evaluation uses Wordnet 1.6 SEMCOR semantic concordance, which has files from the Brown corpus whose words have semantic tags (created by Wordnet's creators) indicating Wordnet sense numbers.
- Evaluation on 6,000 sentences with almost 42,000 nouns compares people tags generated by program with SEMCOR tags

- Total right = 41555, total wrong = 1298, total missing = 0
- Precision = 0.97, Recall = 0.97
- **Confirms that for news at least, people words are usually used in a people sense.**

Rel Clause Extraction Evaluation

- *Leon Panetta, who has been extremely critical of the president,*
- **Evaluation tested how well-formed the relative clauses were**
- The rel clause is judged correct if it has
  - the right extent, the correct coref id.
- 22 docs, 36 instances: 28 correct found, plus 4 spurious finds
- Precision = 0.87, Recall = 0.78, F-measure=.82
- **Although sample is small, results are very promising**

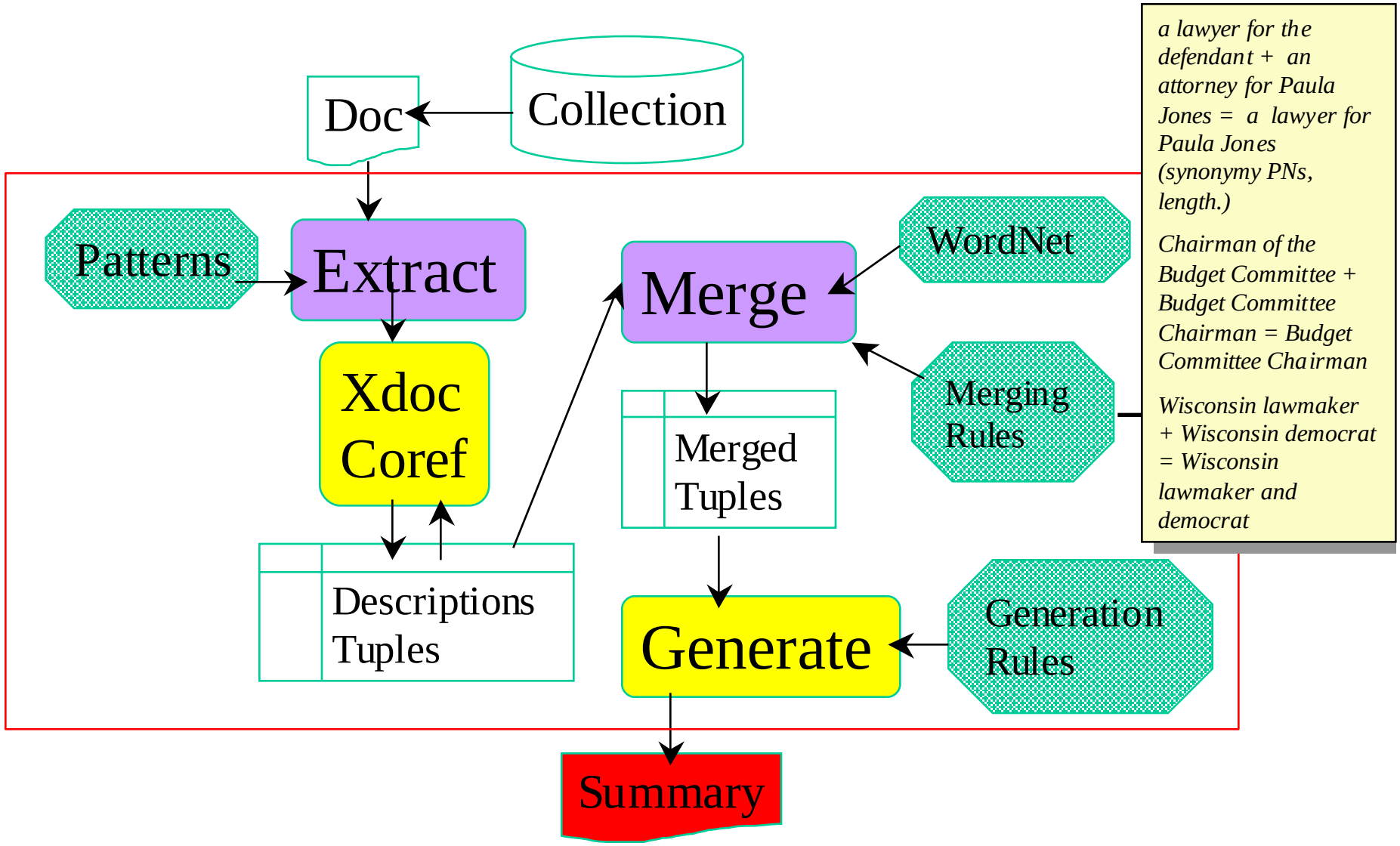
Evaluation of Merging

- **Evaluation tested the system's ability to accurately merge descriptions**
- **Example of Evaluation:**
  - System: E. Lawrence Barcella is a *Washington lawyer, Washington white-collar defense lawyer, former federal prosecutor*
  - System Merge: *Washington white-collar defense lawyer*
  - Human Merge: *a Washington lawyer and former federal prosecutor*
  - *Auto Score: Exact-Description=0; Extra-Words=2; Missed-Words=3*
- Data: 1,300 wire service articles about Clinton impeachment (798, 166 words). System: 1029 'people' (some not), 541 people with descriptions, 315 with unique descriptions, ~90% of which were well-formed. Remaining 226 people, who thus had more than one description, were scored.

Scores on Merging

- System used 815 (total) content words in descriptions of 226 people
- *Correct (Exact-Description)*: 65/226 system descriptions
- *Incorrect:*
- *(Extra-Words)*: 184/815 content words;
- *(Missed-Words):* 258/815 content words
  - About 20% of the content word errors were semantically correct substitutions
- Major problem areas were cross-document coreference, and lack of good descriptions for famous people

Bio Summ Architecture



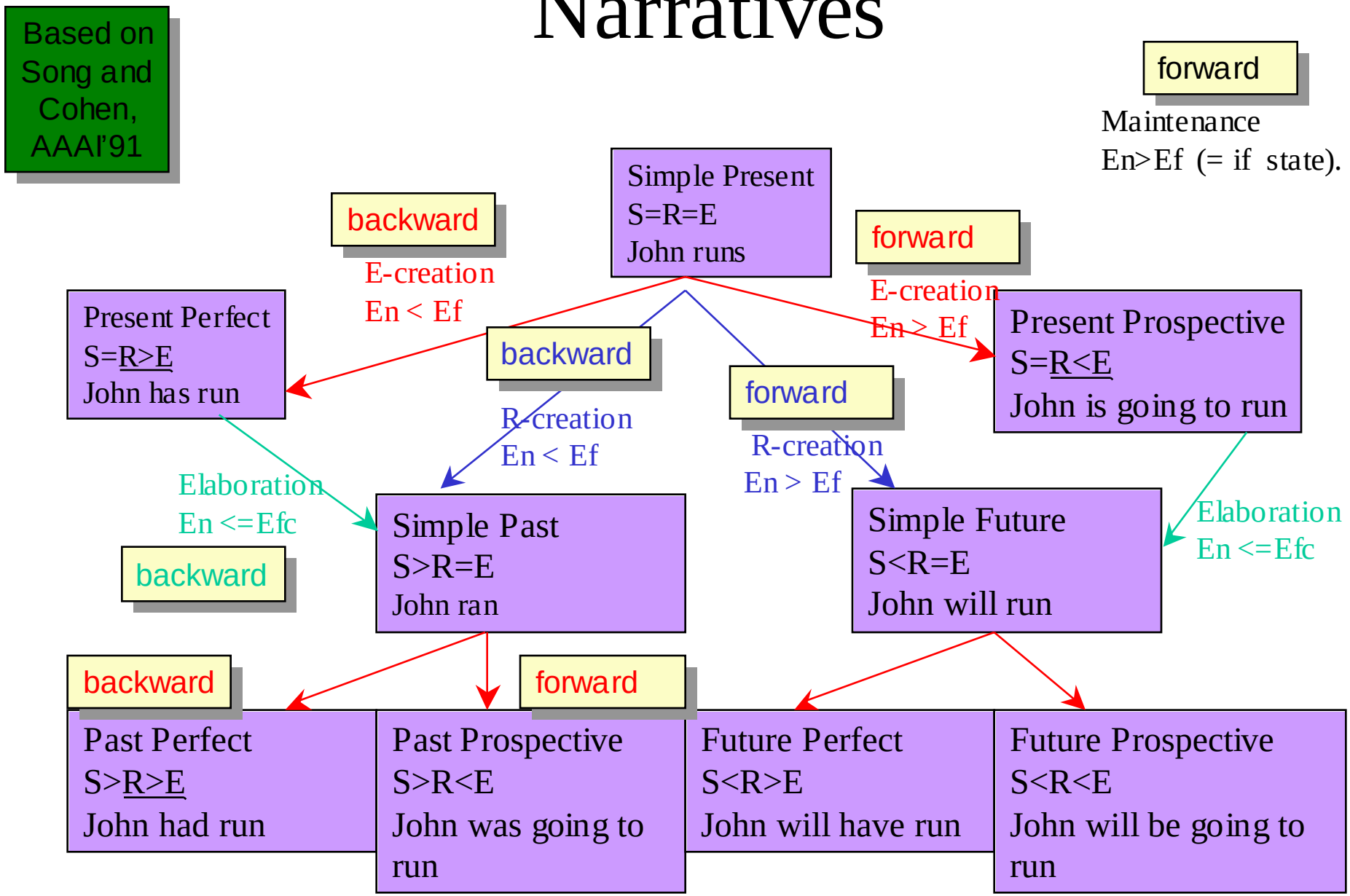
Features for ‘Today’Classification

- ***Poss***: whether “today” has a possessive inflection
- ***Qcontext***: whether “today” is inside a quotation
- ***Said***: presence of “said” in the same sentence
- ***Will***: presence of “will” in the same sentence
- ***Even***: presence of “even” in the same sentence
- ***Most***: presence of “most” in the same sentence
- ***Some***: presence of “some” in the same sentence
- ***Year***: presence of “year” in the same sentence
- ***CCYY***: presence of a four-digit year in the same sentence
- ***DOW***: presence of a day of the week expression (“Monday” thru “Sunday”) in the same sentence
- ***FW***: “today” is the first word of the sentence
- ***POS1***: part-of-speech of the word before “today”
- ***POS2***: part-of-speech of the word after “today”
- ***Label***: specific or non-specific (class label)

Error Breakdowns

	Print	Broadcast	Total
Missing Vals	10	29	39
Extra Vals	18	7	25
Wrong Vals	19	11	30
Missing TIMEX	6	15	21
Extra TIMEX	2	5	7
Bad TIMEX extent	4	12	16
TOTAL	59	79	138

Rules for Temporally Coherent Narratives



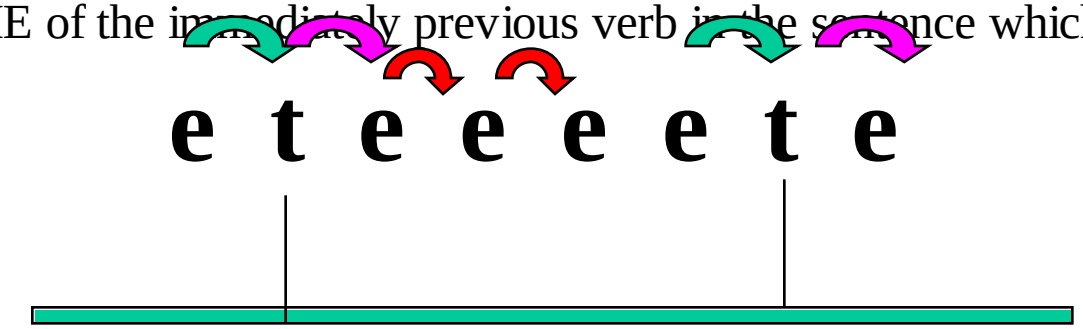
Event Ordering

Based on Song and Cohen, AAAI'91

Event Tense Transition	Type of Move	Event Time Shifts
Same tense	Anaphoric Progression	Forward (except statives and unbounded processes, which keep same time)
Present Perfect to Simple Past; Present Prospective to Simple Future	Anaphoric Elaboration	Backward or keeps same time
Simple Present to Simple Past (Future)	New R-creation	Backward (forward)
Simple to Perfect or Prospective	New E-creation	Backward or forward

Event-Align

- Each taggable verb occurrence lacking a time expression is given the VAL of the immediately previous time expression in the sentence.
- When a time expression is found, the immediately previous verb lacking a time expression is given that expression's VAL
- Maintain temporal focus:
  - Each taggable verb in a sentence lacking a time expression is given the TIME of the immediately previous verb in the sentence which has one



TIMEX Vals

Taggable:

all verbs except modals, auxiliaries, and verbs following “to”, “not”, or a modal verb

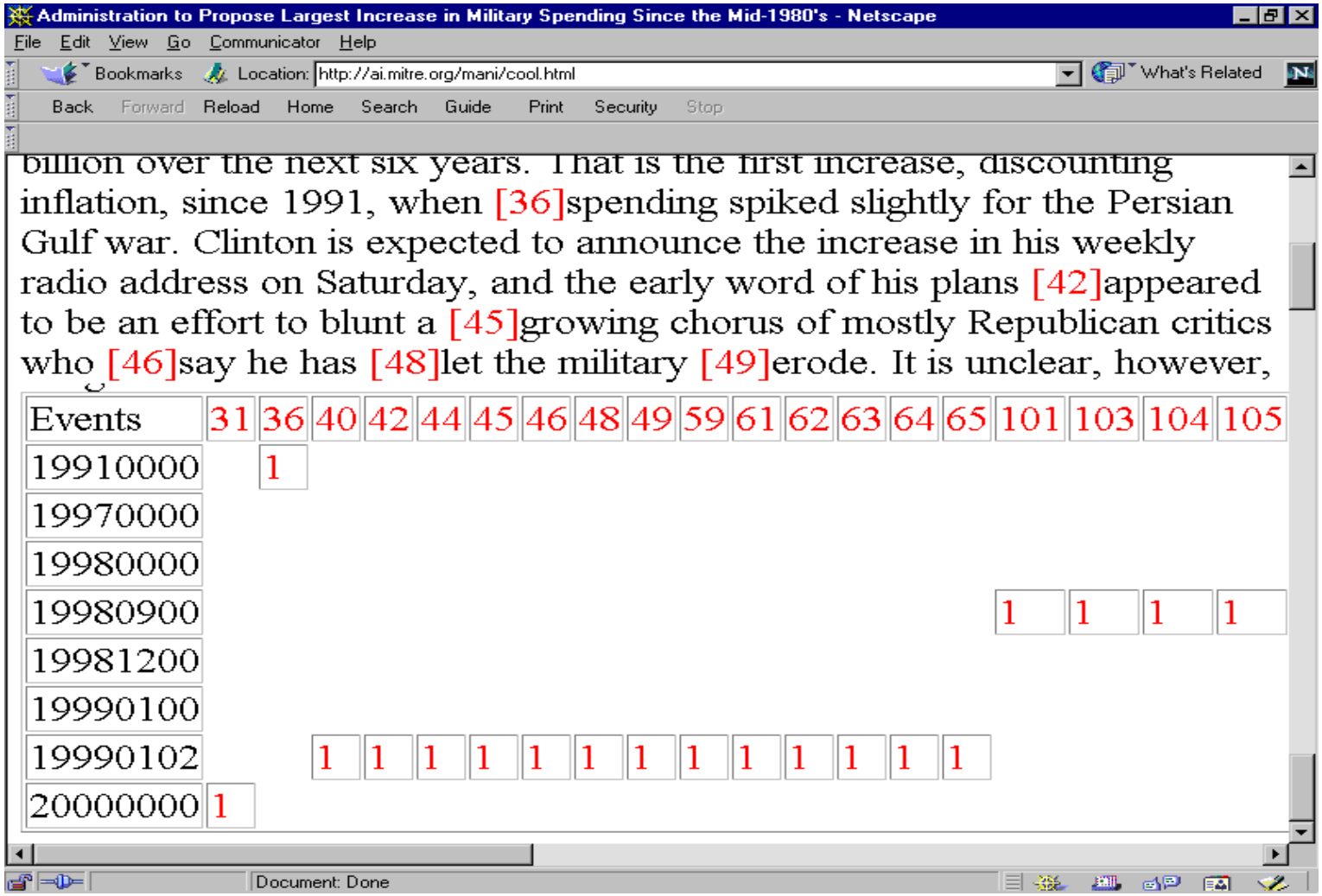
Event-Align Example

In the last step after years of preparation, the countries <lex eindex="9" TIME="19981231">locked</lex> in the exchange rates of their individual currencies to the euro, thereby <lex eindex="10" TIME="19981231">setting</lex> the value at which the euro will begin <lex eindex="11" TIME="19990104">trading</lex> when financial markets open around the world on <TIMEX VAL="19990104">Monday</TIMEX>……

Correctness: *number of correct TIME fills for correctly recognized verbs over total number of correctly recognized verbs.*

Baseline Method Correctness = 394/663 = 59.4% (over tiny sample of 8505 words of text)

Event Chronology Visualization



Events	31	36	40	42	44	45	46	48	49	59	61	62	63	64	65	101	103	104	105
19910000		1																	
19970000																			
19980000																			
19980900																1	1	1	1
19981200																			
19990100																			
19990102			1	1	1	1	1	1	1	1	1	1	1	1	1				
20000000	1																		

Corpus-Based Methodology

- Identify largest sources of errors ==> classes of problem time terms
- For each class of problem time terms
  - Find features which co-occur with time terms in same sentence ==> related features (automatic)
  - Sort related features by frequency in corpus ==> top related features (automatic)
  - Group top related features into feature classes (manual)
  - Refine feature classes as needed (manual)
  - Create vectors for sentences containing problem terms using feature classes (automatic)
  - Train a classifier (automatic)
  - Evaluate (automatic)
  - Incorporate good learnt rules into TEMPEX system (manual)