

# Text Generation for Abstractive Summarization

## A New Approach for Guided Summarization

Pierre-Etienne Genest  
Guy Lapalme

RALI, Université de Montréal

November 15, 2010

# Motivation

- TAC 2010 Guided Summarization Track
  - ▶ *The goal of guided summarization is to encourage a deeper linguistic (semantic) analysis of the source documents instead of relying only on document word frequencies to select important concepts.*
- Our 2009 HexTac Experiment
  - ▶ Showed an empirical limit to the performance of pure extraction
  - ▶ Our conclusion : stay away from extraction, i.e. try abstraction
- Novelty !
  - ▶ A lot has been tried with extraction
  - ▶ We feel that it is time to explore a new direction

# Abstractive Summarization

## 1 Representation of source documents

- ▶ Extract smallest coherent pieces of information found in sentences
- ▶ We call them **Information Items (InIts)**
- ▶ InIts can be a property, a statement, an action, etc.

## 2 Content Selection

- ▶ Pick InIts that answer specific aspects
- ▶ Rank the InIts according to salience

## 3 Summary Generation

- ▶ **Text Planning** : order InIts in a logical way, join them together
- ▶ **Text Generation** : generate sentences to realize the planned output

# 2010 System

- This year's focus : **Text Generation**
  - ▶ Main difference between abstraction and extraction
  - ▶ Main reason for being considered less desirable
  - ▶ And a great research area
  
- Our approach
  - ▶ Relies on parse trees to build Inlts
  - ▶ Uses a NLG Realizer (SimpleNLG)
  - ▶ Does not require generation patterns
  - ▶ Performs some text planning

# 2010 System

- Abstract Representation
  - ▶ Information Items defined as Subject-Verb-Object (SVO) triples
  - ▶ Extracted from each verb in the parse trees
  - ▶ Inlts also have a date and a location
- Sentence Generation
  - ▶ Translates Inlts into NLG format
  - ▶ Rule-based
  - ▶ Uses original context in parse tree
  - ▶ Includes some types of modifiers and complements
  - ▶ Results in short, to-the-point sentences

# Information Item Example 1

## Original Sentence (D1025E-A)

- ▶ At least 25 bears died in the greater Yellowstone area last year, including eight breeding-age females killed by people.

## Information Items

- 1 bear – die – *null* (greater Yellowstone area, last year)
- 2 person – kill – female (greater Yellowstone area, last year)

## Generated Sentences

- 1 25 bears died.
- 2 Some people killed eight breeding-age females.

## Information Item Example 2

### Original Sentence (D1012C-B)

- ▶ The Cypriot airliner that crashed in Greece may have suffered a sudden loss of cabin pressure at high altitude, causing temperatures and oxygen levels to plummet and leaving everyone aboard suffocating and freezing to death, experts said Monday.

### Information Items

- 1 airliner – crash – *null* (Greece, August 15, 2005)
- 2 airliner – suffer – loss (Greece, August 15, 2005)
- 3 loss – cause – *null* (Greece, August 15, 2005)
- 4 loss – leave – *null* (Greece, August 15, 2005)

### Generated Sentences

- 1 A Cypriot airliner crashed.
- 2 A Cypriot airliner may have suffered a sudden loss of cabin pressure at high altitude.
- 3 A sudden loss of cabin pressure at high altitude caused temperatures and oxygen levels to plummet.
- 4 A sudden loss of cabin pressure at high altitude left everyone aboard suffocating and freezing to death.

### Generated Sentence in the Summary

- ▶ On August 15, 2005, a Cypriot airliner may have suffered a sudden loss of cabin pressure at high altitude in Greece.

## 2010 System (cont.)

- Content Selection

- ▶ Ranking of the generated sentences
- ▶ Based only on average document frequency of unique lemmas
- ▶ Run1 : From the generated sentences
- ▶ Run2 : From the original sentences
- ▶ No use of categories and aspects for now

- Summary Generation

- ▶ Order sentences by known or estimated date
- ▶ Categories 1 and 2 : generate date and location modifiers



## Generated Summary

Cluster : D1004A-A

Title : Papua Tsunami

Category : 1. Accidents and Natural Disasters

---

On July 12, 1998, the death toll could rise. An area was spotted with villages in Port Moresby. Earthquake struck the Admiralty Islands. Government station, catholic mission station and the Nimas village in the Sissano area west were destroyed in Aitape. PORT MORESBY was spotted with villages. Some people were confirmed dead. **Jungle materials were built on beaches.** **Tsunami disaster climbed to 599 in CANBERRA.** A 10 meter tsunami engulfed heavily populated villages. Queen Elizabeth II sent a message in Papua New Guinea. Prime Minister Tim Fischer described the events in PNG.

---

Linguistic Quality : 2/5

Overall Responsiveness 3/5

# Pros and Cons

## Pros

- Many short sentences : broad content coverage
- Lower global impact of low-information sentences

## Cons

- Severe linguistic problems, sometimes unavoidable (bad parse)
  - Possibility of unfocused summary
- \* General consideration : better content selection needed

## Results (part A)

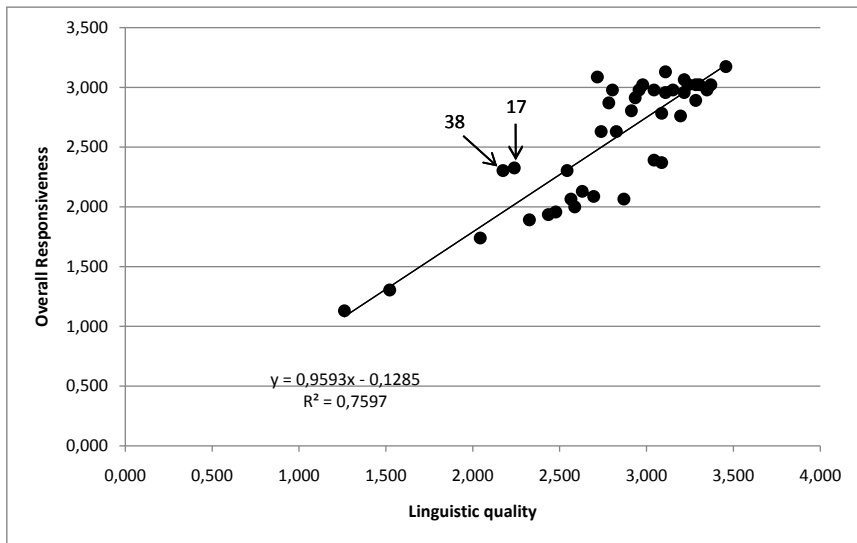
### Scores

	Pyr.	Ling. Q.	Overall R.
Rali1	0.315	2.174	2.304
Rali2	0.282	2.239	2.326
Avg	0.309	2.820	2.576
Best	0.425	3.457	3.174
Models	0.785	4.910	4.760

### Ranks

	Pyr.	Ling. Q.	Overall R.
Rali1	25	40	30
Rali2	29	39	29

## Results (part A)



# Future Work

- Better Information Item definition
  - ▶ Move away from lemmas and towards semantic units
  - ▶ And closer to a predicate-logic semantic representation
  - ▶ Allow for deductions/inferences on InIts
- Content Selection
  - ▶ Rank InIts instead of already-made sentences
  - ▶ Attempt to answer aspects directly
- Text Generation
  - ▶ Improve on the current methodology
  - ▶ Combine InIts in the same generated sentence
- Summary Generation/Text Planning
  - ▶ Exploit the additional knowledge for planning
  - ▶ Order sentences from most generic to least, or by sub-topic

Thank you

Questions ? Comments ?