

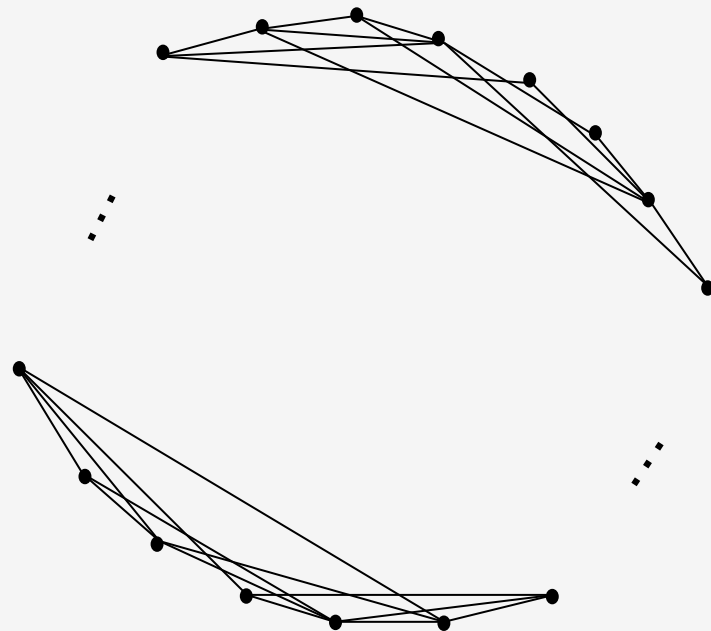
NLP

Introduction to NLP

Summarization Techniques 1/3
(cont'd)

Mitra/Allen/Salton

- Encyclopedia articles
- Semantic hyperlinks
 - Between pairs of paragraphs with lexical similarity above a threshold
 - Paths linking highly connected paragraphs are more likely to contain information central to the topic of the article



NLP

Introduction to NLP

Summarization Techniques 2/3

Mani/Bloedorn (1997, 1999)

- Graph-based method for identifying similarities and differences between documents
- Single event or sequence of events
- Content entities (nodes) and relations (edges)
- Relation types
 - SAME, ADJACENT, ALPHA (WordNet, NetOwl), COREF
- Spreading activation

1.25: "We are clear: the liberation of all our comrades, or we die with all the hostages," a rebel who did not give his name told a local radio station in a telephone call from inside the compound.

...
1.28: Many leaders of the Tupac Amaru which is smaller than Peru's Maoist Shining Path movement are in jail. 1.29: Its chief Victor Polay, was captured in June 1992 and is serving a life sentence, as is his lieutenant, Peter Cardenas.

1.30: Other top commanders conceded defeat and surrendered in July 1993.

Barzilay & M. Elhadad (1997)

- Lexical chains

Mr. Kenny is the person that invented the anesthetic machine which uses micro-computers to control the rate at which an anesthetic is pumped into the blood. Such machines are nothing new. But his device uses two micro-computers to achieve much closer monitoring of the pump feeding the anesthetic into the patient.

Barzilay and M. Elhadad (1997)

- WordNet-based
- Three types of relations:
 - extra-strong (repetitions)
 - strong (WordNet relations)
 - medium-strong (link between synsets is longer than one + some additional constraints)

Barzilay and M. Elhadad (1997)

- Scoring chains:
 - Length
 - Homogeneity index:
= $1 - \# \text{ distinct words in chain}$

$$\text{Score} = \text{Length} * \text{Homogeneity}$$

$$\text{Score} > \text{Average} + 2 * \text{st.dev.}$$

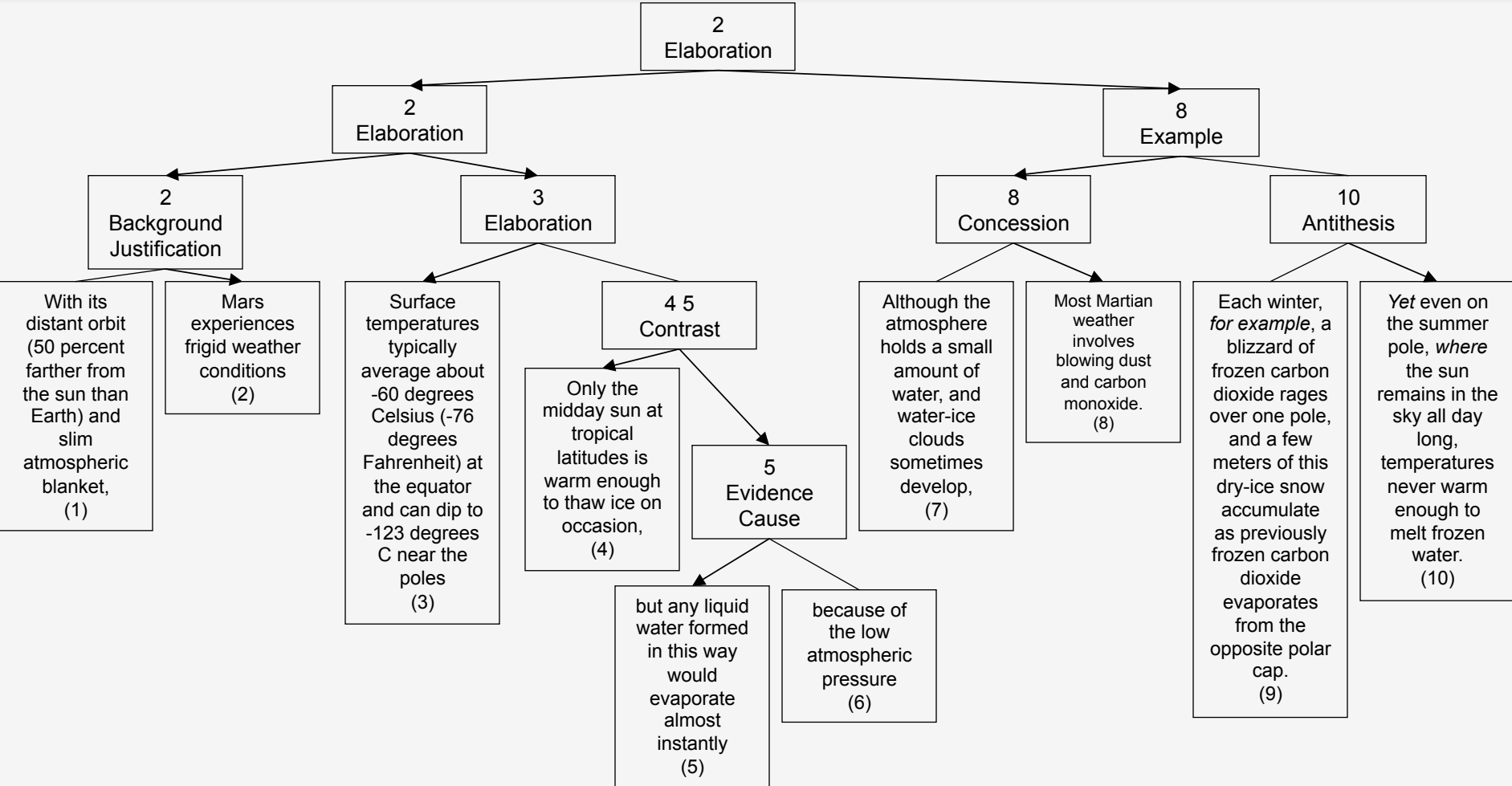
Marcu (1997-1999)

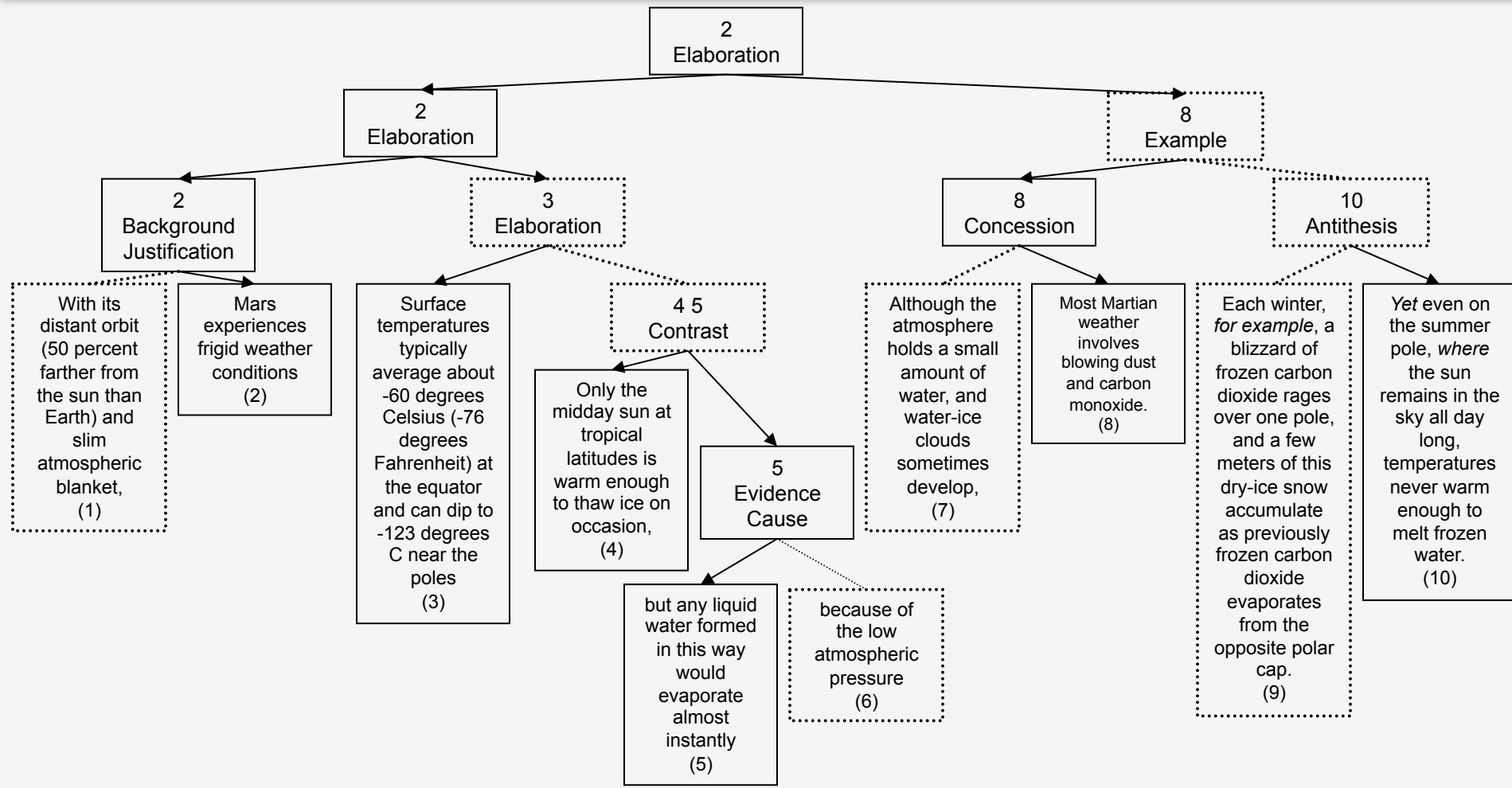
- Focuses on text coherence
- Based on Rhetorical Structure Theory (Mann and Thompson 1988)
 - Nucleus+Satellite
- Example: evidence
 - N+S combination increases reader's belief in N

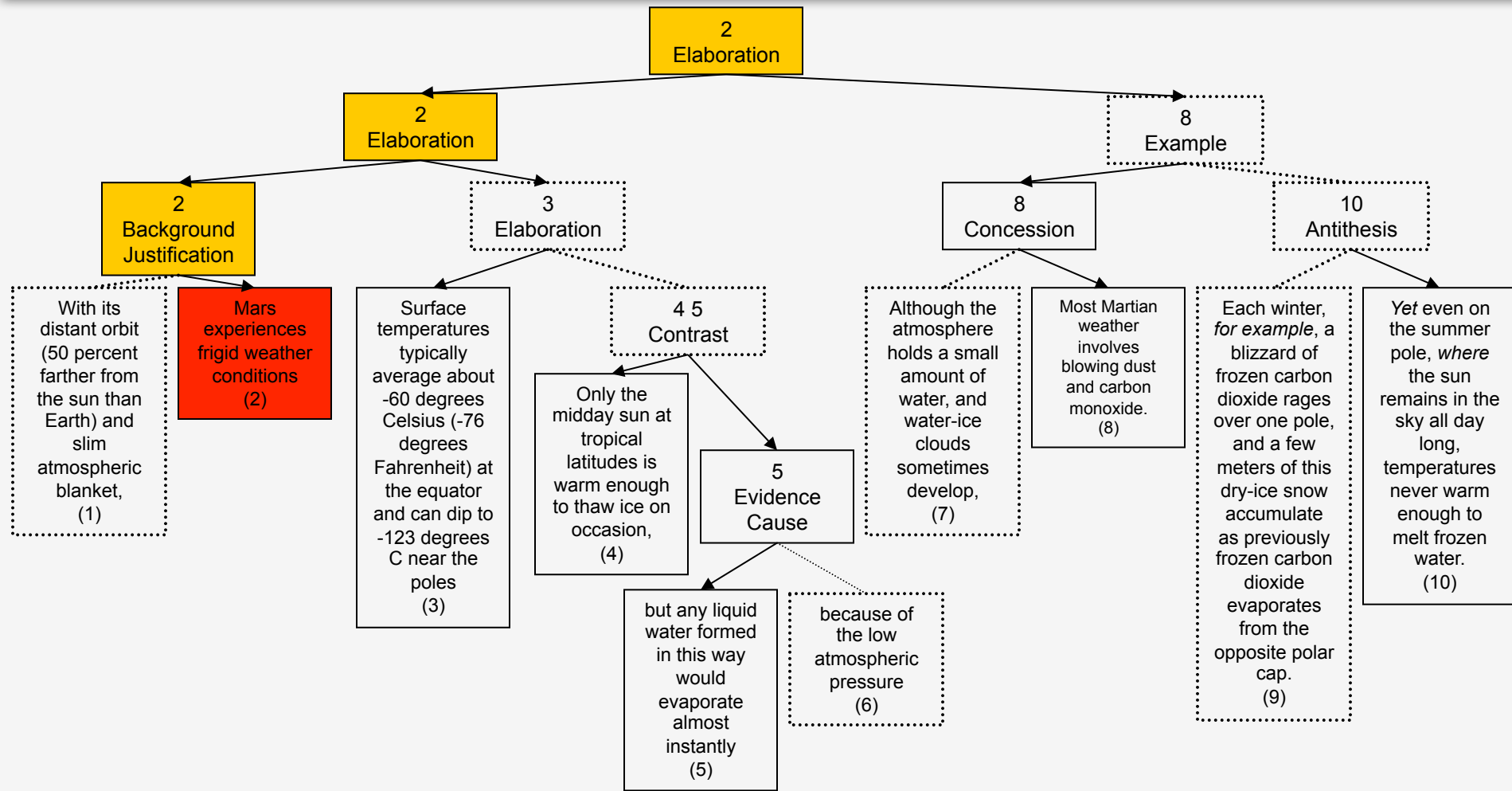
[The truth is that the pressure to smoke in junior high is greater than it will be any other time of one's life:]_N [we know that 3,000 teens start smoking each day.]_S

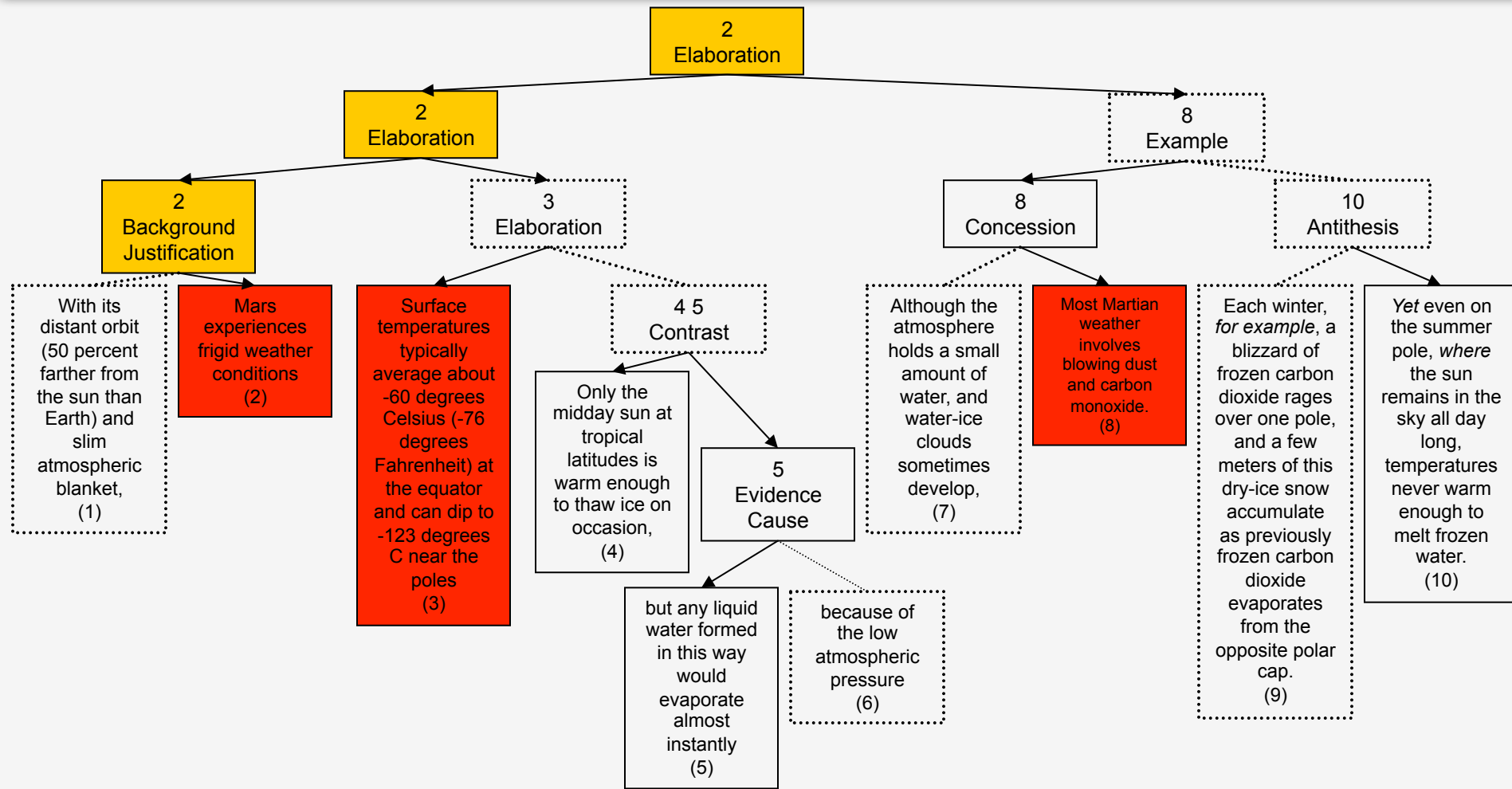


With its distant orbit (50 percent farther from the sun than Earth) and slim atmospheric blanket, (1)	Mars experiences frigid weather conditions (2)	Surface temperatures typically average about -60 degrees Celsius (-76 degrees Fahrenheit) at the equator and can dip to -123 degrees C near the poles (3)	Only the midday sun at tropical latitudes is warm enough to thaw ice on occasion, (4)	but any liquid water formed in this way would evaporate almost instantly (5)	because of the low atmospheric pressure (6)	Although the atmosphere holds a small amount of water, and water-ice clouds sometimes develop, (7)	Most Martian weather involves blowing dust and carbon monoxide. (8)	Each winter, <i>for example</i> , a blizzard of frozen carbon dioxide rages over one pole, and a few meters of this dry-ice snow accumulate as previously frozen carbon dioxide evaporates from the opposite polar cap. (9)	Yet even on the summer pole, <i>where</i> the sun remains in the sky all day long, temperatures never warm enough to melt frozen water. (10)
---	--	---	---	--	---	--	---	---	--



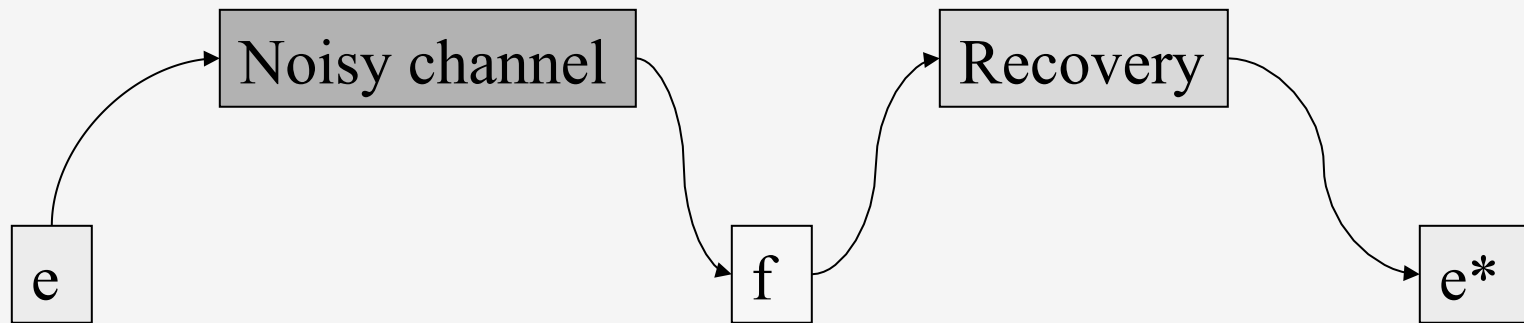






Noisy Channel Models

- Source/target language
- Coding process



Berger and Mittal 2000

- Source language
 - full document
- Target language
 - Summary
- Gisting (OCELOT)

$$g^* = \underset{g}{\operatorname{argmax}} p(g|d) = \underset{g}{\operatorname{argmax}} p(g) \cdot p(d|g)$$

Berger & Mittal 2000

- Training on 100K summary+document pairs
- Testing on 1046 pairs
- Use Viterbi-type search
- Evaluation: word overlap (0.2–0.4)
- No word ordering

Berger & Mittal 2000

Sample output:

Audubon society atlanta area savannah georgia chatham
and local birding savannah keepers chapter of the audubon
georgia and leasing

Carbonell and Goldstein (1998)

- Maximal marginal relevance
- Greedy selection method
- Query-based summaries
- Law of diminishing returns

C = doc collection

Q = user query

R = IR(C,Q,θ)

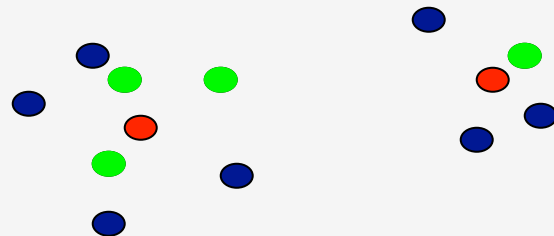
S = already retrieved documents

Sim = similarity metric used

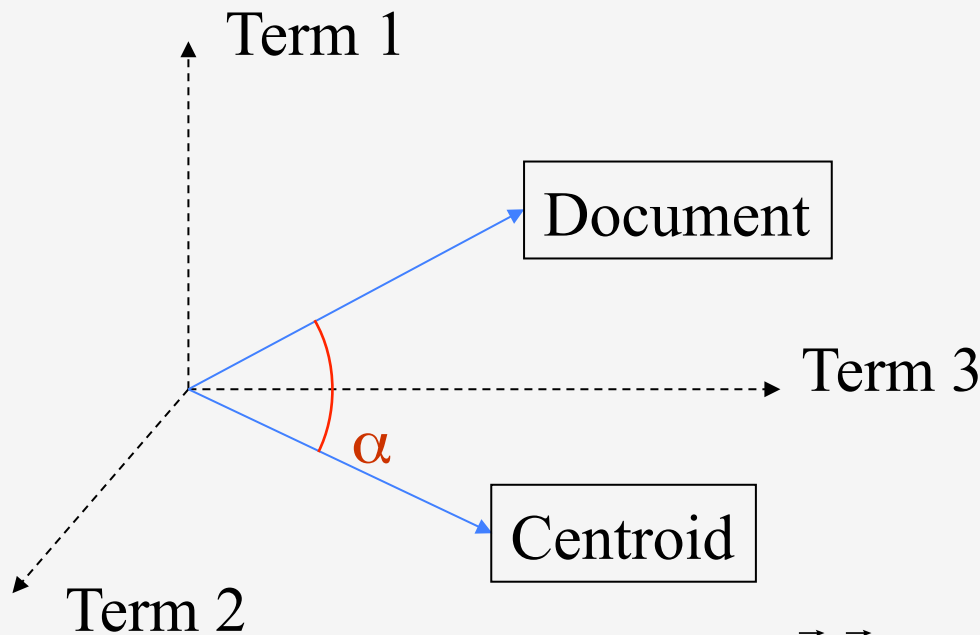
$$\text{MMR} = \underset{D_i \in R \setminus S}{\operatorname{argmax}} \left[\lambda (\text{Sim}_1(D_i, Q)) - (1-\lambda) \max_{D_j \in S} \text{Sim}_2(D_i, D_j) \right]$$

Mead (Radev et al. 2000)

- Salience-based extractive summarization
- In 6 languages
- Centroid-based method
- Single and multi document
- Vector space model
- Features: position, length, centroid
- Cross-document structure theory
- Reranker – similar to MMR
- Open source library (www.summarization.com/mead)



Mead



$$\cos(\vec{x}, \vec{y}) = \frac{\vec{x} \cdot \vec{y}}{|\vec{x}| |\vec{y}|} = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n y_i^2}}$$

Mead

- Input
 - Cluster of d documents with n sentences (compression rate = r)
- Output
 - $(n * r)$ sentences from the cluster with the highest scores

$$SCORE(s) = \sum_i (w_c C_i + w_p P_i + w_f F_i)$$

NewsInEssence (Radev et al. 2001)

- Web-based multi-document news summarization system

NewsInEssence: Web-based News Summarization - Microsoft Internet Explorer provided by AT&T WorldNet Service

Address: http://www.newsinesence.com/nie.cgi

Interactive Multi-source News Summarization

4 Killed In Florida Fireworks Blast July 2, 2003 19:10:48

4 Killed In Florida Fireworks Blast July 2, 2003 19:10:48. BONITA SPRINGS, Fla., July 2, 2003 Investigators and firefighters gather at the scene of a tractor-trailer that exploded as workers were unloading fireworks in Bonita Springs, Fla., Wednesday, July 2, 2003. Kevin McKenzie was mowing a strip of grass at Lover's Key about 300 feet from the tractor trailer when the explosion happened at 2:10 p.m., shooting flames and fireworks from the truck.

[8 Articles from 7 Sources] [4 Summaries]

Recent User Clusters (more)

- 'Liberia's Taylor bans church radio station'**
11 articles, 2 summaries: 07/02, 9:57 PM
- 'Knesset backs Sharon on roadmap'**
7 articles, 3 summaries: 07/01, 11:48 AM
- 'Israel pulls out of Bethlehem'**
5 articles, 4 summaries: 07/01, 11:25 AM

Recent CIDR Clusters (more)

- 'Bush challenge to Iraq attackers: Bring them on'**
25 articles, 2 summaries: 07/02, 7:40 PM
- 'Bill sparks massive Hong Kong protest'**
14 articles, 4 summaries: 07/02, 7:40 PM
- 'Edinburgh Evening News - Top Stories - Palestinian police back in Bethlehem'**
13 articles, 1 summaries: 07/02, 7:40 PM

NewsTroll from URL:
URL must be from CNN, Yahoo!, MSNBC, BBC, or USA Today.
http://www.msnbc.com/

NewsTroll from query:
fireworks bonita flames

Advanced Options

2% Summary

4 Killed In Florida Fireworks Blast July 2, 2003 19:10:48 (4:1) BONITA SPRINGS, Fla., July 2, 2003 Investigators and firefighters gather at the scene of a tractor-trailer that exploded as workers were unloading fireworks in Bonita Springs, Fla., Wednesday, July 2, 2003. (4:2)

NewsInEssence: Web-based News Summarization - Microsoft Internet Explorer provided by AT&T WorldNet Service

Address: http://www.newsinesence.com/nie.cgi?CID=20020830135218

Interactive Multi-source News Summarization

Swede charged with plans to hijack plane

An organizer of an Islamic conference in Birmingham says it is "highly unlikely" that the man charged with trying to hijack a Ryanair flight in Sweden, was due to attend the conference. Security officers at Stockholm's Vasteras Airport say they found a handgun in a toiletries bag when they scanned the 29-year-old man's hand luggage.

FULL SUMMARY

Create A New Cluster

Recent User Clusters (more)

- 'Attacks expose Afghanistan vulnerability'**
3 articles, 1 summary: 09/10, 4:41 PM

Recent CIDR Clusters (more)

- 'Bush Continues to lobby foreign leaders against Iraq'**
29 articles, 4 summaries: 09/10, 8:01 PM

Recent Google Headlines (more)

- 'Fatah says it will prevent attacks on Israeli civilians'**
5 articles, 4 summaries: 09/11, 12:00 AM

User Clusters (Archive) (Create Your Own Cluster)

- 'Attacks expose Afghanistan vulnerability'**
3 articles, 1 summary: 09/10, 4:41 PM
- 'Microsoft hit by hacker attack'**
3 articles, 3 summaries: 09/09, 9:23 PM
- 'Confidence crisis in Canadian ice hockey'**
3 articles: 09/09, 9:06 PM
- 'Ararat: Palestinians on verge of statehood'**
- 'Disarm Iraq, Bush to urge U.N.'**
11 articles, 5 summaries: 09/07, 2:09 AM
- 'Cakewalk to Kabul'**
10 articles, 2 summaries: 09/06, 7:41 PM
- 'Cakewalk to Kabul'**
11 articles: 09/06, 2:39 PM
- 'Karzi attacker from Taliban area'**

Other News Summarization Systems

- Newsblaster (McKeown et al. 2002)
- Google News

Columbia Newsblaster: Summarizing All the News on the Web - Microsoft Internet Explorer provided by AT&T WorldNet Service

Monday, June 30, 2003
Last update: 7:28 AM EST

Columbia Newsblaster
Summarizing all the news on the Web

U.S. World Finance Sci/Tech Entertainment Sports

View Today's Images

View Archives

About Newsblaster

Newsblaster in Press

Academic Papers

Current Sources:
nytimes.com (123 articles)
suntimes.com (63 articles)
dallasnews.com (65 articles)
boston.com

Palestinian militants declare truce; Israel starts pulling troops out of Gaza (33 articles)
JERUSALEM - Indignity in Yasser Arafat's Fatah faction delayed an expected joint truce announcement Sunday by Palestinian militias, officials said, as President George Bush's top foreign policy adviser held talks to promote the U.S.-backed "road map" for peace. Earlier today, The Associated Press reported that the three militant factions had signed an agreement to halt attacks on Israelis, as called for in the beleaguered U.S. "road map" to peace in the Middle East. Seeking to advance a US-backed peace plan, U.S. National Security Adviser Condoleezza Rice held talks with the Palestinian prime minister yesterday, a day ahead of an expected truce announcement by Palestinian militants. The Bush administration welcomed, with reservations, cease-fires declared Sunday by Palestinian militant groups in their long fight against Israel. JERUSALEM Israel agreed Friday to cede security control of much of the Gaza Strip to the Palestinians, and a key leader of the radical Islamic group Hamas said there soon would be a three-month halt to attacks on Israeli targets. As Palestinian officials said they were finishing the details of a three-month suspension of all violence against Israelis, Palestinian militant groups increased their attacks on Thursday, and Israeli forces killed three militants here today.

Once Hailed, Soldiers in Iraq Now Feel Blame at Each Step (28 articles)
The U.S. military has launched a new operation in Iraq in an attempt to crush surviving Saddam Hussein loyalists and "paramilitary elements" believed to be behind a series of lethal attacks, officials said Sunday. American forces carried out an

Top News

Now that court says sodomy's OK, all limits are off (13 articles)
For 15 years, the majority has held sway as the court elevated states' rights, narrowed federal review of state criminal cases and issued the ruling that put George W. Bush in the White House. In a landmark 6-to-3 decision

Google News - Microsoft Internet Explorer provided by AT&T WorldNet Service

Google named best News service by Webby Awards

Top Stories

World
U.S.
Business
Sci/Tech
Sports
Entertainment
Health

Top Stories

Palestinians Resume Control of Bethlehem
Washington Post - 40 minutes ago
BETHLEHEM, West Bank, July 2 -- Israeli troops pulled out of this biblical West Bank town today and turned over control to Palestinian security forces, who raised flags and patrolled the city's historic Manger Square. The hand-over was the latest step ...
US Praises Bethlehem Handover by Israel Reuters
Israel releases eight Palestinian prisoners SABC News
Forward - Guardian - Christian Science Monitor - International Herald Tribune - and 2037 related >

Pentagon readies plans for peace mission in Liberia
Minneapolis Star Tribune - 1 hour ago
WASHINGTON, DC -- The Pentagon has ordered military planners to prepare detailed options for US troops to join an international peace-keeping force in Liberia, two senior military officials said Wednesday.
Bush May Send Troops To Liberia WCCO
US to send troops to Liberia Guardian
CNN - Men's News Daily - CBC News - ABC News - and 758 related >

WorldCom Offers New Stock With Settlement
Washington Post - and 83 related >

3 Top Shuttle Managers Replaced
Washington Post - and 306 related >

Attorney says unconfirmed statements about missing Baylor player feeding media frenzy
News 24 Houston - and 1502 related >

Schwarzenegger takes "T3" act to troops in Gulf
Boston.com - and 247 related >

Food plant maps steps to fight rising obesity
International Herald Tribune - and 108 related >

In the News
New Jersey
Homeland Security
Los Angeles
North Carolina
US Supreme Court
San Diego
Defence Secretary
Civil War
Coca Cola
West Bank

World >

Kenyan Women Accuse British Troops of Rape
Times - 3 hours ago
LONDON, July 2 A group of 60 Kenyan women accuse British

U.S. >

Low-Profile Suburb Braces for Onslaught
Washington Post - 40 minutes ago
PHILADELPHIA, Pa., July 2 At East's Exeter County, Va.

NLP