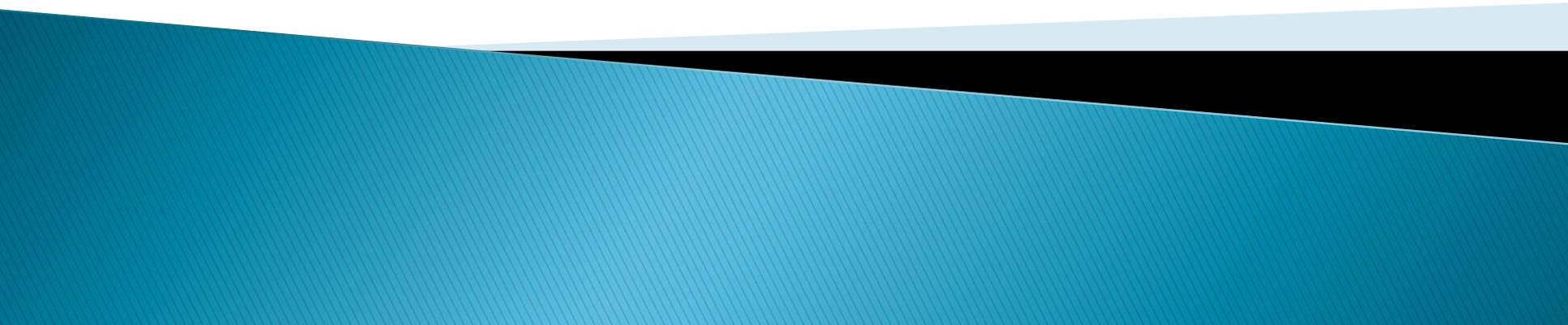


# Résumé automatique de textes

Approche par abstraction



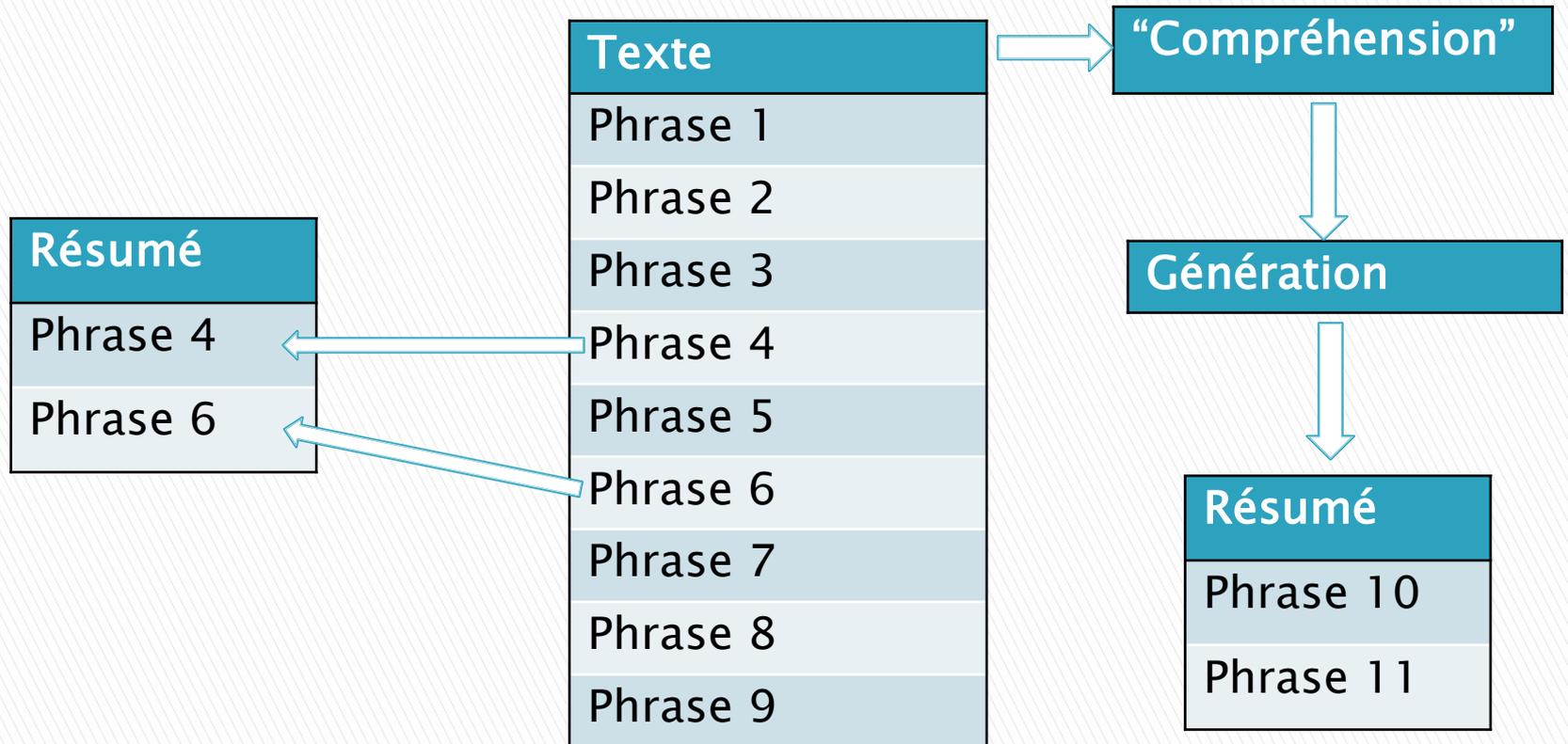
# Enjeux et difficultés

- ▶ Obtenir un résumé clair, concis et non-redondant
  - ▶ Résumer plusieurs documents
  - ▶ Résumer plusieurs opinions au sein d'un même document
- 

# Plan

- I. L'approche par abstraction
  - II. Travaux
  - III. Opinosis : une approche basée sur les graphes
- 

# I. L'approche par abstraction



Extraction

Abstraction

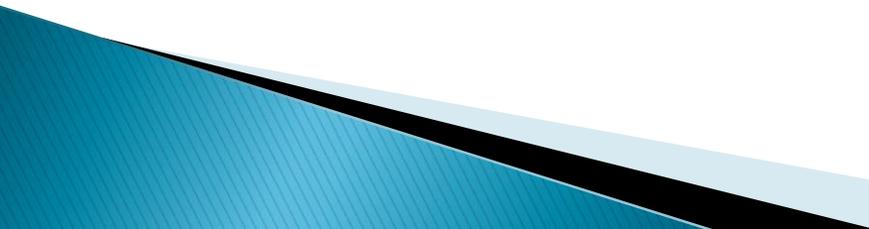
# I. Avantages :

- ▶ Processus similaire au raisonnement d'un humain, donc résultat de meilleure qualité
  - ▶ Adapté au résumé multi-documents
  - ▶ Fonctionne même si le corpus contient plusieurs opinions
- 

# I. Inconvénients

- ▶ Difficile à mettre en place
  - ▶ Difficile à automatiser
  - ▶ Fortement dépendant du contexte
  - ▶ Peu de travaux ont été faits
- 

# II. Travaux

- ▶ FRUMP (*DeJong G., 1982*)
  - ▶ SUMMONS (*Radev D R and McKeown K R., 1998*)
  - ▶ UNL Document Summarization (*Virach S, Potipiti T and Charoenporn T., 2001*)
  - ▶ GISTEXTER (*Harabagiu, S., Lacatusu, F., 2002*)
  - ▶ OPINOSIS (*Ganesan, K. A., C. X. Zhai, and J. Han, 2010*)
- 

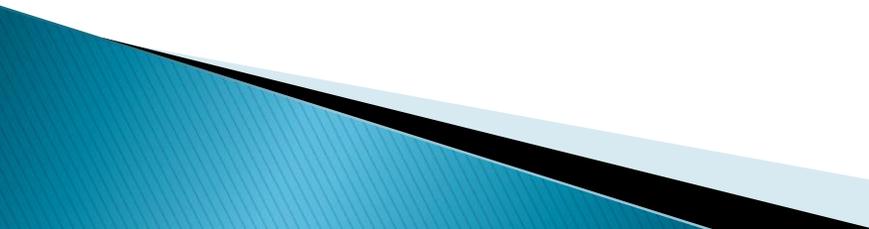
# II. Travaux

- ▶ ... et c'est plus ou moins tout.
- ▶ Autres travaux : plus théoriques, comparatifs, ou restreints à un type de document.

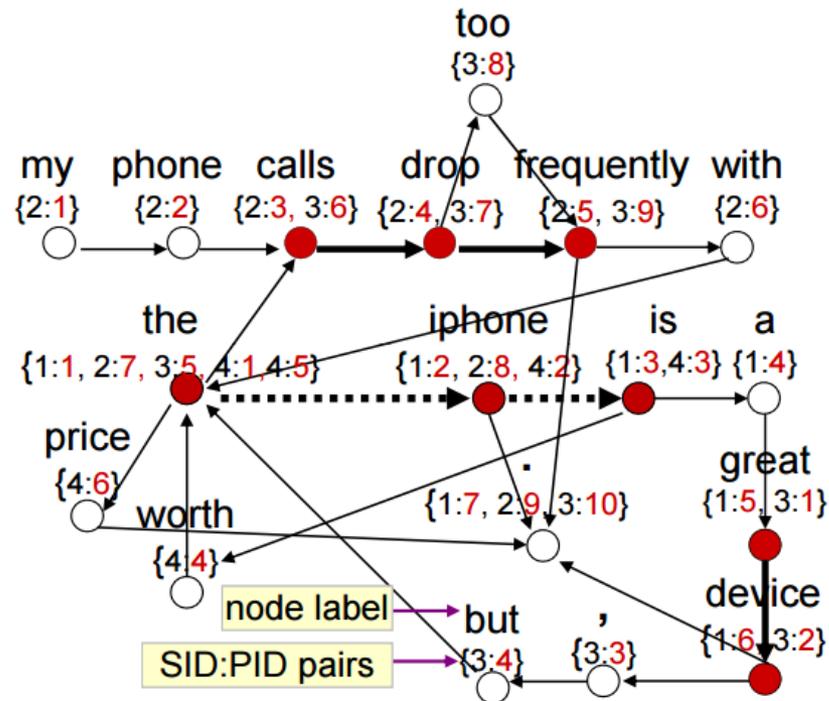
# III. Opinosis

- ▶ Basé sur un graphe représentant la structure des phrases
  - ▶ Sert à résumer des opinions hautement redondantes
  - ▶ Indépendant du domaine, et en partie de la langue
- 

# III. Étape 1 : génération du graphe

- ▶ Entrée = un texte annoté (POS)
  - ▶ Nœud = mot annoté, arêtes (orientées) = adjacence dans une phrase
  - ▶ Les nœuds contiennent une trace de la phrase d'origine et de la position dans celle-ci
- 

# III. Étape 1 : génération du graphe



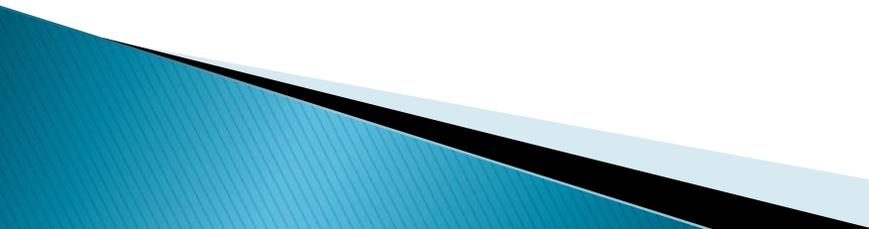
## Input:

SID:1. The iPhone is a great device.  
 SID:2. My phone calls drop frequently with the iPhone.  
 SID:3. Great device, but the calls drop too frequently.  
 SID:4. The iPhone is worth the price.

# III. Étape 1 : propriétés du graphe

- ▶ La redondance est capturée naturellement
  - ▶ Les variations mineures sont ignorées et la structure essentielle apparaît
  - ▶ Des nœuds « hub » apparaissent
- 

# III. Étape 2 : recherche de chemins

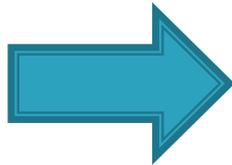
- ▶ Recherche de chemin valide dans le graphe
  - ▶ Chemin valide = qui commence par un mot qui a tendance à se trouver au début des phrases, et finit par un point
  - ▶ Plus quelques contraintes syntaxiques supplémentaires
- 

# III. Étape 2 : recherche de chemins

- ▶ Attribution d'un score à chaque chemin
  - ▶ On cherche des chemins qui représentent un maximum d'opinions redondantes
  - ▶ Certains chemins sont réductibles
- 

# III. Étape 3 : réduction des chemins

Phrases d'origine
The sound quality is really good
The sound quality is clear
Sound quality is great
The iPhone is great
The iPhone is too expensive



Phrases résultantes
The sound quality is really good, clear and great
The iPhone is great but too expensive

# III. Étape 4 : génération du résumé

- ▶ Choix des phrases ayant le meilleur score
  - ▶ Concaténation de ces phrases
- 

# III. Évaluation de la méthode

- ▶ Utilisation de la métrique ROUGE
  - ▶ Comparaison entre Opinosis, une méthode par extraction (MEAD) et des résumés écrits par des humains
- 

# III. Évaluation de la méthode

## *"About food at Holiday Inn, London"*

### **Human summaries:**

[1] Food was excellent with a wide range of choices and good services.

[2] The food is good, the service great. Very good selection of food for breakfast buffet.

### **Opinosis abstractive summary:**

The food was excellent, good and delicious. Very good selection of food.

### **Baseline extractive summary:**

Within 200 yards of leaving the hotel and heading to the Tube Station you have a number of fast food outlets, highstreet Restaurants, Pastry shops and supermarkets, so if you did wish to live in your hotel room for the duration of your stay, you could do.....

## *"What is free at Bestwestern Inn, San Francisco"*

### **Human summaries:**

[1] There is free WiFi internet access available in all the rooms.. From 5-6 p.m. there is free wine tasting and appetizers available to all the guests.

[2] Evening wine reception and free coffee in the morning. Free internet, free parking and free massage.

### **Opinosis abstractive summary:**

Free wine reception in evening. Free coffee and biscotti and wine.

### **Baseline extractive summary:**

The free wine and nibbles served between 5pm and 6pm were a lovely touch. There's free coffee, teas at breakfast time with little biscotti and, best of all, from 5 till 6pm you get a free wine 'tasting' reception which, as long as you don't take.....

# Conclusion

- ▶ Abstraction = beaucoup plus efficace, mais difficile
  - ▶ Opinois en est la preuve
  - ▶ Idée : utiliser une méthode par extraction pour générer un résumé hautement redondant, puis le « purifier » avec Opinois
- 