Sentence Compression

LING 573 — Systems & Applications
May 22, 2018

Begin Recording

Announcements

- I will post the final presentation order this afternoon
  - (Will be day-swapped with D3 presentation order)
  - Please let me know if you have conflicts with that date
- Information Ordering with RNNs (Logeswaran et al, 2018)
  - I may have time to go over it on Thursday, but if not, here is the link

Sentence Compression

- **When to Do Compression**
  - Pre/post-processing — Candidate selection: heuristic/learned
- **Heuristic approaches**
  - Deep vs. Shallow processing
  - Information- vs. readability-oriented
- **Machine-learning approaches**
  - Sequence models
    - HMM, CRF
  - Deep vs. Shallow information

Form	CLASSY	ISCI	UMD	SumBasic+	Cornell
Initial Adverbials	Y	M	Y	Y	Y
Initial Conj	Y		Y	Y	
Gerund Phrases	Y	M	M	Y	M
Relative clause appositives	Y		M	Y	Y
Other adv	Y				
Numeric: ages,	Y				
Junk (byline, edit)	Y				Y
Attributives	Y	Y		Y	Y
Manner modifiers	M	Y	M		Y
Temporal modifiers	M	Y	Y		Y
POS: det, that, MD			Y		
XP over XP			Y		
PPs (w/, w/o constraint)			Y		
Preposed Adjuncts			Y		
SBARs			Y		M
Conjuncts			Y		
Content in parentheses		Y			Y

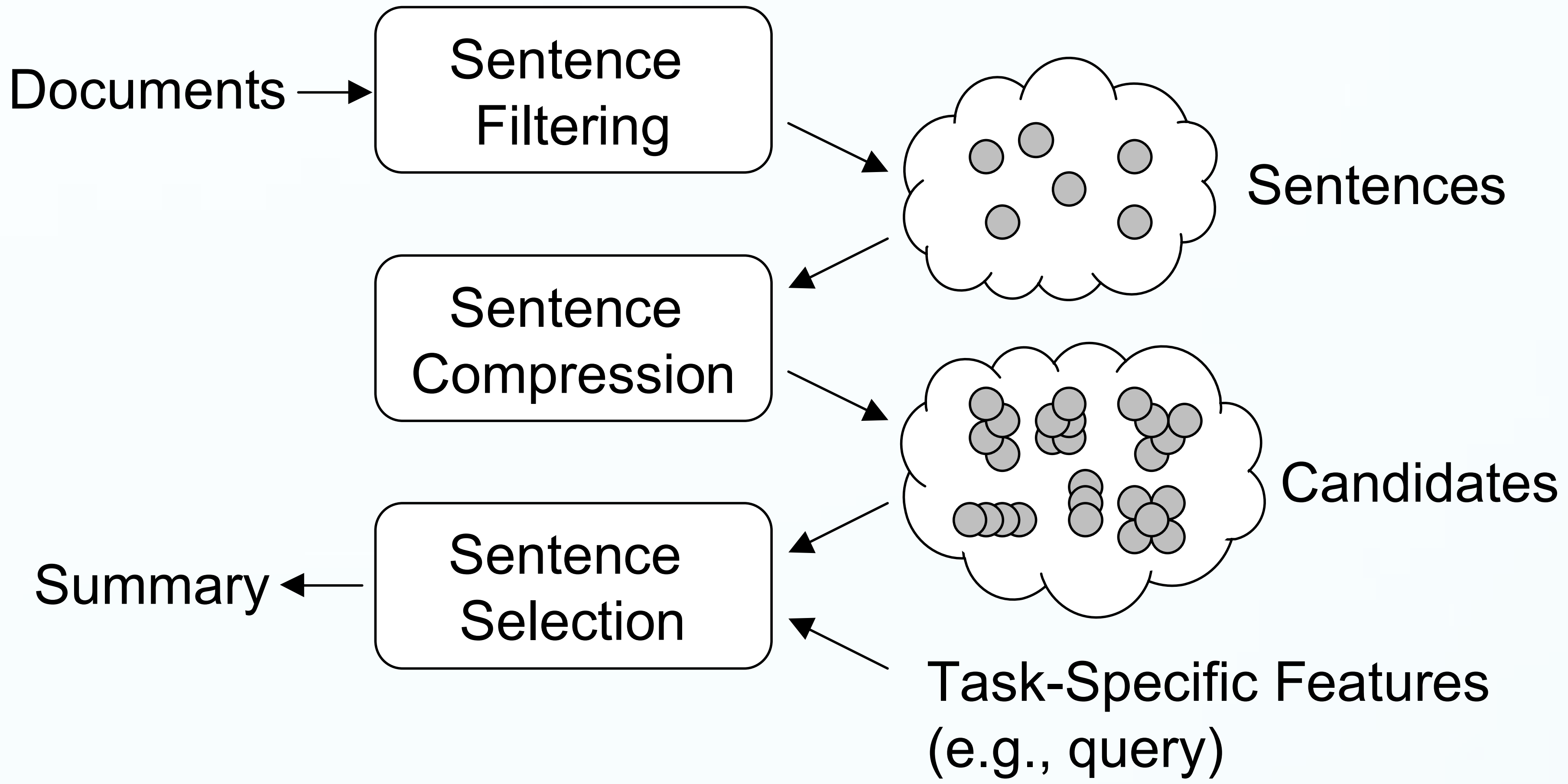
When To Do Compression?

When To Do Compression?

- **Simplest strategy**
  - CLASSY, SumBasic+
  - Deterministic
  - compressed sentence replaces original

When To Do Compression?

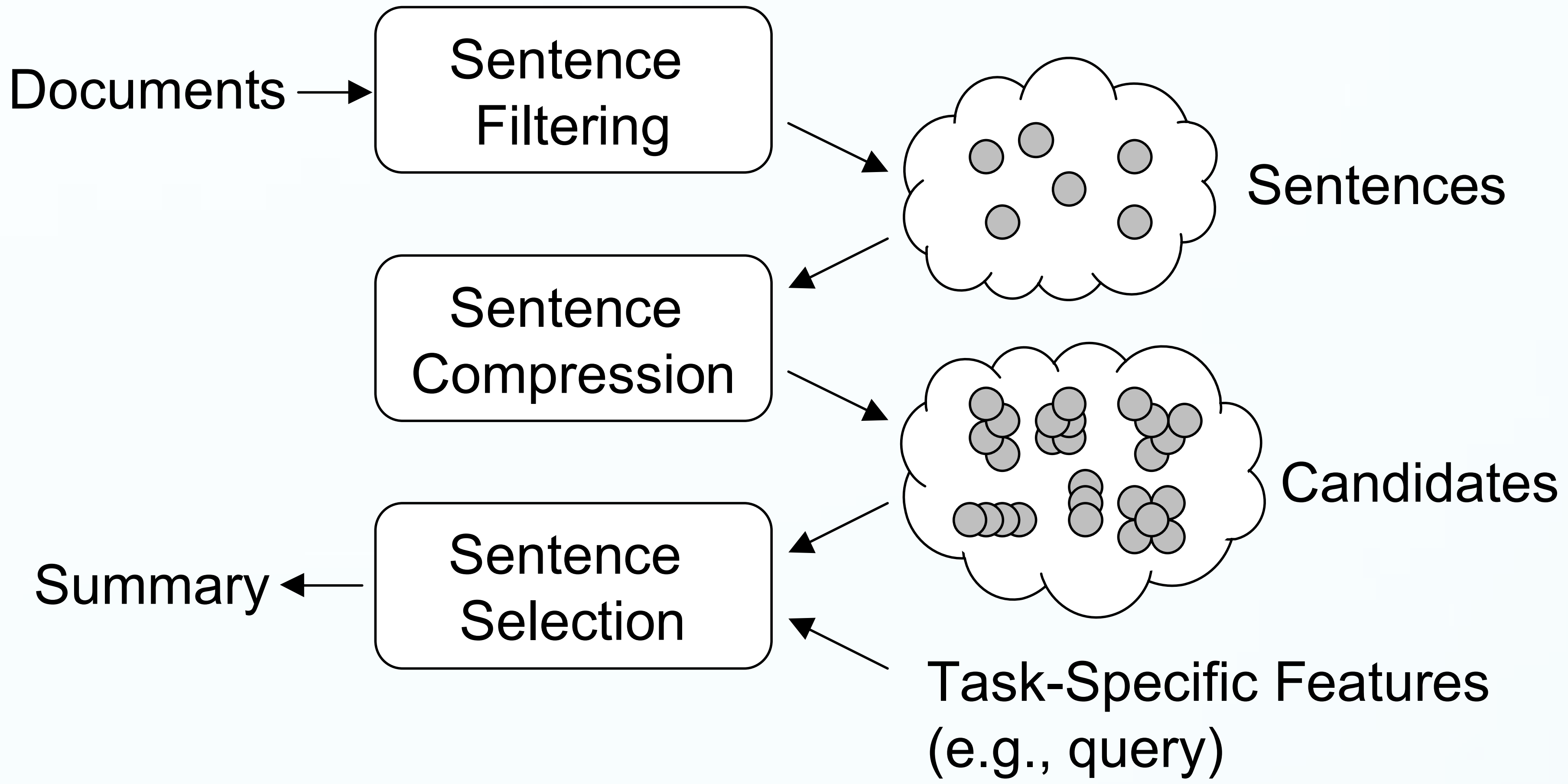
- **Multi-candidate approaches**:
  - most others
  - Generate multiple compressed sentences
    - Possibly constrained by
      - compression ratio
      - minimum length
      - e.g. exclude < 50% original, < 5 words (ICSI)
- Add to original candidate sentences list



Zajic et al (2007), *Fig. 1*

When To Do Compression?

- Use overall content selection procedure
- Limit to single selection per original sentence



Zajic et al (2007), Fig. 1

Shallow, Heuristic:

CLASSY 2006 (Conroy, et. al, 2006)

Shallow, Heuristic:

CLASSY 2006 (Conroy, et. al, 2006)

- **Original Approach**: (Dunlavy et. al, 2003)
  - Rule-based elimination of:
    - sentences that begin with an imperative
    - sentences that contain a personal pronoun at or near the start
    - gerund clauses
    - restricted relative-clause appositives
    - intra-sentential attribution
    - lead adverbs

Offered Trimming Rationale:
Gerunds

- *"Gerunds often comment on, rather than advance, a narration and therefore tend to be incidental"*
  - More than 800 lives were lost when the 21,794 tonne ferry, ~~***sailing from the Estonian capital Tallinn to Stockholm,***~~ sank within minutes early yesterday morning in the Baltic Sea 40 km south west of the Finnish island of Uto.

- *…participial phrase, not gerundive, but okay.*

Offered Trimming Rationale:
Restricted Relative Clause

- *"Restricted relative-clause appositives usually provide background information."*
  - The Menendez family lived in the Princeton Area until 1986, ~~*when they moved to California.*~~

Attribution: Example

- **Attribution**
  - ~~*The federal Government's highway safety watchdog said Wednesday*~~ that the Ford Bronco II appears to be involved in more fatal roll-over accidents than other vehicles in its class and that it will seek to determine if the vehicle itself contributes to the accidents.

Problems with this Approach

- Authors originally used POS tagger
  - Slow part of pipeline
  - Made mistakes
  - Occasionally bugged out and returned no tags

CLASSY 2006 Approach:
Rule-based, Lexicalized Removals

- **Removed**:
  - Sentence-initial adverbs and conjunctions: — *"As a matter of fact," "At this point,"*
  - Ages — *", 51" ", aged 24,"*
  - "Gerund" phrases
  - Relative clause attributives
  - …etc
- **Errors**:
  - **25%** sentences in prior approach made ungrammatical or semantically incorrect
  - **3%** with this approach (though recall likely very low)

Deep, Minimal, Heuristic:

ICSI/UTD (Gillick et al, 2008, 2009)

Original Approach

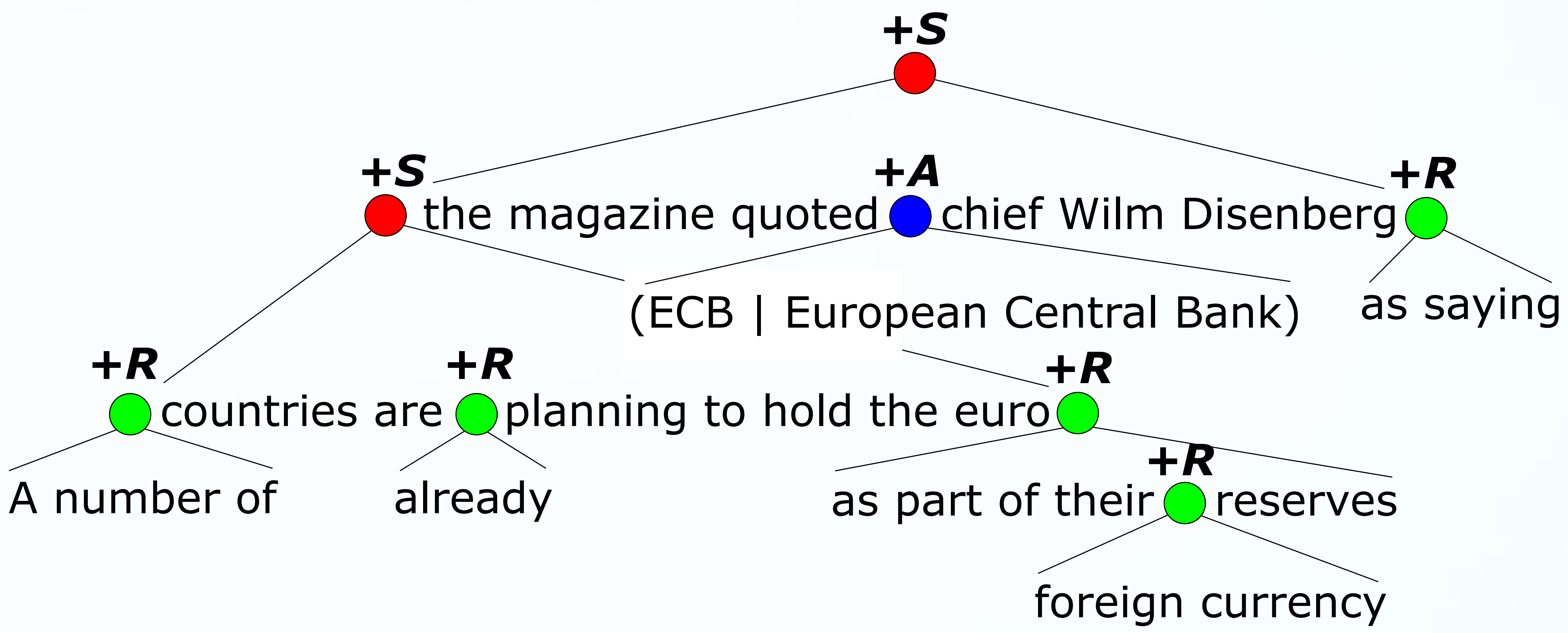
- Original Approach:
  - Parse sentence
  - Analyze sentences for:
    - **+S = clause**
      - NP followed by VP, excluding pronomial subjects
    - **+R = removable node**
      - Temporal phrases, parentheticals, subordinate clauses, subset of prepositional clauses
    - **+A = alternatives**
      - Grouping NPs that occur with the same head (e.g. appositions)

Original Approach

- Propose permutations as alternate sentences
- Allow optimization to account for length, information content, etc.

Gillick et al, 2008, Fig. 1

Example



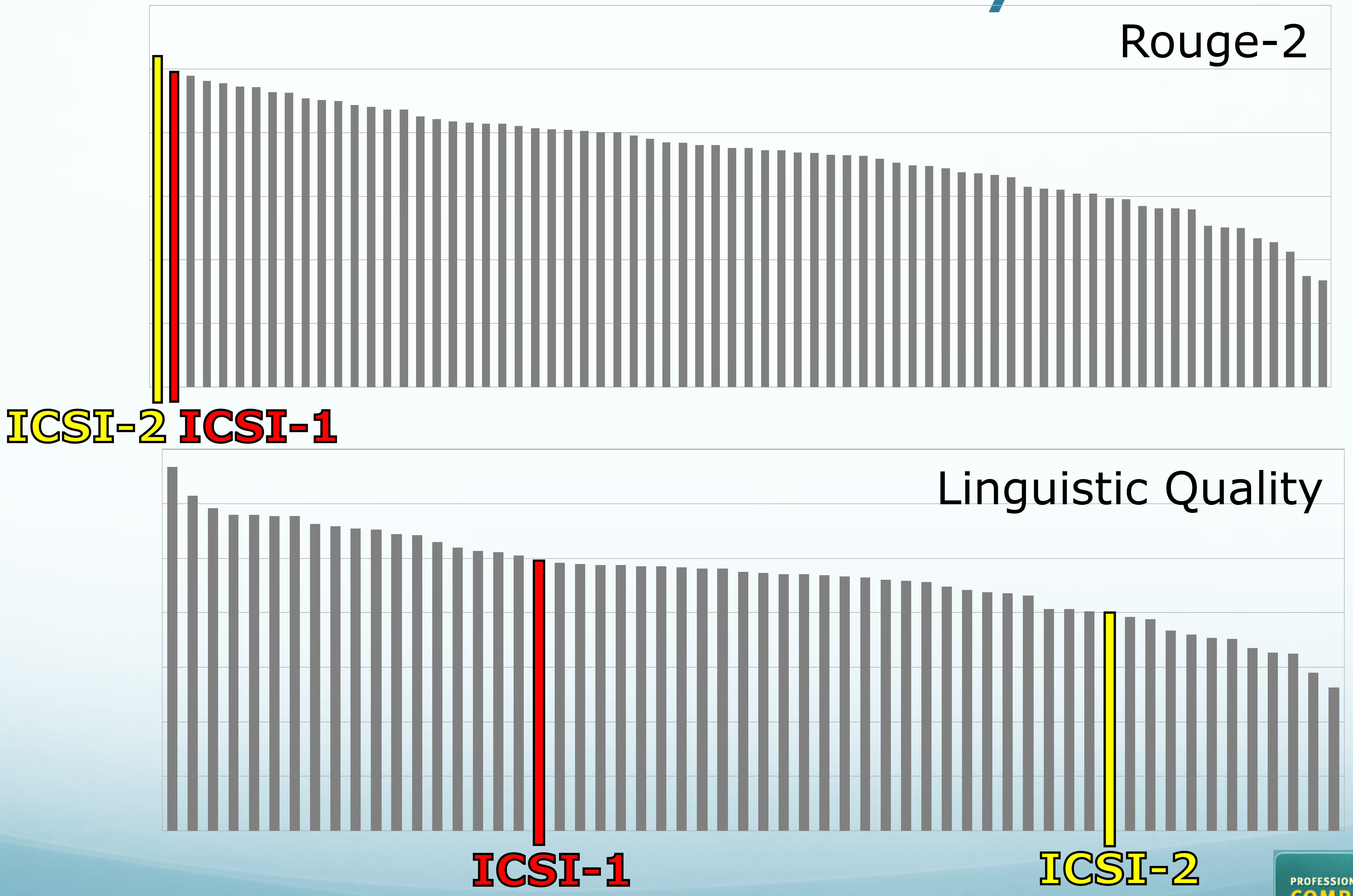
+R = Removable

+S = Subsentence

+A = alternative group

- *Countries are already planning to hold the euro as part of their reserves, the magazine quoted European Central Bank chief Wim Duisenberg as saying.*
- *A number of countries are planning to hold the euro, the magazine quoted ECB chief Wim Duisenberg.*
- *A number of countries are already planning to hold the euro as part of their foreign currency reserves.*
- *A number of countries are planning to hold the euro.*

Results of 2008 System



UNIVERSITY OF WASHINGTON

PROFESSIONAL MASTER'S IN COMPUTATIONAL LINGUISTICS

Results of 2008 System

- Highest reported ROUGE-2 scores
- Less good on Linguistic Quality

2009 Approach

ICSI/UTD (Gillick et al, 2009)

- Walk back the aggressiveness of the 2008 approach
  - Focus on improving readability
  - Still remove temporal expressions, manner modifiers, "said," etc.
- Methodology
  - Automatic SRL labeling over dependencies
    - ARGM-TMP or ARGM-MNR
    - SRL still noisy, so restrict to high-confidence labels

Examples

A ban against bistros providing plastic bags free of charge will be lifted ~~at the beginning of March~~.

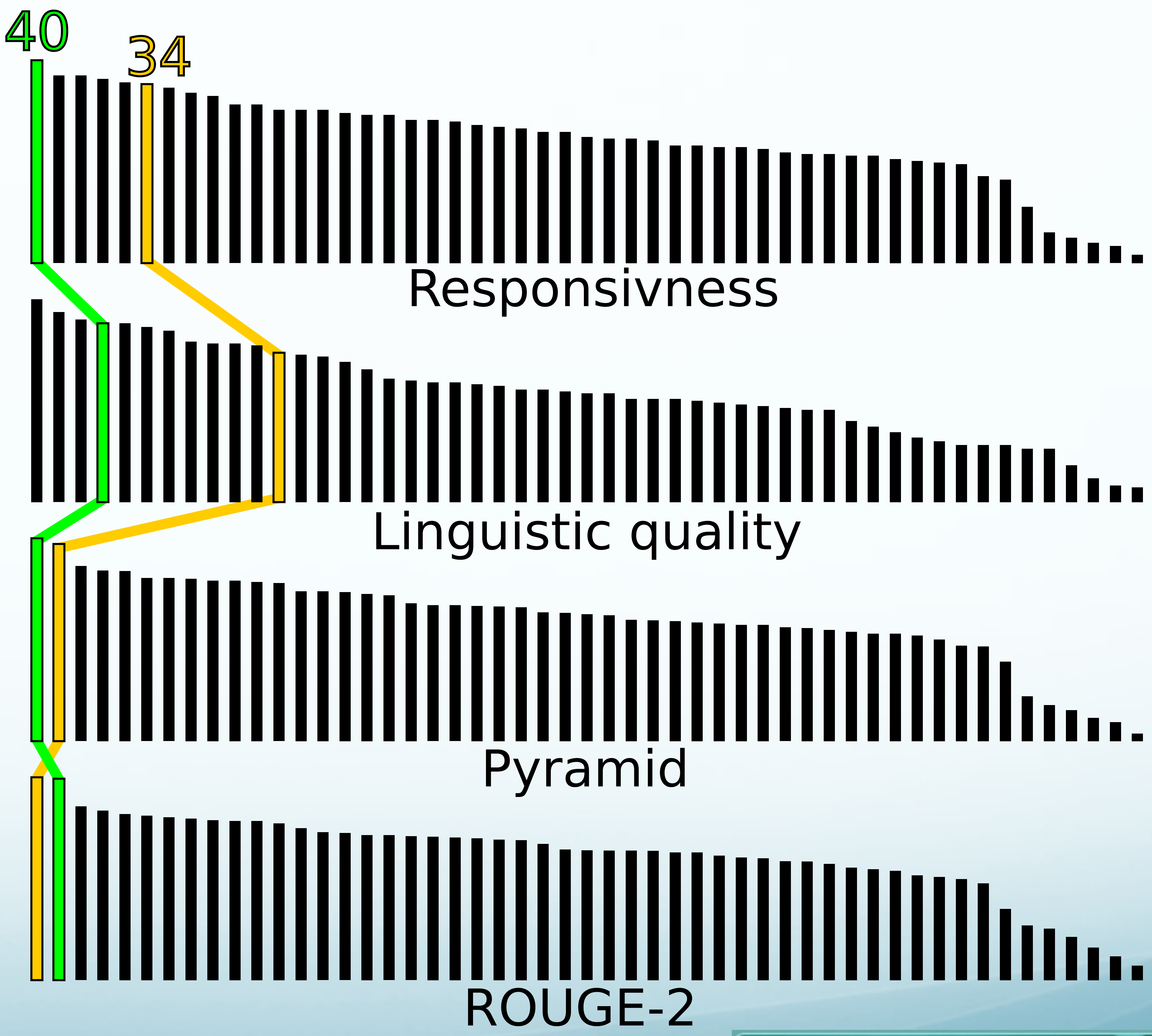
December 19, 2000: Airbus ~~officially~~ launches the plane, calling it the A380.

A ban against bistros providing plastic bags free of charge will be lifted.

December 19, 2000: Airbus launches the plane, calling it the A380.

Results of 2009 System

- 40 = w/Compression
  34 = w/out Compression
- 34% of sentences compressed in final system by some amt.
- More sentences selected
  - (3.9 on avg. vs. 3.8)
- 27% fewer days of week



Results of 2009 System

- Good evidence that compression can help improve multiple factors
  - Allow for inclusion of more sentences → more information
  - Remove extraneous clauses that may impair readability in extractive context

Deep, Extensive, Heuristic

UMD (Zajic et al, 2007) & SumBasic+ (Vanderwende et al, 2007)

Deep, Extensive, Heuristic

UMD (Zajic et al, 2007) & SumBasic+ (Vanderwende et al, 2007)

- Based on output of phrase structure parse
- Goal: Information squeezing, compression to add content

Sentence Compression:
UMD Approaches ([Zajic et al, 2007])

- Two approaches:
  - Syntactic constituent removal — "**Trimmer**" and "**Topiary**"
  - HMM-based Headline Generator — **HMM HEDGE**
    - (Hidden Markov Model HEaDline GEnerator)

UMD Approach #1
Trimmer

- Parse sentences in summary
- Apply ordered cascade of increasingly aggressive rules
  - Adds headline oriented rules
    - (e.g. removing Modals, Determiners)
  - Adds rules to drop large portions of structure
    - halves of AND/OR
    - SBAR/PP deletion

UMD Approach #1
Trimmer

(1) Remove temporal expressions
(2) Select Root S node
(3) Remove preposed adjuncts
(4) Remove some determiners
(5) Remove conjunctions
(6) Remove modal verbs
(7) Remove complementizer *that*
(8) Apply the XP over XP rule

(9) Remove PPs that do not contain NEs
(10) Remove all PPs under SBARs
(11) Remove SBARs
(12) Backtrack to state before step 9
(13) Remove SBARs
(14) Remove PPs that do not contain NEs
(15) Remove all PPs

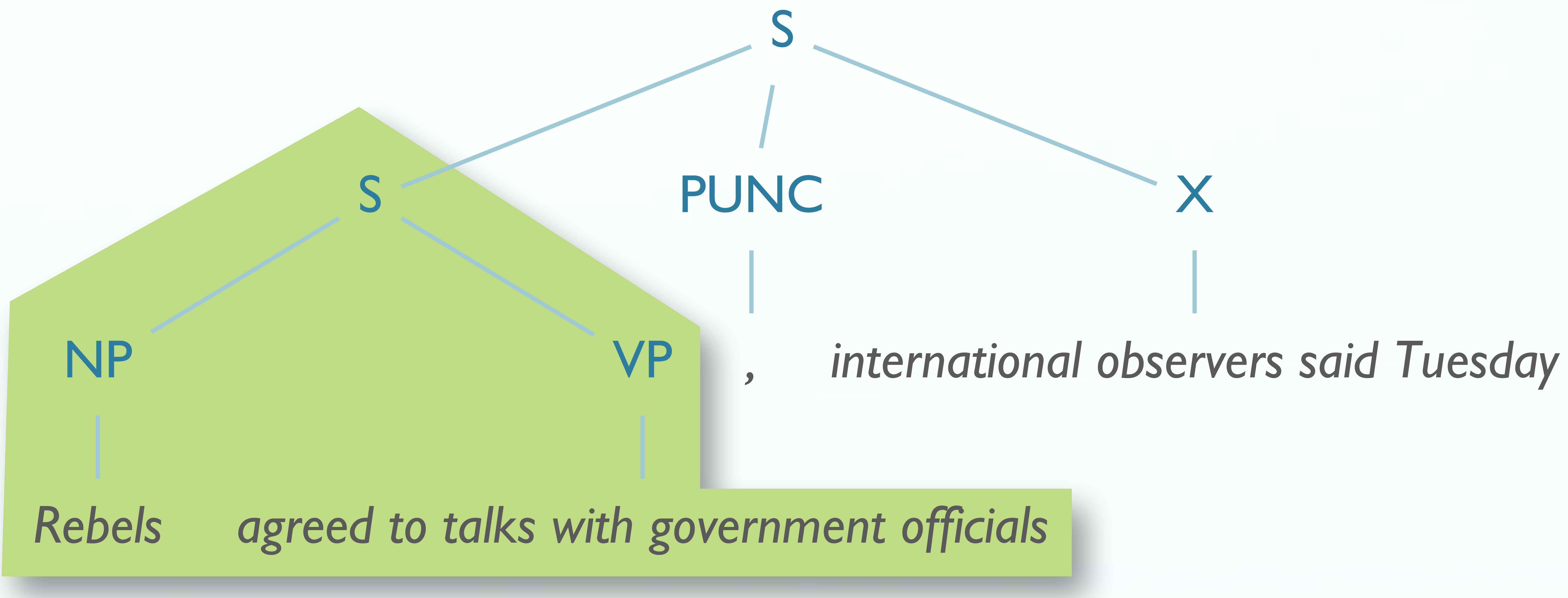
Zajic et. al (2007) – pp. 1553-1554

UMD Approach #1
Trimmer

- Removal of determiners
  - *“[The] Gotti Case Ends With [a] Mistrial for [the] Third Time in a Year”*
  - *“[A] Texas Case Involving Marital Counseling Is [the] Latest to Test [the] Line Between Church and State”*

UMD Approach #1

Trimmer



- Selection of viable "root of compression":
  - First S containing NP VP

UMD Approach #1
Trimmer

- For remaining pruning rules, see paper.
  - *"The remaining steps of the algorithm remove linguistically peripheral material through successive deletions of constituents until the sentence is shorter than a length threshold." p.1555*

UMD Approach #1b
Topiary

- Uses a topic identifier (UTD — Sista et. al, 2002)
  - **TOPICS**: BIN LADEN, EMBASSY, BOMBING, POLICE OFFICIALS, PRISON, HOUSE, FIRE, KABILA
- Use the topics to select compressed candidate:
  - Longest candidate with room to prepend highest-scoring, non-redundant topic terms
- **Example output**:
  - BIN LADEN, EMBASSY, BOMBING: FBI agents this week began questioning relatives.

UMD Approaches:
Adapting to Multi-Documents

- Sentences selected by linear combination of features
- **Static features**:
  - position of sentence in document
  - Relevance of sentence/document
  - Centrality of sentence/document to topic cluster
    - Computed as IDF overlap or (average) Lucene similarity
  - # of compression rules applied

UMD Approaches:
Adapting to Multi-Documents

- **Dynamic Features**
  - Redundancy — how similar candidate sentence is to those already added
  - Sentence-from-doc — # of sentences already taken from same document

UMD Approaches:
Adapting to Multi-Documents

- Redundancy measure:
  - Estimate probability a candidate's words are closer to the existing summary than background corpus (English in general)
  - Simple product of unigram probabilities from different document collections
  - λ is weight to bias estimates (set to 0.3 empirically)
  - D is set of words already in the summary, C is words in background corpus.

$$\text{Redundancy}(S) = \prod_{w_i \in S} \lambda P(w \mid D) + (1-\lambda) P(w \mid C)$$

UM Approaches:
Results

Results on 50 DUC-2006 test topics

	R-1 Recall	R-2 Recall
HMM	0.27311	0.06251
Trimmer	0.29391	0.06718
No Compression	0.27576	0.6126

Zajic et al. (2007) p. 1566, Table 3

Sentence Compression:

SumBasic+ (Vanderwende et al, 2007)

Pattern	Example
Noun appositive	One senior, ~~Liz Parker~~, had slacked off too badly to graduate
Gerundive clause	The Kialegees, ~~numbering about 450~~, are a landless tribe, ~~sharing space in Wetumka, Okla., with the much larger Creek Nation, to whom they are related~~.
Nonrestrictive relative clause	The return to whaling will be a sort of homecoming for the Makah, ~~whose real name which cannot be written in English means "people who live by the rocks and the seagulls"~~
Intra-sentential attribution	~~Seperately, the report said~~ that the murder rate by Indians in 1996 was 4 per 100,000, below the national average of 7.9 per 100,000 and less than the white rate of 4.9 per 100,000
Lead adverbials and conjunctions	

SumBasic+ Results

DUC-2006 results

	R-1 Recall	R-2 Recall
+Compression	0.30753	0.05509
-Compression	0.30026	0.05326

From Vanderwende et al, 2007, p. 1614, Table 4

Sequence-Based Compression

UMD (Zajic et al, 2007) & Wang et al. (2013)

UMD Approach #2

HMM Hedge

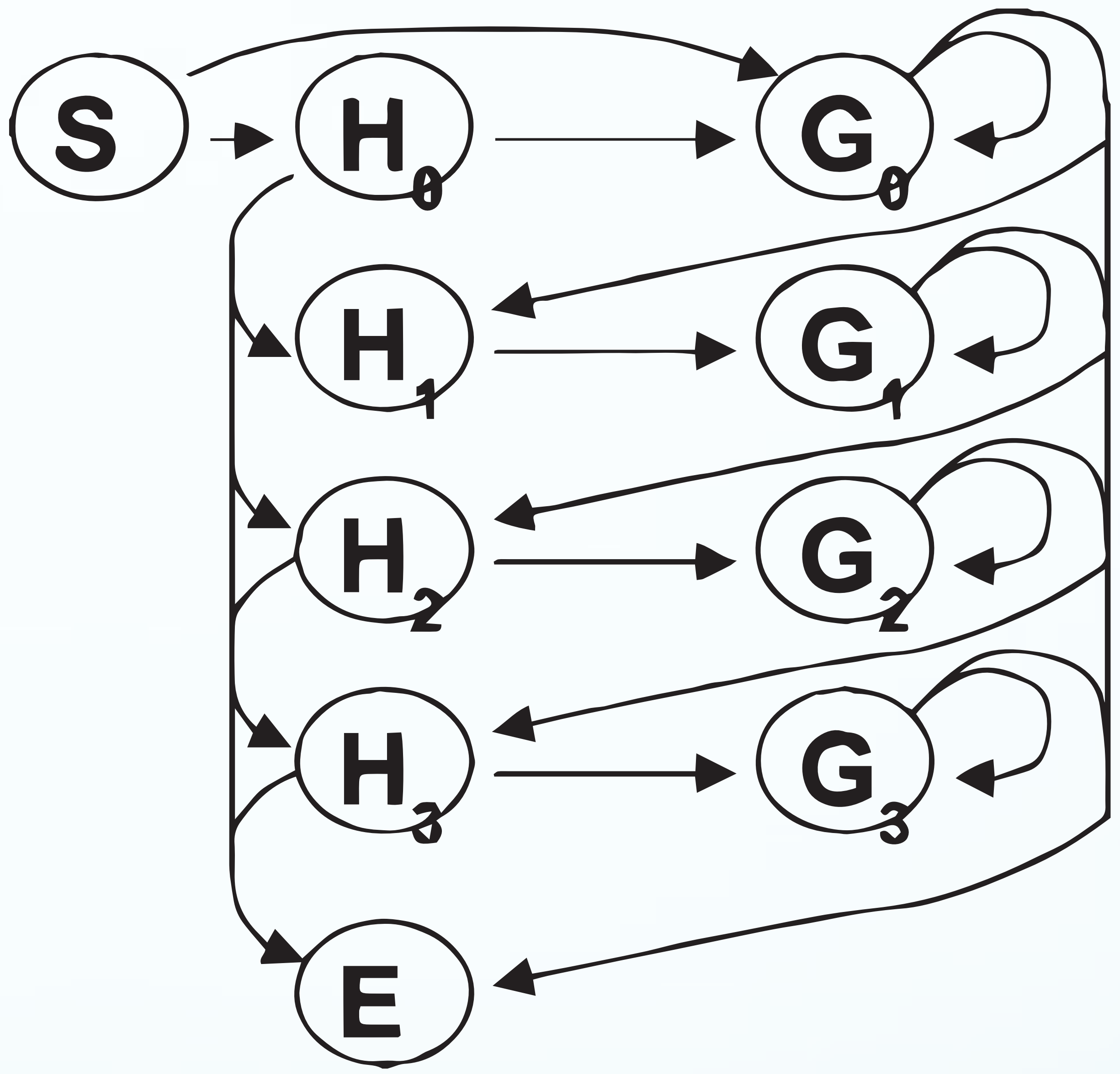
i. ***Pilots not allowed to have guns in cockpits.***

ii. *After months of debate following the September 11 terrorist hijackings, the Transportation Department has decided that airline **pilots** will **not** be **allowed to have guns in** the **cockpits**.*

- "Noisy channel" model:
  - Headline is original sentence.
  - Sentences in article are versions of that original sentence, but with "noise"

UMD Approach #2
HMM Hedge

- From story S that is sequence of N words
  - …find headline H, a subsequence of words from S
  - …that maximizes likelihood that H "generated" the story S $\quad \operatorname{argmax}_H P(H \mid S)$
- **Transition Probabilities**
  - "Headline Word" (H_n) state → Other State (G_n)
  - Modeled on headline collection
- **Emission probabilities:**
  - From H_n: 1.0 → Current Word, else 0



HMM for three-word Story
via Dorr et al. (2003) p. 275

Sequence-based Compression:
Wang et al. (2013)

- Build lienar-chain CRF model for "keep-or-delete" words
- Use {**B**,**I**,**O**} labels
  - **B** = **B**eginning of retention sequence
  - **I** = **I**nside retention sequence
  - **O** = **O**utside retention sequence [remove]

Sequence-based Compression:

Wang et al. (2013) (Table 3, p. 1387)

- **Basic Features**
  - First 1/3/5 tokens
  - last 1/3/5 tokens
  - first letter/all letters capitalized
  - is negation
  - is stopword
- **Dependency Tree Features**
  - relation (deprel)
  - parent/grandparent deprel
  - is root?
  - depth larger than 3/5?
- **Syntactic Features**
  - POS tag
  - parent/grandparent label
  - leftmost child of parent
  - 2nd leftmost child of parent
  - is headword?
  - is NP/VP/ADVP/ADJP chunk?
- **Semantic Features**
  - is predicate?
  - semantic role label
- **Rule-based Features**

Sequence-Based Compression

- So, what is the Wang et. al model trained on, if not headlines?

Learning (Tree-Based) Compression

Wang et al. (2013)

Creating a Compression Corpus:

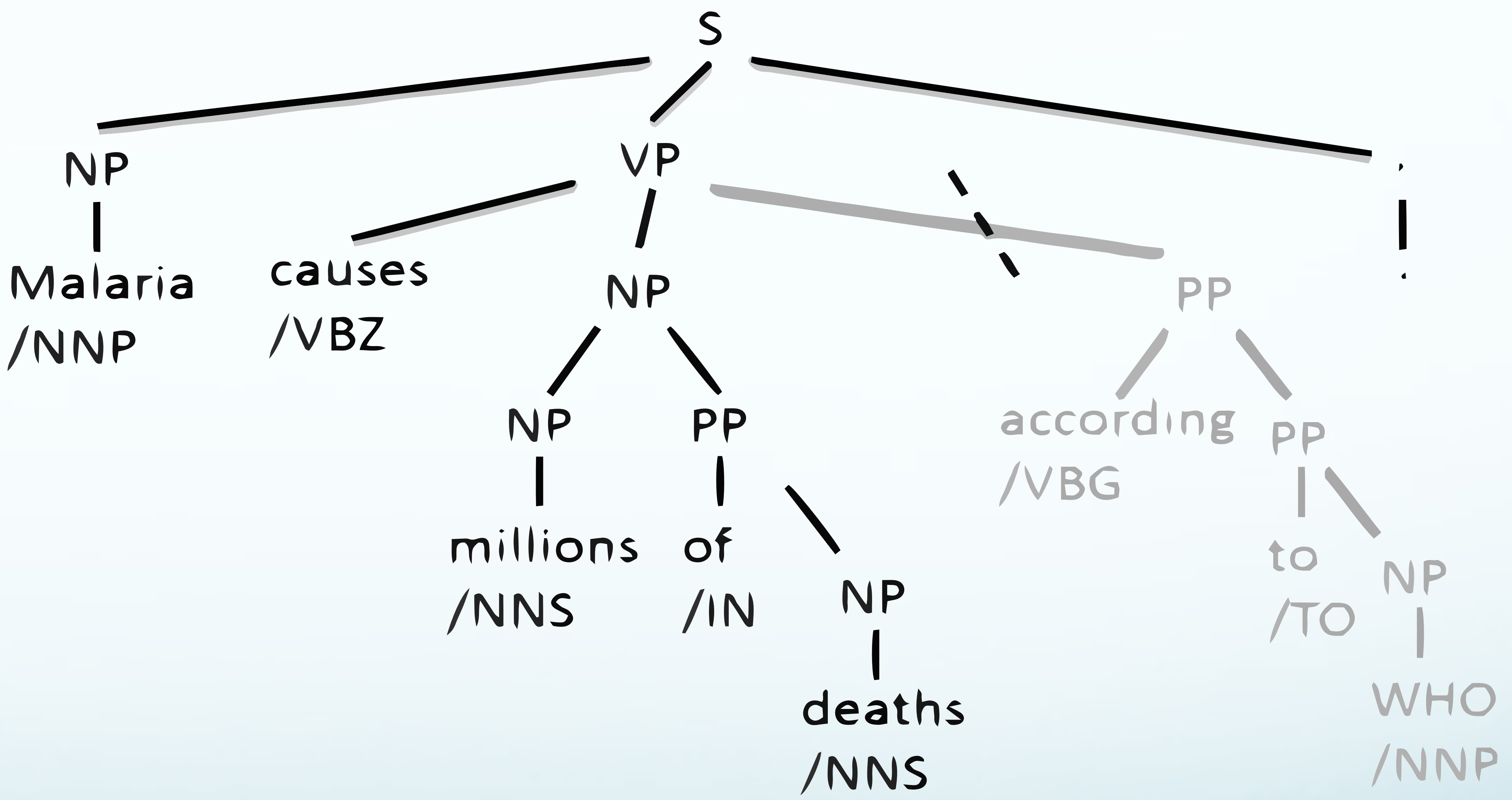
[Clarke & Lapata, 2008]

- Manually created corpus:
  - Written: 82 newswire articles (BNC, ANT)
  - Spoken: 50 stories from HUB-5 broadcast news
- Annotators created compression sentence by sentence
  - Also may mark 'not compressible'
- http://jamesclarke.net/research/resources/
  - On patas: **`compression_corpus/A1G.11`**

Learning Tree-Based Compression

Wang et al. (2013)

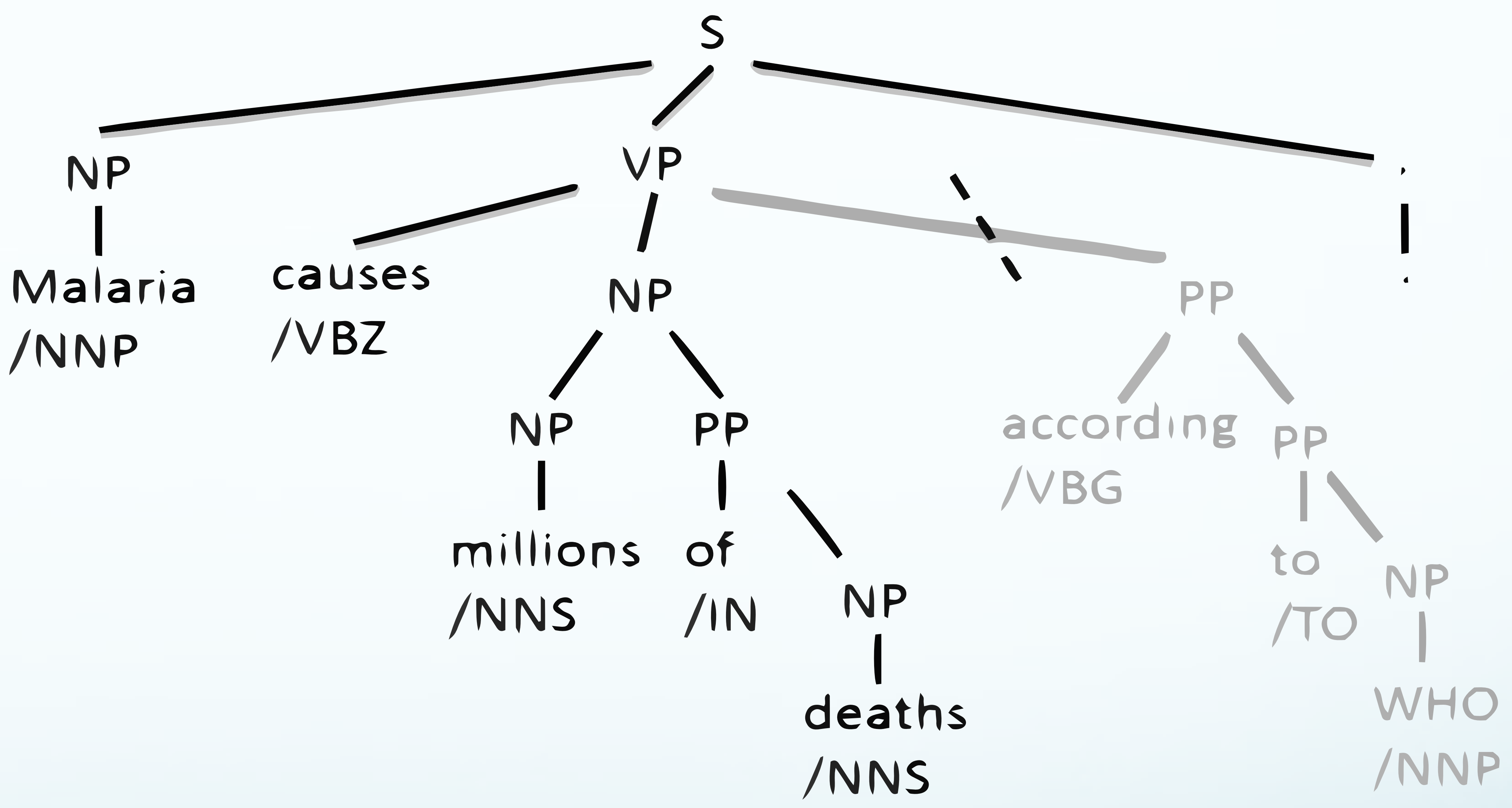
- Using Clarke & Lapata corpus
- …and parses of the parallel sentences
- Determine if each node is:
    - Removed
    - Retained
    - Partial



Learning Tree-Based Compression

Wang et al. (2013)

- Issues:
  - # possible compressions exponential
    - Beam search over candidate labels
  - Need some local way to score node
    - MaxEnt for probability of label
  - Need way to ensure consistency
    - Restrict candidate labels based on context
  - Need way to ensure grammaticality
    - Rerank resulting sentences according to n-gram LM



Tree-based Compression Features:

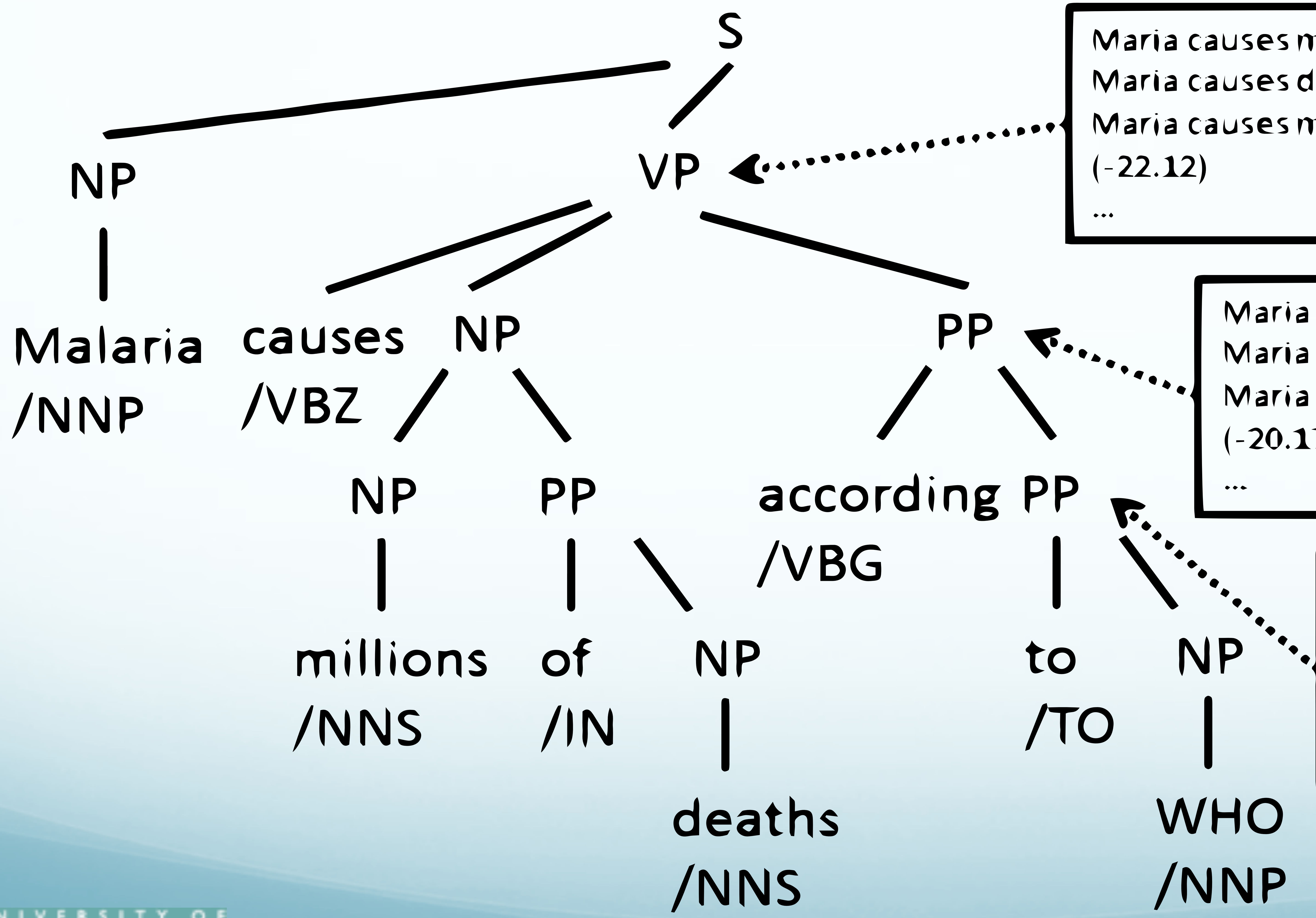
Wang et al. (2013) (Table 4, p. 1389)

- **Basic Features**
  - projection in first/last 1/3/5 toks?
  - subsumes first/last 1/3/5?
  - is leaf node?
  - is root?
  - has capitalization?
  - negation?
  - stopwords?
- **Semantic Features**
  - head node predicate?
  - roles of head node?
- **Syntactic Features**
  - constituent label
  - (grand)parent left/right labels
  - is (second)? leftmost child of parent?
  - is head node of parent?
  - label of head node?
  - Depth > 3? 5? 10?
- **Depdendency Tree Features**
  - deprel of head node
  - deprel of (grand)parent's head node
  - contain root of dep tree?
  - depth larger than 3/5?
- **Rule-based Features**

Feature Breakdown

- Basic Features:
  - Analogous to those for sequence labeling
- Context features:
  - Decisions about child, sibling nodes
- Head-driven search
  - Reorder so head nodes at each level are checked first
  - if head is dropped, don't keep rest.
  - Revise context features for candidates where head is dropped.

Example Beam Search



Maria causes millions of deaths (-21.16)
Maria causes deaths according to WHO (-21.26)
Maria causes millions of deaths according to WHO (-22.12)
…

Maria causes deaths according to WHO (-19.90)
Maria causes millions of deaths (-20.10)
Maria causes millions of deaths according to WHO (-20.17)
…

Maria causes millions of deaths according to WHO (-18.60)
Maria causes deaths according to WHO (-18.92)
Maria causes millions of deaths (-19.09)
…

Final System

- Combined multiple systems
  - Rule-based
  - Sequence model
  - Learned trees
- Used linear combination of:
  - Above systems
  - Language model probability of candidate sentence
  - Redundancy penalty: 1–proportion of overlapping words

Results

on DUC 2006 Data

System	C Rate	R-2	R-SU4
Best DUC	–	0.0956	0.1553
Rule-Based	78.99%	0.1062	0.1573
Sequence	76.34%	0.1049	0.1560
Tree (Basic + $\text{Score}_{\text{Basic}}$)	70.48%	0.1049	0.1586
Tree (Context + $\text{Score}_{\text{Basic}}$)	65.21%	0.1055	0.1610
Tree (Head + $\text{Score}_{\text{Basic}}$)	66.70%	0.1066	0.1618
Tree (Head + Multi)	70.20%	**0.1102**	**0.1625**

Discussion

- Best system incorporates
  - Tree structure
  - Machine learning
  - Summarization features
- Rule-based approach surprisingly competitive
  - Though less aggressive in terms of compression
- Learning based approaches enabled by sentence compression corpus

Alternate views of Summarization

Dimensions of TAC Summarization

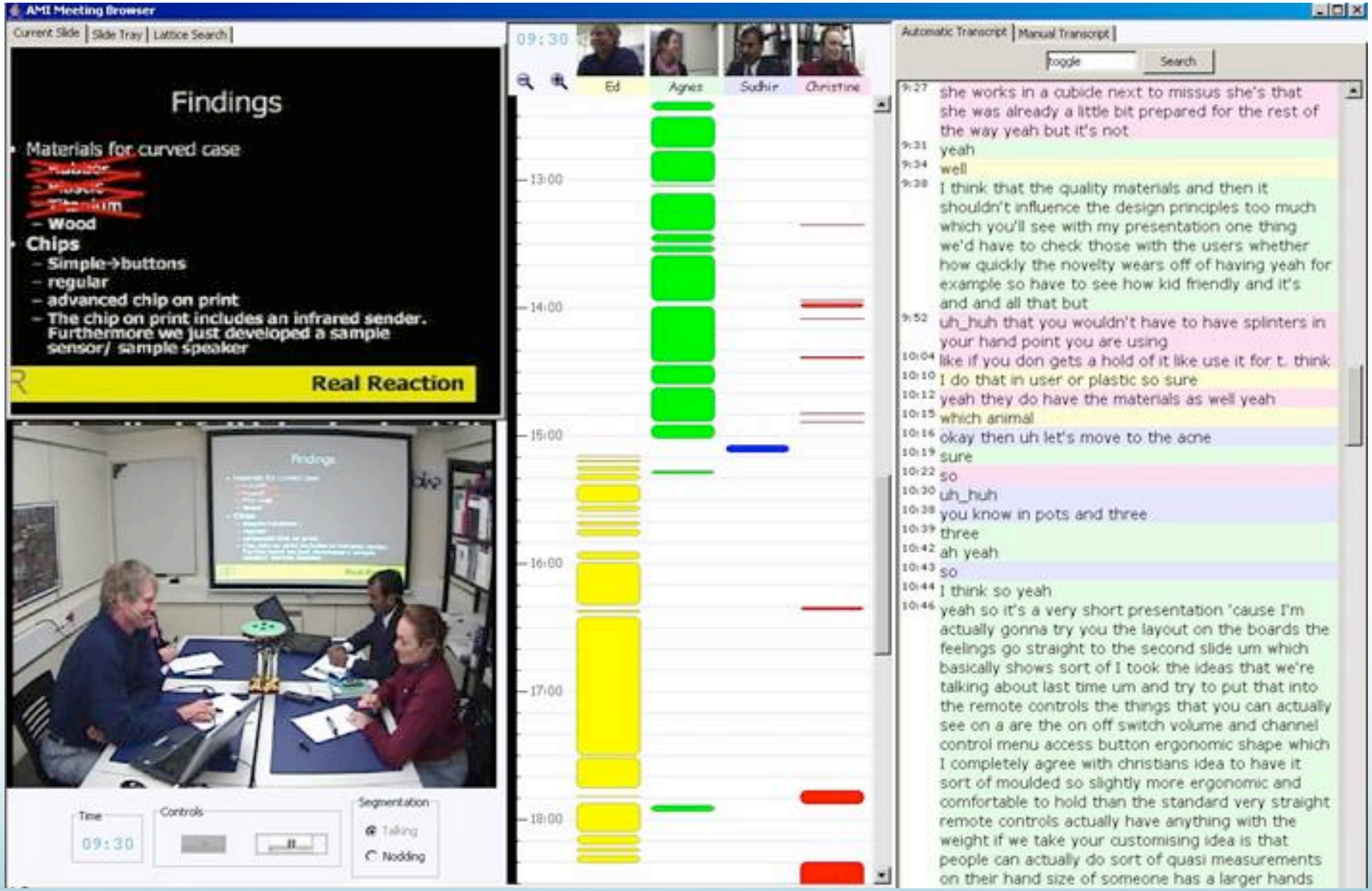
- **Use purpose** — Reflective summaries
- **Audience** — Analysts
- **Derivation (extractive vs. abstractive)** — Largely extractive
- **Coverage (generic vs focused)** — "Guided"
- **Units (single vs. multi)** — Multi-document
- **Reduction** — 100 words

Dimensions of TAC Summarization

- **Input Form Factor**
  - English
  - Newswire text
  - Multiple documents, multiple paragraphs
- **Output Form Factor**
  - English
  - Single paragraph

Other Types of Summaries: Meeting Summaries

- Renals, 2010 & AMI Consortium



Meeting Summaries

- What do you want out of the summary?
  - Minutes?
  - Agenda?
  - To-do list?
    - (Think: automatic sprint management based on team stand-ups!)
  - Points of (dis)agreement?

Dimensions of Meeting Summaries

- Use purpose
  - Catch up on missed meetings
- Audience
  - Ordinary attendees
- Derivation
  - Either abstractive or abstractive

Dimensions of Meeting Summaries

- Coverage (generic vs. focused)
  - Depends on user
    - Team member — "what do ***I*** have to focus on?"
    - PM — How is the team doing as a whole?
- Units (single vs. multi)
  - Single meeting?
  - Recurring problems over project?

Dimensions of Meeting Summaries

- Input/Output form factors
  - Input:
    - Speech
    - …maybe list items, whiteboard diagrams?
  - Output:
    - Lists, bullets, todos
    - 100-200-word summary?

Examples

- Decision summary:
  1. The remote will resemble the potato prototype
  2. There will be no feature to help find the remote when misplaced
  3. Instead, remote will be in bright color
  4. Corporate logo WILL be on the remote
  5. One of the colors will contain the corporate colors
  6. Remote will have six buttons
  7. Buttons will all be one color
  8. The case will be single curve
  9. The case will be rubber
  10. The case will have special color

Examples

- Action items:
  - Each team member receives specific instructions for next meeting by email
  - Team members fill out questionnaire

Examples

- Abstractive summary:
  - When this functional design meeting opens the project manager tells the group about the project restrictions he received from management by email. The marketing expert is first to present, summarizing user requirements data from a questionnaire given to 100 respondents. The marketing expert explains various user preferences and complaints about remotes as well as different interests among age groups. He prefers that they aim users from ages 16-45, improve the most-used functions, and make a placeholder for the remote…