

Une approche mixte-statistique et structurale - pour le résumé automatique de dépêches

Aurélien Bossard
LIPN - UMR 7030
CNRS - Université Paris 13
F-93430 Villetaneuse, France
aurelien.bossard@lipn.univ-paris13.fr

Résumé. Les techniques de résumé automatique multi-documents par extraction ont récemment évolué vers des méthodes statistiques pour la sélection des phrases à extraire. Dans cet article, nous présentons un système conforme à l'« état de l'art » — CBSEAS — que nous avons développé pour les tâches Opinion (résumés d'opinions issues de blogs) et Update (résumés de dépêches et mise à jour du résumé à partir de nouvelles dépêches sur le même événement) de la campagne d'évaluation TAC 2008, et montrons l'intérêt d'analyses structurales et linguistiques des documents à résumer. Nous présentons également notre étude sur la structure des dépêches et l'impact de son intégration à CBSEAS.

Abstract. Automatic multi-document summarization techniques have recently evolved into statistical methods for selecting the sentences that will be used to generate the summary. In this paper, we present a system in accordance with « State-of-the-art » — CBSEAS — that we have developed for the « Opinion Task » (automatic summaries of opinions from blogs) and the « Update Task » (automatic summaries of newswire articles and information update) of the TAC 2008 evaluation campaign, and show the interest of structural and linguistic analysis of the documents to summarize. We also present our study on news structure and its integration to CBSEAS impact.

Mots-clés : Résumé automatique, structure de documents.

Keywords: Automatic summarization, document structure.

1 Introduction

Porté par une large communauté de chercheurs et des campagnes d'évaluation telles que DUC (Document Understanding Conference) et TAC (Text Analysis Conference), le résumé automatique a connu ces dernières années une évolution rapide.

Nous présentons dans ce papier une nouvelle approche pour le résumé automatique. Une nouvelle méthode pour détecter la redondance a été intégrée au coeur de notre système, CBSEAS. Cette méthode vise à mieux prendre en compte la redondance des informations que les systèmes existants afin de créer des résumés *a priori* plus pertinents. Les résultats obtenus lors de la campagne d'évaluation TAC 2008 prouvent que notre système est adaptable et la méthode efficace. Cependant, ils pointent également du doigt les faiblesses de CBSEAS auxquelles nous nous efforçons de trouver des solutions que nous présentons ici.

Cet article est organisé de la manière suivante : en premier lieu, nous présentons les principales approches de résumé automatique par extraction existantes. Nous décrivons ensuite notre système générique, puis les modifications que nous lui avons apportées afin de participer à TAC ainsi que son évaluation lors de cette campagne. Enfin, nous montrons comment nous comptons améliorer CBSEAS sur la tâche particulière du résumé de dépêches en prenant en compte la structure des documents comme une manière de déterminer la pertinence de leurs extraits. Nous présentons également les premières expériences menées dans ce sens.

2 Etat de l'art

L'intérêt pour le résumé automatique a commencé à la fin des années 1950 avec les travaux de (Luhn, 1958) ainsi que ceux (Edmundson & Wyllys, 1961). Les bases du résumé par extraction étaient alors posées. Les travaux de (Luhn, 1958) consistaient à classer les mots du ou des documents à résumer selon un indice de fréquence qui n'est pas sans rappeler le *tf-idf* introduit par (Salton & McGill, 1986). Les phrases contenant le plus de mots proches les uns des autres considérés comme importants par l'indice de fréquence étaient alors sélectionnées pour générer un résumé. Edmundson, en plus de mesures de fréquence des mots, prenait en compte la position des phrases dans leur document et favorisait certaines positions pour certains types de documents —les premières phrases pour des dépêches de presse, les dernières pour un article scientifique. Il prenait également en compte le nombre de mots commençant par une majuscule (à l'époque, on ne parle pas encore d'entités nommées), la présence d'expressions-clé telles que « In conclusion », « hardly ».

Les résumés automatiques sont alors créés uniquement sur la notion de centralité, ou d'importance d'un élément relativement à son contexte. La notion de diversité ou de non-redondance des informations exposées dans le résumé n'était alors pas considérée, la recherche ne s'étant pas encore intéressée au résumé multi-documents.

Plus récemment, des travaux ont intégré cette notion de diversité. Certains en post-traitement, après classement des phrases à extraire par des scores reflétant la centralité. C'est le cas de (Radev, 2004). Un « centroid », groupe de termes dont la fréquence d'apparition dénote l'importance, est créé pour chaque groupe de documents à résumer. Les phrases sont classées suivant le nombre de termes du « centroid » qu'elles contiennent. Radev a également proposé dans (Erkan & Radev, 2004) une méthode inspirée des réseaux sociaux et de la notion de « prestige ». Il

établit un graphe dans lequel les noeuds sont les phrases et les arêtes la similarité des phrases les unes aux autres. Après avoir effectué un parcours aléatoire de ce graphe, il classe les phrases selon le nombre de fois qu'elles ont été visitées. Dans ces deux méthodes, Radev gère la diversité en sélectionnant les phrases dans l'ordre de classement et en éliminant les phrases trop similaires à une phrase déjà sélectionnée pour faire partie du résumé.

D'autres travaux gèrent la diversité dans le même temps que la centralité. Partant du principe que la diversité étant aussi importante que la centralité, leurs auteurs considèrent que ces deux aspects devaient être gérés dans le même temps. Parmi ceux-ci, la méthode MMR de (Carbonell & Goldstein, 1998) combine deux mesures : l'une reflétant la centralité ou l'importance vis-à-vis d'une requête utilisateur, l'autre la diversité. La diversité est fonction de la similarité aux phrases déjà sélectionnées pour le résumé.

On peut également citer (Goldberg, 2007). Sa méthode reprend celle de (Erkan & Radev, 2004). Cependant, au lieu de réaliser un simple parcours aléatoire, il utilise un parcours aléatoire markovien à états absorbants. Ainsi, les noeuds centraux absorbent les scores des noeuds qui les entourent, ce qui permet de gérer la diversité en même temps que la centralité.

(Boudin *et al.*, 2007) a travaillé sur la fusion de différentes métriques. Le système, baptisé Neo-Cortex, tire ainsi les bénéfices de l'utilisation de métriques qui rendent compte de différents types de traits. Ce système s'est très bien classé dans la campagne d'évaluation DUC 2006.

Une dernière approche, celle de (Barzilay, 2003), aborde le problème de manière tout à fait différente : elle détecte dans un premier temps les paraphrases en appliquant une SVM sur les arbres syntaxiques normalisés des phrases des documents à résumer. Les informations les plus importantes sont alors les plus paraphrasées. Cette technique est extrêmement bien adaptée à la problématique du résumé multi-documents, où dû à la multiplicité des sources, la notion de centralité se rapproche de celle de redondance. Cependant, à cause de la quantité de ressources linguistiques qu'elle demande, cette méthode n'est pas adaptable à toutes les langues. Nous voulons développer une méthode dans laquelle, à l'instar de (Barzilay, 2003), la redondance occupe une place majeure, et qui soit assez indépendante des ressources linguistiques pour permettre son adaptation à différentes langues.

3 CBSEAS : « Clustering-Based Sentence Extractor for Automatic Summarization »

Nous supposons que dans une problématique de résumé multi-document, les informations les plus redondantes sont les éléments les plus importants pour produire un résumé pertinent. Par conséquent, les phrases qui portent ces informations sont les phrases à extraire, moyennant l'élimination de la redondance. Le regroupement des phrases qui véhiculent la même information est la première étape de notre approche. L'algorithme développé établit une similarité entre les phrases des documents à résumer puis applique un algorithme de classification — *fast global k-means* (López-Escobar *et al.*, 2006) — sur la matrice de similarité afin de créer des regroupements au sein desquels les phrases véhiculent la même information, ou tout du moins sont les plus proches les unes des autres.

Premièrement, nous sélectionnons n^2 phrases pour créer un résumé de n phrases. L'algorithme d'apprentissage créera n classes pour générer un résumé de n phrases. Il apparaît en effet que *fast global k-means* est plus performant pour créer n classes avec n^2 observations. L'élection de

```

for all  $e_j$  in E
   $C_1 \leftarrow e_j$ 
for i from 1 to k do
  for j from 1 to i
     $\text{centre}(C_j) \leftarrow e_m | e_m \text{ maximise } \sum_{e_n \in C_j} \text{sim}(e_m, e_n)$ 
  for all  $e_j$  in E
    placer  $e_j$  dans  $C_l | C_l \text{ maximise } \text{sim}(\text{centre}(C_l), e_j)$ 
  ajouter un nouveau cluster :  $C_i$  qui contient initialement uniquement l'élément le moins bien représenté par sa classe
done

```

FIG. 1 – Algorithme Fast global k-means

ces n^2 phrases se fait d'après leur similarité à une requête utilisateur dans le cas où il y en a une, ou d'après leur proximité au *centroid* composé des m termes les plus importants, importance reflétée par leur tf-idf.

La similarité entre les phrases est calculée à l'aide de la mesure « Jaccard ». Cette mesure est efficace pour la comparaison d'ensembles. Les phrases sont au préalable étiquetées morpho-syntaxiquement à l'aide de tree-tagger¹ et la comparaison se fait sur les lemmes.

Une fois la matrice de similarité établie, CBSEAS effectue une classification des phrases en utilisant l'algorithme *fast global k-means* (description de l'algorithme en figure 1). Cet algorithme de classification a le double avantage de pouvoir prendre en entrée une matrice de similarité et de s'affranchir de la sélection des centres de classes préalablement à la classification, un défaut majeur de k-means.

La classification établie, CBSEAS sélectionne une phrase par classe afin de générer un résumé qui contienne la plus grande partie des informations/idées pertinentes des documents d'origine. La sélection des phrases s'opère selon une combinaison linéaire de la proximité des phrases au centre de leur classe et de la similarité à la requête/au *centroid* du groupe de documents.

4 Evaluation au sein de TAC 2008 : « Opinion Task »

Afin d'évaluer notre système, nous avons participé à deux tâches de la campagne TAC (*Text Analysis Conference*) 2008. La première des deux tâches, *Opinion Task*, consistait à résumer les opinions contenues dans des blogs. Les résumés étaient orientés par une ou des requêtes utilisateur telles que : « Why do people dislike ... ? ».

Les résumés ont été évalués manuellement par des analystes de NIST² suivant le protocole « Pyramid » (Lin *et al.*, 2006). Le score PYRAMID d'un résumé automatique dépend du nombre d'unités sémantiques qu'il contient et qui sont considérées comme importantes par les annotateurs. L'importance d'une unité sémantique est fonction de son nombre d'occurrences au sein des résumés générés à la main par les annotateurs. Les résumés ont également été notés sur cinq autres critères par les évaluateurs humains : grammaticalité, non-redondance, structure, fluidité et « ré pondance ».

Les organisateurs de TAC offraient aux participants à la campagne d'évaluation la possibilité

¹<http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/>

²National Institute of Science and Technology : <http://www.nist.gov>

Evaluation	CBSEAS	Mean	Best	Worst	Rank
Pyramid	.169	.151	.251	.101	5/20
Grammatic.	5.95	5.14	7.54	3.54	3/20
Non-redund.	6.64	5.88	7.91	4.36	4/20
Structure	3.50	2.68	3.59	2.04	2/20
Fluency	4.45	3.43	5.32	2.64	2/20
Responsiv.	2.64	2.61	5.77	1.68	8/20

FIG. 2 – Résultats de la tâche « Opinion »

d'utiliser des fragments de texte *a priori* pertinents produits par des systèmes de Questions-Réponse (QR). Nous avons pensé qu'il est plus réaliste de considérer la production de résumé indépendamment de la tâche de QR (même si ces deux tâches sont relativement proches) pour deux raisons : suivant le contexte, un tel système n'est pas toujours disponible et surtout cumuler deux systèmes différents brouille l'évaluation, à moins de « tracer » les erreurs. Nous ne présentons en 4.2 que les résultats des systèmes qui n'ont pas utilisé ces fragments de texte.

4.1 Modifications apportées au système

Les blogs ont été passés dans un module de filtrage visant à éliminer les phrases présentant des incohérences lexicales. Les phrases avec un ratio *nombre de mots fréquents/nombre total de mots* en dessous de 0.35 ont été écartées. Les « mots fréquents » sont les cent mots les plus fréquents en langue anglaise ; ils constituent environ la moitié des textes écrits (Fry *et al.*, 2000).

Les requêtes ainsi que toutes les phrases des documents ont été étiquetées selon l'opinion qu'elles véhiculent — positive ou négative — selon une méthode d'analyse d'opinion développée par Michel Génèreux. Plus de détails à ce propos sont disponibles dans (Bossard & Génèreux, 2009).

CBSEAS est alors utilisé sur les données nettoyées afin de produire un premier résumé. Ce résumé est ensuite réorganisé selon l'étiquetage en opinion des phrases qui le composent et de l'ordre des requêtes. Par exemple, si la première requête était « Why do people appreciate Linux ? » et la deuxième « For what reasons do people dislike Linux ? », alors le premier paragraphe du résumé final contiendra les opinions positives et le second les opinions négatives.

4.2 Résultats

Le système proposé pour la tâche Opinion de TAC 2008 s'est très bien comporté (cf figure 2). En effet, il se classe parmi le premier quart des participants excepté sur le score « Responsiveness ». Le champ « Structure » montre que l'intégration du système d'étiquetage d'opinion a constitué un véritable atout. Les scores Pyramid ainsi que le score de non-redondance dénotent quant à eux le bien-fondé de notre approche bâtie sur la détection de la redondance. Le mauvais score relatif en « responsiveness » s'explique en grande partie par la longueur de nos résumés : nous avons choisi d'utiliser le maximum de caractères autorisés par TAC, soit 7000 caractères par requête, ce qui s'est révélé *a posteriori* être une erreur.

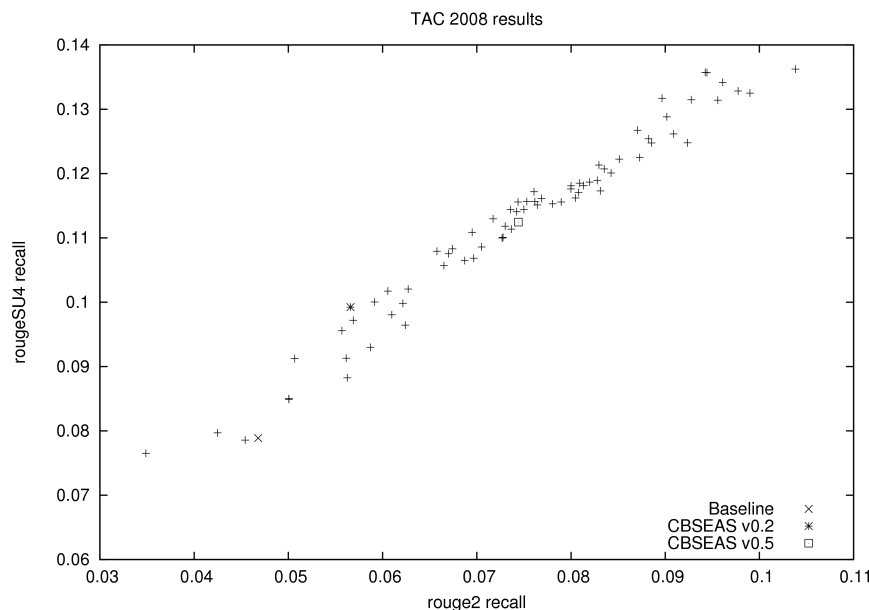


FIG. 3 – Résultats de CBSEAS sur la tâche « Update »

5 Evaluation au sein de TAC 2008 : « Update Task »

Nous avons également évalué CBSEAS dans la tâche « Update ». Cette tâche consistait à fournir des résumés pour différents thèmes. Chaque thème était composé de deux groupes de documents. Les documents étaient uniquement tirés du corpus AQUAINT-2, un corpus regroupant des dépêches de presse tirées des plus grandes agences de presse internationales (*AFP*, *Xinhua*, *NYT*...) Deux résumés devaient être fournis pour chaque thème, le premier synthétisant le premier groupe de documents, le deuxième résumant les informations contenues dans le deuxième groupe de documents et qui constituaient une nouveauté par rapport au premier groupe. Les résumés ne devaient pas excéder 100 mots.

Deux évaluations ont été proposées : la première utilisant PYRAMID, la deuxième utilisant les mesures ROUGE (Lin, 2004), des mesures fondées sur la comparaison de n -grammes entre le résumé automatique et un résumé de référence.

5.1 Modifications apportées au système

Notre système a été modifié de manière à gérer la mise à jour au cours de l'apprentissage de la classification des phrases. Après avoir établi la classification en n classes des phrases du premier groupe de documents et sélectionné une phrase par classe pour générer le résumé, les phrases du second groupe de documents sont ajoutées au système de classification. *Fast global k -means* est alors réitéré jusqu'à obtenir n classes de plus, avec les contraintes suivantes :

- les phrases du premier groupe de documents sont inamovibles ;
- les centres des n premières classes sont fixés et ne peuvent pas être recalculés.

Le résumé du deuxième groupe de documents est alors généré avec les n dernières classes.

5.2 Résultats

Les résultats présentés en figure 3 sont les résultats des mesures ROUGE-2 et ROUGE-SU4 (Lin, 2004). Suite à un mauvais paramétrage de CBSEAS, les résumés envoyés à TAC ne comportaient que 75 mots en moyenne, soit 3/4 des 100 mots autorisés, ce qui a grandement pénalisé le système lors de l'évaluation. Des expériences ont été réalisées après TAC (système CBSEAS v0.5). L'évaluation postérieure à la campagne sur les mêmes données se limite donc aux différentes mesures ROUGE — entièrement automatiques donc reproductibles — avec comme étalon les résumés de référence fournis par NIST.

Malgré une méthode de mise à jour très efficace (le système se classe parmi les premiers sur les résumés des seconds groupes de documents), CBSEAS ne se positionne pas bien sur cette tâche. Cela signifie que les scores de CBSEAS sur les résumés des premiers jeux de documents sont faibles. Nous l'expliquons par un manque d'adaptation du système à la tâche spécifique du résumé de dépêches. En effet, les dépêches de presse comportent des spécificités et adoptent un style d'écriture qui permet d'améliorer considérablement la qualité des résumés si on les prend en compte. La plupart des systèmes de résumé automatique incorporent par conséquent des caractéristiques comme la position d'une phrase dans le document ou la similarité de celle-ci au titre pour calculer son score.

Cependant, ces critères, fondés sur des études anciennes, ne rendent pas entièrement compte des spécificités des dépêches. C'est pourquoi nous proposons ici une manière d'intégrer la structure des dépêches à un système de résumé automatique, afin d'en renforcer la pertinence des résultats.

6 La structure des dépêches et son intégration à CBSEAS

(Lucas, 2005) a mené une étude sur la structure énonciative des dépêches de presse. Il ressort de cette étude que l'aspect temporel prime dans la structuration des dépêches. Elle propose également une catégorisation des dépêches :

- Dépêches commentées (première partie : explication factuelle, deuxième partie : projection), cf fig 4 ;
- Dépêches élaborées (plus d'un événement) ;
- Dépêches d'action (feuilleton : un fil d'événements fortement liés les uns aux autres, dépêches boursières...)

et des méthodes pour les catégoriser ainsi que pour repérer automatiquement les différentes parties qui les composent.

(Lucas, 2005) montre également que les dépêches commentées suivent la même présentation chronologique. L'auteur présente d'abord l'événement présent, puis donne une explication à cet événement en se fondant sur des faits passés. En dernier lieu, l'auteur décrit les conséquences futures, probables ou avérées de l'événement présenté. Identifier ces trois parties peut s'avérer utile : elles suivent toutes la règle classique de l'écriture de dépêches : l'information importante est portée par la première phrase. Les mêmes remarques s'appliquent aux autres types de dépêches.

En parcourant le corpus AQUAINT-2, fourni lors de la campagne d'évaluation TAC 2008, nous avons dégagé d'autres types de dépêches :

Nouvel accrochage entre troupes françaises et rebelles du MPIGO à **Boussou** (PAPIER GENERAL) Un nouvel accrochage a opposé l'armée française et des rebelles du Mouvement populaire ivoirien du grand ouest (MPIGO) au nord de Doudou (200 km à l'ouest de Yamoussoukro), où les militaires français avaient déjà été pris à partie le . "Une patrouille française a été prise à partie par trente à quarante rebelles du MPIGO, à pied, à 500 ou 600 mètres au sud-ouest, sur la route de **Boussou**", a expliqué le porte-parole des troupes françaises en **Abidjan**, le lieutenant colonel **Ange-Antoine Lecia**. La localité de **Boussou**, située à une quarantaine de kilomètres au sud-ouest de Yamoussoukro, a été prise le par le MPIGO. La patrouille française a riposté, en état de légitime défense, aux tirs rebelles et aucun blessé n'a été signalé dans les rangs des militaires français, selon le lieutenant-colonel **Lecia**, qui n'était pas en mesure de communiquer un bilan des éventuelles victimes côté rebelle. Les rebelles du MPIGO, qui descendaient vers le sud lorsqu'ils ont ouvert le feu sur la patrouille française, se sont repliés après l'accrochage. Le , des soldats français déployés à **Boussou** ont déjà ouvert le feu contre des combattants du MPIGO qui progressaient en tirant vers eux de leurs positions, et avaient refusé d'obtempérer à des coups de semonce. Furieux, le MPIGO et les deux autres mouvements rebelles ivoiriens ont accusé les troupes françaises d'avoir outrepassé leur mandat en **Abidjan** et les avaient menacés de représailles, le MPIGO estimant notamment que la **France** s'était engagée dans une "guerre qui n'est pas la sienne" en faisant "le choix" de combattre les rebelles ivoiriens. Réagissant à ces incidents, le chef d'état-major des Armées françaises, le général **Henri Benítez**, avait alors expliqué que ses hommes avaient riposté dans le cadre de la légitime défense et pour protéger les ressortissants étrangers vivant dans la zone, conformément à leur mission. Le MPIGO et le Mouvement pour la justice et la paix (MJP), un autre groupe se réclamant du défunt chef de la junte, le général **Robert Guéi**, avait, le , finalement attribué l'accrochage à "une méprise" et précisé qu'il ne remettait "nullement en cause l'avenir des relations avec la **France**". Le , les militaires français, qui procédaient à l'évacuation de ressortissants étrangers de la région de **Bou** (ouest), s'étaient également heurtés à des combattants du MPIGO sur l'aéroport de la ville. Les échanges de tirs avaient fait un blessé côté français et au moins dix morts chez les rebelles. Le porte-parole du MPIGO, **Guillaume Gbato**, joint par téléphone depuis **Abidjan**, a indiqué ne pas pouvoir confirmer "pour le moment" ce nouvel accrochage, tout en ajoutant attendre un "compte-rendu de son état-major avant de se prononcer". Pendant ce temps, l'ambassadeur de **France** en Côte d'Ivoire, **Gildas Le Lidec**, s'est rendu à **Boussou**, deuxième ville du pays et fief des rebelles du Mouvement patriotique de Côte d'Ivoire (MPCI), qui tiennent la moitié nord du pays, pour y rencontrer certains de ses responsables. Aucune précision n'a pu être obtenue sur les points abordés durant cette rencontre. L'armée française a par ailleurs indiqué que le village de **Boussou**, une position du MPCI à 60 km au sud-ouest de Yamoussoukro, avait été attaquée récemment par un hélicoptère de combat. Des militaires français sont déployés entre **Boussou** et **Boussou** pour contrôler le cessez-le-feu conclu le par le MPCI. La **France** a renforcé mi-son dispositif militaire en Côte d'Ivoire, précisant que ses soldats avaient désormais pour consigne "d'ouvrir le feu" sur "toute personne empêchant manifestement la réalisation" de leur mission ou commettant "des exactions". Les derniers renforts français, 300 soldats et du matériel, doivent débarquer au **Fort d'Abidjan**, ce qui portera à entre 2.300 et 2.500 hommes les effectifs de l'Opération Licorne. bur-ayv/ban/cis

(a) Une dépêche commentée

A timeline from 1974 to present surrounding the BTK case

A timeline from 1974 to present surrounding the BTK case:

- Jan. 15, 1974: Joseph Otero, 38, and his wife, Julie, 34, are strangled in their home along with two of their children, Josephine, 11, and Joseph II, 9.
- April 4, 1974: Kathryn Bright, 21, is stabbed to death in her home. Police later conclude she was a BTK victim.
- October 1974: The Wichita Eagle-Beacon gets a letter from someone taking responsibility for the Otero family killing and including crime scene details.
- March 17, 1977: Shirley Vian, 24, is found tied up and strangled at her home.
- Dec. 8, 1977: Nancy Fox, 25, is found tied up and strangled in her home. The killer's voice is captured on tape when he calls a dispatcher to report the crime.
- Jan. 31, 1978: A poem referring to the Vian killing is sent to The Wichita Eagle-Beacon.
- Feb. 10, 1978: A letter from BTK is sent to KAKE-TV claiming responsibility for the deaths of Vian and Fox, as well as another unnamed victim. Police Chief Richard LaMunyon announces a serial killer is at large and has threatened to strike again.

(b) Une chronologie

FIG. 4 – Exemples de dépêches

- Les revues d'opinion (rapport des réactions de plusieurs intervenants à un même sujet) ;
- Les rapports de discours ;
- Les chronologies (se différencient des dépêches en feuilleton par un style plus concis et un marquage temporel explicite), cf fig 4 ;
- Les dépêches comparatives (comparaison d'un état de fait en divers lieux, époques...) ;
- Les dépêches énumératives.

Les trois dernières catégories sont intéressantes pour un système de résumé automatique. En effet, alors qu'elles ne représentent que 5% du corpus AQUAINT-2, elles contiennent 80% des informations pertinentes dans les données d'entraînement de la tâche « Update » qui utilisent un extrait d'AQUAINT-2. De plus, les dépêches appartenant à ces catégories sont écrites dans un style concis, ce qui fait des phrases qui les composent des candidats parfaits à l'intégration dans le résumé automatique. Elles ont également des caractéristiques bien spécifiques qui les rendent facilement identifiables :

- Les paragraphes des chronologies commencent presque tous par une référence temporelle ;
- les chronologies commencent souvent par une expression-clé telle que « Here is a timeline of events surrounding the election : » ;
- Les dépêches comparatives et énumératives contiennent des listes bien structurées ;
- Les éléments d'une liste issue d'une dépêche comparative débutent par des termes qui appartiennent à un même cadre sémantique (par exemple, des noms de pays) ;

Nous avons implémenté un classifieur qui classe les dépêches en quatre groupes : chronologies, dépêches comparatives, dépêches énumératives et dépêches « classiques ». Nous projetons de développer un système plus complet qui gère tous les types de dépêches que nous avons

identifiés, ainsi que ceux dégagés par (Lucas, 2005).

Nous avons évalué notre classifieur sur une petite partie (300 documents) du corpus AQUAINT-2 annotée à la main. Nous obtenons 100% de précision et 81% de rappel pour les chronologies, 73% de précision et 65% de rappel pour les dépêches comparatives, et 65% de précision et 67% de rappel pour les dépêches énumératives. Nous expliquons les résultats mitigés obtenus sur les deux dernières catégories par la proximité structurelle des dépêches comparatives et énumératives et par la confusion que cela engendre pour notre classifieur.

Nous avons intégré les résultats de notre classifieur au système CBSEAS. Dans un premier temps, les dépêches classées « chronologie », « comparative » ou « énumérative » se voient attribuer un « thème ». Nous définissons le thème comme la concaténation du titre de la dépêche et de la phrase d'introduction de la chronologie ou des comparaisons/énumérations. Si le thème d'une dépêche est assez proche de la requête utilisateur, alors CBSEAS favorise les phrases qui en sont issues, en augmentant leur score de 15%. Nous avons comparé les scores ROUGE-SU4 des résumés des groupes de documents qui contiennent au moins une dépêche classée « non classique » et dont le thème se rapprochait de la requête, et avons constaté une amélioration de 10% par rapport aux résultats de CBSEAS v0.5. Ces résultats nous encouragent à intégrer la structure des dépêches de manière plus globale à notre système de résumé automatique.

7 Perspectives

Notre classifieur de dépêches doit être amélioré : la méthode utilisée ne reconnaît pour le moment que trois catégories, qui sont les plus simples à identifier. Les autres catégories ont également des caractéristiques qui leur sont propres ; la relation structure/importance des phrases diffère d'une catégorie de dépêche à une autre. La structure des dépêches et la temporalité étant fortement liées, nous envisageons d'appliquer des techniques d'apprentissage comme les SVM à des documents annotés temporellement pour détecter les types des dépêches et en dégager la structure automatiquement.

Un deuxième axe de recherche concerne le système de résumé automatique lui-même. En effet, CBSEAS utilise des poids, notamment pour attribuer un score à chaque phrase, qui ont été établis manuellement. Nous voudrions trouver automatiquement les poids qui optimisent la qualité des résumés. Pour cela, nous projetons d'utiliser un algorithme de recherche dans l'espace des paramètres qui fixe les poids de manière à maximiser un score automatique calculé par rapport à un résumé de référence.

8 Conclusion

Dans cet article, nous avons présenté une nouvelle approche au résumé automatique multi-document. Elle utilise une méthode de classification non supervisée pour regrouper les phrases en classes sémantiques. Cette approche peut être comparée aux approches qui utilisent le voisinage des phrases comme critère de sélection (Erkan & Radev, 2004), car les phrases qui sont fortement reliées à un grand nombre d'autres phrases sont celles qui ont la plus grande probabilité d'être extraites. Cependant, notre approche diffère en un point crucial : la sélection des phrases est directement dépendante de la détection de redondance. Un deuxième point de di-

vergence est la méthode d'élimination de la redondance, qui a lieu dans CBSEAS avant la sélection finale des phrases qui constitueront le résumé. De plus, notre méthode permet la détection/élimination de la redondance de manière non supervisée sans définition d'un seuil de similarité entre les phrases au-delà duquel deux phrases sont considérées redondantes.

Nous avons également proposé une manière d'améliorer la qualité des résumés de dépêches, en utilisant la structure spécifique de ce type de documents. Nous avons montré, en intégrant des traits de structure basiques, que la prise en compte de la structure de tels documents augmentait réellement la qualité des résumés générés.

Références

- BARZILAY R. (2003). *Information Fusion for Multidocument Summarization : Paraphrasing and Generation*. PhD thesis.
- BOSSARD A. & GÉNÉREUX M. (2009). R  sum   automatique de textes d'opinion. In *Actes de la 16eme conf  rence sur le Traitement Automatique des Langues (TALN)*, Senlis, France.
- BOUDIN F., BECHET F., EL-BEZE M., FAVRE B., GILLARD L. & TORRES-MORENO J.-M. (2007). The LIA summarization system at DUC-2007. In *Proceedings of DUC 2007 Document Understanding Conference*.
- CARBONELL J. & GOLDSTEIN J. (1998). The use of MMR, diversity-based reranking for reordering documents and producing summaries. In *Proceedings of the 21st annual international ACM SIGIR conference*, p. 335–336, New York, NY, USA : ACM.
- EDMUNDSON H. P. & WYLLYS R. E. (1961). Automatic abstracting and indexing—survey and recommendations. *Commun. ACM*, 4(5), 226–234.
- ERKAN G. & RADEV D. R. (2004). Lexrank : Graph-based centrality as salience in text summarization. *Journal of Artificial Intelligence Research (JAIR)*.
- FRY E. B., KRESS J. E. & FOUNTOUKIDIS D. L. (2000). *The Reading Teachers Book of Lists*. Jossey-Bass, 4th edition.
- GOLDBERG A. (2007). Cs838-1 advanced NLP : Automatic summarization.
- LIN C.-Y. (2004). Rouge : a package for automatic evaluation of summaries. In *Proceedings of the Workshop on Text Summarization Branches Out (WAS 2004)*, Barcelona, Spain.
- LIN C.-Y., CAO G., GAO J. & NIE J.-Y. (2006). An information-theoretic approach to automatic evaluation of summaries. In *Proceedings of HLTC NACACL*, p. 463–470, Morristown, NJ, USA : Association for Computational Linguistics.
- L  PEZ-ESCOBAR S., CARRASCO-OCHOA J. A. & TRINIDAD J. F. M. (2006). Fast global k -means with similarity functions algorithm. In *IDEAL*, volume 4224 of *Lecture Notes in Computer Science*, p. 512–521 : Springer.
- LUCAS N. (2005). The enunciative structure of news dispatches, a contrastive rhetorical approach. In *Proceedings of the ASLA Conference*, p. 154–164.
- LUHN H. (1958). The automatic creation of literature abstracts. *IBM Journal*, 2(2), 159–165.
- RADEV D. (2004). MEAD - a platform for multidocument multilingual text summarization. In *Proceedings of LREC 2004*, Lisbon, Portugal.
- SALTON G. & MCGILL M. J. (1986). *Introduction to Modern Information Retrieval*. New York, NY, USA : McGraw-Hill, Inc.