Analyse de texte / du discours

Brigitte Grau
Master 1 DAC - UPMC

Illustration de TextTiling: Rohit Kate, Univ. of Wisconsin

Sommaire

- **Introduction**
- **Analyse thématique de texte**
  - Approche fondée sur les connaissances
  - Approche fondée sur une analyse de surface
    - Approches locales
      - TextTiling
      - Ajout de connaissances externes
      - Chaines lexicales
    - Approches globales
- **Structuration de texte**
- **Evaluation**
- **Conclusion**

Discours

- **Les productions en langue naturelle sont des successions de phrases**
- **Une phrase souvent non compréhensible de manière isolée**

Marie est allée acheter un manteau. Elle en a trouvé un rouge qu'elle aimait vraiment. Quand elle l'a rapporté chez elle, elle a constaté qu'il allait parfaitement avec sa robe favorite.

2 questions :

Q1 : Qu'est-ce que Marie a trouvé ?
Q2 : Marie a-t-elle acheté quelque chose ?

3

Discours

- **Discours :**
  - Ensemble cohérent et structuré de (groupes de) phrases
- **Tâches pour analyser un discours**
  - Analyse thématique (Segmentation en unités thématiques)
  - Détermination de relations de cohérence (ou relations discursives ou relations rhétoriques)
  - Résolution d'anaphores

4

Analyse thématique de texte

- **Définition informelle**
  - Analyse thématique : rendre compte de ce dont "parlent" les textes
    - exemples : texte "parlant" de la crise du pétrole, ou d'une tentative d'assassinat de Martin Luther King
- **Par rapport à la langue**
  - analyse reposant sur une masse importante d'éléments extra-linguistiques
  - statut non défini de la notion de thème / à la langue

5

Décomposition de l'analyse thématique

- **3 sous-problèmes en interrelation**
  - segmentation thématique
    - définir des segments de texte thématiquement homogènes
      - ➡ définition/délimitation des unités
  - identification thématique
    - associer la représentation d'un thème à chaque segment
      - ➡ typage des unités
  - suivi thématique
    - trouver les relations existant entre les segments (relations de changement de thème, de généralisation, ...)
      - ➡ rôle des unités ; définition d'une structure du texte

6

Une illustration

Segment 1
Séance de dédicace

Il y a quelques années, je me trouvais dans un grand magasin de Harlem, entouré de quelques centaines de personnes. J'étais en train de dédicacer des exemplaires de mon livre "Stride toward Freedom", qui relate le boycottage des autobus de Montgomery en 1955-56. Soudain, tandis que j'apposais ma signature sur une page,

changement de thème

Segment 2
Attentat

je sentis quelque chose de pointu s'enfoncer brutalement dans ma poitrine. Je venais d'être poignardé à l'aide d'un coupe-papier, par une femme qui devait être reconnue folle par la suite. On me transporta d'urgence à l'Hôpital de Harlem

déviation de thème

Segment 3
Hôpital

où je restai de longues heures sur un lit tandis qu'on faisait mille préparatifs pour extraire l'arme de mon corps

7

Applications de l'analyse thématique

- **Segmentation**
  - structuration d'un flot textuel continu (ex.: transcriptions audios)
  - RI et Extraction en contexte (ex. : détection d'événement, indexation, ...)
  - routage de documents
  - classification de documents
- **Suivi**
  - structuration des textes, repérage des éléments importants → résumé, filtrage, visualisation

8

Approche fondée sur les connaissances

- **Années 80-90**
- **Modèle de discours (Grosz 1985)**
- **Travaux sur la compréhension de texte**
  - Travaux de Schank (1977: SAM/PAM), Dyer et al. (1983: Boris)
  - Définition de scripts, schémas représentant des situations
  - Exemple : aller au restaurant, tentative d 'assassinat, réparer une voiture …
- **Structuration des connaissances :**
  - Un schéma fait référence à d'autres schémas pour des événements complexes
  - Hiérarchisation des schémas

9

Approche fondée sur les connaissances

Mécanisme de compréhension

- **Un schéma =** un niveau de description d'un sujet
- **Mécanisme général de compréhension**
  - Identification du schéma, donc du thème
  - Remplissage par appariement avec les événements du texte
  - Recherche du lien entre différents schémas représentatifs de segments de discours différents
- **Illustration**
  - Programmes généraux de compréhension d'histoire
  - Les programmes SAM et PAM de Schank et Abelson (77)

10

Approche fondée sur les connaissances

- **Bilan**
- **Intérêt**
  - Niveau de compréhension fin : segmentation, identification, suivi du thème.
  - Réponse à tout type de question: quoi, qui, comment, pourquoi
- **Inconvénient**
  - Difficile de modéliser ce type de connaissances
    - Application à des domaines restreints
- **Transformation de la tâche**
  - Extraction d'information à partir de « templates »~(~85)
  - Evaluations MUC
- **Connaît un renouveau**
  - Story telling

11

Approche fondée sur des marques de surface

- **Segmentation d'un texte en différents segments thématiques**
- **Non supervisé :**
  - Pas de données d'entrainement
- **Critères**
  - Cohésion lexicale
  - Marques linguistiques
- **2 approches**
  - Globale : regrouper les phrases, passages en sous-thèmes qui forment un bloc cohésif
  - Locale : un changement de thème correspond à une rupture de cohésion
    - Travaux de Salton (1996), Hearst (1997)

12

Définitions

- **Cohésion**
  - Lien entre unités textuelles dues à des procédés linguistiques (anaphores, cohésion lexicale, ellipses, conjonctions, etc.)
- **Cohésion lexicale :**
  - Utilisation des mêmes mots ou de mots similaires ou dans le même champ sémantique pour lier des unités de texte
    - Répétition des mots (distribution) : localiser un thème
    - Répartition des mots (position) : évaluer l'importance du mot/thème
  - Exemple : extrait du vin jaune

13

Segmentation thématique par analyse locale

- **Mécanisme d'analyse : TextTiling**
  - Hearst, 1997
  - Principe : identification des changements de thème
  - Association d'un vecteur de descripteurs à une zone de texte
    - Ex un descripteur : un mot plein
    - Ex. une valeur : nombre d 'occurrences, pondéré par la répartition / différentes zones
    - Ex. zone = 20 mots

14

Segmentation thématique par analyse locale

- **Mécanisme d'analyse : TextTiling**
  - Hearst, 1997
  - Principe : identification des changements de thème
  - Association d'un vecteur de descripteurs à une zone de texte (pseudo-phrase)
    - Ex un descripteur : un mot plein
    - Ex. une valeur : nombre d 'occurrences, pondéré par la répartition / différentes zones
    - Ex. zone = 20 mots

15

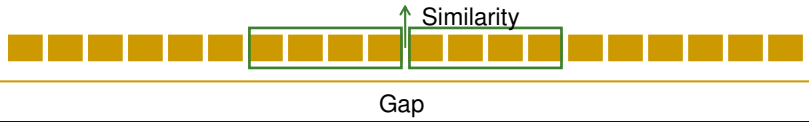
Segmentation thématique par analyse locale

- **TextTiling (suite)**
- **Mesure d'un score entre vecteurs de zones consécutives**
  - ex. produit scalaire, cosinus : élevés quand mots cooccurrent
- **Score de cohésion lexicale**
  - Similarité des descripteurs avant et après le gap (ex. 10 pseudo-phrases avant et après)

16

Segmentation thématique par analyse locale

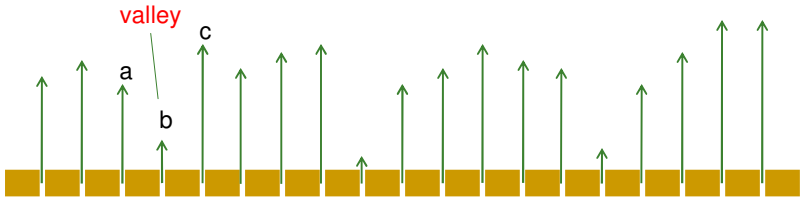
- **TextTiling (suite)**
- **Mesure d'un score entre vecteurs de zones consécutives**
  - ex. produit scalaire, cosinus : élevés quand mots cooccurrent
- **Score de cohésion lexicale**
  - Similarité des descripteurs avant et après le gap (ex. 10 pseudo-phrases avant et après)



17

Segmentation thématique par analyse locale

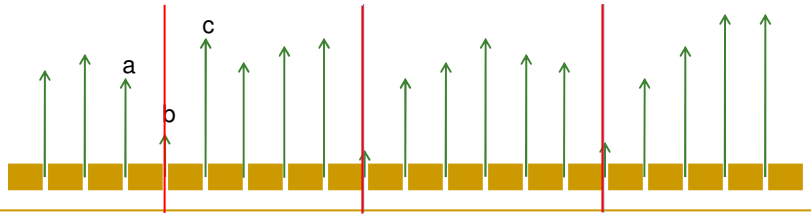
- **TextTiling (suite)**
- **Construction du graphe des valeurs de similarités et calcul du score de rupture des « vallées »**
  - (a-b)+(c-b)
- **Un changement de thème**
  - Si score de rupture > seuil
  - Ex. ecart type > moyenne des scores



18

Segmentation thématique par analyse locale

- **TextTiling (suite)**
- **Construction du graphe des valeurs de similarités et calcul du score de rupture des « vallées »**
  - (a-b)+(c-b)
- **Un changement de thème**
  - Si score de rupture > seuil
  - Ex. ecart type > moyenne des scores



19

Segmentation thématique par analyse locale

- **Application**
  - (Masson et al. 95)
- **Un texte :**
  - un ensemble de paragraphes
- **Un paragraphe :**
  - un vecteur de descripteurs
- **Extraction des descripteurs :**
  - balisage SGML, segmentation, étiquetage-lemmatisation, filtrage des catégories
  - Noms (simples et composés), verbes, adjectifs
- **Pondération des vecteurs par le facteur tf.idf**
  - renforce les descripteurs peu dispersés et inhibe ceux qui le sont
- **Distance entre paragraphes adjacents :**
  - coefficient de Dice (produit scalaire normalisé des vecteurs)

20

Segmentation thématique par analyse locale

Tf.idf : $w_{ij} = \mathrm{tf}_{ij} \cdot \log \frac{N}{df_j}$

W_{ij} : poids du terme T_j dans §i
Tf_{ij} : nombre d'occurrences du terme T_j dans §i
N : nombre total de § dans texte
Df_j : nombre de § contenant T_j

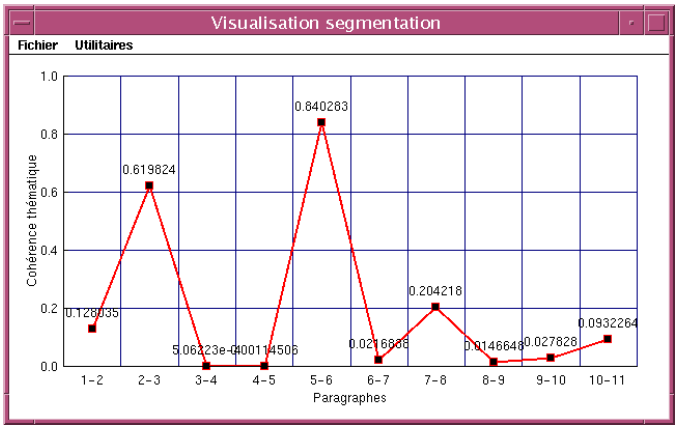
Coefficient de Dice :

$$Dice = \frac{2\sum_{i=1}^{t} w(x_i) \cdot w(y_i)}{\sum_{i=1}^{t} w(x_i)^2 + \sum_{i=1}^{t} w(y_i)^2}$$

$X = (x_1, x_2, \ldots, x_t)$
$Y = (y_1, y_2, \ldots, y_t)$

Segmentation thématique par analyse locale

- **Application : Résultats sur le texte "Vin jaune"**

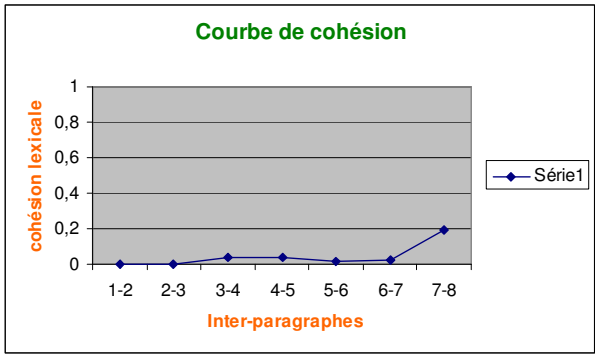


Segmentation thématique par analyse locale

- **Application sur un texte de type journalistique**
  - Article paru dans Le Monde traitant d'un accident de train

Segmentation thématique par analyse locale

- **Application : Résultats sur le texte journalistique**



Calcul de la cohésion inter-paragraphes du texte portant sur l'accident de train

Segmentation thématique par analyse locale avec ajout de connaissances

- **Utilisation de marques de surface**
  - Litman et Passonneau 1997
    - Exemple : Ensuite, parce que, …
  - Problème : définition de ces marques
- **Utilisation de connaissances non dédiées**
  - Kozima (1993), Morris & Hirst (1991), Ferret (1998)
  - Connaissances peu structurées :
    - ex. réseau de co-occurrences, thésaurus, dictionnaire
  - Connaissances disponibles ou pouvant être construite automatiquement
  - Même méthode que précédemment

25

Segmentation thématique par analyse locale avec ajout de connaissances

- **But : améliorer le « rapprochement » des paragraphes thématiquement liés**
- **2 exemples de méthodes :**
  - **M1**: Utilisation des relations sémantiques et pragmatiques entre termes (Masson et al. 95)
    - Pour compléter la représentation vectorielle
      - Estimation des liaisons entre termes par comptage des cooccurrences dans un gros corpus
    - Pour calculer une valeur de cohésion à chaque pas du texte
  - **M2** : Calcul de la cohésion lexicale sur réseau de mots avec des liaisons pondérées acquises automatiquement (Ferret et al. 98)

26

Segmentation thématique par analyse locale avec ajout de connaissances

Construction du réseau

- **Corpus : 24 mois du journal « Le Monde » (entre 1990 et 1994)**
- **Méthode**
  - Pré-traitement des textes
  - Comptage des cooccurrences dans une fenêtre glissante
    - Taille : 20 lemmes (environ 50 mots dans le texte original ⇔ 2 phrases)
    - Pas d'ordre : coo(m1,m2) + coo(m2, m1)
    - Respect des frontières de texte
  - Filtrage des cooccurrences faibles (< 6)
  - Estimation de l'information mutuelle entre termes cooccurrents

⇒ 31 000 lemmes et 7 millions de relations

27

Segmentation thématique par analyse locale avec ajout de connaissances

Extrait du réseau

Lemme1	Lemme2	Nombre occurrences	Valeur de cohésion	Type de lien
imprimante	**ordinateur**	**13**	**0,227**	**pragmatique**
bateau	**voilier**	**125**	**0,224**	**sémantique**
prêtre	**curé**	**44**	**0,209**	**sémantique**
policier	**cambriolage**	**41**	**0,190**	**pragmatique**
chômage	**emploi**	**1985**	**0,167**	**sémantique**
prendre	**racine**	**120**	**0,110**	**lexico-syntaxique**
collision	**franc**	**7**	**0,076**	**bruit**

28

Segmentation thématique par analyse locale avec ajout de connaissances

- **Méthode 1 : Utilisation du réseau lexical**
- **Principe : tenir compte des liaisons dans la représentation vectorielle**
  - P1 = (kmot1, nmot2, 0, …) et P2 = (0, 0, tmot3, …)
- **Renforcement du poids des descripteurs liés**
  - mot1 et mot2 liés par liaison w
  - P1' = ((wn+k)mot1, (wk+n)mot2, 0, …)
- **Ajout de descripteurs dans les vecteurs**
  - mot1 et mot3 liés par liaison z
  - P1' = ((wn+k)mot1, (wk+n)mot2, zkmot3, …)
  - P2' = (ztmot1, 0, tmot3, …)

29

Segmentation thématique par analyse locale avec ajout de connaissances

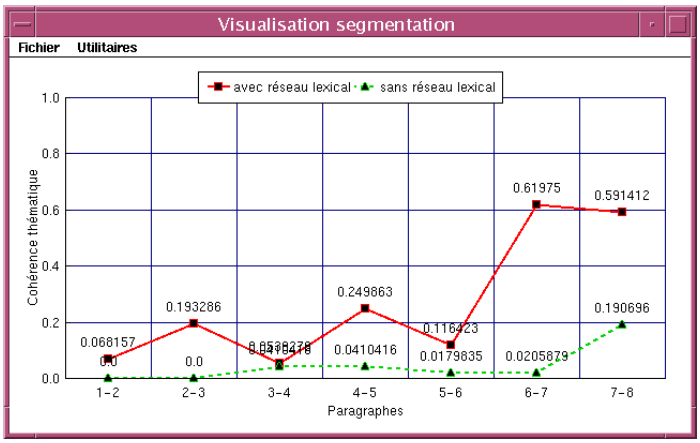
- **Méthode 1 (suite)**
- **Pré-traitement, vectorisation, renforcement et ajout de descripteurs**
- **Pondération de type tf.idf (élimination du bruit dû aux nombreux liens du réseau)**
- **Collocations avec « collision »**

Cooccurrences significatives		Cooccurrences peu significatives	
Accident	0.210329	Pays	0.0684458
Conducteur	0.19373	Président	0.0817475
Train	0.185325	Ministre	0.0890158
Gare	0.181422	Franc	0.0763806

30

Segmentation thématique par analyse locale avec ajout de connaissances

- **Application sur le texte “Accident de train”**

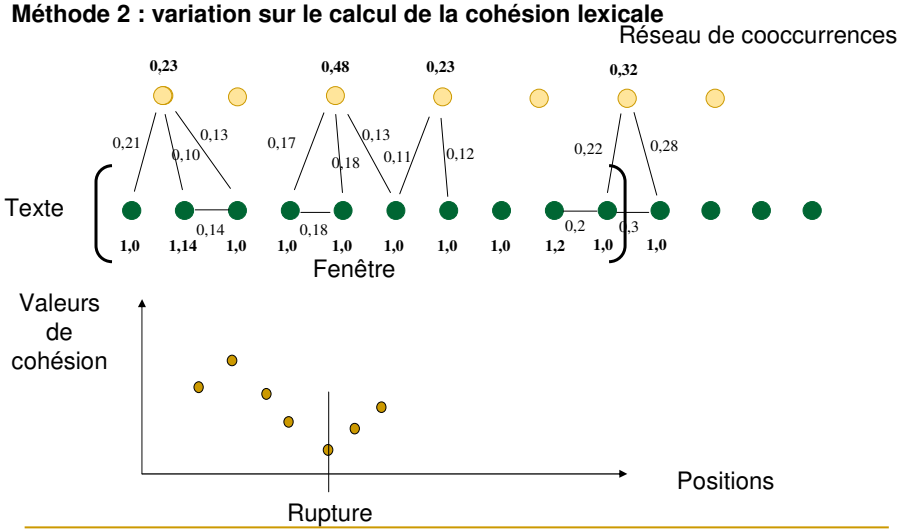


Segmentation thématique par analyse locale avec ajout de connaissances

- **Méthode 2 : variation sur le calcul de la cohésion lexicale**
- **Fenêtre glissante de 20 mots**
  - Pas de découpage a priori en paragraphes
    - Evite le problème de tailles non homogènes
- **Calcul de cohésion pour chaque mot au centre de la fenêtre**

Segmentation thématique par analyse locale avec ajout de connaissances

Méthode 2 : variation sur le calcul de la cohésion lexicale



Segmentation thématique par analyse locale avec ajout de connaissances

Exemple : Attentat de MLK

Segment 1 — Séance de dédicace

Il y a quelques années, je me trouvais dans un grand magasin de Harlem, entouré de quelques centaines de personnes. J'étais en train de dédicacer des exemplaires de mon livre "Stride toward Freedom", qui relate le boycottage des autobus de Montgomery en 1955-56. Soudain, tandis que j'apposais ma signature sur une page,

Segment 2 — Attentat

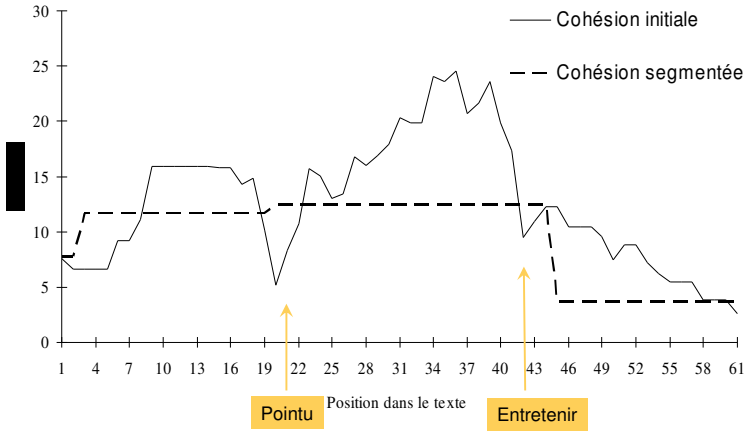
je sentis quelque chose de pointu s'enfoncer brutalement dans ma poitrine. Je venais d'être poignardé à l'aide d'un coupe-papier, par une femme qui devait être reconnue folle par la suite. On me transporta d'urgence à l'Hôpital de Harlem

Segment 3 — Hôpital

où je restai de longues heures sur un lit tandis qu'on faisait mille préparatifs pour extraire l'arme de mon corps. Beaucoup plus tard, quand je fus en état de m'entretenir avec le Dr. Aubrey Maynard, le chirurgien en chef qui exécuta cette délicate et dangereuse intervention, j'appris les raisons de cette longue attente avant l'opération. La lame de l'instrument avait touché l'aorte et, pour l'extraire, il fallait ouvrir toute la cage thoracique.

Segmentation thématique par analyse locale avec ajout de connaissances

Exemple : Attentat de MLK : courbes et segments obtenus



35

Segmentation thématique par analyse locale avec ajout de connaissances

Exemple : Attentat de MLK : segments obtenus

Segment 1
Séance de dédicace

Il y a quelques années, je me trouvais dans un grand magasin de Harlem, entouré de quelques centaines de personnes. J'étais en train de dédicacer des exemplaires de mon livre "Stride toward Freedom", qui relate le boycottage des autobus de Montgomery en 1955-56. Soudain, tandis que j'apposais ma signature sur une page,

Segment 2
Attentat

je sentis quelque chose de pointu s'enfoncer brutalement dans ma poitrine. Je venais d'être poignardé à l'aide d'un coupe-papier, par une femme qui devait être reconnue folle par la suite. On me transporta d'urgence à l'Hôpital de Harlem

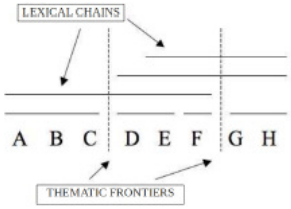
Segment 3
Hôpital

où je restai de longues heures sur un lit tandis qu'on faisait mille préparatifs pour extraire l'arme de mon corps. Beaucoup plus tard, quand je fus en état de m'entretenir avec le Dr. Aubrey Maynard, le chirurgien en chef qui exécuta cette délicate et dangereuse intervention, j'appris les raisons de cette longue attente avant l'opération. La lame de l'instrument avait touché l'aorte et, pour l'extraire, il fallait ouvrir toute la cage thoracique.

36

Segmentation thématique par analyse locale : chaines lexicales

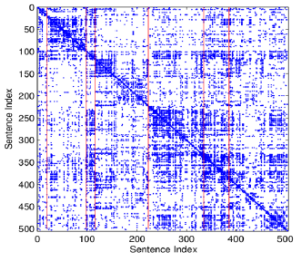
- **Chaine lexicale**
  - (Morris and Hirst, 1991)
  - Relie mêmes mots et mots sémantiquement liés
  - Liaison existe selon la position des mots
    - Distance entre 2 occurrences < seuil
- **Segmentation**
  - LCSeg (Galley et al., 2003) : outil disponible

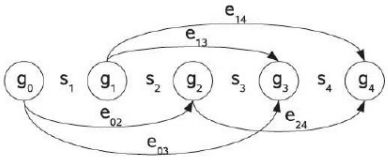


37

Segmentation thématique par analyse globale

- **Maximiser la valeur de cohésion lexicale de chaque segment obtenu**
  - Matrices de similarité des phrases
    - Dotplot (Reynar, 94)
  - Graphe
    - TextSeg (Utiyama & Isahara, 2001)
    - Arc = segment, nœud = frontière
    - Valeur de cohésion probabiliste = capacité d'un ML appris sur le segment à prédire les mots du segment





38

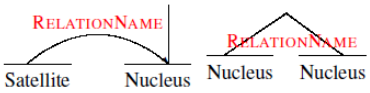
Segmentation thématique
Bilan

- **Niveau de compréhension**
  - Délimite les zones de texte relatives à différents thèmes
  - Pas de structuration (sauf par étude des valeurs de liaison entre toutes les zones)
  - Pas d'identification de thème
- **Intérêt**
  - Pas de connaissances spécifiques
  - Application à de nombreux textes indépendamment du domaine
  - Utilisation en résumé, recherche d'information
    - localiser les informations plus précisément dans le texte
- **Problèmes**
  - Résultats plus ou moins bons selon le type de texte : scientifiques ou dépêches
  - Résultats peu précis et peu fiables
  - Corpus de test

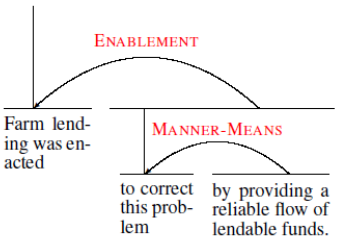
39

Structuration de texte

- **Hiérarchie de segments**
- **Structure par emboîtement de segments**
  - Ex. Projet REGAL : Structure « gros grain »
  - Exemple
- **Structure phrase par phrase : relations discursives**
  - Théorie RST (Rhetorical Structure Theory): Mann & Thomson



Relations Mono-nucléaire – Multi-nucléaire



- **Ressources**
  - Penn Discourse Treebank : http://ww
  - RST-DT (Carlson et al. 2001)

40

Structuration de texte

- **Apprentissage des relations**
  - Système HILDA (Hernault et al. 2010)
    - Décomposition en EDU par apprentissage
      - W -> [0,1}, W = mots, 1 = mot est frontière d'EDU
    - Etiquetage des relations (18)
      - 2 classifieurs : existence d'une relation + label
      - Traits : organisation du texte, marques de discours : 3-gram début et fin EDU, traits syntaxiques de dominance, …
  - Système DST
    - Structure « grain fin » (N. Hernandez, 2000)
    - Apprentissage des relations entre 2 phrases
      - Subordination, coordination, absence de relation
      - Critères :
        - marques linguistiques, cohésion lexicale, suivi thème-rhème, parallélisme syntaxique
    - Exemple

41

Evaluation

- **TDT : Topic Detection and Tracking**
- **Trois tâches :**
  - Segmentation d'histoire : segmentation d'un flot continu
  - Suivi de thème : association d'une histoire avec un thème connu du système
  - Détection de thème : détection et suivi de thèmes inconnus du système
- **Méthodes :**
  - Construction de modèle de langage pour représenter des thèmes à partir d'un corpus d'apprentissage

42

Segmentation thématique
Evaluation

- **Problème avec mesures de précision / rappel**
- **Introduction d'une métrique**
  - WindowDiff (Pevzner & Hearst, 2002)
  - Glissement d'une fenêtre de longueur k sur la référence (correcte) et sur la segmentation évaluée, et comptabilisation de nombre de coupures dans chaque
  - Métrique WindowDiff : moyenne des différences dans le nb de ruptures dans la fenêtre glissante

Conclusion

- **Evaluation difficile**
- **Résultats assez moyens**
  - Place pour amélioration, mais …
- **Questions ouvertes**
  - Définition de la notion de thème
  - Structurations différentes des textes
    - Structure thématique
    - Structure rhétorique
  - Tout passage appartient-il aux 2 structures ?
- **Lecture conseillée :**
  - Semantic structuring of video collections from speech : segmentation and hyperlinking, Anca Simon, thèse de Université de Rennes

<p3>
On soupçonna alors le 4,5 diméthyl 3 hydroxy 2(5H) furanone, ou sotolon, molécule construite autour d'un cycle de quatre atomes de carbone et d'un atome d'oxygène. Comme le sotolon et la solérone sont en concentrations minimes dans les vins de voile et, de surcroît, chimiquement instables, les chimistes dijonnais ont cherché à optimiser leur extraction afin de déterminer la molécule responsable du goût de jaune.

<p4>
L'analyse la plus directe d'extraits de vins est la chromatographie: on injecte un échantillon dans un solvant que l'on vaporise et on fait traverser au mélange une colonne revêtue intérieurement d'un polymère, qui retient les divers composés du mélange à des degrés divers; en bas de la colonne, on détecte la sortie des composés séparés. Le premier travail des chimistes fut la mise au point d'une variante de cette technique pour identifier les composés présents en quantités minimes dans des mélanges complexes.

<p5>
Les chromatogrammes d'échantillons de vin furent alors comparés à ceux de solutions pures de sotolon et de solérone de synthèse: le sotolon est ainsi présent entre 40 et 150 parties par milliard dans les sherrys; la solérone semble moins spécifique, et ses concentrations sont supérieures dans les sherrys, ce qui explique pourquoi on l'a d'abord trouvée dans ces vins.

<p6>
Enfin les dosages, complétés de tests sensoriels des fractions séparées, montrèrent que la solérone, aux concentrations trouvées dans du savagnin (le cépage à partir du quel on fabrique le vin jaune), n'était perçue par les consommateurs ni dans les vins, ni dans des solutions modèles: la solérone n'était pas la molécule caractéristique; le jugement était sans appel.

<p7>
En 1992, les chimistes se consacrèrent alors complètement au sotolon, qui avait été observé dans des molasses de canne à sucre, dans des graines de fenugrec, dans de la sauce de soja, dans du saké...

Extrait de roman

- Extrait du livre *De la Terre à la Lune* de Jules Verne

DST – type de relations

(1) Les résumés par extraction sélectionnent des phrases d'un texte source selon leur importance.

(2) Les critères d'importance incluent la présence de termes fréquents, des mots clefs tels que « en résumé », « meilleur », et la position de la phrase dans le texte.

(3) Cette approche est illustrée par le système ADAM.

(4) Un autre exemple est donné par [2].

(5) Le problème de cette approche est que les phrases extraites ne constituent pas toujours un texte cohérent du fait d'anaphores ambiguës.

47

DST – Types de relations

(1) Les résumés par extraction sélectionnent des phrases d'un texte source selon leur importance.

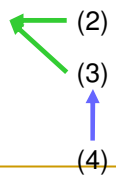
Subordination

(2) Les critères d'importance incluent la présence de termes fréquents, des mots clefs tels que « en résumé », « meilleur », et la position de la phrase dans le texte.

Coordination

(3) Cette approche est illustrée par le système ADAM.
(4) Un autre exemple est donné par [2].

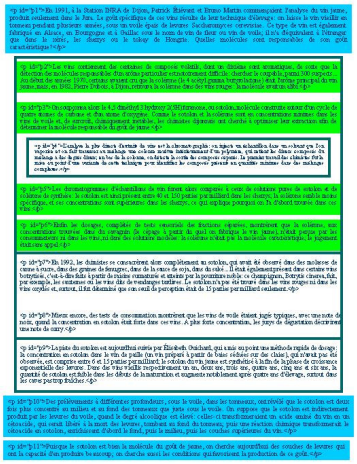
Structure construite



48

REGAL - Principe de structuration

- Repérage de structures emboîtées [Masson, 1998]
  - Digressions, développements d'aspects particuliers
  - Fréquent dans textes expositifs

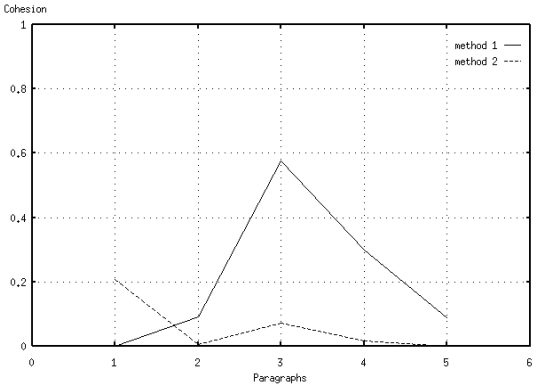


Algorithme

- Recherche des 2 segments non-consécutifs les plus liés
- Application récursive pour les segments englobés ou non englobés restant

Intérêt de l'utilisation du réseau

Relativiser l'importance de certains mots peu significatifs, notamment lorsque les paragraphes sont courts



sans réseau

avec réseau

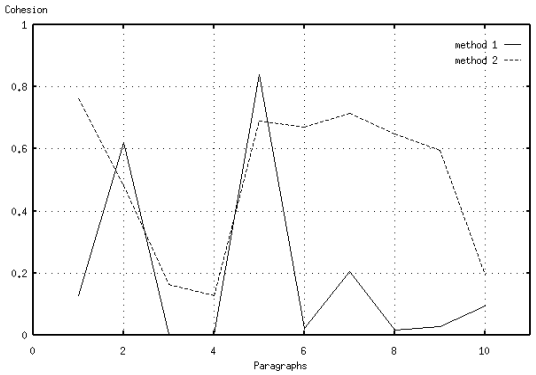
extrait de
"De la Terre à la Lune"
Jules Verne

mots justifiant le rapprochement des paragraphes 3 et 4 (sans réseau) :

gun-club, immense, jour, membre

Problèmes posés par l'utilisation du réseau

Liens entre des mots non significatifs masquant la réitération de mots significatifs absents du réseau



——— sans réseau

..... avec réseau

Vin jaune: regroupement non souhaité des paragraphes
1 et 2, 6 et 7, 8 et 9, 9 et 10, 10 et 11

51

Vers une combinaison des méthodes

Appliquer la méthode fondée sur la récurrence des mots jusqu'à la phase de pondération des termes

Si x % des termes jugés significatifs (termes n'étant pas présents dans tous les paragraphes ; pondération *tf.idf*)
ne sont pas dans le réseau de cooccurrences

alors

continuer avec la méthode fondée sur la récurrence des mots

Sinon

appliquer la méthode utilisant le réseau de cooccurrences

Fin_si

estimation de x : 25 %

52

Une situation

Ses propriétés

Réparer un véhicule

Rôles :
personne, outil, véhicule

Conditions:
Véhicule en panne
Véhicule = rôle (véhicule)
Connaître la mécanique
Agent = rôle (personne)

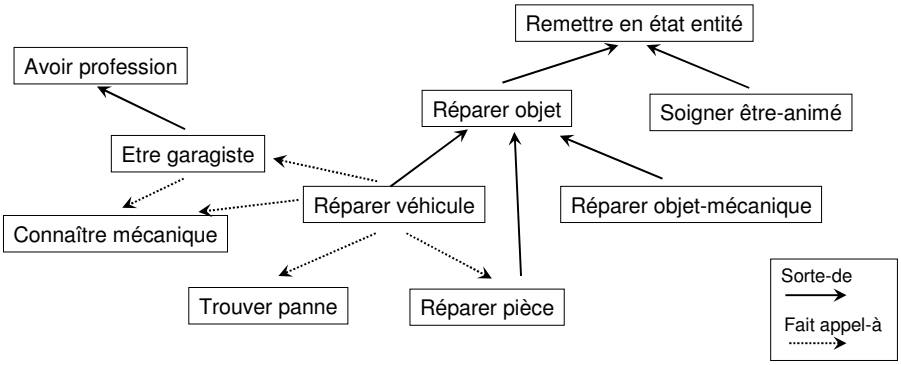
Description
1 Trouver panne
Agent = rôle (personne)
1.Résultat = pièce
2 Démonter pièce
3 Réparer pièce
4 Remonter pièce
Agent = rôle (personne)
Pièce = 1.résultat

Résultat
Véhicule fonctionne
Véhicule = rôle (véhicule)

- Rôles
- Evénements
- Relations causales
- Relations temporelles
- Inférence
- Structuration

53

Graphe de schémas

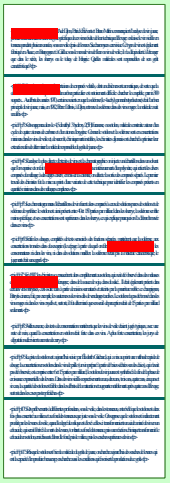


- **Pas de méthode pour élaborer et organiser les connaissances**
  - Quel niveau de granularité, quels regroupements ?
  - Quelle structure hiérarchique ?

54

Exemple d'exploration d'un texte

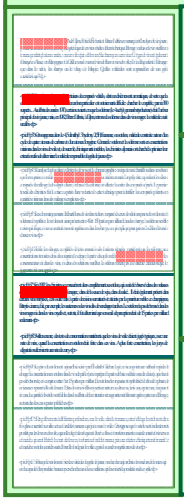
Thème général : vin jaune



55

Exemple d'exploration d'un texte

Thème général : vin jaune



Thème global : vin

Thème local : mélange, composé

Meta-descripteurs : analyse, technique

Thème global : vin

Thème local : goût, noisette

Meta-descripteurs : caractéristique

56

Exemple d'exploration d'un texte

Thème général : vin jaune

Thème global :
vin

Thème local :
mélange, composé

Meta-descripteurs :
analyse, technique

Thème global :
vin

Thèmes local :
goût, noisette

Méta-descripteureurs :
caractéristique

<seg>
L'analyse la plus directe d'extraits de ***vins*** est la chromatographie : on injecte un échantillon dans un solvant que l'on vaporise et on fait traverser au ***mélange*** une colonne revêtue intérieurement d'un polymère, qui retient les divers ***composés*** du ***mélange*** à des degrés divers ; en bas de la colonne, on détecte la sortie des ***composés*** séparés. Le premier travail des chimistes fut la mise au point d'une variante de cette ***technique*** pour identifier les ***composés*** présents en quantités minimes dans des ***mélanges*** complexes.
</seg>

57