Département d'informatique
Institut Galilée

École doctorale Galilée

Contribution au résumé automatique multi-documents

THÈSE

présentée et soutenue publiquement le 12/07/2010

pour l'obtention du

Doctorat de l'Université Paris-Nord – Paris 13
(spécialité informatique)

par

Aurélien Bossard

Composition du jury

Rapporteurs :	Guy Lapalme, Professeur, Université de Montréal
	Juan-Manuel Torres Moreno, Maître de conférences HDR, Université d'Avignon
Examinateurs :	Anne Vilnat, Professeur, IUT d'Orsay
	Céline Rouveirol, Professeur, Université Paris 13
	Daniel Kayser, Professeur, Université Paris 13
Directeur :	Thierry Poibeau, Chargé de recherche CNRS, LaTTiCe

Laboratoire d'Informatique de Paris-Nord – CNRS UMR 7030

Remerciements

Je tiens à remercier tout d'abord Thierry Poibeau, qui m'a encadré depuis le Master 2, soit plus de 4 années, pour ses conseils, sa présence et son suivi qui m'ont permis de suivre ma propre voie tout en veillant à ce que je ne m'égare pas. Il a contribué à faire de mes années de thèse une expérience enrichissante. Je n'oublie pas Daniel Kayser, dont l'expérience, les conseils et les relectures attentives ont été une aide précieuse.

Je remercie Juan-Manuel Torres Moreno et Guy Lapalme pour avoir accepté d'être les rapporteurs de cette thèse, ainsi qu'Anne Vilnat et Céline Rouveirol pour leur participation au jury.

Mes remerciements vont également aux personnes de mon équipe, RCLN, qui ont su me guider depuis mes premiers pas en recherche. De plus, la qualité des formations qu'ils ont su mettre en place dans le master MICR m'ont fourni les outils nécessaires à la réalisation d'une thèse dans un domaine aussi ouvert que le traitement automatique du langage naturel. Je remercie également le LIPN, pour m'avoir fait confiance et permis d'accomplir une thèse dans d'aussi bonnes conditions de travail.

Comment ne pas citer mes collègues de bureau et amis, Christophe et Thibault, qui ont réussi à rendre l'atmosphère de travail à la fois saine et joviale. Merci à eux !

Je tiens aussi à exprimer toute ma gratitude à ma famille, notamment mes parents qui m'ont toujours soutenu durant mes études et dans toutes mes activités, ma petite sœur dont la rigueur a été d'une graide aide lorsqu'elle a relu ma thèse d'un œil extérieur au TAL, et à la dernière venue, Maroussia, dont la patience a été rudement mise à l'épreuve.

Je finirai par remercier mes amis pour leur joie de vivre et les aventures que l'on aura pu vivre ensemble : Jeff, Christophe, Charlotte et tous les autres qui ont contribué et contribueront encore, je l'espère, à rendre la vie plus agréable.

Résumé

Résumer un texte consiste à réduire ce texte en un nombre limité de mots. Le texte ainsi réduit doit rester fidèle aux informations et idées du texte original. Que ce soit pour des professionnels qui doivent prendre connaissance du contenu de documents en un temps limité ou pour un particulier désireux de se renseigner sur un sujet donné sans disposer du temps nécessaire pour lire l'intégralité des textes qui en traitent, le résumé est une aide contextuelle importante. Avec l'augmentation de la masse documentaire disponible électroniquement, résumer des textes automatiquement est devenu un axe de recherche important dans le domaine du traitement automatique de la langue. La production automatique de résumés pose le problème de la détection et de la modélisation des informations contenues dans les textes. Elle suppose également la hiérarchisation de ces informations afin d'intégrer au résumé les plus importantes. Cette thèse de doctorat propose une méthode statistique pour le résumé automatique par extraction ainsi que l'intégration d'analyses linguistiques au processus de sélection de phrases.

La méthode que nous proposons est fondée sur une classification des phrases à résumer en classes sémantiques en utilisant des calculs de similarité entre les phrases. Cette étape nous permet d'identifier les phrases qui risquent de présenter des éléments d'information similaires et ainsi de supprimer toute redondance du résumé généré. Une seconde étape vise à sélectionner une phrase par classe, en tenant compte de la similarité des phrases à une éventuelle requête utilisateur, de la longueur des phrases ainsi que de la centralité dans leur classe. Les résumés ainsi générés doivent maximiser la centralité et la diversité des informations. Cette méthode a été évaluée sur deux tâches de la campagne d'évaluation TAC 2008 : le résumé de dépêches et le résumé d'opinions issues de blogs. Les résultats mitigés sur la première tâche et encourageants sur la deuxième nous ont poussé à prendre en compte des critères de sélection de phrases spécifiques aux types de documents traités. Nous avons alors proposé d'établir une catégorisation des dépêches de presse ainsi que l'annotation automatique de leur structure afin d'améliorer la qualité des résumés générés par notre système. Nous avons également étudié l'apport de l'annotation en entités nommées et de la résolution d'anaphores pour le résumé automatique. Le système et ces trois derniers modules ont été évalués sur la tâche de résumé et mise à jour de résumé de dépêches de la campagne TAC 2009, se classant dans le premier quart des participants. Notre méthode de résumé a également fait l'objet d'une intégration à un système applicatif plus large visant à aider un possesseur de corpus à visualiser les axes essentiels et à en retirer automatiquement les informations importantes.

Abstract

Summarizing a textual document consists in compressing the text in a limited number of words. The compressed text must remain faithful to the information and ideas from the initial text. Professionals who have to peruse documents in a limited amount of time or private individuals who want to be informed about a specific topic without having the time to read all the texts about it both need summaries. The increase in electronic documents available have made the research in automatic summarization an important domain in the field of natural language processing. Producing automatic summaries depends on textual information detection and modelling. Generating good automatic summaries also depends on information hierarchization in order to put only the most important information in the summaries. This PhD Thesis proposes a statistical method to generate automatic extracts, and the integration of linguistic analysis to the sentences selection process.

The method we propose is based on a sentence classification in semantic clusers, using similarity calculation between sentences. This step allows us to identify the sentences which convey the same information and to remove redundancy from the automatically generated summaries. A second step aims to select one sentence per cluster, taking into account the similarity to a user query, the sentences length and the centrality within their cluster. The generated summaries must maximize the centrality and diversity of the information they convey. This method has been evaluated on two different tasks of the evaluation campaign TAC 2008 : news summarization and opinion summarization. The mixed results on the first task led us all the more to take in account sentences selection criterion specific to the documents to summarize, since the results on the second task were encouraging. We then proposed to establish a newswire articles categorization as well as automatic structure tagging in order to improve the quality of the summaries produced by CBSEAS. We also studied the named entity tagging and anaphora resolution contribution to the summaries quality. CBSEAS and the three modules described above have been evaluated on the « Update » summarization task for newswire articles of the TAC 2009 evaluation campaign, ranking itself among the the first quarter of the TAC 2009 participating systems. Our summarization method has also been integrated to a larger application which aims to help the user to visualize the main topics of a corpus and to automatically extract the essential information.

Table des matières

Table des figures

Liste des tableaux

Introduction

Sommaire

Résumer un texte consiste à réduire ce texte en un nombre limité de mots. Le texte ainsi réduit doit rester fidèle aux informations et idées du texte original, et dans la mesure du possible rendre compte du style et de l'intention de l'auteur. Cette discipline, quoique très ancienne, est mal formalisée. Le processus de résumé est en effet dépendant à la fois du type de texte à résumer et de l'utilisation qui en sera faite. Ainsi, un résumé de type rapport d'activités sera dans la forme comme dans le fond radicalement différent d'un résumé d'une oeuvre littéraire, d'un résumé d'ouvrage scientifique, d'un résumé de dépêches ou d'une revue de presse.

Les premiers systèmes de résumé automatique étaient dédiés au résumé mono-document. Cependant, l'explosion de la masse documentaire disponible en ligne et la numérisation de tous types de documents ont conduit les acteurs du TAL à se pencher sur la problématique du résumé multi-documents, c'est-à-dire du résumé des informations contenues dans plusieurs documents. Celui-ci peut en effet être vu comme une couche entre des systèmes de recherche d'information qui renvoient à un utilisateur les documents les plus pertinents par rapport à sa requête, et la prise de connaissance par l'utilisateur des informations contenues dans ces documents. Dans le cadre de tâches comme la veille d'information, la lecture de multiples documents peut se révéler fastidieuse. Un bon système de résumé automatique multi-documents est alors une aide précieuse pour des analystes en quête d'efficacité. La campagne d'évaluation de « *Text Analysis Conference* » organisée par le « *National Institute of Science and Technology* » témoigne de l'intérêt porté au résumé automatique multi-documents.

Le travail de cette thèse se situe dans le cadre du résumé automatique multi-documents et de son application à travers des tâches diverses, comme le résumé d'opinions, le résumé de dépêches ou l'aide à l'analyse de corpus.

Problématique

Malgré le nombre conséquent de travaux sur le domaine et de campagnes d'évaluation telles que TAC, soutenue par le gouvernement américain, résumer au mieux un ou plusieurs textes automatiquement reste un problème ouvert. Synthétiser plusieurs documents donne plus d'importance à un aspect linguistique : la redondance. Une même information aura une probabilité plus élevée d'être énoncée à plusieurs reprises dans plusieurs documents traitant d'un même sujet, que dans un document unique. Cela engendre un risque plus important de sélectionner deux informations identiques pour les intégrer à un résumé automatique, mais permet également d'identifier les informations les plus importantes, celles-ci faisant l'objet de reprises dans les documents à synthétiser. Ce constat a fait naître une série d'approches qui tentent de modéliser au mieux le phénomène, mais n'apportent selon nous que des solutions partielles. Si certaines d'entre elles détectent la redondance et s'en servent à des fins de sélection d'information tout en éliminant le risque de voir apparaître les mêmes éléments plusieurs fois au sein du résumé (par exemple Barzilay et McKeown (2005)), elles sont dépendantes de traitements linguistiques poussés qui en réduisent la portée applicative.

Dans cette thèse, nous avons voulu définir une nouvelle approche de résumé automatique multi-documents la plus générique possible, fondée sur l'observation de la redondance. Cela nous interdit notamment l'utilisation de ressources trop spécifiques à un domaine ou à une langue. Nous avons également voulu en cerner les limites d'application à des domaines spécialisés et proposer des solutions qui permettent de l'adapter à trois tâches différentes :

- le résumé d'opinions,
- le résumé de dépêches,
- l'aide à l'analyse de corpus.

Ces domaines suscitent un intérêt particulier de la part des industriels. Le premier leur permet de suivre *via* le résumé de ressources ouvertes en ligne, telles que les blogs ou les forums, les opinions des participants à ces ressources vis-à-vis de leur produit ou de leur image. Le second, étudié depuis les débuts du résumé automatique (Edmundson, 1969), aide les veilleurs d'information à se tenir au courant des dernières informations concernant un sujet précis, ou à obtenir une synthèse des différents événements de la journée. Le troisième s'adresse aux industriels disposant de masses de données textuelles trop importantes pour en prendre connaissance rapidement, et propose des solutions de visualisation des informations et d'accès au contenu.

Afin d'évaluer au mieux le travail effectué, nous nous sommes fixé pour objectif la participation à une campagne d'évaluation internationale. La campagne TAC [1], qui a

1. Text Analysis Conference

remplacé les campagnes DUC [2] et qui réunit des équipes de recherche du monde entier s'est imposée à nous du fait de l'importance qu'elle accorde au résumé automatique.

Apports

L'apport principal de la présente thèse est la définition et l'implémentation d'une nouvelle approche pour le résumé automatique multi-documents par extraction. Le résumé par extraction consiste à extraire des pans de texte jugés pertinents depuis les documents à synthétiser et à les assembler pour former le résumé. Cette approche se différencie de l'existant par l'utilisation d'une étape préalable à la sélection des phrases à extraire qui véhiculent les mêmes informations. Ainsi, un modèle des documents à résumer est établi qui vise notamment à éviter de sélectionner plusieurs phrases sémantiquement équivalentes et à établir des traitements qui améliorent la cohérence des résumés automatiques.

Ce travail présente d'autres apports, comme l'utilisation de la structure discursive des dépêches de presse pour améliorer la qualité des résumés, le couplage de notre système de résumé, CBSEAS [3], à un outil d'analyse d'opinions et l'utilisation d'un algorithme génétique pour paramétrer au mieux CBSEAS.

Nous introduisons également une nouvelle méthode pour la gestion de la mise à jour de résumé qui se sert du modèle établi par CBSEAS pour détecter les informations nouvelles, ainsi qu'un procédé d'aide à l'analyse de corpus de documents hétérogènes en contenu, et une expérience visant à évaluer l'impact de l'utilisation d'outils de résolution d'anaphores sur la qualité des résumés générés par CBSEAS.

Plan de thèse

La thèse est organisée de la manière suivante : dans un premier temps, nous définissons les différentes formes de résumé puis proposons un état de l'art des approches de génération automatique de résumés. Nous nous sommes concentrés sur la description des approches les plus importantes ainsi que celles qui ont le plus influencé notre travail.

Nous présentons ensuite notre approche générique et le système qui l'implémente, CBSEAS, ainsi que l'intégration à ce dernier de techniques d'analyse discursive pour améliorer la qualité des résumés qu'il produit.

2. Document Understanding Conference
3. Clustering Based Sentence Extractor for Automatic Summarization

La troisième partie est consacrée à l'étude de la production par CBSEAS de résumés automatiques pour trois applications différentes : le résumé et la mise à jour de résumés de dépêches, le résumé d'opinions et l'aide à l'analyse de corpus de documents à contenu informationnel hétérogène. Afin de valider notre approche, les deux premiers cas d'étude ont fait l'objet d'une participation à une campagne d'évaluation internationale dont nous présentons les résultats.

Nous clorons cette thèse en présentant nos conclusions générales sur nos réalisations et en exposant nos perspectives de recherche.

Première partie .

État de l'art

L'objectif de cette partie est de présenter l'arrière plan théorique de notre travail, ainsi que les travaux antérieurs et notre positionnement par rapport à ceux-ci. Le premier chapitre est consacré à la définition de la notion de résumé et à la présentation des travaux les plus influents dans le domaine du résumé automatique. Nous proposons de répondre à une problématique qui nous est apparue mal gérée dans les travaux existants : représenter la redondance dans les documents à résumer afin de l'éliminer du résumé à générer et dans un même temps de sélectionner les éléments les plus pertinents pour le résumé. Dans le deuxième chapitre, nous présentons les méthodes d'évaluation de résumé automatique existantes, et montrons qu'il est encore nécessaire, en l'état actuel des travaux, de procéder à des évaluations manuelles afin d'avoir une mesure concrète et non faussée de la qualité d'un résumé.

1. État de l'art du résumé automatique

Sommaire

Créer automatiquement un résumé d'un ou plusieurs documents compréhensible par un lecteur humain est une problématique étudiée depuis les années 1950 (Mani et Maybury, 1999). Deux types d'approche s'opposent : l'approche par génération et l'approche par extraction, qui consiste à extraire des documents les phrases ou morceaux de phrases les plus pertinents et à les assembler pour créer un résumé.

L'approche par génération nécessite trois étapes :

- la représentation des connaissances contenues dans les textes à résumer ;
- la sélection/condensation de ces connaissances ;
- la génération d'un résumé à partir de ces connaisances.

Passée la première étape d'extraction des connaissances, autrefois improbable du fait du manque de ressources syntaxiques et sémantiques, il est nécessaire de sélectionner les connaissances nécessaires à la construction d'un résumé qui corresponde à l'attente informationnelle du lecteur. L'importance d'une information est fonction du domaine traité. Par exemple, dans le cas d'un résumé de dépêches liées à une catastrophe naturelle,

le système devrait extraire le type de la catastrophe, les dégâts humains et matériels, les régions touchées... Sélectionner les connaissances extraites des textes nécessite donc d'établir une modélisation spécifique du ou des domaines traités. Le champ d'application de telles techniques est par conséquent limité ; ainsi, la recherche sur la création de résumés par extraction a connu un renouveau, en parallèle à l'abandon de l'approche par génération (Minel, 2003; Sparck Jones, 1998).

Dans ce chapitre, nous présentons un aperçu des techniques de résumé automatique par extraction. Dans un premier temps, nous abordons différentes catégories de résumé ainsi que les domaines sur lesquels les outils de résumé automatique sont mis à l'épreuve. Nous traitons ensuite les solutions au problème de la sélection de phrases avant d'exposer les différentes techniques de post-traitement linguistique : compression de phrases et réordonnancement.

1.1. Types de résumés visés dans la thèse

Mettre au point un système de résumé automatique performant nécessite dans un premier temps de savoir ce que l'on entend par résumer et quels sont les objectifs de cette tâche.

Différents travaux, dont (Jones, 1993) et (Minel, 2003) ont tenté de définir une typologie des résumés. Il en ressort trois types définitoires qui peuvent servir de base à la spécification d'un cadre d'application formel d'automatisation de la création de résumés : le résumé indicatif, le résumé informatif et le résumé synthétique. Les définitions qui suivent ne constituent pas une typologie, mais tentent de décrire au mieux différentes catégories de résumés, dont certaines se recoupent.

1.1.1. Le résumé indicatif

Le résumé indicatif d'un document est une présentation abrégée d'un document dans laquelle doivent figurer le ou les thèmes de ce document. Il doit également indiquer la structure du texte, sans détailler l'argumentation. La norme française NF Z 44-004 définit le résumé indicatif comme « un mode de description externe [...] à utiliser essentiellement pour des textes [...] trop détaillés pour permettre la rédaction d'un résumé informatif, par exemple : articles monographiques, synthèse bibliographique. Le résumé indicatif renseigne le lecteur sur les thèmes étudiés. Il s'apparente à une table des matières. Il peut cependant s'enrichir de parties informatives mettant en évidence des éléments significatifs. »(Normes françaises, 1984).

D'après (Aït El Mekki et Nazarenko, 2004), l'index de fin de livre, outil traditionnel d'accès à l'information, s'apparente aux résumés indicatifs.

1.1.2. Le résumé informatif

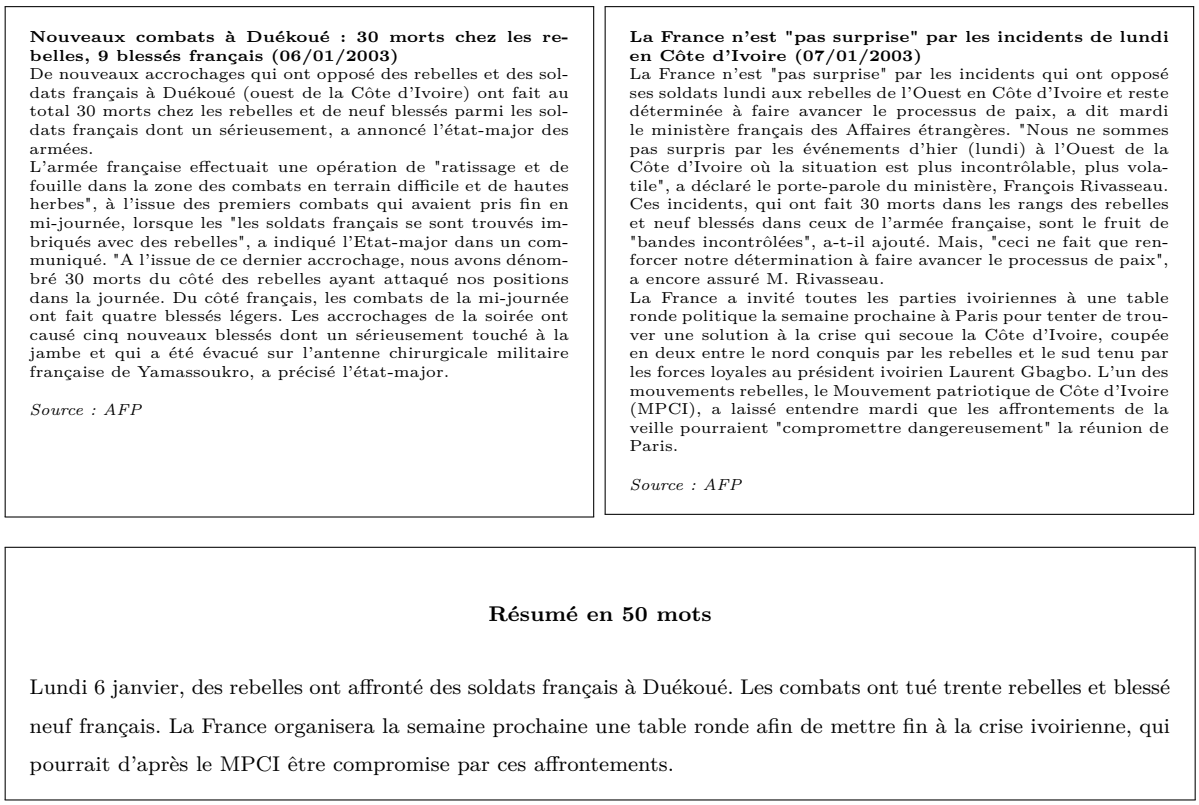
La principale fonction d'un résumé informatif est d'apporter au lecteur les informations essentielles présentes dans le document analysé.

Toujours selon la norme française NF Z 44-004, le résumé informatif est une « représentation abrégée du document, renseignant sur les informations quantitatives ou qualitatives apportées par l'auteur. Ce résumé doit constituer un texte autonome d'une logique rigoureuse. Il forme avec le titre du document un ensemble qui, en principe, ne doit pas être redondant. Les informations retenues pour le résumé sont généralement présentées selon leur ordre d'apparition dans le document. Cet ordre facilite l'exploitation du résumé par le lecteur habitué au plan des articles publiés dans sa spécialité. Généralement, les documents scientifiques et techniques exposent séquentiellement le but de l'étude dans l'introduction, le matériel et les méthodes utilisées, les résultats obtenus, une discussion ou une conclusion évaluant la signification et la pertinence de l'apport. Cependant, en ne négligeant aucune phase du cheminement, les diverses parties du document pourront figurer de façon inégale dans le résumé en fonction de l'importance ou de la nouveauté de l'information. »(Normes françaises, 1984).

1.1.3. Le résumé synthétique

Le résumé synthétique constitue une synthèse de différents documents. Celui-ci existe sous plusieurs formes : résumé critique comme notamment les états de l'art, résumé informatif où l'on cherchera à extraire les informations principales réparties dans plusieurs documents...

La synthèse est le type de résumé traité dans cette thèse. Nous nous intéressons en effet ici au résumé multi-documents, qui peut s'apparenter à la définition de la synthèse selon le TLF (Trésor de la Langue Française) : « Opération consistant à rassembler des éléments de connaissance sur un sujet, une discipline et à donner une vue générale, une idée d'ensemble de ce sujet ».



Nouveaux combats à Duékoué : 30 morts chez les rebelles, 9 blessés français (06/01/2003)
De nouveaux accrochages qui ont opposé des rebelles et des soldats français à Duékoué (ouest de la Côte d'Ivoire) ont fait au total 30 morts chez les rebelles et de neuf blessés parmi les soldats français dont un sérieusement, a annoncé l'état-major des armées.
L'armée française effectuait une opération de "ratissage et de fouille dans la zone des combats en terrain difficile et de hautes herbes", à l'issue des premiers combats qui avaient pris fin en mi-journée, lorsque les "les soldats français se sont trouvés imbriqués avec des rebelles", a indiqué l'Etat-major dans un communiqué. "A l'issue de ce dernier accrochage, nous avons dénombré 30 morts du côté des rebelles ayant attaqué nos positions dans la journée. Du côté français, les combats de la mi-journée ont fait quatre blessés légers. Les accrochages de la soirée ont causé cinq nouveaux blessés dont un sérieusement touché à la jambe et qui a été évacué sur l'antenne chirurgicale militaire française de Yamassoukro, a précisé l'état-major.

Source : AFP

La France n'est "pas surprise" par les incidents de lundi en Côte d'Ivoire (07/01/2003)
La France n'est "pas surprise" par les incidents qui ont opposé ses soldats lundi aux rebelles de l'Ouest en Côte d'Ivoire et reste déterminée à faire avancer le processus de paix, a dit mardi le ministère français des Affaires étrangères. "Nous ne sommes pas surpris par les événements d'hier (lundi) à l'Ouest de la Côte d'Ivoire où la situation est plus incontrôlable, plus volatile", a déclaré le porte-parole du ministère, François Rivasseau. Ces incidents, qui ont fait 30 morts dans les rangs des rebelles et neuf blessés dans ceux de l'armée française, sont le fruit de "bandes incontrôlées", a-t-il ajouté. Mais, "ceci ne fait que renforcer notre détermination à faire avancer le processus de paix", a encore assuré M. Rivasseau.
La France a invité toutes les parties ivoiriennes à une table ronde politique la semaine prochaine à Paris pour tenter de trouver une solution à la crise qui secoue la Côte d'Ivoire, coupée en deux entre le nord conquis par les rebelles et le sud tenu par les forces loyales au président ivoirien Laurent Gbagbo. L'un des mouvements rebelles, le Mouvement patriotique de Côte d'Ivoire (MPCI), a laissé entendre mardi que les affrontements de la veille pourraient "compromettre dangereusement" la réunion de Paris.

Source : AFP

Résumé en 50 mots

Lundi 6 janvier, des rebelles ont affronté des soldats français à Duékoué. Les combats ont tué trente rebelles et blessé neuf français. La France organisera la semaine prochaine une table ronde afin de mettre fin à la crise ivoirienne, qui pourrait d'après le MPCI être compromise par ces affrontements.

FIGURE 1.1.: Exemple d'un résumé informatif

1.1.4. L'Extrait

L'extrait est une forme particulière de résumé. Il consiste non pas en une reformulation d'un ou plusieurs textes d'origine, mais en une compilation de morceaux extraits des textes d'origine. Les revues de presse sont une sorte de résumé par extraction. Le résumé par extraction n'est pas courant pour ce qui concerne les résumés construits manuellement. C'est en revanche un procédé largement utilisé dans les systèmes de résumé automatique (Mani et Maybury, 1999).

1.2. Domaines d'application et enjeux du résumé automatique

Les cadres d'application du résumé automatique sont multiples, et suivent la demande industrielle. Les systèmes de résumé automatique ont longtemps été focalisés sur les dépêches de presse, pour notamment de l'aide à la veille d'information (Mani et Wilson,

2000; Radev *et al.*, 2001b; McKeown *et al.*, 2002) mais également sur les articles scientifiques (Boudin *et al.*, 2008c; Teufel et Moens, 2002) avec des applications évidentes à la veille technologique.

Récemment, l'apparition de nouveaux contenus en ligne, tels que les blogs, permet à tout un chacun de publier ses opinions à propos d'événements récents, de produits culturels et commerciaux, de personnalités... Les industriels, mais également les personnalités médiatiques, ont bien compris qu'extraire de telles ressources accessibles gratuitement les opinions concernant leurs produits ou leur image, et les résumer, peut leur apporter un avantage non négligeable sur la concurrence. La recherche s'est donc récemment portée sur les résumés de blogs (Zhou et Hovy, 2006; Hu *et al.*, 2007; Généreux, 2009a) et d'opinions (Beineke *et al.*, 2003; Hu et Liu, 2004; Stoyanov, 2005; Feiguina et Lapalme, 2007; Généreux, 2009a).

Le résumé automatique d'e-mails ou de fils d'e-mails a également été étudié (Wan et McKeown, 2004; Zajic *et al.*, 2008; Rambow *et al.*, 2004), tout comme le résumé de dialogues écrits ou oraux (Zhou et Hovy, 2005b,a; Gurevych et Strube, 2004; Zechner, 2002), ou le résumé de textes juridiques (Farzindar et Lapalme, 2005)

Ces études montrent que la spécificité des documents à résumer implique l'utilisation de techniques de résumé différentes, ou d'adaptation des techniques existantes. Ce fait a récemment été confirmé dans (Boudin, 2008). Il est nécessaire de mener une étude des types de documents que l'on cherche à synthétiser automatiquement préalablement au développement d'une approche de résumé efficace. Généralement, les développeurs de systèmes de résumé automatique disposent d'un outil de résumé standard sur lequel sont greffés des modules qui permettent d'adapter l'outil à un domaine particulier.

Les prochaines sections visent à décrire les différentes méthodes de sélection de phrases pour la génération de résumé indépendamment du domaine traité.

1.3. Méthodes d'analyse de surface

Dans ce type de méthode, les différentes phrases du ou des documents à résumer se voient attribuer un score qui peut être fondé sur un des critères ou une combinaison des critères suivants, établis par Edmundson (Edmundson, 1969), et repris notamment dans (Kupiec *et al.*, 1995) dans un système de résumé automatique par apprentissage :

– **Fréquence de mots** : si l'on exclut les termes trop fréquents, spécifiques au domaine traité, ainsi que les mots courants de la langue des documents à résumer, les termes et mots les plus fréquents sont représentatifs du thème de ces documents. Plus une phrase contient certains de ces termes, plus elle est représentative du

contenu des documents et plus elle a de chances d'être extraite.
- **Position** : la position d'une phrase dans un document peut, dans certains cas, se révéler un bon indicateur de son importance. Dans un article de presse, la première phrase sert de point d'appui à l'ensemble du document, et est considérée comme la plus importante (*cf* fig.1.2). Dans un article scientifique, ce sont les phrases de la conclusion qui sont les plus représentatives du document (avec les phrases du résumé si celui-ci existe).
- **Similarité au titre** : le titre est censé refléter le contenu d'un document. Une phrase dont le contenu est similaire à celui du titre du document — qui partage des termes similaires à ceux du titre — est donc centrale dans ce document.
- **Termes clefs** : certains termes/expressions rendent compte de la pertinence d'une phrase. Edmundson (Edmundson, 1969) cite parmi ces termes : *hardly*, *impossible*, *greatest*, *significant*.

Kofi Annan présente mardi au Conseil de sécurité l'accord de Marcoussis

Le secrétaire général des Nations Unies, Kofi Annan, présentera mardi au Conseil de sécurité réuni à huis clos l'accord de Marcoussis sur la Côte d'Ivoire intervenu entre les parties aux prises dans ce pays.

« M. Annan, a indiqué lundi son porte-parole, rendra compte au Conseil de sécurité mardi et discutera avec lui des mesures de suivi qu'il pourrait souhaiter prendre ». L'accord, obtenu vendredi dernier à Marcoussis près de Paris, après dix jours de négociations, maintient au pouvoir le président ivoirien Laurent Gbagbo en échange de l'entrée des rebelles dans un gouvernement de réconciliation nationale. Une présence de l'Onu sur le terrain afin de suivre l'application des engagements qui ont été pris serait nécessaire, a estimé une source diplomatique. Cette source, qui parlait sous le couvert de l'anonymat, a également estimé que cet accord nécessitait une présence militaire sur le terrain et que les Nations Unies devaient en être. Les Etats-Unis ont déjà fait savoir leur opposition à toute intervention de l'Onu en Côte d'Ivoire. La France, ancienne puissance coloniale en Côte d'Ivoire, y a actuellement déployé quelques 2.500 soldats.

FIGURE 1.2.: Illustration de l'importance de la première phrase dans les dépêches de presse : la première phrase est surlignée en orange.

Parmi les méthodes utilisant uniquement la fréquence des mots afin d'attribuer un score à une phrase et extraire les mieux classées, on peut citer (Luhn, 1958). Celui-ci construit à partir des documents à résumer un paquet de mots importants dans ces documents. Seuls les mots [1] dont la fréquence est située dans un certain intervalle sont considérés importants et intégrés dans ce paquet (*cf* fig.1.3). Les phrases contenant le

1. Les auteurs ne précisent pas si une lemmatisation préalable à l'analyse a été réalisée ; cependant, au vu de la date de l'étude, nous pouvons supposer qu'ils n'utilisent pas de lemmatisation.

plus de mots appartenant à ce paquet sont sélectionnées. Cette méthode comporte un défaut majeur : un terme est ajouté au paquet des termes importants si sa fréquence est située dans un intervalle donné. L'intervalle optimal dépend des documents à résumer et Luhn ne propose pas de solution pour faire varier cet intervalle selon les documents en entrée du système.

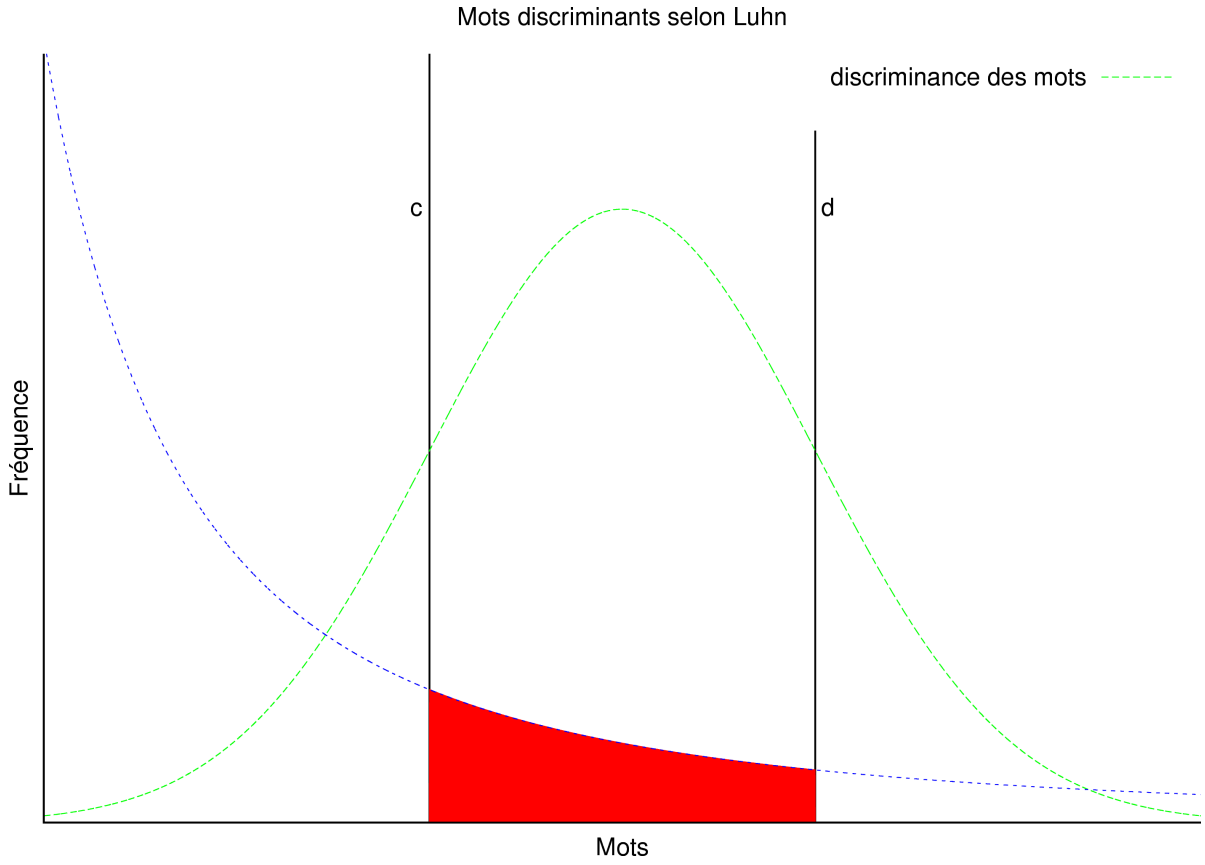


FIGURE 1.3.: Illustration de la discriminance des mots selon leur fréquence d'après Luhn

Les avancées en analyse statistique appliquée au texte ont permis de pallier en partie ce problème : Salton propose dans (Salton et McGill, 1983) une métrique, appelée *tf.idf* (*term frequency, inverse document frequency*), qui permet de déterminer l'importance d'un terme dans un ensemble de documents sans utilisation de ressources exogènes. Cette mesure est fondée sur l'observation que les termes les plus discriminants/importants sont ceux qui apparaissent le plus dans un faible nombre de documents du corpus. Le *tf.idf* d'un terme dans un document est alors la fréquence de ce terme dans le document considéré multiplié par l'inverse de la fréquence des documents dans lesquels le terme apparaît. Il existe plusieurs façons de calculer le *tf.idf* ; la plus courante est détaillée en figure 1.4.

Calcul du tf d'un terme t_i dans le document d_j :

$$\mathrm{tf}_{t_i,d_j} = \frac{|t_i|_j}{\sum_{k=0}^{n} |t_k|_{d_j}}$$

Calcul de l'idf d'un terme t_i au sein d'un corpus D de documents d_j :

$$\mathrm{idf}_{t_i} = \log \frac{|D|}{|d_j : t_i \in d_j|}$$

Calcul du tf.idf du terme t_i pour un document d_j :

$$\mathrm{tf.idf}_{t_i,d_j} = tf_{t_i,d_j} \times idf_{t_i}$$

FIGURE 1.4.: Calcul du *tf.idf*

Cette métrique a été appliquée de différentes manières dans le cadre du résumé automatique :

- Nobata *et al.* (2003) évaluent plusieurs façons de calculer le *tf.idf* au travers d'une application de résumé automatique. Les phrases obtiennent entre autres scores un score égal à la somme des *tf.idf* des mots qui les composent.
- Salgueiro Pardo *et al.* (2002) utilisent la moyenne des *tf.idf* des mots d'une phrase. Ces deux méthodes posent le problème de la taille des phrases. Dans le premier cas, les phrases les plus longues seront favorisées ; dans le deuxième cas, les phrases courtes éventuellement peu informatives mais composées uniquement de mots importants se verront avantagées.
- Radev *et al.* (2001a) créent quant à eux un *centroid*[2] –*i.e.* un vecteur composé des mots les plus importants dans un groupe de documents– en se fondant sur leur nombre moyen d'apparitions. Le score de chaque phrase est calculé en fonction de sa similarité au *centroid*, de sa position et de la similarité au titre. Il utilise, lors d'une étape post-opératoire, une métrique appelée CSIS (*cf* §1.6) afin d'éliminer les phrases considérées redondantes du résumé final.
- Saggion (2005) utilise une approche de sélection proche de (Radev *et al.*, 2001a). La différence principale réside dans le mécanisme final de sélection des phrases : après avoir extrait une liste de phrases candidates qui mises bout à bout créent un texte C, un algorithme itératif de sélection des phrases est appliqué. A chaque itération, l'algorithme établit la liste des textes composés de toutes les phrases moins une : celle qui maximise la similarité entre C et le texte dont elle a été supprimée.
- Boudin et Torres-Moreno (2007) extraient les phrases qui maximisent une combinaison de treize critères différents, dont la similarité au titre du document, la somme

2. Nous utiliserons dorénavant le mot centroïde

de l'entropie des différents termes de la phrase, et d'autres mesures de surface sur les phrases et les termes qui les composent.

1.4. Résumé et apprentissage

Les systèmes de résumé automatique par apprentissage combinent différentes caractéristiques pondérées, souvent issues des travaux décrits en §1.3. L'objectif de ces systèmes est alors de trouver la combinaison de poids pour chacune de ces caractéristiques qui permette de maximiser la qualité des résumés.

Kupiec *et al.* (1995) et Aone *et al.* (1998) utilisent les mêmes descripteurs de phrases qu'Edmundson (Edmundson, 1969) et optimisent le poids de chacun de ces paramètres grâce à un réseau bayésien. L'utilisation d'une telle méthode nécessite l'hypothèse d'indépendance des paramètres.

Yeh *et al.* (2002) ajoutent deux caractéristiques aux phrases : un score de centralité — l'intersection entre les mots-clé de la phrase et les mots-clé de l'ensemble des documents divisé par l'union de ces deux ensembles — et la similarité au titre du document dont la phrase est issue. La meilleure combinaison des paramètres est approximée en utilisant un algorithme génétique. Osborne (2002) utilise une descente de gradient pour optimiser les poids des paramètres.

Le système PYTHY (Toutanova *et al.*, 2007) utilise une approche dynamique étudiée pour trouver une solution au problème du sac-à-dos, décrite dans (McDonald, 2007) afin de trouver la meilleure combinaison de poids pour les différents paramètres qu'ils utilisent.

Toutes les approches présentées ci-dessus tentent de maximiser un score attribué automatiquement aux résumés en fonction de résumés de référence. Les métriques de qualité les plus utilisées sont des comparaisons d'unigrammes et bigrammes entre les résumés de référence et les résumés automatiques, ainsi que les métriques ROUGE, présentées en §2.1.

Ces méthodes fondées sur la fréquence de mots et divers critères de surface n'exploitent qu'une partie des informations contenues dans les documents. Le fait que plusieurs phrases soient fortement liées dans leur contenu, ce qui dénote leur importance, n'est pas exploité. De plus, ce type de systèmes reflète bien la centralité des phrases, mais ne reflète ni ne tire parti de la diversité (le fait que les phrases expriment des informations différentes). De ce constat sont nées d'autres approches, qui tiennent compte des liens que les phrases entretiennent les unes avec les autres.

1.5. Minimiser la redondance tout en maximisant la pertinence

MMR, pour *Maximal Marginal Relevance*, est un algorithme qui permet d'extraire les phrases les plus pertinentes vis-à-vis d'une requête afin de générer un résumé, tout en éliminant la redondance (Carbonell et Goldstein, 1998). Cet algorithme itératif sélectionne à chaque passe la phrase qui maximise la similarité à la requête tout en minimisant la similarité aux phrases déjà sélectionnées, selon la formule :

$$MMR = argmax_{p \in P} [\lambda \times sim_1(p, R) - (1 - \lambda) \times max_{s_i \in S} sim_2(p, s_i)]$$

- P est l'ensemble des phrases à résumer ;
- R la requête ;
- S l'ensemble des phrases déjà sélectionnées ;
- sim_1 la mesure de similarité entre une phrase et la requête ;
- sim_2 la mesure de similarités entre phrases ;
- λ le coefficient permettant de régler l'importance de la redondance vis-à-vis de la pertinence.

FIGURE 1.5.: Formule de MMR

Cet algorithme de sélection de phrases a été dérivé sous diverses formes, notamment pour le résumé multi-documents dans (Carbonell et Goldstein, 1998). Cette variante, *MMR-MD*, prend en compte les spécificités du résumé multi-documents : dans un premier temps, toutes les phrases ou passages considérés comme sémantiquement équivalents sont regroupés dans les mêmes classes. Le critère de redondance n'est alors plus seulement fonction de la similarité aux phrases déjà sélectionnées, mais également fonction de l'appartenance à une classe et/ou à un document dont une phrase a déjà été extraite.

D'autres variantes ont également vu le jour :
- utilisant des similarités différentes de celles de la version d'origine, comme MMR-LSA (Boudin *et al.*, 2007) ;
- faisant varier le facteur λ au cours des itérations afin de favoriser la non-redondance (Murray *et al.*, 2005) ;
- gérant les tâches de mise à jour (*cf* §5.1) (Boudin, 2008).

Ces approches, bien qu'efficaces pour l'élimination de la redondance, ne permettent pas de rendre compte de cette même redondance comme critère d'extraction de phrases. Afin de pallier ce problème, la recherche en résumé automatique a commencé à se tourner vers les applications de la théorie des graphes et des réseaux sociaux. Les réseaux sociaux révèlent en effet les liens que les phrases entretiennent entre elles, et permettent, grâce à des techniques exploratrices, d'analyser dans un même cadre la centralité et la diversité.

1.6. Méthodes à base de graphe

Les méthodes de résumé automatique à base de graphe ont récemment bénéficié des avancées de la recherche dans l'étude des réseaux sociaux. En effet, si l'on considère un corpus comme un réseau social où les différents acteurs du réseau sont les phrases du corpus, extraire les phrases centrales revient à extraire d'un réseau social les noeuds les plus influents, donc ceux qui entretiennent le plus de liens avec les autres. Ce type d'analyse permet de rendre compte de la centralité d'une phrase en tirant parti du contenu global des documents, et non plus uniquement de la requête et de listes de termes, forcément limitatifs.

L'algorithme *LexRank*, développé par Radev (Erkan et Radev, 2004), se fonde sur cette observation. Le but du système est d'extraire les phrases considérées comme « centrales » dans un corpus, en se fondant sur la notion de « prestige » des réseaux sociaux. Un graphe du corpus est établi, chaque noeud étant une phrase et chaque arête recevant comme poids la similarité des deux phrases qu'elle lie. Cette similarité est calculée à l'aide de la mesure *cosinus*. Cette mesure établit une similarité entre deux vecteurs en calculant le cosinus de l'angle de ces deux vecteurs. Pour calculer la similarité entre phrases, il faut établir un vecteur pour chaque phrase. Le vecteur représentant une phrase est généralement rempli avec les *tf.idf* des constituants de cette dernière. Les phrases sélectionnées sont celles considérées comme centrales dans ce graphe : les phrases qui ont les plus fortes composantes. Les résultats obtenus sont alors meilleurs qu'avec la méthode des *centroïdes* (Radev *et al.*, 2001a) utilisée initialement dans *MEAD* (*cf* §*1.3*).

TextRank, développé par Rada Mihalcea et Paul Tarau (Mihalcea et Tarau, 2004), utilise la même méthode de sélection des phrases en leur appliquant une mesure de similarité différente : le nombre de termes partagés par deux phrases normalisé par la taille de ces phrases.

TextRank a été utilisé pour le résumé mono-document, tandis que *LexRank* et *MEAD* ont été conçus pour le résumé multi-documents. Le résumé multi-documents comporte un problème supplémentaire : étant donné la multiplication des sources, le risque de sélectionner des phrases redondantes est accru. Pour résoudre ce problème, les phrases trop similaires à une phrase déjà classée sont éliminées. La métrique *CSIS*, *« Cross-sentence informational subsumption »* (Radev *et al.*, 2004) est utilisée à cet effet. La détection et l'élimination de la redondance, parmi les problématiques majeures du résumé multi-documents, ne sont donc pas gérées au cœur du système de résumé.

Par absence de lactase, beaucoup d'adultes ne peuvent digérer le lait. *Dans les populations consommant du lait, les adultes disposent de plus de lactase, peut-être grâce à la sélection naturelle.*

FIGURE 1.6.: Illustration d'une relation noyau/satellite de contraste. La première possibilité est en gras, la seconde en italique

1.7. Résumé automatique et structure rhétorique

Dans une autre optique, (Mann et Thompson, 1988) ont proposé une théorie : la *Rhetorical Structure Theory* (RST). La RST postule que pour toute partie d'un texte cohérent –exempt d'illogismes et de lacunes–, il existe une raison à sa présence, évidente pour les lecteurs. La théorie des structures rhétoriques tente de formaliser les raisons d'être de chaque partie de texte en décrivant un texte comme un ensemble de structures dont les plus courantes sont les relations noyau/satellite. Le site web « Rhetorical Structure Theory » [3] propose quelques exemples, dont l'exemple de la figure 1.6, qui montre une relation de contraste, *i.e.* une possibilité dans une alternative suivie d'une autre possibilité.

Ces relations noyau/satellite ont fait l'objet d'une typologie. Parmi toutes les relations noyau/satellite répertoriées, la plus parlante est la relation démonstrative. Le noyau, ou cœur du discours, est alors l'affirmation, et le satellite la démonstration de cette affirmation par l'auteur.

La théorie de la structure rhétorique a été appliquée au résumé automatique de textes, notamment par Ono *et al.* (1994) et Marcu (1997). Les relations noyau/satellite dénotent en effet la centralité des phrases ou des clauses. Il est donc possible, à partir d'une représentation de documents en termes de relations noyau/satellite, d'extraire automatiquement les éléments textuels les plus centraux de ces documents.

Teufel et Moens (2002) se fondent également sur le statut rhétorique des phrases pour réaliser des résumés de textes scientifiques. Leurs travaux s'appuient sur la spécificité structurelle des articles scientifiques afin d'extraire les phrases les plus pertinentes d'articles dont le statut rhétorique des phrases a été étiqueté automatiquement.

Ces approches nécessitent toutefois une analyse linguistique fine des documents, et sont confrontées aux problèmes des systèmes utilisant de telles analyses : forte dépendance à la langue et aux types de documents à résumer, mais aussi la difficulté pour un système automatique d'identifier de tels noyaux/satellites.

3. http://www.sfu.ca/rst

1.8. Extraction et fusion d'information

McKeown *et al.* (1999) et Barzilay et McKeown (2005) proposent une méthode novatrice pour le résumé automatique. Celle-ci consiste à identifier dans un premier temps les fragments de texte qui véhiculent la même information, avant de fusionner ces fragments afin d'obtenir une phrase cohérente. Les fragments textuels sont regroupés grâce à la comparaison des arbres syntaxiques de ces fragments. Cette comparaison reconnaît des formulations paraphrastiques, telles que la passivation, la négation et l'utilisation d'un verbe antonyme, la nominalisation d'une phrase, l'utilisation de synonymes... Les arbres syntaxiques qui véhiculent la même information sont ensuite fusionnés afin de créer une nouvelle phrase qui représentera au mieux les informations présentes dans ces arbres.

Cette méthode a obtenu de bons résultats lors de campagnes d'évaluation de résumés automatiques (Nenkova *et al.*, 2003). Elle est cependant dépendante de ressources linguistiques précises, ce qui la rend difficilement portable à d'autres langues que l'anglais. Le coût de son adaptation à une autre langue l'exclut pour toute approche générique indépendante de la langue.

1.9. Post-traitements

1.9.1. Compression de phrases

Les phrases extraites automatiquement par un système de résumé automatique peuvent s'avérer contenir des éléments textuels qui perdent leur sens car ils sont retirés de leur contexte. C'est notamment le cas des marqueurs logiques, temporels, ou de contradiction. Si les références de certains de ces éléments peuvent être résolues afin de leur redonner du sens même une fois extraits de leur contexte (e.g. les marqueurs temporels référentiels : « hier »), les autres doivent être éliminés afin de réduire la taille des résumés et d'en faciliter la lecture. D'autre part, une phrase d'un résumé peut également inclure des informations inutiles, qui devront être supprimées tout en maintenant la cohérence des groupes lexicaux la constituant.

Jing (2000) utilise pour cela le rôle thématique des constituants des phrases à compresser. Le rôle thématique d'un constituant associé au cadre de sous-catégorisation du verbe central de la phrase apporte des informations sur son importance syntaxique ou sémantique. Le rôle thématique indique donc si un constituant peut être éliminé sans que cela ne perturbe ni l'organisation syntaxique ni la sémantique de sa phrase.

Les auteurs de (Knight et Marcu, 2002) considèrent la compression de phrase comme une tâche visant à séparer le noyau de la phrase de ses constituants annexes. Ils utilisent à cette fin deux systèmes. Le premier est fondé sur la reconnaissance des constituants annexes par un réseau bayésien entraîné sur un corpus constitué de duplets (phrase, compression de la phrase par un humain), chaque duplet étant analysé syntaxiquement de manière automatique. Le système renvoie pour une phrase à compresser le schéma de compression (arbre syntaxique d'origine/arbre syntaxique compressé) le plus approprié. Le deuxième système, moins efficace, utilise un arbre de décision. Celui-ci est entraîné sur le même corpus que le premier. Les couples phrase d'origine/phrase compressée sont tous analysés automatiquement afin de trouver le chemin qui permet de passer de la phrase d'origine à la phrase compressée en utilisant des règles de transformation de base sur les arbres syntagmatiques. Les données d'entraînement comprennent 99 variables, telles la catégorie syntaxique des noeuds de l'arbre syntagmatique de l'exemple, le nombre d'arbres dans la pile de transformation, la catégorie syntaxique de la racine de l'arbre au sommet de la pile, ou encore les cinq dernières opérations réalisées à l'instant t. L'algorithme apprend donc la meilleure opération à effectuer sur un arbre à un moment donné de la transformation.

(Yousfi-Monod et Prince, 2005) décrit une méthode de suppression de constituants d'une phrase à l'aide de règles fondées sur la position desdits constituants dans l'arbre syntagmatique et sur leur fonction syntaxique, calculées automatiquement par un analyseur syntaxique du français, SYGFRAN (Chauché, 1984).

(Fernández Sabido et Torres-Moreno, 2009) s'inspire de notions thermodynamiques de la physique statistique afin de supprimer les constituants d'importance secondaire. Les auteurs ont mis en oeuvre une méthode d'apprentissage des scores pour chaque élément de la phrase en fonction de sa catégorie morpho-syntaxique et de celle de ses voisins sur un corpus constitué de paires (phrase d'origine, phrase compressée). L'étape suivante consiste à étiqueter à l'aide d'un algorithme non-déterministe les constituants à garder et ceux à supprimer. Si les résultats en termes d'informativité sont satisfaisants, les auteurs remarquent toutefois que la robustesse du système pour conserver la grammaticalité de la phrase n'est pas satisfaisante.

1.9.2. Réordonnancement

Les résumés créés automatiquement par extraction nécessitent un ordonnancement des phrases. Certaines stratégies visent à exposer à l'utilisateur les phrases des plus pertinentes aux moins pertinentes (Bossard *et al.*, 2008). Cependant, cela pose le risque de désorienter le lecteur en lui fournissant des phrases non cohésives, aussi bien chronologiquement que rhétoriquement parlant. Cet aspect est ignoré de la majorité des systèmes de résumé automatique, et aucune campagne d'évaluation n'a jusqu'à présent porté sur ce thème. Cependant, certains systèmes adoptent des stratégies pour rendre les résumés

qu'ils produisent cohérents et présentables à l'utilisateur.

La modélisation chronologique des documents pour le réordonnancement a été utilisée par (McKeown *et al.*, 1999) et (yew Lin et Hovy, 2001). Selon (Barzilay *et al.*, 2002), celle-ci ne permet toutefois pas d'obtenir des résultats satisfaisants. Les auteurs de (Barzilay *et al.*, 2002) et de (Okazaki *et al.*, 2005) ont proposé de coupler la modélisation chronologique avec une détection des thèmes des phrases, afin de regrouper ensemble les phrases qui portent sur les mêmes sujets.

La RST (*Rhetorical Structure Theory*, *cf* §1.7) a également été appliquée au réordonnancement de phrases (Moore et Paris, 1993; Hovy, 1993).

Ces deux types d'approches nécessitent pour la modélisation des relations chronologiques et rhétoriques entre phrases, des connaissances sur la langue qui les rendent difficilement détachables de la langue des documents à résumer. Elles sont donc difficilement applicables dans la perspective d'une approche générique indépendante de la langue.

Une méthode plus simple consiste à réorganiser le résumé en fonction uniquement de la position des phrases dans le document. Cette méthode aux résultats insatisfaisants, a été utilisée dans des tâches de résumés de dépêches en lui ajoutant la date d'émission des documents dont les phrases sont issues (Saggion, 2005; Goldstein *et al.*, 2000). (Favre *et al.*, 2006) y ont adjoint la provenance géographique des documents.

Ces méthodes nécessitent cependant, afin d'avoir un rendu correct, que les documents à résumer soient dotés de données extra-textuelles précises. Cela pose un problème de généricité : tous les documents à résumer ne disposent pas de ce type de données.

1.10. Conclusion

Dans cette partie, nous avons présenté les différentes méthodes de résumé automatique. Nous avons également dégagé des travaux existants leurs atouts et points faibles, afin de nous en inspirer pour développer une nouvelle approche de résumé automatique. Nous avons constaté que si les systèmes d'extraction statistiques sont les plus utilisés, en particulier pour des systèmes génériques qui ne sont pas seulement appliqués à des domaines spécialisés, de nombreuses approches de compression automatique de phrase ont récemment émergé afin d'améliorer les sorties des systèmes de résumé par extraction. Ceux-ci nécessitent en effet des post-traitements linguistiques afin de rendre en peu de mots le maximum d'information et de les rendre plus agréables à lire. Cependant, nous estimons que les systèmes purement extractifs n'ont pas encore atteint leurs limites et nous proposerons au chapitre 3 notre approche pour résoudre le principal défaut des

systèmes actuels, à savoir la gestion dissociée de la centralité et de la diversité.

2. L'Evaluation de résumés informatifs

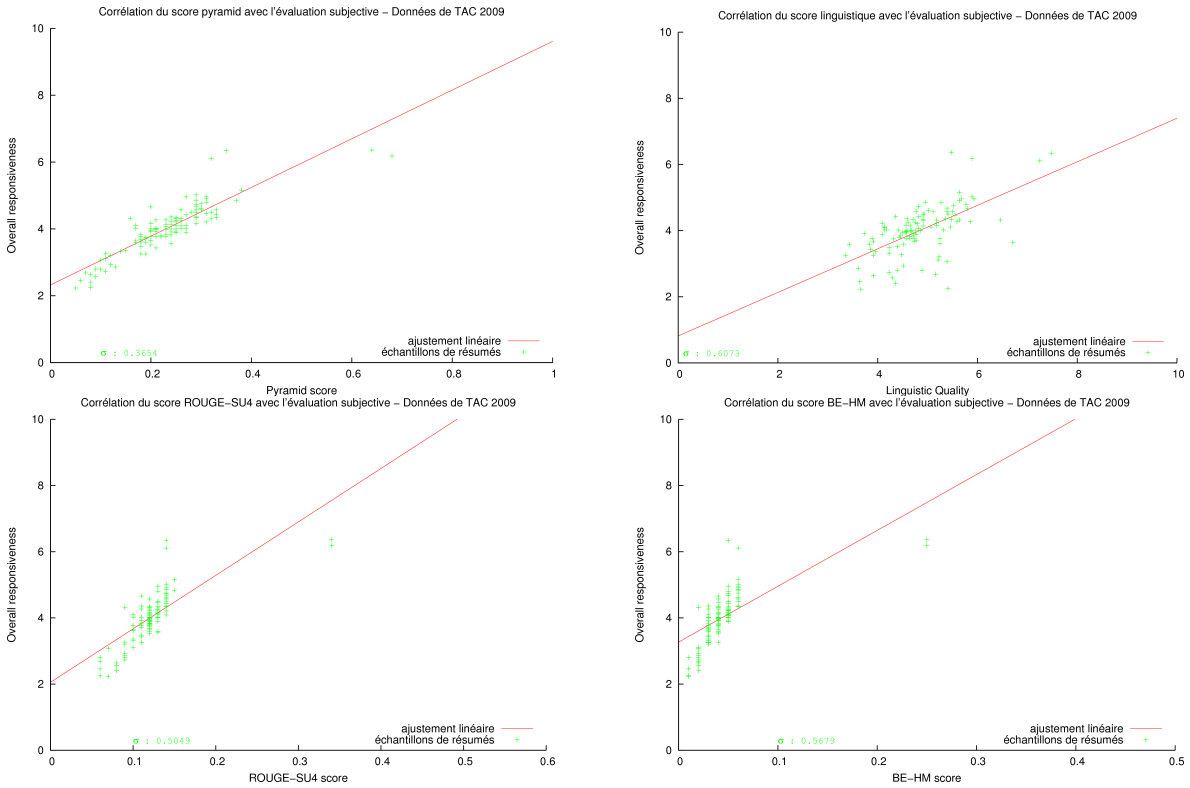
Sommaire

L'évaluation de résumés, qu'ils soient rédigés manuellement ou conçus par un système automatique, est une problématique à part entière. Un protocole d'évaluation de résumé doit faire intervenir deux aspects :

- le fond, ou les informations présentes dans le résumé ;
- la forme, ou la manière de présenter les informations.

Différentes mesures ont été proposées ces dernières années afin d'évaluer le fond, ou l'informativité des résumés de manière automatique. L'évaluation manuelle, telle que formulée dans les campagnes d'évaluation TAC [1] pose un problème évident de subjectivité. Cependant, cette évaluation manuelle s'avère être très corrélée aux résultats des évaluations manuelles guidées de l'informativité (méthode pyramide décrite plus bas), mais peu corrélée aux notes des résumés en qualité linguistique(*cf* fig.2.1). Cela signifie que les juges accordent plus d'importance au fond qu'à la forme des résumés automatiques, et que des mesures automatiques de l'informativité performantes sont essentielles à l'évaluation des résumés automatiques. Cependant, les évaluations de la qualité linguistique ne doivent pas être oubliées car la qualité linguistique est tout de même un critère important, qui permet de départager deux résumés sensiblement identiques du point de vue des informations qu'ils véhiculent.

1. Les évaluateurs devaient noter les résumés en fonction de leur réponse à la question : « Combien paieriez-vous pour ce résumé ? »



	qual. ling.	Pyramide	ROUGE-SU4	BE-HM
Coefficient de corrélation de Pearson	0.6210	0.8818	0.7586	0.6804

FIGURE 2.1.: Corrélation de l'évaluation subjective de résumés avec les évaluations linguistiques et informatives dans TAC 2009 : la mesure pyramide est la plus corrélée avec l'évaluation subjective. Viennent ensuite les mesures automatiques puis la mesure manuelle de la qualité linguistique.

2.1. ROUGE

Les mesures ROUGE, pour *Recall-Oriented Understudy for Gisty Evaluation*, ont été introduites par Lin dans (Lin, 2004). Ces mesures sont fondées sur la comparaison de n-grammes [2] entre un ou plusieurs résumés de référence et un résumé à évaluer. Il n'existe pas un unique résumé de référence, et il est donc essentiel de comparer les résumés automatiques à plusieurs résumés de référence établis manuellement afin d'obtenir des mesures plus précises de la qualité des résumés. Ces mesures nécessitent donc la rédaction de résumés de référence par un ou plusieurs experts au préalable de la mesure de qualité du résumé. Il en existe plusieurs variantes, que nous présentons ci-dessous.

2.1.1. ROUGE-n

Les métriques ROUGE-n sont fondées sur la comparaison simple de n-grammes. Une liste des n-grammes est établie pour chacun des résumés de référence et des résumés cible. Les n-grammes sont composés de n mots consécutifs. Par exemple, pour le texte « ROUGE est une métrique d'évaluation », la liste de n-grammes créée par ROUGE-2 sera « ROUGE est », « est une », « métrique d' », « d' évaluation ». Une fois la liste des n-grammes établie, le score ROUGE est calculé selon la formule en figure 2.1.1.

$$\frac{\sum_{r \in CRef} \sum_{n-grammes \in RC} card_{match}(n-grammes)}{\sum_{r \in CRef} \sum_{n-grammes \in RC} card(n-grammes)}$$

- $CRef$ est la collection des résumés de références,
- RC le résumé cible (le résumé à évaluer).

FIGURE 2.2.: Formule de calcul du score ROUGE-n : la formule revient à diviser le nombre de n-grammes communs entre le résumé à évaluer et les résumés de référence par le nombre total de n-grammes des résumés de référence.

Cette mesure présente un défaut majeur : l'ordre des mots n'influe pas toujours sur le résultat. Ainsi, l'exemple suivant, inspiré de la présentation par Lin de ROUGE au Workshop « Text Summarization Branches Out », montre à quel point les scores ROUGE-n peuvent être décorrélés de la réalité :

- Phrase 1 (référence) : Dr Jekyll tua Hide
- Phrase 2 : Dr Jekyll tue Hide
- Phrase 3 : Hide tue Dr Jekyll

2. Un n-gramme est une sous-séquence de n éléments construite à partir d'une séquence donnée.

Les scores ROUGE-n des phrases 2 et 3 seront similaires. En effet, les deux phrases ont les mêmes n-grammes en commun avec la phrase 1, à savoir (« Dr Jekyll », « Hide »).

2.1.2. ROUGE-L

Afin de pallier en partie un défaut de ROUGE-n, à savoir le fait que l'ordre des mots dans les phrases n'est pas pris en compte, (Lin, 2004) a introduit une autre mesure, ROUGE-L. Etant donné deux séquences S1 (référence) et S2, ROUGE-L est définie comme étant la plus longue sous-séquence commune (LCS) à S1 et S2 divisée par le nombre total d'éléments de S1. Ainsi, les scores des phrases 2 et 3 seront :

- phrase2 = (« Dr Jekyll Hide ») = 3/4
- phrase3 = (« Dr Jekyll ») = 2/4

Cependant, cette mesure n'est pas totalement satisfaisante. En effet, la phrase « Dr Jekyll a été tué par Hide. » obtiendrait avec cette mesure le même score que la « Dr Jekyll tue Hide. ».

2.1.3. ROUGE-SUn

ROUGE-SU, pour *skip-bigram and unigram ROUGE* prend en compte des bi-grammes à trou ainsi que les unigrammes des résumés de référence et du résumé cible. Les bi-grammes à trou (*skip-bigram*) sont définis dans (Lin, 2004) comme n'importe quelle paire de mots dans l'ordre de la phrase, séparés par une distance maximale arbitraire (le n). Ainsi, avec ROUGE-SU4, la phrase 1 comprend6 bi-grammes et 4 unigrammes : « Dr Jekyll », « Dr tua », « Dr Hide », « Jekyll tua », « Jekyll Hide », « tua Hide » et les 4 unigrammes qui composent la phrase. Les phrases 2 et 3 ont ainsi pour score respectivement : 6/10 et 4/10.

Cette mesure permet de rendre compte efficacement des relations de dépendance éloignées dans le texte. La campagne TAC 2008 l'a d'ailleurs confirmé puisque parmi les mesures ROUGE, c'est ROUGE-SU4 qui est la plus corrélée aux jugements humains.

2.2. BE-HM

Cette méthode d'évaluation automatique extrait des résumés de référence des *basic elements »* (BE). Les *Basic Elements* sont les unités sémantiques minimales (Hovy *et al.*,

		ROUGE-2		ROUGE-SU4	
	chaîne	n-grammes	score	n-grammes	score
phrase de référence	Résumer est un art difficile	Résumer est, est un, un art, art difficile		Résumer est, Résumer un, Résumer art, est un, est art, est difficile, un art, un difficile, art difficile + unigrammes	
phrase évaluée 1	Résumer n'est pas facile	Résumer n', n'est, est pas, pas facile	0	Résumer n', Résumer est Résumer pas, n' est, n' pas n' facile, est pas, est facile pas facile + unigrammes	3/15 = 0.2
phrase évaluée 2	Résumer est un art facile	Résumer est, est un, un art, art facile	0.75	Résumer est, Résumer un, Résumer art, est un, est art, est facile, un art, un facile, art facile + unigrammes	10/15 = 0.666

TABLE 2.1.: Exemple d'application des mesures ROUGE

2006) :

- soit la tête du constituant syntaxique majeur,
- soit la relation entre la tête et une unité qui en dépend, exprimée sous la forme d'un triplet (tête | modifieur | relation).

Par exemple, la phrase « Additional test flights are planned. » contient les BE suivants :

- flights|additional|mod
- flights|test|nn
- planned|flights|obj

L'inconvénient de cette méthode réside dans l'extraction automatique des BE des résumés de référence et des résumés à évaluer. D'une part, cette extraction introduit des erreurs puisque les systèmes sur lesquels elle est fondée ne sont pas totalement fiables. De plus, si l'on suppose qu'il existe une fonction permettant de passer de l'ensemble des BE des textes d'origine à l'ensemble des BE des résumés de référence, ou au moins d'approximer cet ensemble, alors cette évaluation favorise les systèmes intégrant le même type d'information. De plus, il a été montré dans (Dang et Owczarzak, 2008a) que cette mesure n'est pas aussi bien corrélée aux évaluations subjectives d'expert que les mesures classiques ROUGE présentées en §2.1.

2.3. Evaluation de résumés et théorie de l'information

La technique d'évaluation présentée dans (Lin *et al.*, 2006) est fondée sur l'hypothèse qu'étant donné un jeu de documents $D = \{d_1, d_2, ...d_i\}$, il existe une distribution de probabilités de paramètres fonction de θ_R qui génère un résumé de référence en fonction de D ; Résumer consiste alors à estimer ce θ_R. Les auteurs supposent que de la même manière, tout résumé réalisé automatiquement est généré par une distribution de probabilités de paramètres fonction de θ_A. Evaluer les résumés revient à calculer la divergence entre les deux distributions θ_R et θ_A. Les auteurs utilisent pour cela la mesure

Jensen-Shannon.

Cette technique d'évaluation a été utilisée avec succès sur le corpus DUC 2002, en résumé mono-document et multi-documents. Les classements des systèmes obtenus sont plus corrélés aux jugements humains que les mesures ROUGE-n pour le résumé multi-documents, et semblables aux mesures ROUGE-n en mono-document. Cependant, les données de DUC 2002 ne suffisent pas à l'évaluation d'une telle mesure : en effet, le nombre de systèmes qui ont participé (12 soumissions en multi-documents, 8 en mono-document). De plus, les mesures auxquelles les auteurs se comparent sont les mesures ROUGE-n, moins efficaces que les mesures ROUGE-SUn. Il est donc difficile de se faire une idée de la réelle efficacité de cette mesure d'évaluation.

2.4. La méthode Pyramide

Les systèmes automatiques d'évaluation de résumés sont tous confrontés au même problème : la reformulation. Les résumés de référence sont en effet écrits en respectant les règles de rédaction de résumé : identifier les informations principales et les reformuler. Par conséquent, le vocabulaire utilisé dans les résumés de référence ainsi que la structure de surface des phrases qui les composent varient des vocabulaire et structure des résumés automatiques, le plus souvent réalisés par des méthodes d'extraction.

L'impossibilité d'ajouter de la sémantique aux systèmes d'évaluation, sous peine de biaiser les résultats en favorisant les systèmes de résumés automatiques qui auront fait les mêmes choix, obligent à mettre au point des protocoles d'évaluation manuelle afin d'évaluer au mieux les résumés automatiques.

Nenkova *et al.* (2007) décrivent un tel protocole. A partir des résumés de référence, une liste de SCUs (*Single Content Units* ou unités de contenu) est établie. Elles sont classées et pondérées selon leur fréquence d'apparition dans les résumés de référence. La liste des SCUs prend alors la forme d'une pyramide (une structure hiérarchisée) reflétant l'importance des SCUs. L'étape suivante consiste à repérer dans les résumés à évaluer les SCUs présentes dans la pyramide. Un score est alors attribué aux résumés : la somme des poids des SCUs qu'ils contiennent, divisée par la somme des poids des SCUs des résumés de référence.

Ce type d'évaluation, peu automatisable, a le défaut d'être extrêmement coûteuse en temps d'expertise. En effet, apprendre à construire une pyramide de SCUs, de même que la construire, sont coûteux en temps. Cependant, cette méthode d'évaluation semble être la plus corrélée aux résultats d'évaluation subjective (sans protocole précis de notation) sur les résultats de la tâche de résumé des campagnes TAC 2008 (Dang et Owczarzak, 2008a) et TAC 2009 (*cf fig. 2.1*).

En attendant des mesures plus précises d'évaluation entièrement automatique, il est donc indispensable de recourir à ce type d'évaluation manuelle lorsque les moyens humains s'y prêtent.

2.5. Évaluation de la forme

L'évaluation de la forme se rapproche de l'évaluation de la lisibilité d'un texte. Un texte lisible répond à différents critères. Il doit être à la fois :

- syntaxiquement correct ;
- sémantiquement cohérent ;
- articulé logiquement ;
- exempt de redondances.

En l'état actuel des recherches en TAL, cette partie de l'évaluation ne peut être réalisée dans sa totalité de manière automatique, surtout en ce qui concerne les évaluations de la cohérence sémantique et de l'articulation logique. Elle est par conséquent extrêmement coûteuse, puisqu'elle fait intervenir des jugements d'experts. Lors des campagnes d'évaluation DUC (Document Understanding Conference) et TAC (Texte Analysis Conference), les experts ont évalué la forme des résumés en se référant à une grille d'évaluation précise, présentée en figure 2.2.

Grammaticality	Grammaticalité	note de 1 à 10
Non-redundancy	Non-redondance	note de 1 à 10
Referential clarity	Auto-référentialité	note de 1 à 10
Focus	Mise en évidence des points principaux	note de 1 à 10 note de 1 à 10
Structure and Coherence	Structure et cohérence	note de 1 à 10

Table 2.2.: Grille d'évaluation linguistique des résumés des campagnes DUC et TAC

A partir de cette grille d'évaluation, les experts attribuent une note de qualité linguistique à chacun des résumés automatiques. Il est intéressant de noter qu'au cours de la campagne d'évaluation TAC 2009, les résumés rédigés par les experts ont obtenu une note linguistique de 8.8/10, tous experts confondus. Malheureusement, les résultats publiés mentionnent uniquement la note linguistique globale ; les notes détaillées des différents critères de la grille d'évaluation n'y apparaissent pas, ce qui nous empêche d'aller plus loin dans la compréhension de ces notes. Cependant, cette grille d'évaluation paraît adéquate. En effet, elle rend parfaitement compte des critères de lisibilité énoncés plus haut.

2.6. Conclusion

Nous avons effectué un panorama des techniques d'évaluation de résumés automatiques. Aucune solution actuelle n'est complètement satisfaisante. En effet, les méthodes entièrement automatiques (ROUGE, BE-HM) manquent de précision. De plus, il est possible d'adapter des systèmes de résumé automatique de manière à ce qu'ils obtiennent de meilleurs scores avec une technique automatique sans pour autant augmenter la qualité du résumé. Cependant l'utilisation de méthodes manuelles qui ne font pas seulement intervenir un expert pour la rédaction d'un ou plusieurs résumés de référence, mais également pour la construction d'un modèle d'évaluation depuis ces résumés, et la confrontation des résumés automatiques à ce modèle, sont particulièrement coûteuses en temps. Une évaluation précise ne peut donc, en l'état actuel des avancées dans ce domaine, être réalisée sur des données trop importantes. Des efforts doivent donc être menés par la communauté afin de développer des méthodes d'évaluation précises et peu coûteuses, qui pourront supporter le passage à l'échelle.

Deuxième partie .

Approche

Cette partie est consacrée à la description de notre système de résumé automatique, CBSEAS. Notre système de résumé est fondé sur la détection de similarité entre phrases afin d'établir un modèle consistant en un graphe et une classification des phrases en classes sémantiques. Ce modèle sert à calculer les scores des phrases, qui détermineront lesquelles doivent être extraites ; la meilleure candidate de chaque classe est sélectionnée, afin de maximiser la diversité informative du résumé tout en minimisant la redondance. Le modèle sert aussi à une approche de réordonnancement qui permet une meilleure lisibilité des résumés générés.
Cette partie présente également un algorithme génétique qui permet d'optimiser les différents paramètres de CBSEAS. Pour finir, elle montre l'intérêt de l'analyse discursive pour le résumé automatique au travers de l'intégration de l'analyse en entités nommées et de l'étiquetage de la structure des dépêches de presse à CBSEAS. Les améliorations de la qualité des résumés générés que nous attendions ont été validées par nos participations aux campagnes d'évaluation TAC.

3. CBSEAS, Une Approche Générique pour le Résumé Automatique

Sommaire

Le résumé automatique est un axe de recherche encore ouvert. Les campagnes d'évaluation récentes montrent que les systèmes actuels de résumé automatique ont encore du chemin à parcourir avant d'égaler les performances des résumés manuels. Nous situons notre travail dans le cadre du résumé automatique multi-documents. C'est à l'heure actuelle la branche la plus explorée du résumé automatique ; les besoins industriels dans ce domaine sont en effet importants et ont conduit à un renouveau de la recherche en génération automatique de synthèses multi-documents.

Les systèmes de génération de résumés par extraction se fondent soit sur la centralité seule et l'élimination dans un second temps de la redondance afin de maximiser la diversité (Salgueiro Pardo *et al.*, 2002; Radev *et al.*, 2001a; Saggion, 2005; Boudin et Torres-Moreno, 2007), soit sur la similarité à une requête utilisateur associée à l'élimina-

tion de la redondance (Carbonell et Goldstein, 1998; Boudin *et al.*, 2008b). Nous estimons que la gestion de la diversité est importante dans le cadre du résumé multi-documents, et nous présentons ici une approche qui vise à établir un modèle de représentation du corpus qui rend compte à la fois de la diversité et de la centralité avant de sélectionner les phrases à extraire.

3.1. Intuitions

Générer automatiquement des résumés multi-documents ajoute une problématique supplémentaire à la génération de résumés mono-document : l'élimination de la redondance. La redondance d'information est en effet plus présente sur plusieurs documents que dans un document unique. Le risque d'extraire plusieurs phrases véhiculant la même information est donc plus élevé.

Cette redondance d'information apporte cependant des informations supplémentaires pour la réalisation d'un résumé, qu'elle soit automatique ou non. En effet, les informations centrales d'un groupe de documents ont de fortes chances d'être reprises sur plusieurs documents (*cf* fig. 3.1). Par conséquent, réussir à identifier les passages redondants permet non seulement l'élimination de la redondance du résumé final – un des critères de qualité des résumés (*cf* §2) –, mais également de mieux sélectionner les phrases à extraire dans le résumé.

De plus, nous considérons que dans un groupe de documents à résumer, il n'existe pas une information la plus centrale, mais une multitude d'informations qui constituent la diversité informationnelle des documents. Chaque information est véhiculée par une ou plusieurs phrases dont une peut être identifiée comme la plus représentative. Le résumé idéal est alors constitué de chaque phrase la plus représentative des informations les plus importantes.

Des systèmes de résumé automatique récents se sont concentrés sur la détection de redondances et la fusion d'information (Barzilay et McKeown, 2005) en utilisant des techniques linguistiques fortement dépendantes de la langue. D'autres systèmes fondés sur l'analyse statistique ont tenté de rendre compte du phenomène de dépendance entre la centralité d'une information et son occurence en corpus (Erkan et Radev, 2004). Nous voulons aller plus loin que ces systèmes statistiques, tout en restant assez indépendant de la langue pour permettre la généricité de notre approche. Notre système pourra alors, sans avoir à effectuer de changements majeurs, être opéré sur différentes langues tout en gardant le même niveau de performances. Ceci exclut donc tout traitement linguistique poussé.

Notre système doit donc être fondé sur la détection de redondances, et l'extraction

La France n'est "pas surprise" par les incidents de lundi en Côte d'Ivoire

La France n'est "pas surprise" par **les incidents qui ont opposé ses soldats lundi aux rebelles de l'Ouest en Côte d'Ivoire** et reste déterminée à faire avancer le processus de paix, a dit mardi le ministère français des Affaires étrangères. "Nous ne sommes pas surpris par les événements d'hier (lundi) à l'Ouest de la Côte d'Ivoire où la situation est plus incontrôlable, plus volatile", a déclaré le porte-parole du ministère, François Rivasseau. **Ces incidents, qui ont fait 30 morts dans les rangs des rebelles et neuf blessés dans ceux de l'armée française**, sont le fruit de "bandes incontrôlées", a-t-il ajouté. Mais, "ceci ne fait que renforcer notre détermination à faire avancer le processus de paix", a encore assuré M. Rivasseau. La France a invité toutes les parties ivoiriennes à une table ronde politique la semaine prochaine à Paris pour tenter de trouver une solution à la crise qui secoue la Côte d'Ivoire, coupée en deux entre le nord conquis par les rebelles et le sud tenu par les forces loyales au président ivoirien Laurent Gbagbo. L'un des mouvements rebelles, le Mouvement patriotique de Côte d'Ivoire (MPCI), a laissé entendre mardi que les affrontements de la veille pourraient "compromettre dangereusement" la réunion de Paris. Interrogé sur ce point, M. Rivasseau a rappelé que toutes les parties avaient été invitées à se rendre à Paris où elles étaient les "bienvenues", y compris les rebelles de l'Ouest que l'ambassadeur de France à Abidjan Gildas le Lidec doit rencontrer à partir de mercredi. "Ils sont invités (à Paris) et nous voulons qu'ils viennent", a déclaré le porte-parole du quai d'Orsay. Lundi, à deux reprises, des combattants rebelles de l'ouest ivoirien se sont heurtés aux militaires français positionnés à Duékoué et **le bilan des affrontements de la journée est de 30 morts chez les rebelles et de neuf soldats français blessés**, dont un grièvement, selon l'état-major des armées françaises.

Nouveaux combats à Duékoué : 30 rebelles tués, neuf blessés français (bilan global)

De nouveaux accrochages ont opposé lundi soir des rebelles et des soldats français à Duékoué (ouest de la Côte d'Ivoire), et **le bilan des affrontements de la journée est de 30 morts chez les rebelles et de neuf soldats français blessés**, dont un grièvement, a annoncé l'état-major des armées françaises. Il s'agit de la plus grande attaque lancée par les rebelles contre les soldats français depuis leur déploiement en Côte d'Ivoire, commencé à la mi-septembre, immédiatement après le début de l'insurrection militaire. Le bilan de neuf blessés est également le plus lourd depuis le début de l'opération française. Jusqu'à lundi, seul un soldat français avait été blessé dans le cadre de l'opération Licorne en Côte d'Ivoire. L'armée française effectuait dans la soirée de lundi une opération de "ratissage et de fouille dans la zone des combats en terrain difficile et de hautes herbes", à l'issue des premiers combats qui avaient pris fin en mi-journée à la sortie nord de Duékoué, lorsque "les soldats français se sont trouvés imbriqués avec des rebelles", a indiqué l'état-major dans un communiqué. "A l'issue de ce dernier accrochage, nous avons dénombré 30 morts du côté des rebelles ayant attaqué nos positions dans la journée. Du côté français, les combats de la mi-journée ont fait quatre blessés légers. Les accrochages de la soirée ont causé cinq nouveaux blessés, dont un sérieusement touché à la jambe et qui a été évacué sur l'antenne chirurgicale militaire française de Yamoussoukro" (200 km à l'est de Duékoué), a ajouté l'état-major, précisant que les familles des blessés avaient été prévenues.

Nouveaux combats à Duékoué : 30 morts chez les rebelles, 9 blessés français

De nouveaux accrochages qui ont opposé lundi soir des rebelles et des soldats français à Duékoué (ouest de la Côte d'Ivoire) ont fait au total 30 morts chez les rebelles et de neuf blessés parmi les soldats français dont un sérieusement, a annoncé lundi soir l'état-major des armées. L'armée française effectuait une opération de "ratissage et de fouille dans la zone des combats en terrain difficile et de hautes herbes", à l'issue des premiers combats qui avaient pris fin en mi-journée, lorsque "les soldats français se sont trouvés imbriqués avec des rebelles", a indiqué l'Etat-major dans un communiqué. "A l'issue de ce dernier accrochage, nous avons dénombré 30 morts du côté des rebelles ayant attaqué nos positions dans la journée. Du côté français, les combats de la mi-journée ont fait quatre blessés légers. Les accrochages de la soirée ont causé cinq nouveaux blessés dont un sérieusement touché à la jambe et qui a été évacué sur l'antenne chirurgicale militaire française de Yamassoukro", a précisé l'état-major.

source des dépêches : AFP

FIGURE 3.1.: Rapport entre la centralité et la redondance : en gras, les deux informations majeures de ces trois dépêches. Ces informations sont les plus redondantes dans les trois dépêches.

des phrases les plus redondantes pour la production automatique d'un résumé. Cela permettra une élimination efficace de la redondance et la création de résumés riches en diversité, tout en garantissant l'extraction des phrases les plus centrales.

Etant donné que nous voulons notre système le plus générique possible, celui-ci devra pouvoir établir des résumés guidés par une requête, mais également générer des résumés indépendants de toute requête utilisateur.

L'indépendance vis-à-vis de la langue nous prive de traitements linguistiques qui rendraient les résumés plus fluides. Nous privilégions donc la quantité d'informations véhiculée par nos résumés à leur qualité linguistique. Nous n'excluons toutefois pas, dans un second temps, d'utiliser des ressources ou outils linguistiques spécialisés dans une langue donnée afin de faciliter à l'utilisateur la lecture des résumés de CBSEAS.

3.2. Le système CBSEAS

Cette partie présentera l'architecture globale du système CBSEAS (*Clustering-Based Sentence Extractor for Automatic Summarization*) ainsi que les différents modules du système en détail. CBSEAS est le système au coeur de cette thèse, puisqu'il servira à implémenter les différentes expérimentations présentées dans les chapitres suivants. Il a fait l'objet de deux publications, l'une à « SIGIR 2009, Consortium doctoral » et l'autre à « EACL 2009 », dont les références complètes se trouvent dans l'annexe C.

3.2.1. Architecture

CBSEAS a été pensé de manière à être aisément adaptable aux différentes tâches du résumé automatique. Nous l'avons découpé en quatre modules :

1. Préparation des documents et des éventuelles requêtes utilisateur (*cf* §3.2.2) ;
2. Construction d'un modèle pour représenter les phrases des documents et prendre en compte centralité et diversité ;
3. Extraction des phrases d'après le modèle (*cf* §3.2.5) ;
4. Post-traitements utilisant éventuellement le modèle des phrases construit en 2 (*cf* §3.2.6).

Notre modèle de représentation des phrases est composé :

– d'un ensemble de phrases et de leur score « requête » ou « centroïde » (*cf* §3.2.2) ;

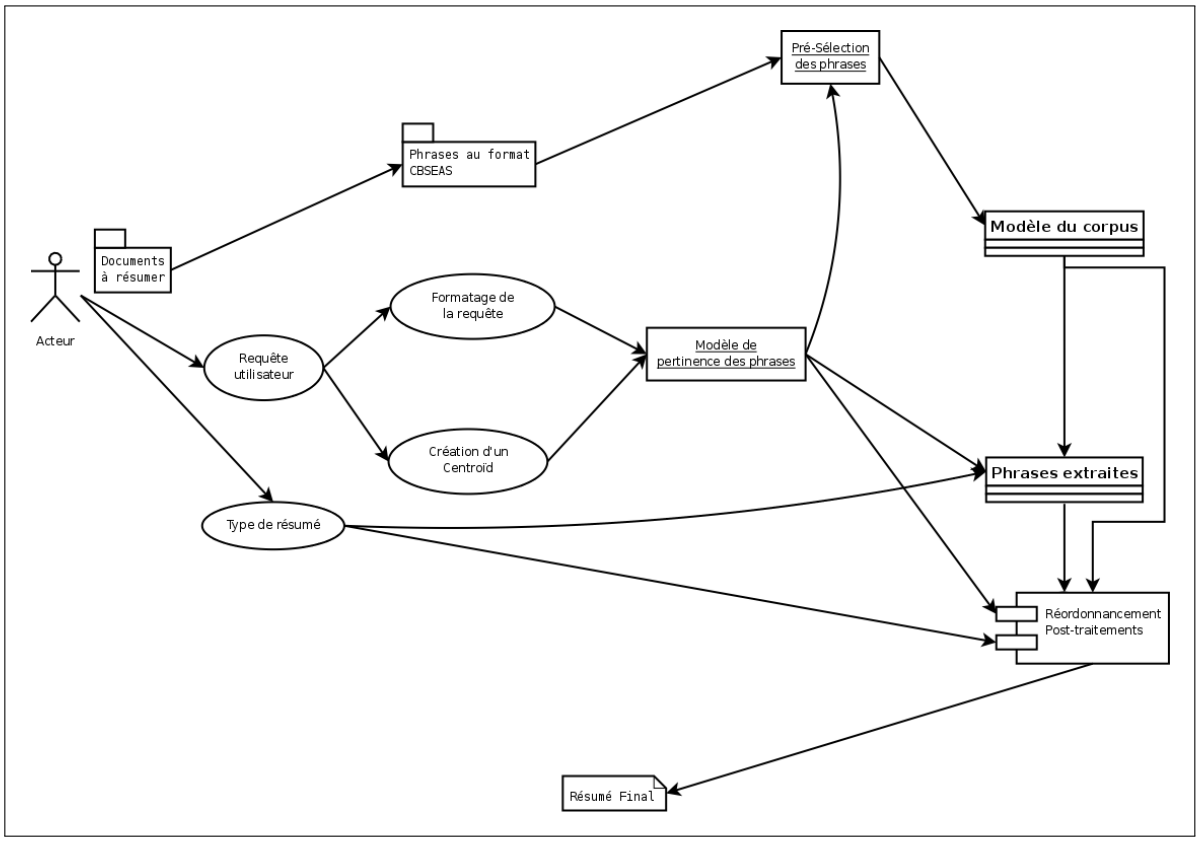


FIGURE 3.2.: Architecture de CBSEAS

– des similarités entre phrases (*cf* §3.2.3) ;
– du regroupement des phrases en « classes sémantiques » (*cf* §3.2.4) ;

Ce modèle nous permet d'obtenir une vue sur la redondance *via* les classes sémantiques. La similarité entre les phrases d'une même classe nous permet d'évaluer la centralité des phrases dans leur classe. Plus une phrase sera centrale, plus la probabilité sera grande qu'elle véhicule l'information essentielle de cette classe, sur laquelle le regroupement aura été réalisé. Associer cette notion de centralité à d'autres paramètres utilisateur permettra au système de sélectionner les phrases à la fois centrales et correspondant aux besoins de l'utilisateur, comme une requête ou des demandes de résumés orientés (*e.g.* des résumés d'opinion).

3.2.2. Préparation des documents

Les documents en entrée du système de résumé automatique subissent un traitement préalable à leur prise en compte par CBSEAS. Nous présentons ici les différents aspects de ce pré-traitement.

Annotation morpho-syntaxique

Les documents sont analysés morpho-syntaxiquement, et les unités textuelles annotées par tree-tagger [1] (Schmid, 1994). Cela permet la différenciation des différents types morpho-syntaxiques dans le calcul de similarité avec la requête, et d'affiner le calcul de similarité entre phrases. Le tree-tagger a été utilisé pour annoter onze langues différentes et peut être adapté à la majorité des langues indo-européennes, à condition de disposer d'un corpus annoté morpho-syntaxiquement ; son utilisation n'est donc pas problématique quant à l'indépendance de CBSEAS vis-à-vis de la langue, d'autant qu'il n'est pas indispensable au fonctionnement de CBSEAS et peut être désactivé.

Découpage des documents en phrases

Certains systèmes de résumé automatique font le choix de ne pas travailler sur les phrases, mais sur des structures plus petites. Ils travaillent alors à l'extraction de groupes de mots reliés syntaxiquement, en découpant les phrases en propositions (Marcu, 1997). D'autres systèmes ont choisi d'extraire des paragraphes entiers afin de garder une cohérence linguistique au sein du résumé. Le risque est alors plus élevé d'extraire des phrases

1. Page web de tree-tagger : http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/

non pertinentes puisque c'est le paragraphe dans son ensemble dont la pertinence est évaluée.

Afin d'éviter cet écueil mais également d'obvier à l'extraction de propositions mal découpées qui perdraient alors leur sens et leur grammaticalité, nous avons fait le choix d'extraire uniquement des phrases entières.

Annotation en entités nommées

Annoter en entités nommées permet d'affiner le calcul de similarité entre phrases. En effet, l'annotation permet de prendre en compte la co-occurence d'un groupe lexical tel que « Dr. Henry Jekyll » de la même manière qu'une entité nommée simple, étant donné qu'un tel groupe lexical sera reconnu comme étant une entité nommée composée. Il sera donc considéré non comme trois termes distincts, mais comme un seul et même terme. Beaucoup de langues disposent aujourd'hui d'annotateurs d'entités nommées performants ; il est cependant possible de désactiver la prise en compte des entités nommées dans le calcul de la similarité des phrases afin de répondre à la contrainte d'indépendance vis-à-vis de la langue. Les entités nommées sont annotées par GATE et le système ANNIE (Cunningham *et al.*, 2002).

Calcul d'un score requête

Si l'utilisateur fournit au système une requête, les phrases des documents à résumer se verront attribuer un score fonction de cette requête. Le score d'une phrase est la similarité de celle-ci avec la requête. Nous privilégions les requêtes utilisateurs en langue naturelle. Celles-ci sont exprimées sous la forme de phrases. Nous calculons donc la similarité à la requête de la même manière que la similarité entre phrases, décrite en 3.2.3.

Calcul d'un score centroïde

L'utilisateur peut également choisir de ne pas guider la création d'un résumé par une requête. Dans ce cas, le score attribué à chaque phrase sera un score centroïde. Nous calculons ce score d'après la méthode décrite dans (Radev *et al.*, 2001a). Le centroïde est un groupe de mots composé des n mots les plus importants des jeux de documents à résumer. L'importance des mots est définie par leur *tf.idf*. Le score centroïde d'une phrase est la somme des *tf.idf* communs à la phrase et au centroïde. Ce score est fonction de la centralité d'une phrase vis-à-vis du contenu global des documents à résumer.

Pré-sélection de phrases

Calculer le résumé en prenant en compte toutes les phrases des documents source peut entraîner la sélection de phrases « hors sujet » du fait de notre méthode de regroupement des phrases en classes sémantiques et de notre méthode d'extraction. En effet, les regroupements créent parfois une classe « fourre-tout » dans laquelle sont regroupées toutes les phrases qui n'avaient pas leur place dans une autre classe. Ces phrases sont le plus souvent marginales dans les documents à résumer, et l'extraction de l'une d'elles dans le résumé est préjudiciable aux performances de CBSEAS. L'utilisateur peut donc déterminer le nombre de phrases en entrée du système. Les phrases les plus similaires à la requête et au centroïde sont sélectionnées, évitant ainsi d'extraire dans le résumé des phrases non pertinentes.

3.2.3. Calcul des Similarités entre Phrases

Nous faisons l'hypothèse que la similarité entre les phrases devra tenir compte du type des documents que CBSEAS aura à résumer. Par exemple, les caractéristiques qui détermineront si deux phrases sont similaires dans le cadre de l'analyse d'opinion diffèrent des caractéristiques qui détermineront la similarité de deux phrases issues d'un corpus boursier. Dans le premier cas, les adjectifs, adverbes et verbes de sentiments seront discriminants, tandis que dans l'autre cas ce seront les devises et montants, les verbes d'action et les entités nommées de type COMPAGNIE ou GROUPE.

Nous voulons rendre compte de ce fait au travers d'une mesure de similarité paramétrable, adaptable aux différentes tâches qui peuvent être proposées dans le large domaine du résumé automatique. Nous avons opté par une mesure de comparaison d'ensembles du type Jaccard (Jaccard, 1901), pondérée par une valeur dépendant du type des termes comparés.

Le caractère discriminant des termes dans le cadre d'une mesure de similarité entre phrases ne dépend pas seulement de leur catégorie morpho-syntaxique, mais également de leur importance relative à l'ensemble des documents. Certaines approches, comme (Saggion, 2005) utilisent pour cela la pondération *idf (inverse document frequency)* calculée sur un corpus comme le *BNC (British National Corpus)*, assez large pour que les mesures *idf* soient valides. Cependant, une telle méthode, qui est viable pour la langue générale, devra être révisée pour chaque application à un domaine de spécialité, ou à une langue différents. Pour cette raison, nous avons choisi de pondérer les termes par leur *tf.idf* (*cf fig. 1.4*) dans les documents à résumer, qui reflète bien leur importance vis-à-vis de ceux-ci (Gerard Salton, 1988).

La mesure de similarité qui résulte de ces deux pondérations est présentée en figure

3.3.

$$sim(p_1, p_2) = \frac{\sum_{t_m \in T} poids(t_m) \times fsim(p_1, p_2)}{\frac{fsim(p_1, p_2) + gsim(p_1, p_2)}{\sum_{t_m \in T} poids(t_m)}}$$

$$fsim(p_1, p_2) = \sum_{n_1 \in p_1} \sum_{n_2 \in p_2} tsim(n_1, n_2) \times \frac{tfidf(n_1) + tfidf(n_2)}{2}$$

$$gsim(p_1, p_2) = card\left((n_1 \in p_1, n_2 \in p_2) \mid tsim(n_1, n_2) < \delta\right)$$

où T est la liste des types morpho-syntaxiques et des types d'entités nommées, p_1 p_2 les phrases, tsim la fonction de similarité entre deux termes, et δ le seuil de similarité fixé empiriquement en dessous duquel deux termes sont considérés comme n'ayant aucun lien.

FIGURE 3.3.: Mesure de similarité entre les phrases utilisée par CBSEAS

3.2.4. Classification des Phrases en Classes Sémantiques

Une fois la matrice de similarité établie, CBSEAS regroupe les phrases similaires. Cette étape s'effectue grâce à l'algorithme *fast global k-means* (Likas *et al.*, 2001) (*cf fig. 3.4*), une variante itérative de *k-means*. Cet algorithme de classification a l'avantage d'être itératif, et de pouvoir être facilement adapté à des tâches de résumé avec ajout de groupes de documents à la volée, comme les tâches de mises à jour des campagnes d'évaluation TAC (*cf* §5.2). Il permet également de s'affranchir de la sélection des centres de classes préalablement à la classification, un défaut majeur de *k-means*. De plus, *fast global k-means* fonde les regroupements sur les similarités ou dissimilarités entre les éléments. Cela nous permet d'utiliser aisément différentes mesures de similarité. L'algorithme *k-means* (Forgy, 1965a; MacQueen, 1967), ou *k*-moyennes, réalise une classification en k classes en minimisant la variance intra-classe. Cet algorithme se déroule en quatre étapes :

1. Choisir aléatoirement k objets qui seront les centres de k classes ;
2. Parcourir tous les objets, et les affecter ou les réaffecter à la classe qui minimise la distance entre l'objet et le centre de la classe ;
3. Calculer les barycentres de chaque classe, ils deviennent les nouveaux centres ;
4. Répéter les étapes 2 et 3 jusqu'à convergence. La convergence est atteinte lorsque les classes deviennent stables.

pour tous les e_j *dans* E
$C_1 \leftarrow e_j$
pour i de 1 à k faire
pour j de 1 à i
centre$(C_j) \leftarrow e_m | e_m$ *maximise* $\sum_{e_n in C_j} sim(e_m, e_n)$
pour tous les e_j dans E
placer e_j dans $C_l | C_l$ maximise $sim(centre(C_l, e_j)$
si i<>k alors ajouter un nouveau cluster : C_i qui contient initialement uniquement l'élément le moins bien représenté par sa classe
fin pour

FIGURE 3.4.: Algorithme Fast global k-means

L'algorithme *fast global k-means* commence par créer une classe, dans laquelle sont placés tous les éléments (les phrases dans notre cas). A chaque itération, l'algorithme crée une nouvelle classe dont le centre sera l'élément le moins bien représenté par sa classe. Chaque élément est alors placé dans la classe dont il est le plus proche du centre ; pour finir, les centres des classes sont recalculés.

Les regroupements étant établis d'après la matrice de similarité, elle même uniquement fonction du lexique des phrases, utiliser l'adjectif « sémantiques » pour décrire ces regroupements pourrait paraître abusif. Cependant, étant donné que le sens d'un mot est fortement dépendant de son contexte, la probabilité est élevée pour que deux phrases qui partagent les mêmes mots partagent également le même sens. La figure 3.5 donne un aperçu d'un regroupement réalisé par CBSEAS. Les mots partagés par au moins la moitié des phrases de la classe ont été surlignés afin de pouvoir identifier les raisons du regroupement.

3.2.5. Sélection des Phrases

Après avoir construit la représentation informatique des documents à résumer, le système CBSEAS extrait les phrases qu'il juge les plus pertinentes pour constituer un texte qui servira de base au résumé final. Afin d'obtenir un résumé qui maximise la diversité informationnelle, CBSEAS extrait une phrase par classe sémantique. La phrase sélectionnée doit maximiser un score issu des trois caractéristiques suivantes :

1. la centralité locale (la centralité d'une phrase dans sa classe sémantique) ;
2. la centralité globale (la centralité d'une phrase vis-à-vis d'une requête utilisateur, ou d'un *centroïde* de l'ensemble des documents) ;

- p1 : Chess legend Bobby Fischer , who faces prison if Bobby returns to the United States , can only avoid deportation from Japan if Iceland upgrades its granting of residency to full citizenship , a Japanese lawmaker said Wednesday.
- p2 : " Fischer , 62 , has been detained in Japan awaiting deportation to the United States , where Bobby is wanted for violating economic sanctions against the former Yugoslavia by playing a highly publicized chess match there in 1992.
- p3 : " Fischer , 62 , has been detained in Japan awaiting deportation to the United States , where Bobby is wanted for violating economic sanctions against the former Yugoslavia by playing a highly publicized chess match there in 1992.
- p4 : But during Thursday 's meeting the committee said it was awaiting more documents from Japan to be sure Bobby would be let go if given citizenship in Iceland.
- p5 : Palsson said Fischer , 62 , would be informed of the vote Tuesday morning in Japan , where he has been detained awaiting deportation to the United States on a warrant for violating economic sanctions against the former Yugoslavia by playing a chess match there in 1992.
- p6 : Iceland recently issued Fischer a special passport and residence permit to help resolve the standoff over Bobby 's status.
- p7 : Iceland had previously sent Fischer a passport which showed residency status but not full nationality , a document not deemed sufficient to save him from deportation to the United States.
- p8 : Iceland had already agreed to issue a resident 's permit and special passport to Fischer , but that had not proved enough for the Japanese authorities.

Source : Données de TAC 2009, tâche « Update »

FIGURE 3.5.: Exemple d'une classe sémantique générée par CBSEAS. Les mots partagés par au moins la moitié des phrases sont surlignés afin de visualiser la raison du regroupement de ces phrases.

3. la taille de la phrase par rapport à la taille idéale voulue et paramétrée par l'utilisateur.

Centralité locale

La mesure de la centralité locale vise à déterminer si une phrase est représentative du contenu sémantique de sa classe, tout comme les mesures de centralité globale rendent compte de la représentativité du contenu global des documents. L'idée derrière la mesure de la centralité locale est la suivante : chaque phrase exprime une ou plusieurs « informations atomiques » [2]. L'ensemble des phrases d'une classe exprime ainsi un ensemble I d'informations atomiques. Ainsi, la phrase la plus centrale de cette classe I sera celle qui exprime les informations les plus importantes de I. Nous travaillons d'après l'hypothèse que les informations les plus redondantes sont les plus importantes. Ainsi, la phrase qui contient les informations les plus répétées dans sa classe sera la plus centrale.

Une de nos contraintes consiste à nous affranchir de tout traitement linguistique poussé qui rendrait l'approche dépendante de la langue. Par conséquent, nous ne pouvons fonder

2. Information atomique : information ne pouvant être découpée en plusieurs informations plus petites

cette approche sur l'extraction des informations atomiques comme réalisé dans (Hovy *et al.*, 2005) pour mesurer la centralité globale. La mesure de centralité que nous utilisons est cantonnée à des calculs sur la comparaison d'éléments lexicaux et d'entités nommées.

La similarité entre toutes les phrases d'une même classe est calculée comme décrit en §3.2.3. Nous avons testé deux variantes du *tf.idf* pour mesurer l'informativité des éléments lexicaux, qui travaillent à des niveaux de granularité différents. La première utilise les documents comme base pour le calcul du *tf.idf*, tandisque la seconde, que nous avons appelé *tf.icf*, utilise les classes sémantiques générées dans l'étape précédente (*cf* §3.2.4). L'apport de ces deux mesures est discuté en §5.5.

La phrase p_{max} qui maximise la somme des similarités aux autres phrases reçoit un score de 1, tandis que les autres phrases se voient attribuer un score égal à la similarité à p_{max}.

La figure 3.6 illustre ce processus, avec comme données les phrases de la classe sémantique de la figure 3.5. Afin de simplifier l'illustration, la mesure de similarité utilisée ici est une simple mesure Jaccard. Nous avons comparé notre mesure automatique de centralité locale à une mesure manuelle. Pour cela, nous avons utilisé la méthode Par conséquent, la méthode d'évaluation « Pyramide », décrite en §2.4, qui permet de rendre compte efficacement de la centralité d'une phrase dans sa classe : plus son score « pyramide » est élevé, puisque le score d'une phrase est alors fonction du nombre d'occurencesdes informations qu'elle contient au sein de sa classe. Nous avons établi une liste des informations atomiques contenues dans la classe sémantique. Cette liste est sujette à débat, la notion d'information atomique étant floue, et a fait l'objet d'une validation croisée. Nous avons donné à chaque information atomique un poids égal au nombre de phrases dans lesquelle elle apparaît. Pour finir, nous avons calculé le score « Pyramide » de chaque phrase, en sommant le poids des informations atomiques qu'elle contient.

Centralité globale

La mesure de centralité locale permet de gérer le problème de la diversité en attribuant des scores élevés aux phrases les plus centrales dans leur classe. Cependant, il faut tout de même rendre compte de la centralité de la phrase soit vis-à-vis du contenu global des documents à résumer, soit vis-à-vis d'une requête utilisateur.

Les scores utilisés pour refléter la centralité globale des phrases sont ceux calculés lors de la préparation des documents (*cf* §3.2.2). La figure 3.1 illustre la centralité globale vis-à-vis d'une requête. Les phrases sont issues de la figure 3.5 et la requête est celle fournie dans les données de TAC 2009.

Liste des informations atomiques (SCUs) contenues dans les phrases de la figure 3.5

Informations	**Poids**
Fischer faces deportation	5
Fischer chess player	4
Iceland issued Fischer a residence permit	4
Fischer violated economic sanctions against Yugoslavia	3
Iceland issued Fischer a passport	3
Fischer played a chess match in Yugoslavia in 1992	3
Fischer is 62	3
Fischer has been detained in Japan	3
Fischer faces prison	1
Fischer could avoid deportation	1
Iceland special passport can not avoid him deportation	1
Iceland wants to be sure Fischer would be let go if given citizenship	1

Phrase	p1	p2	p3	p4	p5	p6	p7	p8
Score Pyramide	15	21	21	1	21	7	8	7

La mesure de la centralité locale d'après la méthode pyramide montre que les phrases centrales sont p2, p3 et p5.

	p1	p2	p3	p4	p5	p6	p7	p8	Somme
p1	1.0	0.171	0.171	0.156	0.150	0.1	0.133	0.133	**2.014**
p2	0.171	1.0	1.0	0.091	0.821	0.031	0.094	0.064	**3.272**
p3	0.171	1.0	1.0	0.091	0.821	0.031	0.094	0.064	**3.272**
p4	0.156	0.091	0.091	1.0	0.083	0.075	0.069	0.077	**1.642**
p5	0.15	0.821	0.821	0.083	1.0	0.027	0.108	0.055	**3.065**
p6	0.1	0.031	0.031	0.075	0.027	1.0	0.167	0.315	**1.746**
p7	0.133	0.094	0.094	0.069	0.108	0.167	1.0	0.115	**1.78**
p8	0.133	0.064	0.053	0.077	0.055	0.315	0.115	1.0	**1.812**

D'après la méthode de calcul de centralité locale, les scores de centralité locale seront :

Phrase	p1	p2	p3	p4	p5	p6	p7	p8
Score	0.171	1.0	1.0	0.091	0.821	0.031	0.094	0.064

FIGURE 3.6.: Illustration du concept de centralité locale : mesure d'après la méthode pyramide et mesure automatique réalisée par CBSEAS.

Requête : « Bobby Fischer
Describe efforts to secure asylum in Iceland for chess legend Bobby Fischer. »

phrase	p1	p2	p3	p4	p5	p6	p7	p8
score	0.38	0.178	0.178	0.295	0.178	0.148	0.110	0.068

TABLE 3.1.: Illustration de la centralité globale vis-à-vis d'une requête

Taille des phrases

Un utilisateur peut préférer des résumés composés de phrases plus ou moins longues, pour des raisons de lisibilité. Privilégier des phrases longues augmentera le nombre d'informations véhiculées par le résumé au détriment de la lisibilité tandis que privilégier des phrases courtes augmentera la lisibilité du résumé. Nous avons choisi de laisser l'utilisateur paramétrer cet aspect du système afin que les résumés générés correspondent à ses attentes. Les phrases reçoivent alors un score en fonction de leur taille en nombre de mots et de la taille désirée par l'utilisateur :

$$score_{taille} = \frac{1}{e^{(|taille(phrase) - taille_{requise}|)}}$$

FIGURE 3.7.: Score fonction de la taille des phrases dans CBSEAS

3.2.6. Réordonnancement

Après avoir sélectionné les phrases qui vont constituer le résumé, il est nécessaire de les ordonnancer afin que le texte soit cohérent. Il existe aujourd'hui deux façons d'aborder le problème : en adoptant des méthodes simples fondées uniquement sur l'ordre des phrases et les informations contextuelles des documents telles que la date et le lieu d'émission (Saggion, 2005; Goldstein *et al.*, 2000; Favre *et al.*, 2006), ou en utilisant une modélisation des relations logiques (Moore et Paris, 1993; Hovy, 1993) ou temporelles (McKeown *et al.*, 1999; yew Lin et Hovy, 2001; Barzilay *et al.*, 2002; Okazaki *et al.*, 2005) des textes d'origine.

Ces méthodes, comme nous l'avons vu en §1.9.2, posent les problèmes de la généricité applicative (le type de documents auxquels elles s'appliquent) et de la dépendance à la langue. C'est pourquoi nous proposons une nouvelle approche générique fondée sur la position des phrases et sur notre modèle de représentation des documents. Les classes

```
pour toutes les classes C_i dans C
    pour toutes les classes C_j dans C
        lien (C_i, C_j) <– 0
        pour toutes les phrases p_i dans C_i
            pour toutes les phrases p_j dans C_j
                si doc (p_i) = doc (p_j) alors
                    lien (C_i,C_j) <- lien (C_i, C_j) + (position (p_i)-position(p_j) )/|(position(p_i)-position(p_j)|
                fin si
            fin pour
        fin pour
    fin pour
fin pour
rang <– 0
pour toutes les composantes connexes CC_i du graphe de C
    tant que CC_i n'est pas vide
        C_min ← argmin_{C_j ∈ CC_i} (∑ a_k ∈ aretes(C_j) a_k)
        C_min ← rang
        CC_i ← CC_i \ C_min
        rang ← rang + 1
    fin tant que
fin pour
```

FIGURE 3.8.: Algorithme de réordonnancement

sémantiques regroupant des phrases sémantiquement similaires, nous faisons l'hypothèse qu'établir un ordre global entre les classes fondé sur la position des phrases qu'elles contiennent revient à établir l'ordre des phrases qui en sont extraites.

Notre algorithme de réordonnancement crée un graphe dans lequel les noeuds sont les classes sémantiques, et les arêtes des relations d'ordre entre les classes. La relation d'ordre entre deux classes C_1 et C_2 est la somme des différences entre les positions de toutes les phrases de C_1 et C_2 qui appartiennent au même document. Le graphe est alors parcouru en commençant par le noeud dont la somme des poids des arêtes est la plus faible. Les noeuds du graphe se voient ensuite attribuer un rang lors du parcours du graphe, selon l'ordre dans lequel le graphe est parcouru. Le rang d'un noeud détermine la position de ses phrases dans le résumé généré. Si le graphe comporte des composantes non connexes, alors les ensembles connexes les plus peuplés sont parcourus prioritairement.

Le résumé est alors réordonné en donnant pour position à chaque phrase le rang du noeud de la classe dont elle est tirée.

3.3. Apprentissage Automatique de Paramètres pour le Résumé Automatique

Dans la section précédente, nous avons introduit une méthode de résumé automatique ainsi que le système qui l'implémente, CBSEAS. Ce système utilise des paramètres dont certains ont des valeurs standards ou que nous avions fixées empiriquement. Ces paramètres sont importants car ils influent pour certains directement sur la sélection des phrases qui seront extraites dans le résumé *via* leur intégration dans le calcul des scores. Les autres influencent le regroupement des phrases en classes sémantiques, les scores fondés sur des mesures pondérées, ou le calcul de similarité entre les phrases et à la requête. Dans cette section, nous proposons une méthode pour déterminer les poids les plus adaptés à une tâche donnée.

3.3.1. Problématique

Adapter les poids des différentes composantes des mesures utilisées dans les systèmes de résumé automatique est une problématique peu abordée dans la littérature. Certains travaux laissent supposer que les poids ont été fixés empiriquement (Saggion, 2005; Radev *et al.*, 2004). D'autres, comme (Boudin, 2008), déterminent manuellement les poids en fonction de l'efficacité des différents paramètres isolés. Vu l'étendue de l'espace des paramètres, ces méthodes, si elles permettent de trouver des combinaisons de paramètres efficaces, ne permettent pas d'approximer la solution optimale pour une tâche donnée avec une précision suffisante.

Des approches, présentées en §1.4, permettent d'ajuster automatiquement les combinaisons les plus efficaces. Une bonne optimisation des paramètres a une influence notoire sur les résultats des systèmes de résumé automatique. Osborne (2002) l'a montré en comparant les résultats obtenus après deux méthodes d'optimisation différentes. Nous avons donc voulu développer notre approche d'optimisation des poids, en tenant compte des spécificités de notre système de résumé : absence d'hypothèse d'indépendance des variables, et besoin de parallélisation des calculs. En effet, les calculs d'un résumé automatique et de son score sont coûteux en temps ; explorer l'espace des combinaisons de poids et confronter ces combinaisons à une évaluation automatique nécessite une grosse puissance de calcul.

3.3.2. Choix d'un algorithme d'optimisation

Dans notre cas, nous ne pouvons pas prouver la régularité ni la continuité d'une fonction de l'espace des hypothèses au score de résumé. Les paramètres que nous cherchons à optimiser ne sont en effet pas uniquement des poids utilisés pour des combinaisons linéaires de caractéristiques. La continuité de cette fonction est un pré-requis à l'utilisation de méthodes à base de descente de gradient. De plus, certains paramètres opèrent à des étapes différentes de CBSEAS, et interviennent sur différents aspects de la sélection des phrases. Construire un modèle probabiliste de l'espace des hypothèses est une tâche très difficile pour notre système, eu égard aux relations de dépendance compliquées entre les différents paramètres. L'utilisation d'un algorithme génétique, qui peut s'affranchir de ces deux contraintes, apparaît donc la méthode la plus appropriée afin d'apprendre la meilleure combinaison de paramètres pour notre système.

Les algorithmes génétiques ont été introduits par John Holland (Holland, 1975). Son but était d'utiliser la métaphore de l'adaptation naturelle des espèces afin d'adapter de manière optimale des paramètres à leur environnement. L'idée principale est d'engendrer une génération composée d'individus caractérisés par les paramètres à optimiser, et grâce à des opérateurs de mutation et de croisement entre certains individus de cette génération, d'engendrer une nouvelle génération qui sera mieux adaptée à son environnement que la précédente.

3.3.3. Notre algorithme génétique

Dans cette section, nous présentons en détails l'algorithme génétique que nous avons implémenté.

Méthode de sélection des individus

La sélection des individus comme parents d'une nouvelle génération est une tâche problématique des algorithmes génétiques. En effet, sélectionner uniquement les individus les plus forts (eugénisme) fait converger l'algorithme plus rapidement, mais comporte le risque de tomber dans un maximum local. Il faut donc préserver la diversité de chaque génération en sélectionnant également des individus plus faibles.

Pour ce faire, nous avons sélectionné une méthode largement utilisée : la sélection par tournoi. Pour chaque génération composée de γ individus, μ tournois entre λ individus sont organisés. Le gagnant de chaque tournoi est sélectionné pour faire partie des parents de la prochaine génération. L'ajustement des paramètres μ et λ permet de faire varier

la vitesse de convergence de l'algorithme et sa capacité exploratoire.

Opérateur de mutation

Nous ne savons pas quels paramètres sont dépendants les uns des autres. Nous voulons donc que notre opérateur de mutation change plusieurs paramètres par individus. Afin d'éviter une variation trop élevée des individus due à la mutation simultanée de paramètres trop nombreux, nous avons choisi de limiter la quantité de variation de chaque paramètre. Nous affaiblissons la probabilité d'obtenir une trop forte variation des individus en réalisant une variation logarithmique des poids des paramètres.

Opérateur de croisement

L'opérateur de croisement sert à créer un nouvel individu fils à partir de n individus parents. L'individu fils héritera de caractéristiques issues de chacun de ses parents. Nous avons choisi de créer un individu à partir de deux parents. Chacune de ses caractéristiques sera copiée d'un des deux parents, avec une probabilité équivalente pour chacun des parents, quel que soit leur score. Ne pas favoriser les meilleurs individus permet de réduire le risque de tomber dans un maximum local.

Création d'une nouvelle génération

Nous avons choisi empiriquement les paramètres de l'algorithme génétique, de manière à obtenir un bon compromis entre exploration de l'espace des données (complétude) et rapidité de convergence de l'algorithme. Chaque génération est composée de 200 individus. L'algorithme organise vingt tournois impliquant dix individus sélectionnés aléatoirement. Chaque nouvelle génération est composée des vingt gagnants, de quatre-vingt-dix individus créés par mutation des gagnants, et de quatre-vingt-dix individus issus de croisements entre les gagnants.

3.3.4. Paramètres expérimentaux

Nous avons utilisé les données de la tâche Update de TAC 2008 (*cf* §5.2) afin de constituer un corpus d'entraînement et un corpus de test. La génération automatique d'un grand nombre de résumés est une tâche coûteuse en temps, c'est pourquoi nous avons dû réduire au maximum le corpus d'entraînement. Celui-ci est composé de cinq jeux de données issues de la tâche Update de TAC 2008, soit dix résumés à créer au total

(cinq résumés classiques et cinq résumés « mise à jour ») par individu, et donc près de 2000 résumés à créer par génération.

Chaque individu est composé de 14 paramètres. Certains de ces paramètres correspondent à des caractéristiques décrites ultérieurement.

- Nombre de phrases en sortie de CBSEAS
- Longueur moyenne des phrases en sortie
- Nombre de phrases pré-sélectionnées par CBSEAS (résumé standard)
- nb de phrases pré-selectionnées par CBSEAS (résumé update) (*cf §5.3*)
- poids des différentes catégories morpho-syntaxiques (noms communs, noms propres, verbes, adjectifs, adverbes, nombres)
- poids du score de similarité au centre de la classe
- poids du score de similarité à la requête
- poids du score de longueur de phrase
- Réduction de la similarité « update » (*cf §5.3*)

3.3.5. Résultats

L'algorithme a été lancé en parallèle sur 17 machines qui calculent en moyenne un résumé en 45 secondes quand la puissance allouée à CBSEAS est au maximum. En raison des déconnexions de machines et de l'utilisation des machines pour d'autres tâches, le temps de calcul a été plus que doublé puisque le calcul de 270 générations a pris près de 10 jours. Le temps de calcul aurait pu être considérablement diminué. En effet, l'accès à la ressource WordNet pour le calcul de similarité sémantique entre *synsets* est coûteuse en temps. Sans accès à cette ressource, la génération d'un résumé dure environ 3 secondes. Mettre en cache les similarités entre *synsets* après la première génération afin de les charger ultérieurement sans devoir accéder à WordNet aurait été une stratégie d'optimisation payante.

La figure 3.9 montre la convergence de l'algorithme. Elle présente les résultats moyens de chaque génération ainsi que le résultat obtenu par le meilleur individu de chaque génération. L'algorithme a convergé à la $222^{\text{ème}}$ génération.

La figure 3.2 présente les paramètres de CBSEAS obtenus après optimisation par l'algorithme génétique. Les résultats témoignent du fait que donner un poids faible aux noms propres a une influence positive sur les scores ROUGE. Les noms communs, adjectifs et verbes semblent être les types morpho-syntaxiques les plus importants pour le système.

Le poids des noms propres est faible car la plupart des phrases pré-sélectionnées contiennent les mêmes noms propres. Ceci est dû au fait que les phrases pré-sélectionnées

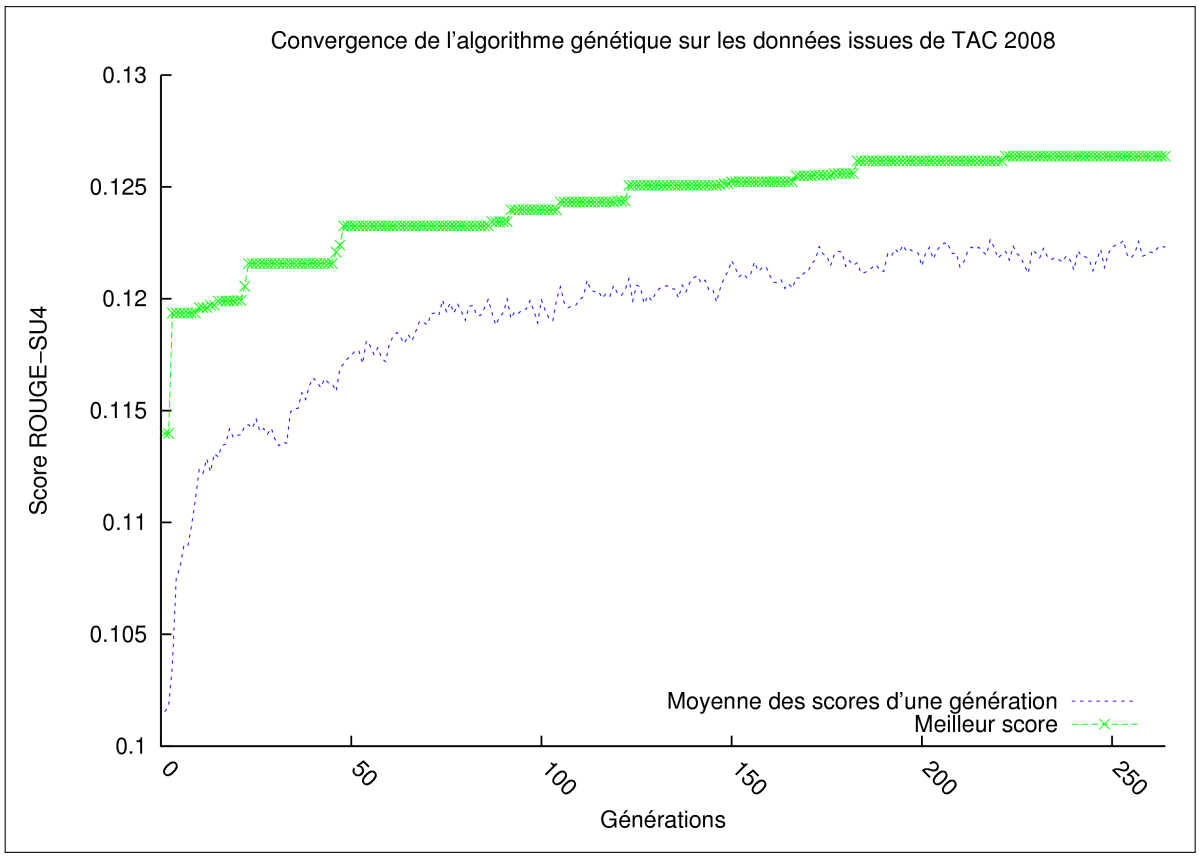


FIGURE 3.9.: Convergence de l'algorithme génétique

paramètre	valeur
nb de phrases	14
longueur moyenne	8
nb de phrases pré-selectionnées	47
nb de phrases pré-selectionnées update	83
poids des noms communs	171
poids des noms propres	29
poids des verbes	207
poids des adjectifs	270
poids des adverbes	12
poids des nombres	66
poids de sim. au centre du cluster	7
poids de sim. à la requête	25
poids score longueur de la phrase	72
Réduction de la sim. « update »	0.87

TABLE 3.2.: Combinaison gagnante de paramètres

sont celles qui sont proches de la requête utilisateur, qui est souvent articulée autour d'entités nommées. Ainsi, faire jouer un rôle important aux noms propres dans le calcul de similarité entre les phrases apporte du bruit à la mesure de similarité, et affecte négativement l'algorithme de classification. D'une manière plus générale, cela confirme les observations d'Aone *et al.* (1998), selon lesquelles décroître l'impact des noms propres dans la méthode de sélection des phrases pour le résumé automatique de dépêches accroît la qualité des résumés.

La variable « réduction de la similarité *update* » est une variable spécifique à la tâche « Résumé et mise à jour » de la campagne d'évaluation TAC 2008. L'objectif de cette variable est expliqué plus avant dans le document, en §5.3.3, dans le chapitre consacré à cette tâche. Elle augmente les scores des résumés quand elle est fixée de manière à ce qu'elle augmente les similarités entre les phrases issues du jeu de documents d'origine (D1) et celles issues du jeu de documents destiné à la mise à jour (D2) (*cf §5.2*). Cela signifie qu'affaiblir la probabilité qu'une phrase de D2 se retrouve liée à une classe de la mise à jour améliore la qualité de la gestion de la mise à jour [3].

Le paramètre « Similarité au centre de la classe » reçoit le poids le plus faible des paramètres de la dernière étape de CBSEAS. Etant donné que les travaux récents ont prouvé la pertinence des méthodes à base de graphes pour le résumé automatique, cela tend à prouver que notre mesure de similarité n'est pas adaptée à un tel score. D'autres mesures de similarité doivent être testées afin d'améliorer l'impact de ce paramètre.

3.3.6. Evaluations

Nous avons choisi d'évaluer l'apport de l'algorithme génétique de deux manières : en comparant dans un premier temps les résultats des métriques ROUGE obtenus sur les données de TAC 2008 avec et sans algorithme génétique, puis en réalisant une évaluation subjective en aveugle de différents systèmes sur les mêmes données.

Evaluation automatique

L'évaluation automatique a porté sur les données de la tâche *Update* de TAC 2008 dont la partie qui a servi à l'apprentissage des paramètres de CBSEAS a été enlevée. L'échantillon destiné à l'évaluation est composé de 43 jeux de documents à résumer. La métrique utilisée est ROUGE-SU4, fondée sur la comparaison de bigrammes à trous, décrite en §2.1.3.

3. Pour plus de précision sur la gestion de la mise à jour et l'influence de cette variable, cf 5.3

La figure 3.10 présente les résultats des différents participants à la tâche Update de TAC 2008 ainsi que les résultats d'une version de CBSEAS sans optimisation des paramètres, et d'une version de CBSEAS avec optimisation des paramètres.

On constate une nette amélioration des résultats de CBSEAS, malgré un échantillon dédié à l'apprentissage assez faible, puisque composé uniquement de cinq jeux de documents à résumer.

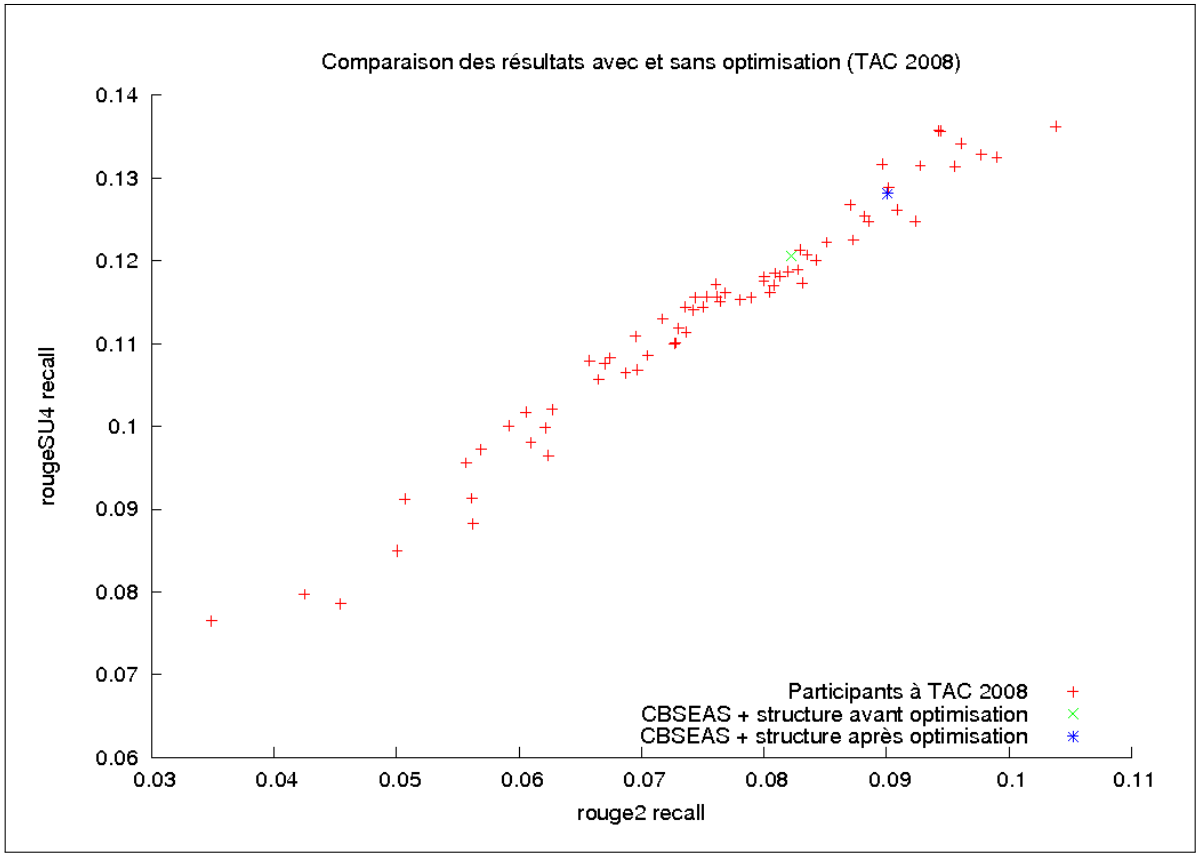


FIGURE 3.10.: Résultats de l'évaluation automatique de l'algorithme génétique (données TAC 2008)

Cette évaluation automatique de l'apport de l'algorithme génétique, réalisée avec une mesure ROUGE-SUn qui est assez corrélée aux jugements humains, nécessite tout de même une validation par une évaluation humaine. En effet, la seule évaluation d'un résumé par une méthode automatique ne permet pas de rendre compte de la lisibilité dudit résumé.

Evaluation manuelle

Nous avons décidé de procéder à une évaluation manuelle comparative. Les différents systèmes évalués le sont les uns par rapport aux autres. Pour ce faire, nous avons sélectionné le meilleur système de résumé automatique de TAC 2008, et CBSEAS avant et après optimisation des poids. Afin de faciliter le travail de l'évaluateur, nous avons limité le nombre des résumés à évaluer. Nous avons sélectionné aléatoirement 15 paires de jeux de documents, soit 15 paires de résumés. Les résumés des différents systèmes se sont vu attribuer des identifiants aléatoires, afin d'éviter que l'évaluateur puisse identifier les systèmes dont les résumés proviennent, et ainsi obtenir une véritable évaluation en aveugle.

Après lecture des trois paires de résumés à évaluer et des jeux de documents correspondant, les évaluateurs répondent aux deux questions suivantes, en dissociant les résumés standard des résumés « mise à jour » :

- Lequel de ces trois résumés reflète le mieux le contenu des documents ? (ce résumé reçoit un score de 6)
- En les comparant à la meilleure paire de résumés, donnez une note entre 1 et 5 aux deux autres paires de résumés :
  - 1 : aucune comparaison n'est possible, le meilleur résumé est bien trop au-dessus ;
  - 2 : ce résumé est beaucoup moins bon que le meilleur ;
  - 3 : ce résumé est moins bon que le meilleur ;
  - 4 : ce résumé est un peu moins bon que le meilleur ;
  - 5 : ce résumé est quasiment aussi bon que le meilleur.

Afin de juger la qualité d'un résumé, les évaluateurs ont dû s'appuyer sur les trois points suivants :
- lisibilité ;
- informativité ;
- non-redondance.

Les résultats de l'évaluation manuelle, présentés dans le tableau 3.3, montrent également une nette amélioration du système CBSEAS, et viennent confirmer les évaluations automatiques ROUGE. Cependant, malgré les améliorations, CBSEAS optimisé par notre algorithme génétique n'arrive toujours pas au niveau des tout meilleurs systèmes qui ont concouru à TAC 2008, bien qu'il s'en rapproche beaucoup en se situant parmi les dix meilleurs systèmes. Ceci s'explique par deux points : notre mesure de similarité à la requête, trop peu performante. La plupart des systèmes utilisent en effet des modules d'extension de requête fondés sur une analyse distributionnelle des documents à résumer. Celle-ci permet d'étendre la requête selon le champ lexical utilisé dans les documents, plutôt qu'en se fondant sur les données d'une ressource lexicale générale telle que WordNet. De plus, une telle méthode est facilement adaptable à n'importe quelle

	Meilleur système à TAC 2008	CBSEAS sans optimisation	CBSEAS avec optimisation
Résumés standards			
Nombre de fois gagnant	9	2	4
Note moyenne	4.7	3.9	4.3
Résumés « mise à jour »			
Nombre de fois gagnant	8	2	5
Note moyenne	5	3.7	4.5
Tous Résumés			
Nombre de fois gagnant	17	4	9
Note moyenne	4.8	3.8	4.4

TABLE 3.3.: Evaluation manuelle de l'apport de l'algorithme génétique à CBSEAS

langue, au contraire de celle que nous utilisons, qui nécessite le développement préalable de ressources lexico-sémantiques. Par ailleurs, les meilleurs systèmes utilisent des méthodes de filtrage de phrases fondés sur leur structure ainsi que des post-traitements visant à supprimer toute ou partie d'une phrase jugée inutile, qui permettent à la fois d'intégrer plus d'informations et de faciliter la lecture du résumé (Gillick *et al.*, 2009).

3.3.7. Conclusion

Dans ce chapitre, nous avons présenté un algorithme d'optimisation de paramètres d'un système de résumé automatique. Notre choix s'est porté sur un algorithme génétique, plus adapté à notre tâche en raison des relations trop variées de dépendance entre les différents paramètres à optimiser. Cet algorithme utilise les évaluations automatiques ROUGE, qui ont montré une forte corrélation aux jugements humains, comme fonction de *fitness*.

La génération automatique d'un résumé et son évaluation étant des tâches particulièrement coûteuses en temps, les jeux de données destinés à l'apprentissage ont été restreints au maximum. Malgré cela, l'utilisation de cet algorithme d'apprentissage a conduit à une nette amélioration des résultats de notre système de résumé automatique, toujours en deça des meilleurs systèmes qui ont participé à TAC 2008.

3.4. Discussion

Le principal intérêt de notre méthode de résumé est le regroupement des phrases en classes sémantiques pour détecter la redondance, afin de sélectionner les phrases les plus

centrales des documents à résumer et de maximiser la diversité informationnelle au sein du résumé généré. Il ressort des campagnes d'évaluation sur lesquelles notre méthode a été évaluée que le regroupement en classes sémantiques est efficace, puisque CBSEAS se classe parmi les trois meilleurs systèmes en diversité (*non-redundancy*) à la tâche « Résumé d'opinions » de TAC 2008 (*cf §6.4.2*).

Les résultats de l'algorithme génétique montrent que la centralité locale n'est pas le trait le plus important dans le calcul du score des phrases. Il permet cependant de trancher efficacement entre deux phrases qui ont des scores de requête proches l'une de l'autre.

3.5. Bilan

Dans cette section, nous avons présenté une approche novatrice pour le résumé automatique. A travers cette approche, nous avons voulu rendre en compte du fait qu'il n'y a pas une information centrale dans les documents à résumer, mais plusieurs informations importantes que certaines phrases sont le plus à même de présenter. Cette approche se fonde sur le regroupement des phrases en classes sémantiques, et applique une méthode de sélection des phrases à extraire qui est fonction de la centralité intra-classe et de la similarité à une requête utilisateur – ou à un centroïde dans le cas de résumés non guidés par une requête.

Nous avons également présenté une méthode d'optimisation des paramètres de CBSEAS, le système qui implémente cette approche, que nous avons évaluée sur les données de la campagne d'évaluation TAC 2008. Cette optimisation, bien qu'imparfaite en ce qui concerne les jeux de données d'apprentissage – trop restreints – entraîne une amélioration conséquente des résultats de CBSEAS, aussi bien pour l'informativité des résumés que pour la lisibilité, en raison de facteurs décrits en §4.2.6.

4. Analyse discursive de documents pour le résumé automatique

Sommaire

Nous avons créé un système de résumé automatique générique, facilement adaptable à de nouvelles langues car fondé en grande partie sur des statistiques. Notre système utilise également un étiqueteur morpho-syntaxique, le *tree-tagger*, qui prend en charge onze langues différentes. Son utilisation permet de traiter plus efficacement la majorité des langues européennes. Nous avons également implémenté l'utilisation d'une ressource lexico-sémantique, *WordNet*, qui est inactive dès que l'on traite une langue autre que l'anglais.

Cette généricité a un coût : CBSEAS ne peut pas rivaliser avec des systèmes qui traitent spécifiquement une langue en réalisant des traitements linguistiques poussés. C'est pourquoi nous proposons dans ce chapitre deux méthodes pour améliorer la qualité des résumés générés : d'un côté, l'intégration d'un outil de reconnaissance des entités nommées et de résolution d'anaphore, de l'autre la gestion fine de types de documents particuliers *via* l'identification de leur structure rhétorique. Cette deuxième partie a fait l'objet de deux publications, dans « TALN 2009 » et « RANLP 2009 », dont les références

complètes se trouvent en annexe C.

4.1. Reconnaissance des entités nommées, résolutions d'anaphore et de co-référence

4.1.1. Enjeux

Les entités nommées « correspondent traditionnellement à l'ensemble des noms propres présents dans un texte, qu'il s'agisse de noms de personnes, de lieux ou d'organisations, ensemble auquel sont souvent ajoutées d'autres expressions comme les dates, les unités monétaires, les pourcentages et autres »(Ehrmann et Jacquet, 2006). Elles sont généralement très fréquentes en corpus et composent près d'un tiers des mots des corpus de dépêches que nous avons étudiés (corpus de la tâche « Résumé et mise à jour » de TAC 2008, corpus « Côte d'Ivoire » du projet Infom@gic).

Reconnaître et traiter spécifiquement les entités nommées constitue une problématique importante du fait de leur fréquence d'apparition et de l'intérêt que leur portent les utilisateurs de systèmes d'extraction d'information. De plus, les entités nommées sont souvent composées de plusieurs mots. Leur variabilité influence négativement leur pondération dans les différents calculs de CBSEAS. Reconnaître ces entités et regrouper les différents mots qui composent une même entité nommée pourrait permettre d'améliorer les calculs de CBSEAS sur les phrases.

Les entités nommées sont sujettes à des variations morphologiques ainsi qu'à la pronominalisation. Une même entité nommée peut donc avoir plusieurs réalisations différentes, comme illustré en figure 4.1. L'entité « Mark Felt » est exprimée par « Mark », « Felt », « Mark Felt Sr », et « Deep Throat ». Identifier les trois dernières réalisations comme référant à Mark Felt peut se faire *via* des systèmes à base de règles portant sur la forme des syntagmes. En revanche, la dernière réalisation pose problème, puisque l'on ne peut résoudre sa référence que par une analyse syntaxique et sémantique de l'extrait de phrase « *Felt told he was "Deep Throat"* ». Relier les différentes réalisations d'une entité nommée à un unique référent serait positif pour le calcul de similarité entre phrases. De plus, cela permettrait de compresser au maximum les résumés générés, en remplaçant toutes les entités nommées par leur réalisation la plus courte.

Résoudre les co-références améliorerait non seulement les différents calculs de similarité de notre système, mais permettrait également de rendre les résumés générés plus lisibles. Un résumé doit en effet se suffire à lui-même, et doit donc être exempt de tout élément

A retired FBI official, Mark Felt, was the "Deep Throat" source who leaked secrets to two Washington Post reporters about the Watergate scandal that brought down former president Richard Nixon, his family said Tuesday.
Felt's grandson, Nick Jones, made the claim in a statement read to reporters outside the family home in Santa Rosa, California, following an article in Vanity Fair in which Felt told the magazine he was "Deep Throat." The identity of the person who provided Washington Post reporters Bob Woodward and Carl Bernstein with crucial information about the Watergate cover-up has been one of journalism's most enduring mysteries.
Woodward and Bernstein have said they will not reveal the identity of "Deep Throat" until his death. "My grandfather is pleased that he is being honored for his role as Deep Throat with his friend Bob Woodward," said the statement read by his grandson Jones. "The family believes my grandfather, Mark Felt Sr., is a great American hero who went well and above the call of duty at much risk to himself," the statement said. "We all sincerely hope the country will see him this way, as well.
"Mark had expressed reservations in the past about revealing his identity and about whether his actions were appropriate for an FBI man," the statement said. "But as he recently told my mother, I guess people used to think Deep Throat was a criminal. But now they think he's a hero."

The statement also called for reporters to respect the privacy of Felt, who is 91.

FIGURE 4.1.: Illustration des différentes réalisations d'une même entité nommée au sein d'une dépêche tirée du corpus de la tâche « résumé et mise à jour » de TAC 2008.

non auto-référentiel [1] dont l'antécédent n'y figure pas également. Des systèmes comme celui de (Edmundson, 1969) ou (Luhn, 1958) pénalisaient les phrases contenant des pronoms afin d'éviter de les extraire dans le résumé. Les récentes avancées dans la résolution d'anaphores et de co-références permettent aujourd'hui d'extraire ces phrases si elles sont jugées les plus pertinentes, en remplaçant les pronoms par leur référent. La figure 4.2 présente des résumés automatiques extraits des résultats de TAC 2008. Nous avons choisi parmi les résultats soumis par les participants à la campagne des résumés qui illustrent bien l'importance de la résolution d'anaphore, ou de l'élimination d'éléments non auto-référentiels. Les résumés deviennent incompréhensibles en raison du trop grand nombre d'entités non auto-référentielles non résolvables.

Nous présentons dans la section suivante nos réalisations dans CBSEAS concernant l'intégration de la reconnaissance d'entités nommées et d'un outil de résolution d'anaphore et de co-référence, ceci afin d'améliorer la sélection des phrases et la lisibilité des résumés générés.

1. Nous entendons par là tout élément qui n'est pas son propre référent. Exemple : les pronoms, l'utilisation du prénom seul pour référer à une personne...

D0806 : Describe the legal developments in the Terry Schiavo case.

The US House of Representatives passed legislation late Wednesday that would allow such a case to be moved to federal courts, which have twice turned down Schiavo's request to intervene.
He criticized the DCF move, saying it "reeks of the intervention of politics into the case".
Some doctors have testified that Terri Schiavo is in a persistent vegetative state with no hope for recovery, but the Schindlers have countered with other medical opinions that she might improve with rehabilitation.

D0808 : Describe the events related to Christian minorities in Iraq and their current status.

He was abducted and taken off in a car.
On Oct. 16, bomb attacks targeted five churches in Baghdad _ which damaged buildings but caused no casualties.
Samaan said Christians are vulnerable in predominantly Muslim Iraq.
He said throughout the years, Iraq's Christians had kept to themselves trying to keep out of trouble.
Christians hold only one portfolio in the interim government of Iyad Allawi.

D0801 : Describe developments in the production and launch of the Airbus A380.

and some others have expressed frustration at program delays.
Probably we won't.
"Tests were rather good.
Federal Aviation Administration will decide what if any modifications are required, she said.
opened 1.2 percent lower Friday on Euronext, but had recovered to euro31.58 (US$37.45) by early afternoon, up 1.5 percent.
Singapore Airlines has ordered 10 A380s with an option for another 15.
It was to depart for Asia on Wednesday.
By 2008, 38 airports will be ready and 60 by 2010, Airbus said in a statement.
"We have not yet received a delivery schedule from Airbus.

FIGURE 4.2.: Illustration de l'importance de la résolution d'anaphores et de l'élimination d'éléments non auto-référentiels dans les résumés. Textes extraits de résumés produits par des systèmes ayant participé à TAC 2008. Les entités non auto-référentielles sont surlignées. Les liens de co-référence que l'on pourrait établir sont erronés (« Saaman » et « He »)

4.1.2. Réalisations

Etiquetage d'entités nommées

Il existe aujourd'hui bon nombre d'outils d'analyse d'entités nommées performants et disponibles gratuitement. Ceux-ci ne sont certes pas indépendants de la langue, et leur intégration est au prix de la généricité de notre système ; celle-ci est cependant indispensable afin de permettre à notre système d'être concurrentiel, mais peut être désactivée pour le traitement de langues qui ne disposent pas de tels outils. Nous avons donc choisi d'utiliser l'existant plutôt que de développer un nouveau système. Parmi les systèmes libres et gratuits, on peut citer TextPro (Emanuele Pianta et Zanoli, 2008), Supersense (Ciaramita et Altun, 2006), et ANNIE (Cunningham *et al.*, 2002). A notre connaissance, seul le troisième système propose un outil de résolution d'anaphore et de co-référence en plus de la détection d'entités nommées. Dans un souci de modularité, nous avons donc choisi d'utiliser le système ANNIE développé pour l'anglais et dont les performances sont au niveau des deux premiers (Marrero *et al.*, 2009).

Le module d'étiquetage en entités nommées du système ANNIE utilise un module nommé *gazetteer*, qui consiste en un dictionnaire. Celui-ci est composé d'entités nommées – noms de villes, de pays, d'organisations, jours de la semaine–, mais également d'indices textuels tels que les désignateurs typiques de compagnies (e.g. *Cie*, *Ltd.*), de titres (e.g. « monsieur », « Président »). Le module d'étiquetage en entités nommées applique ensuite des règles établies manuellement, qui décrivent les patrons à reconnaître et les annotations qui en résultent. Ces règles utilisent les annotations réalisées par d'autres modules du système ANNIE, notamment le *tokenizer* et l'étiqueteur morphosyntaxique. Les entités nommées sont catégorisées en quatre types différents : personne, organisation, lieu et date.

Résolution d'anaphore et de co-référence

Nous avons ajouté à la chaîne de traitement pour l'étiquetage d'entités nommées du système ANNIE le module optionnel d'appariement orthographique d'entités nommées. Ce module rapproche des entités nommées sur la base des sous-chaînes qu'elles partagent. Il utilise également un dictionnaire composés de prénoms et de leur surnom (e.g. « *Christopher* », « *Chris* »), de noms d'entreprises et de leur accronyme (e.g. « Banque Nationale de Paris », « BNP »). Une entité nommée citée uniquement par son prénom sera rattachée à l'entité complète « prénom, nom » si elle existe. Cependant, afin d'éviter les rattachements erronés, des règles empêchent de rattacher « *John* » à « *John Smith* » s'il existe une séquence dans le texte comme « John Smith ... John ... John Anderson ». D'autres règles évitent de rapprocher des entités morphologiquement proches en se fondant sur le *middle name* ou sur le genre : ainsi, « John C. Smith » ne sera pas rattaché

à « John Q. Smith », ni « John Smith » à « Mrs Smith ». Toutes les entités nommées reçoivent alors une liste des entités nommées avec lesquelles elles co-réfèrent.

Deep Throat has a book deal and a movie deal, and he could end up being played by Tom Hanks.
the family of 91-years-old Mark Felt , who revealed his role as the Washington Post's key Watergate source two weeks ago, has choosen PublicAffairs book to publish a combination of autobiography and biography, publisher and CEO Peter Osnos said Wednesday night.
Osnos said that Universal Pictures has optioned Felt's life story and the book for a movie to be developped by Hanks ' production company , Playtone.
He said the book will blend Felt's own writing about his life, including his out-of-print 1979 memoir , "the FBI Pyramid : from the Inside ," and some unpublished material, with contribution from Felt's family and from lawyer John O'Connor , who has been advising the Felts.
O 'Connor wrote the Vanity Fair article that named Felt as Deep Throat after the secret had been kept for 33 years.
The additional material from Felt , Osnos said, includes discussion of how he decided to provide guidance to Post reporter Bob Woodward, and why.
The book is to be published next spring.
Its work title be "a G-Man 's Life : the FBI , being 'Deep Throat ' and the Struggle for Honor in Washington.
"David Kuhn, the New York-based agent who has been representing the family in conjunction with Calif.-based Creative Artists Agency, said Wednesday night that "Hanks' company was interested in the rights to the story within a day or two" of the revelation of Deep Throat's identity.
He said the movie deal was concluded Tuesday night.
Kuhn would not comment on the sum paid in either the book or the movie deal, except to say that the family's decision on the book "is not based on money" but rather on the vision for its publication put forth by Osnos, a former Washington Post reporter and editor who helped cover the Senate Watergate hearing.
PublicAffairs generally doesn't pay advance of more than $75,000.
Neither Felt's daughter , Joan, nor his son , Mark Jr.
, returned phone call Wednesday night.
Kuhn said they would be "interviewed for the book and would participate in the storytelling.
"During the two weeks the Felt project was being shopped , it met with considerable skepticism from publisher.
The two reasons usually cited were the health of the elderly Felt, whose physical and mental deterioration appeared to preclude new contribution from him, and the presence of a compete book from Woodward, who had already written his own version of the deep Throat story.
Woodward's "the secret man : the story of Watergate's Deep Throat" is being rushed into print.
Robert's publisher , Simon & Schuster , has set a publication date of July 6.
Jamie Raab said last week that Schuster had heard the Felt pitch and decided not to bid.
"The book is not a Watergate book per se ," Raab said.
"It is not going to answer some of the linger questions.
... They are sort of written around Mark Felt's life.
" Little, Geoff Shandler expressed skepticism as well, but Shandler didn't rule out the success of a Felt book.
" Traveling in Woodward's wake could be a profitable place to be , depending on what you pay and when you publish ," Shandler said.
"There is an equation that would work.
"But publishing executive agreed that the real money was on the Hollywood end.

FIGURE 4.3.: Illustration des sorties de l'étiquetage en entités nommées du système ANNIE. Les entités nommées sont surlignées. Les entités nommées appariées par le système ANNIE ont le même code couleur. Les entités en rose pâle sont annotées uniques par ANNIE, et ne sont donc pas reliées ensemble.

Un exemple de sortie du module d'étiquetage d'entités nommées et de résolution de co-référence est présenté en figure 4.3. Les balises xml ont été remplacées par des codes couleur pour rendre le texte plus compréhensible. Toutes les entités nommées rattachées ensemble sont surlignées de la même couleur. Nous avons choisi de surligner d'une même couleur – rose pâle – tous les éléments reconnus comme des entités nommées et dont aucun co-référent n'a été identifié par le système ANNIE. On peut constater

quelques erreurs d'étiquetage, comme « *O' Connor* », « *Simon & Schuster* » et « *Deep Throat* » séparés en deux entités distinctes, ou encore « *Struggle* », « *Honor* », « *Little* » et « *Traveling* » étiquetés à tort comme des entités nommées.

Les traitements décrits ci-dessus permettent d'établir si deux entités nommées ont le même référent grâce à une analyse de leur forme. Cependant, ceci n'est pas suffisant, puisque nous voulons également résoudre les anaphores et les co-références pronominales afin d'améliorer le calcul de similarité entre les phrases et les résumés générés, mais également supprimer des résumés les entités non auto-référentielles.

L'outil de résolution d'anaphores et de co-référence pronominale est décrit de manière détaillée dans la documentation en ligne de GATE[2]. Il exclut d'abord les pronoms non référentiels (e.g. « *It is raining.* »). Il sélectionne ensuite pour chaque pronom personnel et adjectif possessif une liste de candidats référents, qui partagent le genre de l'entité référentielle considérée. Est alors sélectionné le candidat référent anaphorique le plus proche s'il existe, le candidat référent cataphorique le plus proche sinon.

Cette approche n'utilise pas d'analyse syntaxique et se fonde uniquement sur la distance d'un pronom ou d'un adjectif possessif à ses candidats référents et leurs positions relatives. Or il est rare que l'entité la plus proche d'un pronom soit son référent. La résolution du genre du pronom et de celui des entités nommées par ANNIE permet cependant de résoudre certains cas. La figure 4.4 présente une sortie du système de résolution d'anaphores. Les entités nommées et les pronoms identifiés comme co-référents sont surlignés de la même couleur. Les mauvaises associations sont rayées.

Ainsi, CBSEAS ne comparera plus les chaînes de caractères des entités nommées, mais considèrera une entité $E1$ comme égale à $E2$ si $E1$ apparaît dans la liste d'entités qui co-réfèrent avec $E2$. Par ailleurs, nous remplaçons dans les résumés générés tous les pronoms et adjectifs possessifs par leur référent identifié par le système ANNIE.

4.1.3. Evaluation

L'évaluation des deux apports à CBSEAS présentés ci-dessus a été réalisée de deux façons différentes. L'intégration de la reconnaissance des entités nommées a été évaluée *a posteriori* sur les données de la tâche « Résumé et mise à jour » de la campagne d'évaluation TAC 2008, et la même version de CBSEAS a été proposée sur la même tâche en 2009 pour une évaluation manuelle. Le même système intégrant la résolution d'anaphores a été évalué uniquement manuellement sur les données de TAC 2009. Le système précédent sert alors de point de point de comparaison. Nous avons jugé l'évaluation de l'intégration de la résolution d'anaphores de manière automatique inutile, car l'évalua-

2. http ://gate.ac.uk/sale/tao/splitch6.html

FIGURE 4.4.: Illustration des sorties de l'étiquetage en entités nommées et de la résolution d'anaphores du système ANNIE. Les entités nommées appariées par le système ANNIE ont le même code couleur (à l'exception des entités en rose pâle). La résolution de co-référence des entités rayées et suivies d'un * est erronée.

tion de l'apport d'une telle technique doit faire intervenir la compréhension des textes à résumer et des résumés par les juges. Les protocoles d'évaluation et les données utilisés pour les campagnes TAC sont détaillés en §5.5.

Le tableau 4.1 présente les résultats de CBSEAS avec et sans la gestion des entités nommées sur les données de TAC 2008 avec une mesure d'évaluation automatique (ROUGE-SU4). La prise en compte d'un étiquetage précis des entités nommées et de la résolution orthographique de la co-référence permet une amélioration des résultats. Celle-ci peut paraître minime, mais elle permet de faire progresser notre système de résumé de la 30^{eme} à la 24^{eme} place sur 72 systèmes.

	Score ROUGE-SU4
CBSEAS	0.1167
CBSEAS + EN	0.1183

TABLE 4.1.: Evaluation de l'apport de l'étiquetage en entités nommées sur les données de TAC 2008.

Nous avons donc proposé à TAC 2009 notre meilleure version de CBSEAS (avec la gestion de la structure et des entités nommées) ainsi qu'une copie de cette version qui implémente la gestion des anaphores présentée en §4.1.2. Nous avons extrait des résultats de TAC 2009 les deux évaluations les plus intéressantes pour nous : l'évaluation de performance générale (une évaluation manuelle subjective) et l'évaluation pyramidale (*cf §2.4*), que nous présentons en figure 4.5.

La version qui implémente la résolution d'anaphores, notée v0.5 dans les résultats, se comporte moins bien que la version qui ne l'implémente pas (v0.4a). Ceci est valable quel que soit le type d'évaluation examiné [3]. Cela signifie que la résolution des anaphores perturbe à la fois l'extraction des phrases et la facilité de lecture des résumés. Nous pouvons en conclure que la qualité de l'outil utilisé, parmi les meilleurs disponibles (Marrero *et al.*, 2009), n'est pas suffisante pour qu'il soit intégré tel que nous l'avons fait à un système de résumé automatique.

4.1.4. Conclusion

Nous avons présenté dans cette section l'intégration et la validation expérimentale de l'analyse en entités nommées, de résolution d'anaphores et de co-référence. L'intégration d'un outil d'étiquetage en entités nommées et de résolution orthographique de co-référence influence positivement les résultats de notre système de résumé automatique. Les résultats de tels outils, dont les étiquetages ont une précision et un rappel de l'ordre de 90% (Marrero *et al.*, 2009) sont en effet assez bons pour en tirer profit en les intégrant à des systèmes opérationnels.

En revanche, l'intégration d'un outil de résolution d'anaphores parmi les meilleurs disponibles actuellement a affecté négativement les résultats de notre système. Sans

3. nous présentons des résultats plus détaillés en §5.5.3

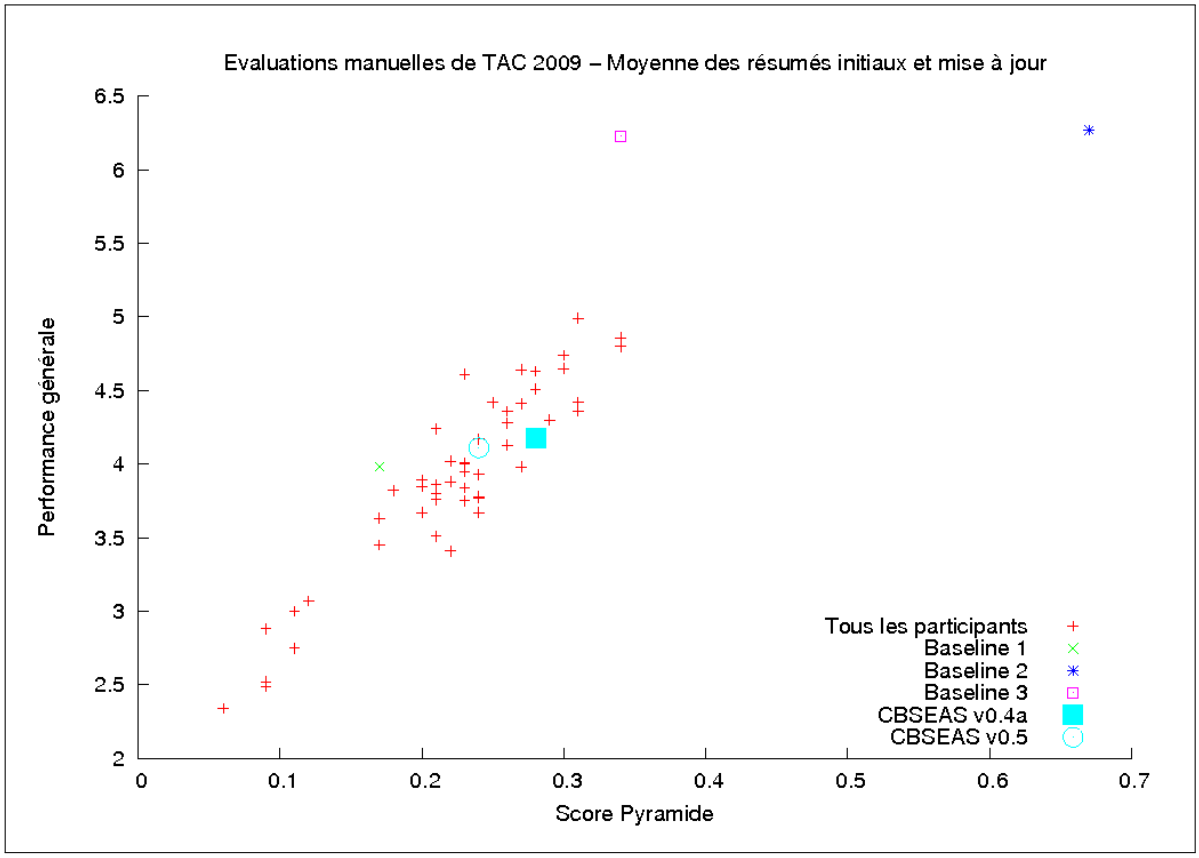


FIGURE 4.5.: Evaluation de l'apport de l'étiquetage en entités nommées et de la résolution d'anaphores : campagne TAC 2009.

doute la qualité des résolutions d'anaphores proposées en l'état actuel des recherches dans le domaine ne permet pas d'intégrer efficacement de tels outils à des systèmes plus larges de traitement automatique de la langue.

4.2. Utilisation de la Structure Rhétorique pour le Résumé Automatique

4.2.1. Introduction

Certains types de textes répondent à des critères d'écriture bien définis. Identifier leur structure peut être précieux pour des applications comme la recherche d'information. En effet, selon le type d'information recherché par un utilisateur, il est possible de déterminer dans quelle partie du texte il sera plus probable de trouver l'information. De plus, au sein d'un même document, certains fragments de texte seront plus denses en informations que les autres. Dans les articles scientifiques par exemple, la conclusion et le résumé sont les parties les plus à même de contenir les informations essentielles.

4.2.2. Etat de l'Art

Les systèmes de résumé automatique implémentent depuis longtemps des éléments de structure dans la méthode de sélection des phrases, comme la position des phrases dans le texte (Edmundson, 1969; Baxendale, 1958). Edmundson (1969) a étudié la meilleure manière de sélectionner les mots représentatifs des documents, qui seront discriminants dans l'extraction des meilleures phrases pour constituer un résumé. Il en a conclu que les mots les plus représentatifs sont ceux du titre et des première et dernière phrases des documents (les auteurs précisent qu'ils travaillent sur des documents techniques d'une longueur approximative de 400 mots).

Baxendale (1958) est arrivé aux mêmes observations. En analysant un échantillon de 200 paragraphes, il a conclu que dans 85% des paragraphes, la première phrase était la phrase centrale [4] contre 7% pour la dernière.

Les études suivantes témoignent de différences importantes dans la rhétorique des auteurs selon le genre textuel auquel appartiennent leurs écrits. Dans le cadre de l'implémentation de critères structurels dans un système de résumé automatique, il convient

4. la phrase centrale d'un paragraphe est la plus représentative du contenu de ce dernier

donc de ne généraliser aucune conclusion sur un genre textuel à d'autres types de documents.

Braddock (1974) et Dolan (1980) ont étudié la relation entre la centralité des phrases et leur position dans leur paragraphe. Braddock (1974) a déterminé que dans les discours écrits de présentation d'articles concernant l'Histoire américaine, 13% seulement des paragraphes commencent par une phrase centrale.

Dolan (1980) a lui conclu que dans les discours écrits d'Histoire, les phrases centrales peuvent se situer n'importe où dans le paragraphe. Selon une étude psychologique menée par Kieras (1985) concernant des écrits techniques, la position d'une phrase témoigne bien de son importance.

Popken (1987) apporte une nuance à l'étude de Braddock : se fondant sur l'analyse de corpus de textes académiques plus hétérogènes, Popken montre que la variabilité de la longueur des paragraphes est un biais important de l'étude de Braddock : les paragraphes les plus courts n'ont pas toujours de phrase centrale. Celle-ci se trouve alors dans les paragraphes précédents. 78% des paragraphes sont influencés directement par une phrase centrale.

Smith (2008) fait la jonction entre ces quatre dernières études en étudiant des discours écrits académiques : selon lui, les phrases centrales sont majoritairement situées en début de paragraphe. Une phrase centrale peut cependant se situer en fin de paragraphe si l'auteur utilise un raisonnement inductif plutôt que déductif. Une phrase centrale peut également être placée au milieu du paragraphe si elle suit des informations d'arrière plan qui débutent traditionnellement le paragraphe.

Brandow *et al.* (1995) ont réalisé une expérience sur des dépêches de presse en confrontant un système de résumé automatique fondé sur la fréquence des mots en corpus à un système qui compose le résumé d'un document en extrayant ses n premières phrases. Le deuxième système avait alors obtenu de meilleurs résultats que le premier. Les systèmes qui réalisent des résumés de cette manière sont aujourd'hui en dessous de l'état de l'art, cependant cela ne doit pas faire oublier que la position d'une phrase est un critère valable pour évaluer sa pertinence pour ce qui est des dépêches de presse. Lin et Hovy (1997) ont également montré qu'il était possible d'apprendre une méthode de sélection de phrases par leur position pour un genre textuel déterminé, et que cette méthode était efficace sur les dépêches de presse.

Au vu des conclusions contradictoires de ces études, il apparaît important de prendre en compte le type et la structure des documents étudiés lorsque l'on souhaite générer automatiquement des résumés en fonction de critères positionnels.

A notre connaissance, les travaux qui intègrent des informations structurelles prennent en compte soit la position d'une phrase dans le document, soit la structure rhétorique

des phrases et les liens rhétoriques qu'elles entretiennent entre elles (*cf §1.7*). L'identification des liens rhétoriques reste un problème ouvert. En effet, les solutions existantes ne sont pas portables pour tous types de documents. Les travaux de (Teufel et Moens, 2002) utilisent la structure rhétorique, mais sont cantonnés aux articles scientifiques qui arborent une structure régulière et un formatage rhétorique normalisé. Nous avons donc choisi de ne pas utiliser la structure rhétorique pour le résumé automatique.

Cependant, les travaux sur l'aspect positionnel des phrases nous paraissent incomplets. En effet, ils prennent seulement en compte la position absolue des phrases sans observer le contexte rhétorique dans lequel elles s'inscrivent. Nous avons quant à nous choisi de nous intéresser à cet aspect en structurant les documents en parties. Il faut au préalable réaliser un étiquetage des documents en sous-genres. Un genre textuel n'est en effet pas composé de documents tous semblables d'un point de vue rédactionnel, et il convient de pouvoir identifier les grandes composantes d'un genre textuel déterminé afin d'augmenter la précision des traitements réalisés sur des documents qui en font partie.

Nous avons choisi de nous intéresser à un type de documents en particulier, les dépêches de presse. Beaucoup d'applications de résumé automatique visent en effet à résumer ce type de documents, et des campagnes d'évaluation récentes (les campagnes DUC et TAC) portent sur ce genre textuel, ce qui nous permettra d'évaluer l'apport de l'intégration de la structure de ces documents à notre système de résumé automatique.

4.2.3. Structure des Dépêches

Les dépêches de presse présentent une structure particulière. Leur formatage est identique quelle que soit la langue, « les lois du genre s'appliquant prioritairement » (Lucas, 2004). Le titre de la dépêche vient en premier, suivi de l'en-tête qui comprend la source (l'agence qui délivre la dépêche) ainsi que les paramètres d'émission (lieu, date). Vient ensuite le texte à proprement parler, éventuellement suivi d'un pied de page comprenant des références à des dépêches antérieures ou à des ressources extérieures à l'agence à l'origine de la dépêche (*cf fig. 4.6*).

Le corps du document (le texte en lui-même) présente le plus souvent les informations de manière hiérarchique, d'après la méthode dite de la pyramide inversée (Agnès, 2008). Cela consiste à présenter les informations selon leur importance, dans l'ordre décroissant. La ou les premières phrases, appelées « l'accroche », énoncent succinctement les informations essentielles. C'est la raison pour laquelle les méthodes de résumé privilégiant les premières phrases de dépêches de presse obtiennent de si bons résultats.

Cependant, si les dépêches de presse qui adoptent cette présentation, que nous appellerons ici « dépêches classiques », sont majoritaires au sein des dépêches (Agnès, 2008), il existe d'autres types de dépêches qu'il est intéressant d'étudier. En parcourant deux cor-

Hurricane Katrina sends Louisiana unemployment rate to highest point since 1980s

APW
NEW ORLEANS, Oct 25 - Job losses after Hurricane Katrina sent Louisiana's unemployment rate soaring to 11.5 percent, the highest level since the 1980s oil crash, officials said Tuesday.
The state Labor Department said in its monthly report that 234,000 jobs were lost in September _ the majority in southeastern Louisiana _ after Katrina struck on Aug. 29. The figures represent only some of the fallout from Katrina, and none of the damage done by Hurricane Rita's foray into Louisiana on Sept. 24.
The September unemployment rate was the highest since January 1987, when Louisiana's economy was reeling from a major oil price crash that resulted in a deep recession. The jobless rate was 5.8 percent in August and 5.7 percent in September 2004.
Labor officials said in an early estimate last week that 296,365 people were left unemployed by Katrina and Rita. It said that in a seven-week period that stretched into October, it processed 271,846 eligible unemployment claims for Katrina and 24,519 for Rita. By comparison, there were 12,710 initial claims for unemployment in August and 11,946 in September of last year.
On the Net : http ://www.smallbusinessarena.com/unemployment-rises-after-hurricane-katrina/

FIGURE 4.6.: Structure globale d'une dépêche de presse : titre (en jaune), en-tête (bleu clair), corps (sans couleur), pied de page(en rose)

pus composés uniquement de dépêches de presse : AQUAINT-2 [5], composé de dépêches de presse en anglais issues de diverses agences de presse, et le corpus « Côte d'Ivoire » du projet Infom@gic [6], composé de dépêches de presse en français issues de l'AFP, nous avons identifié les types de dépêches suivants :

- les micro-trottoirs ;
- les chronologies ;
- les rapports de discours ;
- les fiches techniques ;

Les sections suivantes présentent ces différents types de dépêches et la façon d'exploiter leur structure afin d'identifier les zones les plus informatives.

5. http://www.nist.gov/tac/data/data_desc.html
6. Infom@gic, projet du pôle de compétitivité Cap Digital, vise à étudier les domaines scientifiques qui peuvent permettre à court terme de proposer des solutions techniques facilitant l'analyse et la recherche de données (http://www.capdigital.com/infomagic/).

Dépêches « classiques »

La majorité des dépêches de presse, que nous nommons « dépêches classiques », doivent répondre à six questions lorsqu'elles présentent un événement : quel est l'événement ? Qui y a pris part ? Quand et où a-t-il eu lieu ? Comment et pourquoi s'est-il produit ? (Agnès, 2008) La réponse à cette dernière question faisant intervenir le point de vue de l'auteur, celle-ci est séparée de la présentation factuelle de l'événement. Il est possible de l'identifier en s'aidant notamment d'indices temporels (Lucas, 2004). Cette réponse prend souvent la forme d'un commentaire en apparence objectif qui permet de replacer l'événement dans un contexte plus large, et peut être suivie d'un retour à l'événement et d'une projection dans le futur. La figure 4.7 illustre la structure de ce type de dépêches. On peut constater que les faits ne sont jamais relatés dans l'ordre chronologique, y compris dans la deuxième partie qui relate des faits similaires au sujet de la dépêche en utilisant un retour en arrière. Les événements sont à l'inverse présentés du plus au moins frappant.

Cette hiérarchisation de l'information dans toutes les parties des dépêches montre que favoriser uniquement les premières phrases d'une dépêche pour générer un résumé automatique est trop restrictif. Il faut au contraire favoriser les premières phrases de chacune des parties pour réaliser un résumé cohérent. La figure 4.8 présente deux résumés de quatre phrases, l'un composé des quatre premières phrases, l'autre de la première phrase de chacune des parties de la dépêche en figure 4.7. Le premier résumé présente l'accrochage entre troupes françaises et troupes du MPIGO de manière exhaustive, mais ne fournit au lecteur que peu d'informations sur le contexte : les antécédents, les réactions, ainsi que les conséquences probables ou les événements à venir. Le second résumé est moins complet quant à la description de l'événement-même, mais fournit des informations essentielles à sa compréhension. Cette constatation ne vaut pas pour toutes les dépêches ; cependant, le formatage typique de celles-ci la rend majoritaire.

Micro-trottoirs

Les micro-trottoirs (wikipedia; Agnès, 2008), ou revues d'opinion, consistent en une succession de réactions à un événement ou à une situation, ou de réponses à une même question. Le journaliste organise généralement la dépêche en présentant dans son accroche un résumé des réactions, puis en énumérant les réactions en commençant par celle qui vient appuyer ou expliquer le mieux son accroche. La figure 4.16 [7] présente une de ces dépêches [8].

7. Pour des raisons de lisibilité, cette figure est placée en fin de chapitre.

8. Celle-ci pourrait également être vue comme un reportage traditionnel. Cependant, étant donné la forte importance donnée aux citations et la recherche par le journaliste de personnes à interroger, nous l'avons classée comme micro-trottoir

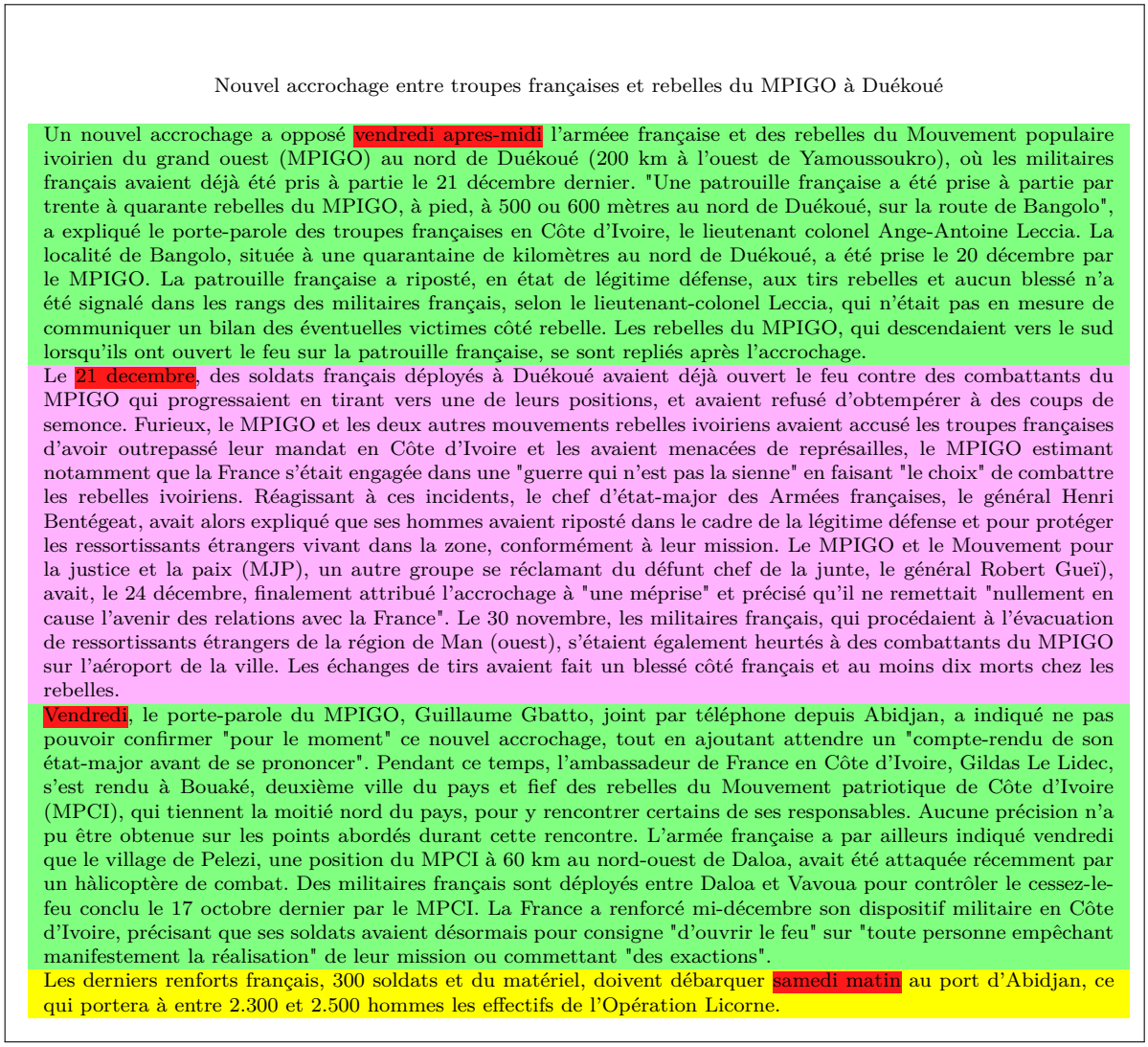


FIGURE 4.7.: Illustration des différentes zones d'une dépêche d'un événement commenté. En vert clair : présentation factuelle de l'événement ; en rose : les événements antérieurs qui y sont liés ; en jaune : la projection vers le futur ; en rouge : les indices temporels

Quatre premières phrases	Première phrase de chaque partie
Un nouvel accrochage a opposé vendredi apres-midi l'armée française et des rebelles du Mouvement populaire ivoirien du grand ouest (MPIGO) au nord de Duékoué (200 km à l'ouest de Yamoussoukro), où les militaires français avaient déjà été pris à partie le 21 décembre dernier. "Une patrouille française a été prise à partie par trente à quarante rebelles du MPIGO, à pied, à 500 ou 600 mètres au nord de Duékoué, sur la route de Bangolo", a expliqué le porte-parole des troupes françaises en Côte d'Ivoire, le lieutenant colonel Ange-Antoine Leccia. La localité de Bangolo, située à une quarantaine de kilomètres au nord de Duékoué, a été prise le 20 décembre 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 Leccia, qui n'était pas en mesure de communiquer un bilan des éventuelles victimes côté rebelle.	Un nouvel accrochage a opposé vendredi apres-midi l'armée française et des rebelles du Mouvement populaire ivoirien du grand ouest (MPIGO) au nord de Duékoué (200 km à l'ouest de Yamoussoukro), où les militaires français avaient déjà été pris à partie le 21 décembre dernier. Le 21 decembre, des soldats français déployés à Duékoué avaient déjà ouvert le feu contre des combattants du MPIGO qui progressaient en tirant vers une de leurs positions, et avaient refusé d'obtempérer à des coups de semonce. Vendredi, le porte-parole du MPIGO, Guillaume Gbatto, 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". Les derniers renforts français, 300 soldats et du matériel, doivent débarquer samedi matin au port d'Abidjan, ce qui portera à entre 2.300 et 2.500 hommes les effectifs de l'Opération Licorne.

FIGURE 4.8.: Illustration de deux méthodes de résumé différentes : l'une sélectionne les quatre premières phrases d'une dépêche, l'autre la première phrase de chaque partie de la dépêche

Etre capable d'identifier ces dépêches présente un intérêt non négligeable. Leur organisation est en effet différente de celle des dépêches explicatives. L'accroche contient généralement un résumé des réactions/citations, et nous avons observé que la citation la plus représentative des opinions exprimées, ou la plus marquante, est choisie par le journaliste pour apparaître en première position. Au-delà de cette première citation, il n'est pas possible de hiérarchiser les informations. Les phrases ont alors toutes la même importance. De plus, pour générer certains types de résumés comme des résumés d'opinion, donner l'avantage aux citations issues de ces dépêches peut améliorer la pertinence du résumé vis-à-vis de la tâche.

Il existe un autre type de dépêche que nous ne traitons pas ici comme un type à part : le reportage. La journaliste dresse un panorama d'un lieu, d'une situation ou d'un événement dans une démarche d'investigation. Le discours du journaliste est alors agrémenté de réactions des personnes rencontrées lors de l'investigation, et prend souvent l'aspect d'une revue d'opinion à laquelle sont rajoutés des éléments narratifs. Nous ne ferons pas ici de distinction entre ces deux catégories de dépêches et considérerons les reportages comme des micro-trottoirs. Les micro-trottoirs sont également souvent agrémentés d'éléments narratifs, c'est pourquoi la distinction entre ces catégories est difficile à faire y compris pour un humain.

Chronologies

Les chronologies sont minoritaires dans les corpus de dépêches. Cependant, elles permettent d'obtenir à moindre coût des informations cruciales concernant un événement. En effet, elles sont écrites dans un style très concis et l'auteur n'y fait figurer que les événements importants. Ces dépêches sont organisées par ordre chronologique et ne présentent donc pas les informations par ordre d'importance. Contrairement aux types de dépêches précédents, l'accroche d'une chronologie n'est généralement pas la phrase la plus pertinente du document. Elle ne fait qu'expliciter le fait que le lecteur a bien affaire à une chronologie d'événements. Cependant, il existe des cas dans lesquels l'auteur choisit de présenter un résumé de la chronologie (*cf fig. 4.18*). La figure 4.17, située en fin de chapitre, est une chronologie tirée du corpus Côte d'Ivoire du projet Infom@gic.

Rapports de discours

Ce type de dépêche résume un discours ou une conférence de presse d'une personnalité ainsi que son contexte. Les dépêches de ce type peuvent également comprendre des réactions au discours par d'autres personnalités, voire par des personnes de l'auditoire. La fin de la dépêche fournit généralement des informations concernant les raisons de l'intervention rapportée, et utilise un temps de narration différent de la partie contenant la citation. L'accroche, quant à elle, fournit généralement un résumé du discours. Cette structuration qui différencie les faits, relatés en fin de dépêche et les opinions et discours rapportés peut être utile à un système de résumé automatique qui cherche à adapter au mieux sa réponse à une requête utilisateur.

Fiches Techniques

Les fiches techniques sont écrites lors d'un événement pour lequel il apparaît nécessaire à l'auteur d'informer précisément les lecteurs sur le contexte géo-politique ou historique. L'accroche d'une fiche technique décrit l'événement en raison duquel la fiche est publiée. La figure 4.20 présente une fiche technique sur la Côte d'Ivoire, publiée à l'occasion de l'explosion des violences en Côte d'Ivoire. L'auteur décrit la situation géopolitique de la Côte d'Ivoire et l'arrière-plan historique afin d'aider le lecteur à comprendre les raisons de la crise en cours.

Les fiches techniques semblent être assez rares (1% environ des dépêches du corpus Côte d'Ivoire). Cependant, étant donné leur statut de résumé descriptif, toutes les informations qu'elles présentent sont pertinentes. En ce sens, les phrases qui en sont issues constituent de bons candidats à l'extraction dans un résumé descriptif.

Bien que les informations principales de ce type de dépêches se situent dans la partie énumérative, l'accroche ne doit pas être oubliée lors de l'analyse. En effet, elle fournit un résumé de l'événement en raison duquel la dépêche est publiée. Si cet événement est celui sur lequel porte la requête utilisateur, l'accroche est alors un candidat parfait à l'extraction dans le résumé.

Synthèse de l'analyse typologique et structurelle des dépêches

Dans cette section, nous avons montré que le genre des dépêches de presse comprend plusieurs catégories distinctes. Si les dépêches dites classiques (les dépêches explicatives) sont les plus nombreuses, il n'en reste pas moins que les autres catégories de dépêches sont intéressantes pour un système de résumé automatique. Le tableau 4.2 présente une synthèse de l'analyse de la structure de chacun des types de dépêches que nous avons identifiés. La plupart des systèmes de résumé qui intègrent des critères positionnels dans le choix des phrases à extraire favorisent les premières phrases des dépêches. Cette stratégie fonctionne dans la majorité des cas, mais n'est pas valable pour toutes les dépêches. De plus, selon le type d'information recherché par l'utilisateur d'un système de résumé, les phrases les plus pertinentes peuvent être situées à d'autres endroits qu'au début des dépêches, y compris des dépêches explicatives. Notre objectif consiste donc à prendre en charge la position d'une phrase non dans l'ensemble du document analysé, mais dans sa structure afin d'améliorer la précision des critères positionnels pour la génération de résumés de dépêches de presse par extraction.

Type	Hiérarchisation	Importance des premières phrases	Autres caractéristiques
Dépêche classique	Pyramide inversée	Oui	Fréquemment structurée en parties : – événement présent – explicitation du contexte – plongée dans le futur
Chronologie	Chronologique	Sous condition	Concision
Fiche technique	Pyramide inversée	Sous condition	Concision
Rapport de discours	Chronologique	Sous condition	Subjectivité dûe au discours rapporté
Micro-trottoir	Aucune	Oui	Subjectivité dûe au discours rapporté

TABLE 4.2.: Tableau de synthèse des traits structurels des différents types de dépêches dégagés en §4.2.3

4.2.4. Reconnaissance des types de dépêches

Il est nécessaire d'annoter la structure des documents afin de la prendre en compte dans un système de résumé automatique. La première étape consiste à détecter le type des dépêches présentées au système de résumé. Ces dépêches présentent des caractéristiques

régulières dont certaines sont discriminantes pour leur classement dans les différentes catégories dégagées en §4.2.3.

Nous présentons en figure 4.9 la liste des catégories de dépêches ainsi que les caractéristiques facilement appréhendables par un système automatique qui permettent de les catégoriser. Nous les tirons de l'analyse de deux corpus de dépêches différents, l'un en français et l'autre en anglais, rassemblant des dépêches issues de plusieurs agences de presse différentes. Ces catégories et leurs caractéristiques ont été observées quelle que soit l'origine des dépêches et sont donc généralisables à tout corpus composé de dépêches de presse. Ceci s'explique par le fait que le style journalistique s'applique avant tout (Lucas, 2004).

- les dépêches classiques (n'ont pas de signe distinctif) ;
- les rapports de discours : comportent beaucoup de citations directes et indirectes (*cf fig. 4.14*) ;
- les micro-trottoirs : comme les rapports de discours, comportent beaucoup de citations, mais les intervenants sont multiples) ;
- les chronologies : la plupart des phrases débutent par une date et la fin de l'accroche est systématiquement marquée par deux-points ;
- les fiches techniques : les énumérations débutent par des symboles non alphanumériques afin de marquer un nouvel élément de l'énumération.

FIGURE 4.9.: Liste des catégories de dépêches ainsi que leurs caractéristiques identifiées en corpus.

Toutes les caractéristiques citées en figure 4.9 doivent être lues au regard de la longueur des documents. En effet, la longueur des dépêches dans les corpus que nous étudions est très variable, de 1 à 130 phrases pour le corpus de la tâche « Résumé et mise à jour » de TAC 2008 (*cf fig. 4.10*).

Partant de ces caractéristiques, nous utilisons les indices présentés en figure 4.11 afin de catégoriser les dépêches.

Les citations directes sont identifiées grâce à une méthode à base d'expressions régulières et un lexique de verbes et expressions de citation que nous avons établi, présenté en 4.12. Celui-ci nous permet généralement de retrouver le locuteur, qui est l'entité nommée la plus proche du verbe de citation et suivant celui-ci dans le cas où la citation précède le verbe, ou l'entité nommée située avant le verbe dans le cas contraire. Dans le cas de l'expression « According to », le locuteur est l'entité nommée qui suit l'expression. Si aucun locuteur n'est utilisé, alors nous supposons que le journaliste l'a cité précédemment, et nous attribuons à la citation le premier locuteur trouvé en amont.

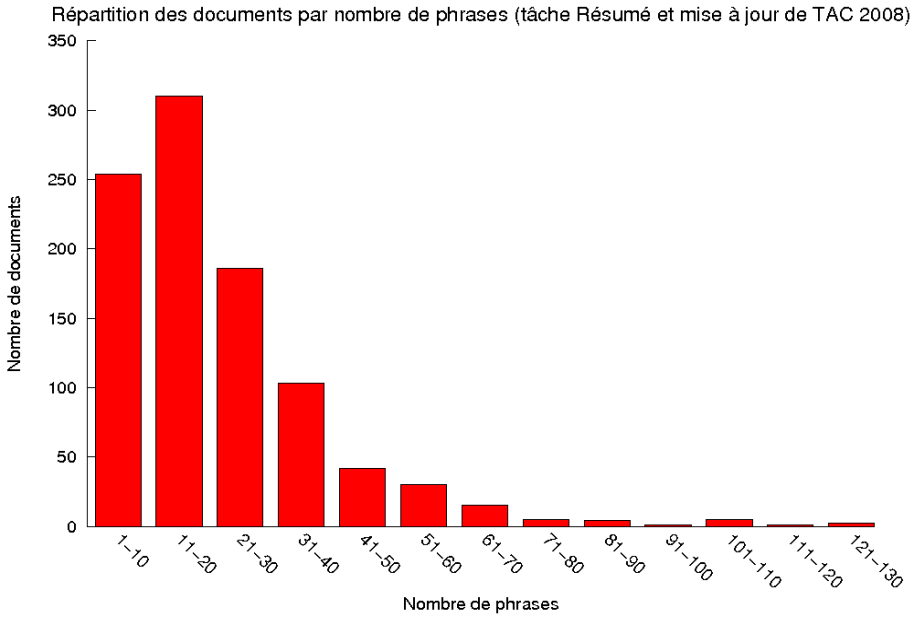


FIGURE 4.10.: Répartition des dépêches de la tâche « Résumé et mise à jour » de TAC 2008 en nombre de phrases

- pc : pourcentage de phrases identifiées comme citations ;
- ic : nombre de locuteurs différents identifiés divisé par le nombre de citations ;
- pa : pourcentage de phrases précédées d'un caractère non alphanumérique ;
- pd : pourcentage de phrases qui débutent par une date ou comportent un nombre dans les trois premiers mots ;
- in (intro) : parmi les 5 premières phrases, si l'une comporte un « : » suivi d'un retour à la ligne, 1 sinon 0 (correspond à l'accroche des chronologies).

FIGURE 4.11.: Indices utilisés pour catégoriser les dépêches

Les citations indirectes sont plus problématiques (*cf fig. 4.13*). Certains verbes utilisés pour citer sont en effet polysémiques. Ainsi, « *to declare* » n'introduit pas forcément de citation. Nous avons choisi d'adopter un biais qui exclut tous les cas problématiques, mais fait baisser le taux de rappel de détection des citations. Si le verbe de citation est suivi d'une conjonction de subordination (*that*, *wether*...) en excluant les éventuels adverbes situés entre ces deux éléments, la phrase est considérée comme une citation. Si le verbe de citation n'est pas suivi d'une conjonction de subordination, tournure spécifique à l'anglais, la phrase est considérée comme une citation seulement si un verbe est trouvé dans les 5 mots qui suivent le verbe de citation. Cette méthode exclut quelques phrases,

add + that	propose + that
announce	say
argue	stress
ask	tell
claim + that	warn
complain	wonder + if
denounce	write + that
deny	according + to
point + out	in words of
predict	

FIGURE 4.12.: Liste des verbes et expressions utilisés par notre système pour repérer les citations.

comme la dernière phrase de la figure 4.13 mais permet d'augmenter la précision de l'étiquetage des citations.

- The republic declared its independence from Georgia after winning a war 11 years ago. *
- Gregoire was declared the winner Thursday by Secretary of State Sam Reed, a Republican.
- On Nov. 24, Reed, following state election law, declared Rossi the winner, by 42 votes.*
- Shays wrote a letter to Michael Chertoff, secretary of the Homeland Security Department, asking whether the federal government has any plan to save the thousands of pets[...].
- ~~Cooper wrote a similar story a few days after also talking with Rove.~~
- ~~Woodward warned him off.~~
- More than 10 countries have warned their citizens of vigilance in their trips to France.*

FIGURE 4.13.: Exemples de verbes de citation polysémiques et problématiques dans le cadre de la détection automatique de citations. Les phrases marquées d'une * sont ambiguës quant à leur aspect citatif, les phrases rayées ne sont pas des citations. Le verbe est surligné en orange.

Il est alors possible de fixer des règles sur les indices présentés en figure 4.11 afin de typer les dépêches. Plutôt que d'établir nous-même les règles et seuils permettant d'obtenir le meilleur étiquetage possible, nous avons choisi d'utiliser un algorithme d'apprentissage supervisé qui trouve automatiquement les règles les plus adaptées à notre problème de classification. Nous avons choisi le classifieur à base d'arbre de décision J48, fondé sur l'algorithme C4.5(Quinlan, 1993). Il a pour avantage de permettre une lecture humaine des critères de classification. Son utilisation, contrairement à l'algorithme C4.5,

- "The worst thing in my view is not to release one's soul to peace when that person has declared that's what they want," said Smith, the only gubernatorial candidate with a vote on the measure.
- In 1972, the administration of Baath, the ruling party at that time, declared officially that other nationalities are Assyrian, Chaldean and Syrian.
- Even though Deep Throat's identity had been one of Washington's most enduring and fascinating mysteries, Novak declared that just about everyone in the Capital had known all along.
- Forgeard said "we are well above our market plan," and added that his aim was to bring in two new airlines as customers each year.
- He said a study was carried out which indicated each airport would have to spend about 80 million dollars to rebuild or widen taxiways and runway bridges for the A380.
- "Yes, it's true that we have a parachute on our backs and an evacuation hatch," he said in the radio interview.

FIGURE 4.14.: Exemple de citations directes et indirectes tirées du corpus de la tâche « Résumé et mise à jour » de TAC 2008. Le locuteur est surligné en jaune, le verbe de citation en orange, les citations directes en vert et les citations indirectes en rose.

ne requiert pas de discrétiser les variables continues données en paramètre. Nous avons utilisé son implémentation dans le projet Weka [9].

Un tel algorithme nécessite des données d'entraînement. Nous avons donc constitué un corpus d'entraînement en sélectionnant 500 documents du jeu de données de la tâche « Résumé et mise à jour » de TAC 2008 (*cf* §5.2). Nous avons étiqueté manuellement le type des documents et calculé automatiquement les différents paramètres fournis en entrée de l'arbre de décision. Le modèle ainsi établi obtient une précision moyenne de 81% en le projetant sur 200 autres documents également annotés manuellement.

4.2.5. Etiquetage de la Structure des Dépêches

Pour chaque dépêche, nous identifions les différentes parties qui la composent, décrites en §4.2.3. Les dépêches de presse de nos corpus sont toutes formatées en paragraphes. Cependant, ce formatage n'est pas régulier. Certaines dépêches ont une phrase par paragraphe, d'autres comportent des paragraphes à réelle valeur sémantique. Il est impossible de se fonder sur ce formatage pour une analyse de la structure des dépêches.

Les chronologies et les fiches techniques comportent des indices de structure très ré-

9. http ://www.cs.waikato.ac.nz/ ml/index.html

guliers ; nous avons donc choisi d'étiqueter leur structure de manière *ad hoc*. La fin de l'accroche est identifiée grâce au caractère « : » suivi d'un retour à la ligne. Notre système étiquette tout ce qui suit le dernier paragraphe précédé d'un mot en majuscule, d'une date ou d'un caractère spécial comme la partie dédiée aux commentaires.

Les dépêches de type « rapport de discours » et « micro-trottoir » sont différentes. Elles ne présentent en effet pas d'éléments de structures apparents. Nous avons choisi de faire courir l'accroche de la première phrase à la première citation reconnue (en dehors de la première phrase qui contient souvent une citation, résumé du reste du document). L'accroche ainsi étiquetée contient les phrases les plus importantes de la dépêche. Le système ne gère pas encore l'étiquetage des phrases explicatives insérées dans la partie principale.

Nous n'avons pas réalisé la détection des différentes parties des dépêches classiques. Ces dépêches représentant largement plus de 50 pour cent des dépêches de presse, l'étiquetage de ces différentes parties est un axe prioritaire de nos perspectives.

4.2.6. Utilisation de la Structure des dépêches dans CBSEAS

La position des phrases dans les différents types de dépêches identifiés témoigne de leur importance. Afin de valider cette hypothèse, nous avons implémenté la gestion de la position dans CBSEAS pour notre participation à la tâche « Résumé et mise à jour » de la campagne d'évaluation TAC 2009. Cependant, en raison de la proximité de la date limite de soumission pour TAC 2009, seule une implémentation basique a été réalisée. Dans la tâche « Résumé et mise à jour », les résumés sont guidés par une requête utilisateur.

Dans un premier temps, les dépêches classées « chronologie » ou « fiche technique » 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 de la fiche technique. Si le thème d'une dépêche est assez proche de la requête utilisateur [10], alors les éléments extraits de cette dépêche doivent être favorisés. En cas d'absence de requête utilisateur, on peut imaginer favoriser les éléments extraits de cette dépêche si et seulement si son thème est assez proche du centroïde (*cf §3.2.2*) de l'ensemble des documents à résumer. Cependant, la phrase d'introduction de la partie centrale ne doit pas figurer dans le résumé généré par CBSEAS. En effet, celle-ci n'est généralement pas informative (*cf fig. 4.17*) et contient des éléments pouvant perturber la lecture du résumé. Elle contient cependant des mots centraux dans le document qui augmentent la probabilité qu'elle soit extraite. La figure 4.15 présente un exemple issu du corpus « Résumé et mise à jour » de TAC 2008 qui illustre bien pourquoi ces phrases introductrices doivent être éliminées avant le lancement de CBSEAS.

10. Nous utilisons ici la mesure de similarité présentée en §3.2.2 associée à un seuil fixé empiriquement.

Key dates in the North Korea nuclear standoff
Key dates in the standoff over North Korea's nuclear program :
_1993 : North Korea shocks the world by announcing it is withdrawing from the Nuclear Nonproliferation Treaty amid suspicions that it is developing nuclear weapons. Pyongyang suspends that decision three months later following negotiations with the United States[...]

FIGURE 4.15.: Extrait d'une chronologie issue du corpus de la tâche « Résumé et mise à jour » de TAC 2008. La phrase en vert clair, qui introduit la chronologie, doit être éliminée afin d'éviter à CBSEAS de la sélectionner dans le résumé.

CBSEAS supprime donc la phrase d'introduction de la partie principale et favorise les phrases qui sont issues de la dépêche en augmentant leur score de 15%.

Si la dépêche contient un résumé qui précède la phrase d'introduction, nous avons choisi d'affecter empiriquement un bonus décroissant de 20%, 15% et 10% aux trois premières phrases. Nous procédons de la même manière pour les phrases issues de dépêches classiques, de type « rapport de discours » et « micro-trottoirs ».

Ces bonus ont été fixés empiriquement. La meilleure solution aurait été de nous servir de l'algorithme génétique décrit en §3.3 afin de trouver les valeurs des bonus les plus adaptées à notre système de résumé automatique. Le temps de convergence de l'algorithme était cependant trop important pour que nous puissions le lancer avant la campagne d'évaluation TAC 2009, et nous avons choisi de définir nous-même des valeurs que nous estimions acceptables.

Les dépêches de type « rapport de discours » et « micro-trottoirs » ne sont pas encore exploitées de manière optimale. On peut imaginer relier les requêtes utilisateur à des configurations du système de résumé qui permettraient, selon le type de résumé souhaité par l'utilisateur, d'apporter la réponse la plus précise possible en extrayant prioritairement les phrases de certains types de dépêches, comme proposé en §4.2.3.

4.2.7. Evaluation

Nous avons évalué l'apport de l'utilisation de la structure sur les données de TAC 2008, en comparant les scores de CBSEAS avec et sans l'utilisation de la structure.

Afin de différencier les améliorations dues à l'utilisation classique de la position des phrases dans les documents des améliorations dues à l'intégration plus fine d'éléments

de structure à CBSEAS, nous avons séparé le corpus de TAC 2008 en deux. Les jeux de documents dans lesquels au moins une dépêche classée « chronologie » ou « fiche technique » dont le thème se rapproche de la requête sont classés dans « Jeux 2 », les autres dans « Jeux 1 ». Comparer les résultats de résumés obtenus pour ces deux groupes de jeux de documents nous permet donc d'évaluer l'apport de l'intégration de la structure des types de dépêches visés.

Nous avons donc comparé les scores ROUGE-SU4 des résumés des groupes de documents classés dans « Jeux 2 », et avons constaté une amélioration de près de 10% par rapport aux résultats de CBSEAS sans la gestion de la structure (*cf tab 4.3*). L'amélioration des résultats des jeux de documents de « Jeux 1 » est moins flagrante, de l'ordre de 3%. Cela montre que si favoriser les premières phrases des dépêches améliore la qualité des résumés, prendre en compte des éléments de structure plus précis permet d'aller plus loin dans l'amélioration des résumés générés.

	Jeux 1	Jeux 2	score total
CBSEAS	0.1161	0.1191	0.1167
CBSEAS + Struct	0.1185	0.1291	0.1206

TABLE 4.3.: Résultats de CBSEAS avec et sans utilisation de la structure des dépêches (données de TAC 2008. Jeux 1 concerne tous les résumés pour lesquels aucune chronologie ou fiche technique n'est donnée en entrée, Jeux 2 concerne les résumés pour lesquels au moins une chronologie ou fiche technique est donnée en entrée.)

4.2.8. Conclusions

Dans cette section, nous avons montré les approches existantes qui se servent de la structure des documents afin d'améliorer la génération de résumés par extraction. Nous avons également montré que les études sur le rapport entre la pertinence d'une phrase et sa position dans les documents étaient contradictoires.

Nous avons réalisé une étude des dépêches de presse en corpus et trouvé que la position absolue d'une phrase au sein d'une dépêche n'était pas le seul indicateur de la pertinence pour son intégration à un résumé. Selon les types de dépêches et la requête utilisateur, les phrases les plus pertinentes se trouvent à des positions différentes.

Nous avons implémenté un système de typage des dépêches et d'étiquetage de leur structure. Celui-ci n'est pas encore complet : il reste encore des catégories de dépêches dont la structure n'est pas repérée automatiquement. Nous l'avons intégré à notre système de résumé automatique, CBSEAS, et obtenu des résultats encourageants.

Il reste à finaliser le système d'étiquetage de la structure pour les dépêches classiques, les rapports de discours et les micro-trottoirs, ainsi qu'à réaliser une intégration plus poussée de la structure à notre système de résumé automatique, comme relier celle-ci aux requêtes utilisateur afin d'améliorer la qualité de la réponse apportée par CBSEAS.

Accords de Marcoussis : manifestations de joie dans le bastion rebelle de Bouaké

Des manifestations de joie, ponctuées de coups de feu en l'air, ont éclaté samedi soir à Bouaké, bastion rebelle du centre de la Côte d'Ivoire, après l'annonce à Paris de la nomination du nouveau Premier ministre et de la répartition des portefeuilles ministériels, a constaté un journaliste de l'AFP.

"Gbagbo vient d'être désarmé, et complètement alors", a souligné l'un des porte-parole de la rébellion, Chérif Ousmane, dans son quartier général de Bouaké. Entouré de ses troupes, il félicite par téléphone satellitaire ses camarades encore en France : "Bon boulot, nous vous réservons un accueil digne de ce nom mardi à votre retour", lance-t-il devant un journaliste de l'AFP.

"Nous avons gagné, nous avons gagné", scandent des rebelles, tout en tirant en l'air dans le centre de la deuxième ville de Côte d'Ivoire, pendant que quelques autres klaxonnaient au volant de leurs véhicules. Joyeux, deux rebelles ont enlevé un poste de contrôle à l'est de Bouaké, deuxième ville du pays et quartier général du principal mouvement rebelle, le Mouvement patriotique de Côte d'Ivoire (MPCI).

"C'est fini, la paix est de retour en Côte d'Ivoire", affirme un autre, visiblement éméché.

"Il (le président ivoirien, Laurent Gbagbo) n'a plus de pouvoir, c'est fini", hurle le caporal "Sinclair", pour qui "le chemin de la réconciliation est en marche".

M. Gbagbo a accepté samedi à Paris un accord de réconciliation pour mettre fin à quatre mois de violences dans son pays et annoncé la nomination d'un Premier ministre de consensus, Seydou Diarra. Celui-ci devra former un gouvernement comprenant notamment neuf ministres d'Etat représentant à égalité les trois grandes forces politiques du pays et les rebelles. Le secrétaire général du MPCI, Guillaume Soro, a affirmé à l'AFP à Paris que son mouvement avait obtenu les postes "de la Défense et de l'Intérieur" dans ce nouveau gouvernement.

Assis sur de la ferraille au sud de la ville, d'autres mutins pensaient déjà à leur avenir. "Si la guerre se termine vraiment", certains iront au village, voir leur mère, ou leur fiancée.

D'autres évoquent les quatre mois de conflit : "Pourquoi tout ça ? Pourquoi Gbagbo a refusé, dès le début, la négociation. Pourquoi il a tué nos parents ?", sanglote Salif, affirmant être devenu rebelle "par nécessité".

Chassant de la main la fumée d'une cigarette, un autre rebelle demande à ses compagnons de ne pas crier victoire trop vite : "Gbagbo a plus d'un tour dans son sac. Il est capable de revenir pour dribbler tout le monde".

Les populations civiles, à cause du couvre-feu théoriquement en vigueur, n'ont pas pu exprimer leur sentiment dans la soirée dans les rues de la ville.

Cependant certaines personnes, interrogées devant leur domicile, se sont déclarées "très satisfaites" des accords.

"Nous faisons entièrement confiance au nouveau Premier ministre, c'est un homme d'expérience", affirme Noutié, fonctionnaire.

"Nous tenons à remercier sincèrement la France", déclare un autre.

"La répartition des portefeuilles ministériels est très, très, équilibrée", estime Joël, un handicapé.

Fatoumata Mariana, secrétaire général du Mouvement national de soutien et de défense (MNSD), association proche du MPCI, s'apprête à organiser des manifestations de soutien aux rebelles. Une bande d'amis, installés dans un "maquis" de la ville rebaptisé "Marcoussis", du nom de la commune française où s'est déroulée la table ronde inter-ivoirienne, ont quant à eux décidé de fêter l'événement en mettant un mouton entier au menu.

sd/at/eg/mt eaf.tmf

FIGURE 4.16.: Exemple d'une dépêche de type « micro-trottoir », ou « revue d'opinion ». L'accroche est surlignée en vert, les opinions citées sont surlignées en orange.

La crise politico-militaire en Côte d'Ivoire

Voici une chronologie de la crise politico-militaire en Côte d'Ivoire, confrontée depuis trois mois à un soulèvement de militaires contre le régime du président Laurent Gbagbo :

–SEPTEMBRE 2002–

- 19 : Des attaques violentes à Abidjan visent le coeur du pouvoir. Un des hommes forts du régime, le ministre de l'Intérieur, Emile Boga Doudou, est tué. La deuxième ville, Bouaké, et la principale ville du nord, Korhogo, sont sous contrôle rebelle. Le général Robert Gueï, ancien chef de la junte (déc 1999-oct 2000), est tué.
- 20 : Le président Laurent Gbagbo, rentré à Abidjan, évoque à mots couverts une possible implication étrangère.
- 22 : Arrivée des premiers renforts militaires français pour assurer la sécurité des étrangers, dont quelque 2.700 seront évacués. - 23 : L'ancien Premier ministre Alassane Ouattara accuse des gendarmes d'avoir voulu l'assassiner le 19.
- 24 : Le journal du parti au pouvoir accuse le président burkinabè, Blaise Compaoré, d'être le "seul et unique déstabilisateur de la Côte d'Ivoire". Le Burkina Faso dément toute implication. - 28 : Abidjan, qui avait évoqué une "agression extérieure", affirme avoir actionné les accords de défense avec la France.
- 29 : La Communauté économique des Etats d'Afrique de l'Ouest (CEDEAO) crée un "groupe de contact" et décide de l'envoi d'une force de paix.

–OCTOBRE–

- 1er : Les mutins veulent renverser le régime et invitent la France à une "stricte neutralité".
- 6-7 : Combats intenses dans Bouaké, les forces loyalistes repoussées.
- 16 : Les loyalistes reprennent Daloa (ouest).
- 17 : Les rebelles signent à Bouaké un accord de cessation des hostilités. Le président Gbagbo annonce qu'il l'accepte et demande à la France de contrôler le cessez-le-feu.
- 20 : Les militaires français se déploient sur une ligne médiane traversant le pays d'est en ouest.
- 22 : La France envoie des renforts pour assurer sa mission de contrôle de cessez-le-feu, qui s'ajoute à celle de sécurisation des étrangers.
- 23 : Le président togolais Gnassingbé Eyadéma est désigné coordinateur de la médiation.
- 24 : Le président Gbagbo déclare qu'Alassane Ouattara est une vraie "pomme de discorde" entre Paris et Abidjan.
- 29 : Amnesty International dénonce les massacres de dizaines de civils à Daloa et appelle loyalistes et rebelles à cesser les exactions. Le secrétaire général du Mouvement patriotique de Côte d'Ivoire (MPCI, branche politique de la rébellion), Guillaume Soro, affirme que des généraux soutiennent la rébellion.
- 30 : Début à Lomé des premières négociations directes entre gouvernement et rebelles.

–NOVEMBRE–

- 1er : Le gouvernement accepte le principe d'une amnistie et d'une réintégration des mutins dans l'armée.
- 27 : Visite du chef de la diplomatie française, Dominique de Villepin, marquée par l'exfiltration de Côte d'Ivoire d'Alassane Ouattara. Les deux camps s'engagent à Lomé à oeuvrer pour une solution politique.
- 28 : Offensive de forces gouvernementales accompagnées de mercenaires dans la région de Vavoua (ouest). Le Mouvement populaire ivoirien du grand ouest (MPIGO) et le Mouvement pour la justice et la paix (MJP), deux nouveaux groupes, revendiquent la prise de Man et Danané (extrême ouest).
- 30 : Touba, au nord de Man, passe sous contrôle du MJP.

–DECEMBRE–

- 1er : Les troupes françaises évacuent quelque 160 étrangers depuis l'aéroport de Man lors d'une opération marquée par les premiers affrontements meurtriers entre militaires français et rebelles : une dizaine de morts chez les rebelles et un blessé français.
- 3 : Rencontre à Bamako des présidents Gbagbo et Compaoré, qui condamnent les exactions contre les civils et les "tentatives de déstabilisation" de la Côte d'Ivoire.
- 4 : L'armée lance une offensive contre Toulépleu (extrême ouest).
- 5 : Des soldats français découvrent un charnier près du village de Monoko-Zohi (ouest), où selon des témoignages d'habitants, 120 personnes, en majorité des immigrés ouest-africains, sont ensevelies ou ont été jetées dans des puits.
- 6 : L'armée déclare poursuivre des opérations de "ratissage" à Man, après l'avoir reprise aux rebelles.
- 7 : L'agence missionnaire catholique Misna fait état de la découverte de cadavres d'au moins 86 gendarmes et soldats tués lors des affrontements avec la rébellion, ensevelis dans un charnier près de Bouaké.
- 8 : Attaque du MPCI contre les forces gouvernementales dans l'est.
- 9 : Entretien Eyadéma-Gbagbo à Yamoussoukro. Les rebelles du MPCI affirment être "prêts à reprendre l'offensive au cas où rien n'est clarifié sur le charnier" dans l'ouest. Ils reconnaissent l'existence d'une "fosse commune" renfermant des cadavres de quelque 90 militaires loyalistes à Bouaké, insistant sur la différence avec un "charnier de civils" découvert dans l'ouest. Le MPCI décrète un couvre-feu à Bouaké et ses environs.
- 10 : Des milliers de jeunes désireux de s'enrôler dans l'armée affluent au ministère de la Défense, répondant à l'appel à la "mobilisation générale". Le début de l'opération commencera le 12. La Grande-Bretagne et la Belgique demandent à leurs ressortissants de quitter le pays. Nouveaux combats à Bloléquin, près de la frontière du Liberia.

FIGURE 4.17.: Illustration d'une dépêche de type « Chronologie ». L'accroche est surlignée en vert.

Les relations tendues entre le Burkina et la Côte d'Ivoire (CHRONOLOGIE)

Les relations entre le Burkina et la Côte d'Ivoire se sont envenimées ces dernières années, en raison des soubresauts politiques en Côte d'Ivoire et des accusations d'ingérence portées par Abidjan contre Ouagadougou. Un peu moins de 3 millions de Burkinabès, pour beaucoup des ouvriers agricoles, vivent en Côte d'Ivoire, qui compte 15 millions d'habitants.

–1999– En novembre, un litige foncier entre autochtones Kroumens et Burkinabè dans la région de Tabou (sud-ouest) fait cinq morts.

En quelques semaines, quelque 20.000 Burkinabè sont chassés des plantations par les autochtones.

–2000– - 21 jan : Le président burkinabè Blaise Compaoré émet "le souhait de voir la Côte d'Ivoire réussir le processus de transition" engagé par la junte au pouvoir depuis le renversement, fin 1999, du président Henri Konan Bédié, après une rencontre avec le général Robert Gueï, qui dirige cette junte.

- août-sept : Au moins 13 morts dans des affrontements entre autochtones Kroumens et Burkinabè dans le sud-ouest de la Côte d'Ivoire.

En octobre, de nouvelles violences font cinq morts.

- En octobre et décembre, des immigrés sont battus ou tués en Côte d'Ivoire lors d'affrontements entre partisans d'Alassane Ouattara d'une part et forces de l'ordre et militants soutenant le pouvoir de l'autre, en marge des élections présidentielle et législatives.

- 6 nov : Visite du ministre ivoirien de l'Intérieur au Burkina, la première d'un membre du gouvernement du nouveau président Laurent Gbagbo.

–2001– - 18 jan : Le parti au pouvoir au Burkina Faso condamne le coup d'Etat avorté des 7-8 janvier en Côte d'Ivoire.

Des dizaines d'étrangers, notamment Burkinabè, ont fui, se disant l'objet de violences après que le gouvernement eut affirmé que des étrangers étaient impliqués dans la tentative.

FIGURE 4.18.: Exemple d'une chronologie dont l'accroche (surlignée en vert) est un résumé du reste du document.

Polynésie française : François Hollande demande un rendez-vous à Jacques Chirac

17/10/200419 : 16

PARIS (AP) –

François Hollande a proposé dimanche à «l'ensemble de l'opposition» de «solliciter un rendez-vous avec Jacques Chirac» pour lui demander de dissoudre l'assemblée de Polynésie française.

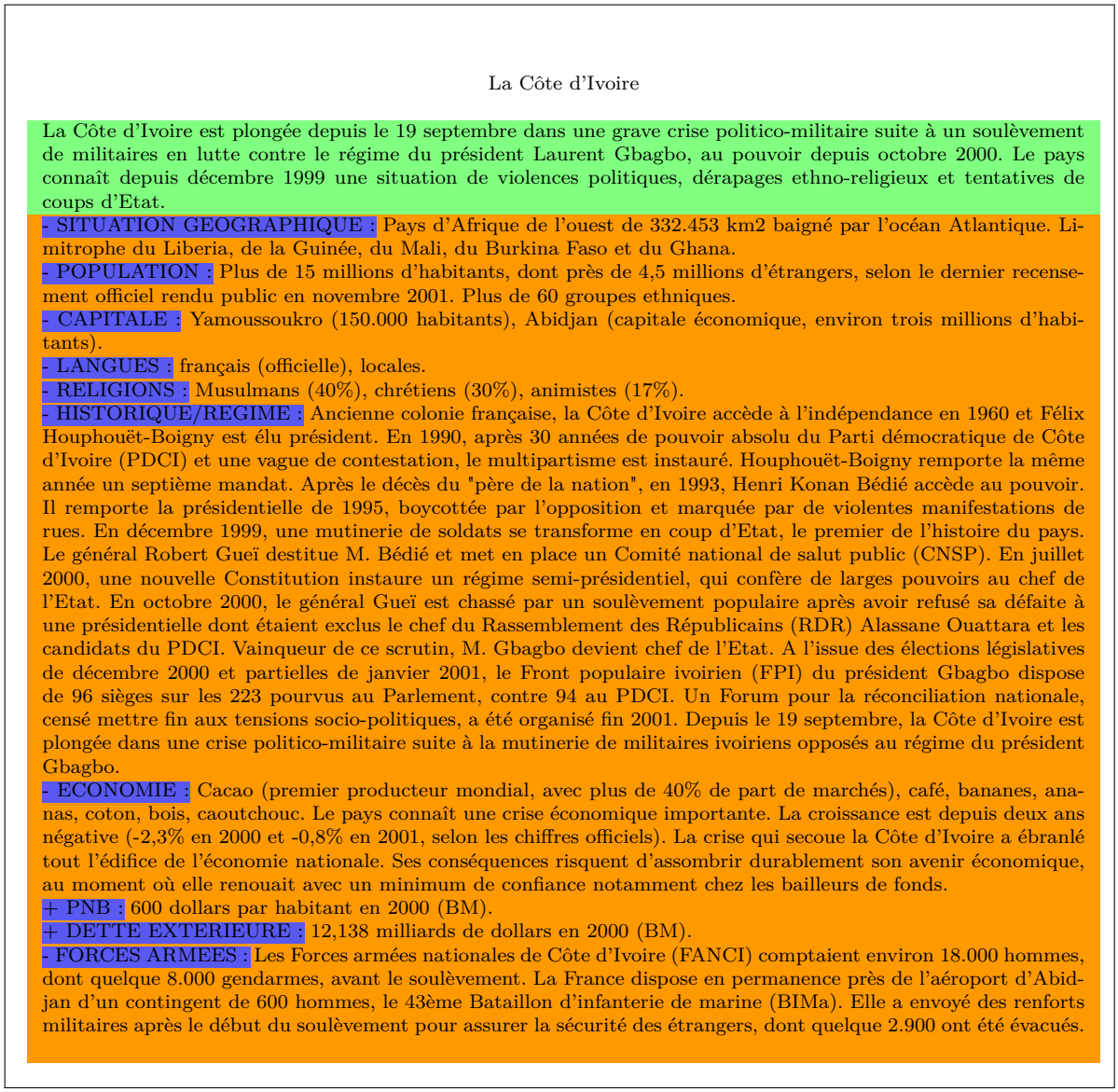
«Ce que nous voulons, c'est la dissolution de l'assemblée polynésienne», a déclaré le Premier secrétaire du PS lors du «Grand rendez-vous» sur Europe-1.

François Hollande, qui a précisé avoir déjà écrit au chef de l'Etat pour formuler cette demande, a rappelé «l'étroitesse» des liens entre Jacques Chiracet l'ancien président de la Polynésie française Gaston Flosse.

«Ce qui se passe est extrêmement grave. Depuis quatre mois, il y a eu une entreprise de déstabilisation du gouvernement polynésien, par Gaston Flosse d'abord, et ensuite par les amis de Gaston Flosse, le gouvernement de la République», a accusé le Premier secrétaire du PS.

Plus de 15.000 personnes ont manifesté samedi à Papeete pour réclamer la dissolution de l'assemblée et la tenue de nouvelles élections. Le 9 octobre, le gouvernement de l'indépendantiste Oscar Temaru -élu à la présidence grâce à la courte majorité acquise aux élections de mai 2004- avait été renversé à la suite de l'adoption d'une motion de censure à l'assemblée.

FIGURE 4.19.: Illustration d'une dépêche de type « Rapport de discours » et de ses zones. En orange : les citations ; en rose : l'explicitation du contexte du discours.



La Côte d'Ivoire

La Côte d'Ivoire est plongée depuis le 19 septembre dans une grave crise politico-militaire suite à un soulèvement de militaires en lutte contre le régime du président Laurent Gbagbo, au pouvoir depuis octobre 2000. Le pays connaît depuis décembre 1999 une situation de violences politiques, dérapages ethno-religieux et tentatives de coups d'Etat.

- SITUATION GEOGRAPHIQUE : Pays d'Afrique de l'ouest de 332.453 km2 baigné par l'océan Atlantique. Limitrophe du Liberia, de la Guinée, du Mali, du Burkina Faso et du Ghana.
- POPULATION : Plus de 15 millions d'habitants, dont près de 4,5 millions d'étrangers, selon le dernier recensement officiel rendu public en novembre 2001. Plus de 60 groupes ethniques.
- CAPITALE : Yamoussoukro (150.000 habitants), Abidjan (capitale économique, environ trois millions d'habitants).
- LANGUES : français (officielle), locales.
- RELIGIONS : Musulmans (40%), chrétiens (30%), animistes (17%).
- HISTORIQUE/REGIME : Ancienne colonie française, la Côte d'Ivoire accède à l'indépendance en 1960 et Félix Houphouët-Boigny est élu président. En 1990, après 30 années de pouvoir absolu du Parti démocratique de Côte d'Ivoire (PDCI) et une vague de contestation, le multipartisme est instauré. Houphouët-Boigny remporte la même année un septième mandat. Après le décès du "père de la nation", en 1993, Henri Konan Bédié accède au pouvoir. Il remporte la présidentielle de 1995, boycottée par l'opposition et marquée par de violentes manifestations de rues. En décembre 1999, une mutinerie de soldats se transforme en coup d'Etat, le premier de l'histoire du pays. Le général Robert Gueï destitue M. Bédié et met en place un Comité national de salut public (CNSP). En juillet 2000, une nouvelle Constitution instaure un régime semi-présidentiel, qui confère de larges pouvoirs au chef de l'Etat. En octobre 2000, le général Gueï est chassé par un soulèvement populaire après avoir refusé sa défaite à une présidentielle dont étaient exclus le chef du Rassemblement des Républicains (RDR) Alassane Ouattara et les candidats du PDCI. Vainqueur de ce scrutin, M. Gbagbo devient chef de l'Etat. A l'issue des élections législatives de décembre 2000 et partielles de janvier 2001, le Front populaire ivoirien (FPI) du président Gbagbo dispose de 96 sièges sur les 223 pourvus au Parlement, contre 94 au PDCI. Un Forum pour la réconciliation nationale, censé mettre fin aux tensions socio-politiques, a été organisé fin 2001. Depuis le 19 septembre, la Côte d'Ivoire est plongée dans une crise politico-militaire suite à la mutinerie de militaires ivoiriens opposés au régime du président Gbagbo.
- ECONOMIE : Cacao (premier producteur mondial, avec plus de 40% de part de marchés), café, bananes, ananas, coton, bois, caoutchouc. Le pays connaît une crise économique importante. La croissance est depuis deux ans négative (-2,3% en 2000 et -0,8% en 2001, selon les chiffres officiels). La crise qui secoue la Côte d'Ivoire a ébranlé tout l'édifice de l'économie nationale. Ses conséquences risquent d'assombrir durablement son avenir économique, au moment où elle renouait avec un minimum de confiance notamment chez les bailleurs de fonds.
+ PNB : 600 dollars par habitant en 2000 (BM).
+ DETTE EXTERIEURE : 12,138 milliards de dollars en 2000 (BM).
- FORCES ARMEES : Les Forces armées nationales de Côte d'Ivoire (FANCI) comptaient environ 18.000 hommes, dont quelque 8.000 gendarmes, avant le soulèvement. La France dispose en permanence près de l'aéroport d'Abidjan d'un contingent de 600 hommes, le 43ème Bataillon d'infanterie de marine (BIMa). Elle a envoyé des renforts militaires après le début du soulèvement pour assurer la sécurité des étrangers, dont quelque 2.900 ont été évacués.

FIGURE 4.20.: Illustration d'une dépêche de type « fiche technique ». En vert, l'accroche ; en orange, la partie énumérative ; en bleu : les intitulés des éléments listés.

Troisième partie .

Cas d'étude

Cette partie est dédiée à l'étude de l'application de notre système de résumé automatique, CBSEAS, présenté dans la partie précédente, à trois tâches différentes. La première concerne le résumé incrémental de dépêches de presse (et donc le résumé de nouveauté), la deuxième le résumé d'opinions, et la troisième l'aide à l'analyse de corpus. Nous avons participé à deux campagnes d'évaluation internationales portant sur les deux premières tâches. Les résultats de notre système –qui se place dans le premier quart du classement de chacune des campagnes– montrent sa versatilité et la généricité de notre approche. De plus, les différentes expériences que nous avons menées prouvent que l'intégration d'analyses linguistiques performantes améliore la qualité des résumés générés. Cependant, les résultats témoignent également du manque de post-traitements de CBSEAS. L'utilisation de techniques de compression de phrases, notamment, pourrait rendre notre système plus compétitif en éliminant des résumés les informations les moins pertinentes, laissant plus de place à l'intégration de nouvelles phrases qui apporteraient leur contribution à l'informativité des résumés. Le dernier cas d'étude montre l'intégration de notre système de résumé à une application d'aide à l'analyse de corpus dont l'intérêt a été validé dans le cadre d'un projet collaboratif, Infom@gic.

5. Utilisation de CBSEAS pour du résumé incrémental

Sommaire

5.1. Introduction

Le résumé automatique s'est longtemps intéressé aux dépêches de presse, car elles constituent un terrain d'expérimentation plus tolérant que d'autres types de documents : elles sont en effet synthétiques, et ont un style qui facilite leur analyse automatique. Aujourd'hui, la multiplication des sources de dépêches et la publication de nouvelles in-

formations quasiment en direct empêchent les personnes (qu'il s'agisse de professionnels ou de particuliers) désireuses de se renseigner sur un sujet donné d'avoir une vision immédiate et claire de ce sujet. Il est donc important de pouvoir résumer automatiquement de tels documents en utilisant des techniques qui leur sont adaptées. La tâche « Résumé et Mise à Jour » (*Update*) de TAC 2009 se place dans cette optique.

Cette campagne d'évaluation annuelle a pour nous été l'occasion de valider notre approche. Menée par le NIST [1], elle innove chaque année en proposant des tâches en relation avec les demandes des industriels dans le domaine de l'analyse automatique de textes. De plus, cette campagne mobilise un grand nombre de chercheurs liés au traitement automatique des langues, et permet donc de se comparer aux avancées les plus récentes. Participer à cette campagne nous permet donc d'obtenir des données quantitatives permettant d'évaluer notre système sur une tâche réaliste.

5.2. La Tâche « Résumé et mise à jour » dans TAC 2008 et TAC 2009

5.2.1. Description de la Tâche

La tâche « Update » de TAC 2008 est inspirée d'un scenario dans lequel un utilisateur pose une même question à un système d'extraction d'information/résumé automatique à deux instants différents. La première fois, le système renvoie à l'utilisateur des articles de presse pertinents que l'utilisateur lit dans son ensemble. Plus tard, l'utilisateur revient au système afin de prendre connaissances des nouvelles informations concernant sa question. De nouveaux articles sont parus et le système doit générer un résumé de « mise à jour » de ces articles sachant que l'utilisateur a déjà lu le résumé lié à sa première demande.

La tâche « Résumé et mise à jour »(*Update Task*) est restée inchangée en 2009, afin que le NIST puisse évaluer les avancées réalisées en un an par les différents participants. Dans la continuité de la tâche de mise à jour de la campagne d'évaluation DUC 2007 [2], elle consiste à réaliser deux types de résumés :

- Initial : un résumé initial de 100 mots d'un jeu de 10 dépêches qui traitent d'un sujet particulier ;
- Mise à jour : un résumé de 100 mots d'un nouveau jeu de 10 dépêches qui traitent du même sujet. Ce résumé doit être généré en tenant compte du fait que le lecteur a déjà pris connaissance des informations contenues dans les 10 dépêches du jeu

1. National Institute of Science and Technologies
2. Les campagnes TAC ont succédé aux campagnes DUC, également menées par le NIST : http://www-nlpir.nist.gov/projects/duc/

de documents utilisé pour le résumé « initial ». L'objectif de ce second résumé est d'informer le lecteur des nouvelles informations à propos du sujet traité.

5.2.2. Description du corpus

Les dépêches de presse à résumer sont issues du corpus AQUAINT-2 (Dang et Owczarzak, 2008b). Ce corpus est un recueil de dépêches et comprend environ 907.000 documents constituant 2.5Go de texte, soit environ 40 millions de mots. Ces dépêches couvrent la période d'octobre 2004 à mars 2006 et proviennent de sources variées : AFP (en anglais), CNA (Taiwan), *Xinhua News Agency*, *LA Times-Washington Post*, *NY Times* et *Associated Press*.

Les évaluateurs de NIST ont créé 48 sujets et sélectionné 20 documents issus d'AQUAINT-2 pour chacun d'eux. Les documents extraits ont été ordonnés chronologiquement et divisés en deux jeux de 10 documents chacun, de telle manière que les documents d'un des jeux soient tous postérieurs à ceux de l'autre jeu, et constituent les sources du résumé de « mise à jour ». Le tableau 5.1 montre les titre, origine et date des documents qui constituent un sujet issu de TAC 2008.

Les dépêches issues des jeux de documents à résumer sont de longueur variable. Dans TAC 2008, le nombre de mots par dépêche varie entre 46 pour la plus petite et 3049 pour la plus grande. Elles font en moyenne 550 mots. Les données de TAC 2009 sont sensiblement identiques : la dépêche la plus petite fait 52 mots, la plus grande 4325 et la moyenne est de 580 mots par dépêche.

5.3. Adaptation de CBSEAS pour la tâche de résumé et mise à Jour

Les corpus de TAC 2008 et 2009 sont constitués de dépêches de taille moyenne. Ces dépêches ne sont pas majoritairement des dépêches courtes, qui relatent un événement en quelques phrases, sans le développer. Les dépêches de TAC replacent généralement l'événement dont elles traitent dans son contexte, en relatant notamment les événements majeurs qui ont conduit à la situation décrite. Ainsi, sur plusieurs dépêches, les mêmes informations essentielles reviennent à plusieurs reprises. Notre approche de résumé automatique, CBSEAS, se fonde sur le regroupement des phrases en classes sémantiques. Elle paraît tout indiquée ici, au vu du nombre de redondances identifiables dans les jeux de documents à résumer. Cependant, l'aspect particulier de la mise à jour ainsi que le traitement de dépêches de presse uniquement impliquent des adaptations afin de rendre notre système efficace pour cette tâche. Nous décrivons ci-dessous les adaptations que

Topic D0848 : Airbus A380 Describe developments in the production and launch of the Airbus A380.		
Jeu de documents « initial »		
16/01/2005	AFP	The Airbus A380 : from drawing board to runway-ready in a decade
16/01/2005	AFP	A380 'superjumbo' will be profitable from 2008 : Airbus chief
16/01/2005	APW	Airbus prepares to unveil A380 « superjumbo », world's biggest passenger jet
17/01/2005	LTW	Can Airports Accomodate the Giant Airbus A380 ?
19/01/2005	AFP	After fanfare, Airbus A380 now must prove it can fly
25/01/2005	AFP	Airbus mulls boosting A380 production capacity
10/04/2005	AFP	While US government moans, airports ready for Airbus giant
27/04/2005	AFP	Paris airport neighbors complain about noise from giant Airbus A380
27/04/2005	NYT	Giant Airbus 380 makes maiden flight
04/05/2005	AFP	Airbus A380 takes off on second test flight
Jeu de documents « mise à jour »		
01/06/2005	AFP	Airbus announces delay in delivering new superjumbo A380
03/06/2005	AFP	German wing of Airbus denies superjumbo A380 parts delay
05/10/2005	AFP	US aviation officials to study A380 turbulence
15/10/2005	AFP	Airbus says it cannot meet demand for A380 superjumbo
18/10/2005	APW	Second Airbus A380 makes maiden flight
13/11/2005	APW	Airbus executive says company will pay millions in compensation for late A380 deliveries
17/02/2006	APW	Airbus sees no delay to A380 after wing ruptured during test
22/02/2006	AFP	Airbus confident of A380 certification
26/03/2006	APW	33 people injured in evacuation drill for A380 super-jumbo
29/03/2006	APW	Airbus A380 superjumbo passes emergency evacuation test

TABLE 5.1.: Exemple d'un jeu de documents tiré de la tâche « Résumé et mise à jour » de TAC 2008. Les systèmes doivent résumer les documents du jeu de documents « initial » avant de résumer les nouveautés contenues dans le jeu de documents « mise à jour ».

nous avons effectuées pour notre participation à la tâche « résumé et mise à jour » de TAC 2008 et TAC 2009.

5.3.1. Filtrage des documents

Les dépêches de presse comportent des parties de texte indésirables dans un résumé automatique. Elles nécessitent donc un pré-traitement afin d'éliminer toute partie que l'on ne veut pas voir figurer dans le résumé généré par CBSEAS. Après avoir analysé le corpus d'apprentissage fourni par NIST, nous avons trouvé deux catégories de textes problématiques. Premièrement, certaines dépêches présentent également un récapitulatif des dépêches déjà publiées et en cours de publication sur un sujet précis, avec leur titre, leur auteur, et leur date de publication. Les descriptions de ces dépêches ont de fortes probabilités d'être intégrées dans le résumé car les requêtes utilisateur reprennent souvent les mêmes entités que celles contenues dans les titres des dépêches. Ensuite, certaines dépêches comportent des notes de pied de page. Celles-ci ne sont malheureusement pas annotées dans les documents de TAC, et constituent un réel problème pour un système de résumé automatique.

Ces notes figurent en bas d'un tiers des dépêches. Elles sont constituées en très grande majorité par des renvois à des sources d'information sur internet débutant généralement par les expressions : « On the Web » et « On the Net ». La figure 5.1 présente des exemples de ce type d'éléments. Ils correspondent à des motifs réguliers et facilement identifiables automatiquement et se révèlent particulièrement gênants pour notre système. Ils doivent être éliminés en amont de la génération du résumé. En effet, ces motifs répétitifs seraient regroupés par CBSEAS dans une seule et même classe, et l'un d'eux serait extrait dans le résumé car central dans sa classe et bien noté en similarité avec la requête. En effet, ces motifs contiennent généralement des éléments du titre de la dépêche. Au vu de la méthode de génération des jeux de documents et des requêtes, ils contiennent également par extension des éléments de la requête. Pour éviter ce biais, nous avons établi des filtres permettant d'éliminer ce type d'expressions.

5.3.2. Analyses linguistiques fines

Notre système repose en grande partie sur le calcul de similarité entre les phrases. Améliorer ce calcul de similarité permet d'obtenir un regroupement en classes sémantiques ; ceci est la clé pour gérer plus finement la diversité et la centralité locale telle qu'elle est définie en §3.2.5. Il existe trois solutions à cela : améliorer la méthode de calcul de similarité en utilisant des outils statistiques plus performants, prendre en compte des critères linguistiques plus fins, ou encore travailler sur la sémantique des phrases afin de reconnaître les reformulations, comme l'ont fait (Barzilay et McKeown, 2005) pour

SENATOR ASKS BUSH TO REVEAL SOURCE OF CIA LEAK

WASHINGTON - Sen. Charles Schumer, D-N.Y., called on President Bush Wednesday to reveal who leaked the name of a CIA operative to the press after reports that columnist Robert Novak said he is "confident the president knows who the source is."
A special prosecutor has been probing whether administration officials leaked the name of CIA agent Valerie Plame to Novak and others in an organized effort to retaliate against her husband, administration critic Joseph Wilson. As a covert officer, Plame is protected under a statute that makes it a crime under certain circumstances to reveal her identity.
"In keeping with your stated desire to root out leaks, I urge you to share with the American people whatever knowledge you have about which senior administration officials were responsible for disclosing Ms. Plame's identity as a CIA operative," Schumer wrote Bush in a letter.
The White House would not discuss the letter, saying its policy is to not comment on the leak while it is under investigation.
On the Web : Documents regarding the CIA leak investigation : www.dcd.uscourts.gov/04ms407.html
Eunice Moscoso's e-mail address is emoscoso@coxnews.com

Retreat of Antarctic ice shelves is not new, report says

The current retreat of ice shelves in the Antarctic due to global warming is nothing new _ but this time the problem is manmade and therefore potentially more serious, according to research released Wednesday.
Writing in the latest issue of the journal "Geology," British scientists said a survey had shown that ice shelves had retreated thousands of years ago as a result of rising air and ocean temperatures.
"What this tells us is that ice shelves don't just break up because they get too big _ as the global warning skeptics argue," said Dominic Hodgson, a scientist with the British Antarctic Survey and one of the leading investigators.
He said previous periods of warming _ about 9,500 years ago and some 2,000 years to 4,000 years ago _ were caused by natural causes, including the ending of ice ages, rather than man's emissions and the ice shelves had been able to reform.
"This time, the problem is man-made and if we don't take steps, the damage will be worse," he said. "There is no room for complacency."
The study, by scientists from the universities of Durham, Edinburgh and from the British Antarctic Survey, or BAS, said the George VI Ice Shelf on the Antarctic Peninsula is the first to show that a currently 'healthy' ice shelf experienced an extensive retreat about 9,500 years ago, more than anything seen in recent years.

On the Net : British Antarctic Survey, www.antarctica.ac.uk

FIGURE 5.1.: Exemples de notes de pied de page qui perturbent un système de résumé automatique, issues du corpus de la tâche « Résumé et mise à jour » de TAC 2008.

l'anglais. Cependant, cette dernière solution est difficilement adaptable à des langues différentes. Nous avons choisi la deuxième option, qui nous permet d'explorer les avancées récentes du TAL afin d'améliorer notre calcul de similarité entre phrases et présentons ici les différentes solutions que nous avons mises en œuvre à cet effet.

Analyse en entités nommées

Les entités nommées « correspondent traditionnellement à l'ensemble des noms propres présents dans un texte, qu'il s'agisse de noms de personnes, de lieux ou d'organisations, ensemble auquel sont souvent ajoutées d'autres expressions comme les dates, les unités monétaires, les pourcentages et autres »(Ehrmann et Jacquet, 2006). Les syntagmes qui composent ces objets sont généralement composés de plusieurs mots. Parce qu'ils réfèrent majoritairement à une entité nommée, nous avons choisi d'inclure les groupes nominaux définis (*e.g.* « le Président de la République française », « l'Université Paris 13 ») dans notre définition des entités nommées .

Prendre en compte de tels syntagmes comme une unité sémantique dans nos calculs de similarité entre phrases ou de similarité à la requête permet d'améliorer ces calculs. En effet, les entités nommées comportent souvent des titres communs. Deux entités nommées différentes qui partagent le même titre et apparaissant dans deux phrases distinctes augmentent la similarité entre celles-ci. Ce phénomène est très répandu. En effet, en analysant les dépêches de TAC 2008, nous avons trouvé que, dans un même jeu de 20 documents (nous présentons ici les analyses du jeu de documents D0805 donné en annexe A.2.1), le titre « Président » peut annoncer dix entités nommées différentes (*cf fig.5.2*).

De plus, une entité nommée composée de deux unités lexicales compterait deux fois dans le calcul de la similarité entre deux phrases dans lesquelles elle apparaît. Il est donc nécessaire d'analyser les documents en entités nommées afin de regrouper ces unités lexicales au sein d'un même élément. Le poids d'une entité nommée dans les calculs sera alors indépendant du nombre d'unités lexicales qui la composent. Les calculs de similarité ainsi que nos regroupements en classes sémantiques seront de fait plus précis. Pour ce faire, nous avons utilisé la méthode décrite en §4.1.2.

Résolution d'anaphore et de co-référence

L'une des règles de la rédaction de textes est de minimiser la redondance lexicale en ne répétant pas un même mot deux fois, mais en utilisant des synonymes. Cette règle s'applique également aux entités nommées. La règle est alors d'utiliser un syntagme nominal différent ou un syntagme pronominal pour référer à une entité nommée déjà

Entités nommées issues du jeu de documents D0805 de TAC 2008 et comportant le titre « *president* »

- Jacklyn Kosecoff, executive vice **president** of specialty companies at Cypress-based PacifiCare
- Susan Rawlings, **president** of senior services of WellPoint
- Helen Darling, **president** of the National Business Group on Health
- Larry Boress, vice **president** of the Midwest Business Group on Health
- Scott M. Spencer, senior vice **president** of the National Rural Electric Cooperative Association
- Robert Hayes, **president** of the Medicare Rights Center
- **President** Bush
- Drew. E. Altman, **president** of the Kaiser Family Foundation
- James L. Firman, **president** of the National Council on the Aging

FIGURE 5.2.: Illustration du nombre d'entités différentes annoncées par un même titre dans un jeu de documents issu de TAC 2008 (D0805)

citée (cf fig. 5.3).

Les dépêches de presse courtes font exception à la règle, car elles sont rédigées de manière à ce que leur compréhension soit immédiate. Les co-références et anaphores, qui demandent un effort cognitif de la part du lecteur, sont donc évitées au maximum. Cependant, les corpus de TAC 2008 sont composés majoritairement de dépêches de taille moyenne. Il y a donc des risques de rencontrer des expressions non auto-référentielles. Or l'amélioration du calcul de similarités entre phrases passe par la résolution de la référence de telles expressions. De même, un résumé doit se suffire à lui-même ; il ne faut donc pas intégrer à un résumé des phrases dans lesquelles des expressions pronominales n'ont pas été remplacées par leur référent.

Nous avons donc utilisé un outil de résolution d'anaphore et de co-référence (*cf §4.1*), le système ANNIE intégré à GATE (Cunningham *et al.*, 2002), que nous avons appliqué aux entités nommées uniquement, afin d'améliorer les calculs de similarité. Nous avons également utilisé ce système pour remplacer dans les résumés les expressions pronominales par leur référent, et remplacé les entités nommées longues par des entités équivalentes plus petites afin de réduire la taille des résumés. La méthode utilisée est décrite plus en détails en §4.1.2.

Bobby Fischer can escape US if Iceland makes him citizen
Chess legend Bobby Fischer, who faces prison if he returns to the United States, can only avoid deportation from Japan if Iceland upgrades its granting of residency to full citizenship, a Japanese lawmaker said Wednesday.(...) Fischer, 62, known for his maverick behavior and outspoken criticism of his country, may face 10 years in prison if convicted of violating US sanctions imposed on the former Yugoslavia by playing there in 1992.(...)Fukushima met Masaharu Miura, chief of the justice ministry's immigration bureau, who had told parliament Tuesday that the chess genius could only be deported to his country of origin – the United States.

Cheney lacked license when he accidently shot another hunter
US Vice President Dick Cheney did not have a legal permit to hunt quail when he accidentally shot a 78-year old hunting partner over the weekend, the White House said on Monday. The Texas Parks and Wildlife Department informed Cheney that he lacked a stamp for hunting « upland game birds » in the state when the shooting incident occurred, the vice president's office said in a statement.(...) An interview with Whittington had confirmed Cheney's account, the sheriff's office said.(...) While awaiting a warning from game wardens, "the Vice President has sent a 7 dollar check to the Texas Parks and Wildlife Department, which is the cost of an upland game bird stamp," it said.(...)

FIGURE 5.3.: Illustration de reformulations d'entités nommées dans les dépêches de presse (dépêches tirées du corpus AQUAINT-2).

Utilisation d'une ressource lexicale sémantique

Afin d'obtenir un calcul de similarité encore plus précis entre les phrases, il est nécessaire de pouvoir encoder les relations de synonymie et d'hyperonymie entre les termes. Ainsi, en prenant les trois phrases :

- « Un lion a été aperçu dans le bois de Boulogne. »
- « Un gros félin a été repéré aux alentours du bois de Boulogne. »
- « Treiber a été aperçu dans le bois de Boulogne. »,

les deux premières seraient plus proches l'une de l'autre que la première ne le serait de la troisième en prenant en compte les relations d'hyperonymie entre « lion » et « félin », et de synonymie entre « apercevoir » et « repérer ». En revanche, en ne prenant pas en compte de telles relations, la première phrase serait évaluée comme étant plus proche de la troisième que de la seconde.

Les tâches « Résumé et Mise à Jour » de TAC 2008 et TAC 2009 portent sur l'anglais. Il existe des ressources accessibles qui encodent de telles relations. WordNet, une ressource lexicale sémantique organisée hiérarchiquement, encode les relations de synonymie et d'hyperonymie que nous cherchons à utiliser (Fellbaum, 1998). Largement utilisée par des systèmes de TAL pour des applications variées, cette ressource a été à la base de méthodes de mesures de similarité sémantique, notamment celles définies dans (Wu et

Wu, 1994; Resnik, 1995; Leacock et Chodorow, 1998).

WordNet repose sur la notion de *synset*, un groupe de mots dénotant un sens particulier (Wikipedia). Les différents mots d'un même synset (pour *synonym set*) peuvent être considérés comme synonymes. Toutes les mesures de similarités énoncées ci-dessus sont fondées sur la distance entre deux *synsets* dans l'arbre de WordNet. La mesure JCn (Jiang et Conrath, 1997) se démarque des autres en complétant l'utilisation de WordNet par une analyse distributionnelle des *synsets* en corpus. Nous avons utilisé cette mesure, qui obtient de plus une corrélation très forte avec les résultats humains (0.828 contre 0.885 pour deux humains sur la même tâche) (Jiang et Conrath, 1997) sur une tâche de mesure de la similarité sémantique entre mots.

Utilisation de la structure des dépêches

Nous avons montré en chapitre 4.2 que si les dépêches de presse partagent certaines caractéristiques, elles peuvent être classées en plusieurs catégories. Nous avons également montré que parmi les différentes catégories de dépêches, certaines sont constituées de phrases plus propices à une extraction directe pour générer un résumé.

Lors de la participation à TAC 2009, notre système d'étiquetage des dépêches et d'analyse de leur structure distinguait cinq types différents de dépêches, définis en §4.2.3 :

- chronologies,
- fiches techniques,
- Rapport de discours,
- Revues d'opinion (ou micro-trottoirs),
- dépêches « classiques »,

et analysait complètement la structure des dépêches de type « chronologie », « fiche technique » et partiellement la structure des dépêches « classiques », des micro-trottoirs et des rapports de discours. La structure des dépêches de ces trois dernières catégories est en effet moins explicite et nécessite une analyse poussée qui fera l'objet de travaux ultérieurs.

Nous avons intégré ce système à CBSEAS de la manière suivante : les phrases issues d'une dépêche d'une chronologie ou d'une fiche technique reçoivent un bonus (Bs) lors du calcul de leur score si le contenu de leur dépêche correspond à la requête utilisateur (*cf §4.2.6*). Ces phrases sont en effet écrites dans un style concis, ce qui en fait des candidats parfaits à l'intégration dans le résumé automatique.

Les premières phrases des dépêches classiques sont, du fait de l'écriture pyramidale des dépêches de presse, porteuses des informations les plus centrales. Celles-ci sont, depuis

(Edmundson, 1969), privilégiées dans la plupart des systèmes de résumé par extraction. En attendant de pouvoir utiliser notre système d'analyse de la structure des dépêches classiques, nous avons également empiriquement privilégié les trois premières phrases, en leur attribuant un bonus dégressif : Bs pour la première, $\frac{2\times Bs}{3}$ pour la seconde, et $\frac{Bs}{3}$ pour la troisième.

Dans un premier temps, Bs a été fixé empiriquement (*cf §4.2.6*). Il fait partie des paramètres d'entrée de CBSEAS, et sa valeur peut être calculée en même temps que la valeur des autres paramètres par notre algorithme génétique.

5.3.3. Gestion de la mise à jour

La tâche de mise à jour consiste à reconnaître ce qui, dans le deuxième jeu de documents, constitue une information nouvelle par rapport au jeu de documents « initial ». Certains systèmes, comme (Galanis et Malakasiotis, 2008) ont géré la mise à jour en supprimant du deuxième jeu de document toute phrase dont la similarité à une phrase des documents initiaux est supérieure à un seuil. D'autres ont choisi de déplacer le biais que constitue l'instauration d'un seuil à la similarité entre les jeux de documents eux-mêmes. Ainsi, les phrases du second jeu de documents sont supprimées itérativement jusqu'à ce que la similarité entre le premier et le second jeu de documents soit inférieure à un certain seuil (He *et al.*, 2008). C'est alors un seuil sur une similarité entre ensembles qui est utilisé, et la limite en dessous de laquelle une phrase est considérée comme apportant de la nouveauté est adaptée automatiquement à chaque nouveau corpus en fonction de ce seuil. D'autres systèmes ont simplement vu leur méthode de résumé automatique adaptée sans que le principe en soit changé : par exemple, les auteurs de (Boudin *et al.*, 2008a) ont augmenté dans leur variante de MMR (*cf §1.5*) l'importance de la dissimilarité aux phrases déjà sélectionnées pour le calcul du score des phrases issues des documents de mise à jour ; ceux de (Varma *et al.*, 2008) ont ajouté à leur modèle de langage la similarité entre les phrases des documents de mise à jour et les phrases des documents initiaux.

Nous avons voulu aller plus loin que les premières méthodes citées en proposant un algorithme qui ne nécessite pas de fixer *a priori* de seuil de similarité, tout en conservant notre procédé initial. Notre algorithme de résumé modélise la diversité informationnelle des documents dont il dispose grâce à la classification en classes sémantiques des phrases de ces documents. Cette modélisation peut être à la base de la détection de la nouveauté. En effet, après avoir modélisé la diversité informationnelle d'un premier jeu de documents, détecter la nouveauté dans un second jeu de documents consiste à repérer les phrases qui ne se s'intègrent pas dans le modèle alors établi.Les phrases qui seront intégrées aux modèle du premier jeu de documents seront considérées comme ne véhiculant pas une information nouvelle, tandis que les phrases qui ne s'y intègreront pas seront des candidates au résumé de « mise à jour ».

Identifier les éléments d'information nouveaux dans les jeux de documents « mise à jour » consiste alors à trouver ceux qui ne se regroupent pas avec les éléments d'information des documents initiaux. Nous avons pour cela utilisé l'incrémentalité de notre algorithme de clustering, fast global k-means (*cf §3.2.4*). Après avoir réalisé le résumé des documents initiaux, nous sauvegardons le résultat de la classification des phrases. Lors du résumé des documents de mise à jour, nous initialisons l'algorithme de classification à la sauvegarde précédemment réalisée, en y ajoutant les phrases des documents de mise à jour. L'algorithme poursuit alors jusqu'à obtenir n classes supplémentaires [3] avec les contraintes suivantes :

- les phrases du groupe de documents initial sont inamovibles. Permettre à ces phrases de changer de classe pourrait perturber les classes censées porter de la nouveauté en y ajoutant des phrases issues du premier jeu de documents.
- les centres des classes sauvegardées sont fixés et ne peuvent pas être recalculés. La proximité aux centres des classes est en effet le seul critère de regroupement de *fast global k-means*. Recalculer les centres des classes qui modélisent la diversité informationnelle du premier jeu de documents augmenterait donc la portée sémantique de ces classes, ce qui affecterait négativement notre procédé de détection de la nouveauté.

Ainsi, les phrases des documents de mise à jour ne modifieront pas la modélisation que CBSEAS avait réalisé des phrases des documents initiaux. De plus, seules les phrases de mise à jour considérées les plus éloignées des phrases initiales seront sélectionnées pour créer une nouvelle classe et ainsi pouvoir être extraites dans le résumé.

Au cours d'expériences que nous avons menées, nous nous sommes aperçus que les phrases issues des seconds groupes de documents avaient, pour des raisons que nous n'expliquons pas, tendance à se regrouper avec les phrases du premier groupe de documents, générant des classes peu peuplées pour le résumé de mise à jour. Nous avons choisi de limiter ce phénomène en introduisant un nouveau paramètre, compris entre 0 et 1, qui réduit les similarités entre les phrases du second groupe et les phrases du premier groupe de documents en les multipliant par sa valeur. Dans la section §3.3.3 consacrée à l'optimisation de notre système par un algorithme génétique, ce paramètre porte le nom de « Réduction de la similarité Update ».

Cette méthode possède un inconvénient majeur : en l'absence d'informations nouvelles, le système sélectionnera tout de même les phrases de mise à jour les plus éloignées des centres des classes initiales pour faire partie des nouvelles classes. Des phrases des documents de mise à jour seront alors extraites même si elles ne présentent aucun élément d'information inédit. Il existe des solutions pour remédier à cela : l'utilisation d'indices de validité de classification, par exemple l'indice de Davies-Bouldin (Vesanto *et al.*, 2000), permet de déterminer si ajouter de nouvelles classes ne réduit pas la qua-

3. n est donné en paramètre de CBSEAS

- Over the last year, the shares of Evergreen **Solar**,(...) – all small domestic companies that make equipment for converting **solar power** into electricity – have more than doubled in price.
- Many didn't sleep on account of the **rain** – and concerns about **leaks** or **floods**.
- The deal to free **Fischer** came after **Iceland** – a **chess**-loving nation that hosted his historic Cold War-era victory over Soviet Boris Spassky
- in 1972 – granted **Fischer** citizenship this week in a move to help him evade trial in the United States.
- (...) restrictive policies on (...) **emergency contraception** (the **morning after pill**) and medical care for the working poor force many women (...).
- The success of the Weyburn Project could have incredible implications on reducing CO2 (**carbon** dioxide) emissions and increasing America's oil production.

FIGURE 5.4.: Illustration de l'importance des incises dans le calcul des similarités à la requête - En gras : les mots liés à la requête

lité de la classification. Cependant, ce type de techniques est superflu dans notre cas, puisque les documents sélectionnés pour la mise à jour sont tous des dépêches postérieures aux dépêches sélectionnées pour le résumé initial. Or l'écriture d'une dépêche suppose l'apparition d'un élément d'information nouveau. Nous pouvons donc utiliser ici notre approche sans risque de sélectionner des phrases indésirables dans le résumé de mise à jour.

5.3.4. Post-traitements spécifiques aux dépêches

Les dépêches contiennent de nombreuses incises. Ces précisions, apportées par l'auteur pour clarifier son propos ou pour définir des termes ou des entités qui ne sont pas communs sont censées pouvoir être supprimées ou éludées par le lecteur sans perturber la lecture de la dépêche. Cependant, les supprimer en pré-traitement priverait notre système de résumé – qui est incapable, contrairement à l'humain, d'associer des notions à des termes – d'informations lui permettant de relier des phrases porteuses de ce type d'éléments à la requête utilisateur. La figure 5.4 illustre par quelques exemples tirés des données de TAC l'importance que peuvent avoir les incises sur le calcul de la similarité à la requête.

Nous avons coupé des résumés toutes les incises délimitées par des tirets ou des parenthèses. Cependant, les parenthèses sont également utilisées dans les discours rapportés afin de compléter grammaticalement ou sémantiquement une phrase rendue incomplète suite à la citation partielle du discours. Nous avons donc laissé intacts les groupes parenthésaux inclus dans une citation. Le module d'élimination des incises a été développé

	Paramétrage manuel	Algorithme génétique	WordNet	Entités nommées	Structure	Résolution d'anaphore	tf.icf
v0.1a	+	-	-	-	-	-	-
v0.1b	+	-	+	-	-	-	-
v0.2	-	+	+	-	-	-	-
v0.3a	+	+	+	-	+	-	-
v0.3b	-	+	+	-	+	-	-
v0.4a	-	+	+	+	+	-	-
v0.4b	-	+	+	+	+	-	+
v0.5	-	+	+	+	+	+	-

TABLE 5.2.: Récapitulatif des différentes versions de CBSEAS pour la tâche de « Résumé et mise à jour »

à l'aide d'expression régulières écrites en perl.

5.4. Systèmes soumis aux campagnes TAC

Nous présentons ici les différentes versions des systèmes que nous avons évaluées grâce aux campagnes TAC ou à leurs données. Les différentes versions du système correspondent à des états d'avancement des travaux différents, et n'intègrent pas toutes les mêmes modules. Le tableau 5.2 présente un récapitulatif des caractéristiques des différentes versions de CBSEAS.

5.4.1. TAC 2008

La tâche de résumé et mise à jour de TAC 2008, présentée plus en détails en §5.2, requiert la création de deux types de résumé différents, un résumé initial et un résumé de mise à jour guidés par une requête utilisateur. Seule la première version de CBSEAS a pu être soumise à TAC 2008. Les autres versions, qui y apportent des corrections et intègrent des modules supplémentaires, ont été évaluées en dehors de la campagne en utilisant les données pour l'évaluation automatique fournies par le NIST, l'organisateur de TAC. Cela nous a permis de mesurer l'apport précis de chacun des modules dans le système global.

Version TAC 2008 officielle (v0.1a)

Cette version de CBSEAS est celle que nous avons officiellement soumise à la campagne TAC 2008. Notre système n'était pas encore finalisé, et cette version a servi comme *base-*

line afin d'évaluer les changements apportés à CBSEAS. Elle porte l'identifiant (v0.1a) dans les résultats présentés dans cette section.

Cette version n'intègre ni l'analyse en entités nommées, ni la résolution d'anaphore, ni l'utilisation de WordNet. Le calcul de la similarité à la requête correspondait au nombre de mots communs entre une phrase et la requête après lemmatisation et élimination des mots vides [4]. Ce calcul servait uniquement à filtrer les phrases en entrée du système de résumé. Nous avions alors fait le choix d'utiliser uniquement la centralité locale afin de sélectionner une phrase par classe. En évaluant les résultats de TAC 2008, nous avons constaté que certaines phrases étaient trop éloignées de la requête pour figurer dans le résumé et avons donc choisi d'ajouter au score des phrases un score de similarité à la requête.

Les poids liés aux différentes catégories morpho-syntaxiques ont été fixés empiriquement, en donnant aux constituants principaux des phrases (noms propres, noms communs et verbes) des poids supérieurs à ceux des constituants, adjectifs et adverbes [5]. Le système a été paramétré pour générer des résumés longs de cinq phrases.

TAC 2008 non officielle : paramétrage du système et utilisation de WordNet (v0.1b)

Immédiatement après la campagne d'évaluation, nous avons modifié le système afin de pallier les différents problèmes identifiés en §5.5.4. Le trop faible nombre de mots des résumés générés provoquait une perte de l'ordre du tiers en précision ; le nombre de phrases en sortie a donc été augmenté. Nous avons repris le score de similarité entre phrases et l'avons appliqué au score de similarité à la requête. La ressource lexicale WordNet a également été utilisée de la manière décrite en §5.3.2 afin d'améliorer la précision du calcul de similarité entre phrases.

TAC 2008 non officielle : optimisation avec un algorithme génétique (v0.2)

Cette version fait suite à la version soumise à TAC 2008 et porte l'identifiant 0.2. Elle corrige les défauts majeurs de la première version, à savoir l'absence d'un score de similarité à la requête et les poids des différents paramètres fixés empiriquement. Le score de similarité à la requête est calculé de la même manière que la similarité entre phrases.

4. Nous avons utilisé la liste des mots vides pour l'anglais issue de http://www.textfixer.com/resources/common-english-words.txt.

5. Nous avons montré en §3.3.5 que c'était une erreur.

Nous nous sommes aperçus que les requêtes de TAC présentent des régularités dans le vocabulaire utilisé. Ainsi, des mots tels que « *describe* », « *provide* », « *developments* », « *report* », « *include* », ou « *details* » reviennent fréquemment et interviennent dans la mesure de similarité à la requête au même titre que les autres, tandis que leur intérêt pour cette mesure est limité. Afin d'éviter de perturber le calcul de similarité, nous avons éliminé les 10 mots les plus fréquents de l'ensemble des requêtes de la tâche « Résumé et mise à jour » de TAC 2008.

Afin d'obtenir une approximation des paramètres de CBSEAS (*cf §3.3.4*) les plus adaptés à la tâche proposée par NIST, un algorithme génétique détaillé en §3.3.3 a été utilisé avec comme données d'apprentissage une partie du corpus de TAC 2008.

TAC 2008 non officielle : gestion de la structure (v0.3a et v0.3b)

Nous présentons ici deux versions différentes. Celles-ci reprennent les caractéristiques de la version 0.2, à laquelle nous avons ajouté la gestion de la structure des dépêches afin de prendre en compte la position d'une phrase vis-à-vis de la structure du document dont elle est tirée lors du calcul de son score (*cf §5.3.2*). Dans la version notée v0.3a, les poids liés à la prise en compte de la structure des documents ont été fixés empiriquement et les autres laissés inchangés par rapport à la version précédente. Ceci nous permet d'évaluer l'apport de l'intégration de la structure seul. Dans la version v0.3b, l'algorithme génétique a été relancé de manière à ajuster les poids, y compris ceux liés à l'utilisation de la structure des dépêches en vue d'une nouvelle évaluation dans TAC 2009.

5.4.2. TAC 2009

TAC 2009 nous permettait de soumettre seulement deux systèmes. Nous avons choisi d'évaluer les apports de l'utilisation d'une annotation en entités nommées ainsi que de la résolution d'anaphores. Nous avons également testé *a posteriori* une dernière version de CBSEAS sur les données de TAC 2009. Celle-ci utilise une méthode de calcul du score de centralité locale utilisant le *tf.icf*[6] à la place du *tf.idf* afin d'en améliorer la précision.

5.4.3. TAC 2009 : soumission officielle 1 (v0.4a)

La première version que nous avons soumise à TAC 2009 reprend les caractéristiques de la version TAC 2008 avec gestion de la structure. Nous lui avons ajouté la gestion des entités nommées décrite en §5.3.2. Le système doit ainsi être plus précis dans le calcul

6. Le tf.icf est décrit en §3.2.5

des similarités à la requête, des similarités entre phrases, et dans le calcul du score de centralité locale. Cette version porte l'identifiant v0.4a dans les résultats de TAC 2009.

5.4.4. TAC 2009 : évaluation *a posteriori* du tf.icf

Nous avons voulu différencier ici le calcul de similarité entre phrases servant au regroupement des phrases en classes sémantiques et le calcul de similarité entre les phrases qui font partie d'une même classe sémantique, qui donne une mesure de la centralité locale à une classe. Nous avons voulu que ce deuxième calcul reflète mieux cet aspect. Pour cela, nous avons utilisé la même méthode de calcul de similarité en changeant la pondération des mots. Nous avons utilisé le *tf.idf* pour le calcul de similarité globale, et le *tf.icf* (*cf §3.2.5*) pour le calcul de similarité locale (à l'intérieur d'une classe sémantique).

Nous pouvions présenter seulement deux systèmes à TAC 2009 et avons choisi d'évaluer prioritairement les apports de l'utilisation de techniques de TAL. Par conséquent, nous avons évalué ce système *a posteriori*, sur les données de TAC 2009. Elle porte l'identifiant v0.4b dans les résultats présentés en figure 5.8.

5.4.5. TAC 2009 : soumission officielle 2 (v0.5)

Cette dernière version de CBSEAS, soumise à TAC 2009, intègre tous les modules décrits en §5.3. Elle est similaire à la soumission officielle 1 de TAC 2009, à laquelle nous avons intégré la résolution d'anaphore et de co-référence, présentée en §4.1, afin d'évaluer indépendamment l'apport de ce module à notre système de résumé.

5.5. Evaluation

Nous décrivons ici les protocoles d'évaluation utilisés par NIST lors des campagnes TAC 2008 et 2009 pour la tâche de résumé et mise à jour, ainsi que les résultats obtenus par nos différents systèmes, soit directement évalués lors de la campagne, soit évalués *a posteriori* sur les données de TAC 2008.

5.5.1. Protocole

NIST a choisi d'évaluer les résumés de deux manières :

– automatiquement via les mesures ROUGE [7],
– manuellement en utilisant la méthode Pyramide et des évaluations subjectives.

Les évaluations automatiques sont moins précises que les évaluations manuelles. Cependant, elles sont essentielles à toute campagne d'évaluation. Elles permettent en effet de se comparer *a posteriori* aux systèmes qui ont participé à une campagne d'évaluation, avec des mesures somme toute assez corrélées aux jugements humains (*cf §2.1*).

Les jugements humains sont néanmoins plus fiables et donc nécessaires lorsqu'il s'agit de classer officiellement des résultats, quelle que soit la tâche. La méthode Pyramide décrite en §2.4 fournit un procédure objective pour évaluer l'informativité (la quantité d'informations essentielles) d'un résumé. Afin de la compléter, il était nécessaire de fournir une évaluation de la qualité linguistique. NIST l'a fait en demandant à ses juges de donner à chaque résumé une note de lisibilité (*readability)* en tenant compte des critères suivants :

– grammaticalité,
– non-redondance,
– clarté des références (referential clarity),
– précision de la réponse apportée à la requête de l'utilisateur (*focus*) [8],
– cohérence

NIST a également demandé aux juges de donner une note qui corresponde plus à la réaction d'un utilisateur devant un résumé. La note de cette évaluation subjective, que nous nommerons « performance générale » (*overall responsiveness*) évalue à quel point un résumé répond aux besoins en information exprimés par une requête en considérant à la fois le contenu du résumé et sa qualité linguistique. Les scores en performance générale et en qualité linguistique sont exprimés de 1 à 5 sur l'échelle suivante :

1. Très mauvais (*Very Poor*)
2. Mauvais (*Poor*)
3. Acceptable (*Barely Acceptable*)
4. Bon (*Good*)
5. Très bon (*Very Good*)

5.5.2. Baselines

Afin de donner un élément de comparaison fiable et d'évaluer l'état actuel des recherches en résumé automatique, le NIST a fourni les résultats d'un système « Baseline

7. Les mesures d'évaluation automatiques ROUGE sont présentées en §2.1
8. Nous utiliserons le terme *focus* dans le reste du document

pour TAC 2008 et 2009. Ce système extrait simplement les n premières phrases du document le plus récent des jeux de documents initiaux et de mise à jour (la $n+1$ est celle qui si elle était extraite, ferait dépasser le maximum de 100 mots imposé par la tâche). Cette *baseline*[9] fournit une approximation du score minimum que devrait obtenir un système de résumé automatique.

Pour TAC 2009, le NIST a ajouté deux *baselines*. Celles-ci, contrairement à la première, ne correspondent pas à ce que l'on appelle communément une *baseline*. En effet, une *baseline* correspond traditionnellement à la borne inférieure que doit atteindre un système automatique. Les deux *baselines* dont il est question ici fournissent plutôt une borne supérieure. La *baseline* numéro 2 est un système qui choisit aléatoirement un des résumés de référence et en mélange les phrases. Elle permet d'évaluer l'impact de la qualité linguistique (en tous cas du réordonnancement des phrases) sur la perception d'un utilisateur de la qualité d'un résumé (performance globale). La *baseline* numéro 3) consiste en des résumés générés manuellement et composés uniquement de phrases extraites en intégralité des documents sources. Réalisée par une équipe de l'Univeristé de Montréal, elle fournit une borne supérieure de ce que l'on pourrait obtenir de systèmes de résumé fonctionnant par extraction uniquement (Genest *et al.*, 2009).

5.5.3. Résultats

Nous présentons dans cette section les résultats des évaluations des systèmes CBSEAS dédiés à l'update que nous avons présentés en §5.4. Nous expliquons le choix de présentation des résultats, qui seront discutés en §5.5.4.

NIST a décidé de présenter deux types de résultats : des résultats manuels et des résultats automatiques. Ces résultats automatiques, ainsi que les données fournies, permettent de comparer les résultats obtenus par des sytèmes développés après la campagne d'évaluation aux résultats de ceux qui ont participé à la campagne. Cependant, nous avons montré en chapitre 2 que si les évaluations automatiques sont de plus en plus corrélées aux évaluations manuelles, celles-ci ne sont pas encore d'une qualité suffisante pour être le seul critère utilisé pour rendre compte de la valeur d'un résumé.

Nous présentons donc les évaluations manuelles des campagnes TAC, ainsi que des évaluations automatiques *a posteriori* que nous avons validées par des évaluations manuelles.

9. la seule baseline de TAC 2008, et la baseline 1 dans TAC 2009

TAC 2008

Le système présenté à TAC 2008 (v0.1) a obtenu de mauvais résultats, se classant dans les moins bons systèmes de la campagne d'évaluation. La figure 5.5 présente les résultats automatiques des différentes versions de CBSEAS auxquelles nous avons ajouté différents modules afin de les rendre compétitives. Elle montre une nette amélioration des résultats lors de l'utilisation de l'algorithme génétique pour adapter les paramètres de CBSEAS à la tâche « Update ». Un paramétrage correct du système semble donc primordial afin d'obtenir des résumés d'une qualité suffisante. Nous avons voulu nous assurer que des évaluateurs humains étaient également sensibles à cette amélioration. Nous avons utilisé le protocole décrit en §3.3.6 à cette fin. Le tableau 5.3 montre que l'amélioration obtenue grâce à l'algorithme génétique n'est pas perçue uniquement par l'évaluation automatique, mais également par des juges humains. Le système paramétré à l'aide de l'algorithme génétique (CBSEAS v0.3) obtient en effet de meilleures notes moyennes, et 33% des résumés qu'il a générés ont été perçus comme les meilleurs des trois systèmes évalués contre 13% pour la version de CBSEAS sans optimisation des paramètres.

	Meilleur système de TAC 2008	CBSEAS v0.2	CBSEAS v0.3
Résumés standards			
Nombre de résumés gagnants	9	2	4
Note moyenne	4.7	3.9	4.3
Résumés « mise à jour »			
Nombre de résumés gagnants	8	2	5
Note moyenne	5	3.7	4.5
Tous Résumés			
Nombre de résumés gagnant	17	4	9
Note moyenne	4.8	3.8	4.4

TABLE 5.3.: Évaluation manuelle de l'apport de l'algorithme génétique à CBSEAS sur les données de TAC 2008

TAC 2009

Afin d'obtenir une évaluation plus objective et précise que celle que nous aurions pu obtenir par nos propres moyens, nous avons décidé de valider l'impact des changements apportés au système par une seconde participation à TAC.

La figure 5.6 présente les résultats des évaluations manuelles de TAC 2009 (score subjectif de « performance générale » et score pyramide), tous résumés confondus : résumés initiaux et mise à jour. La figure B.1 en annexe présente également les résultats des

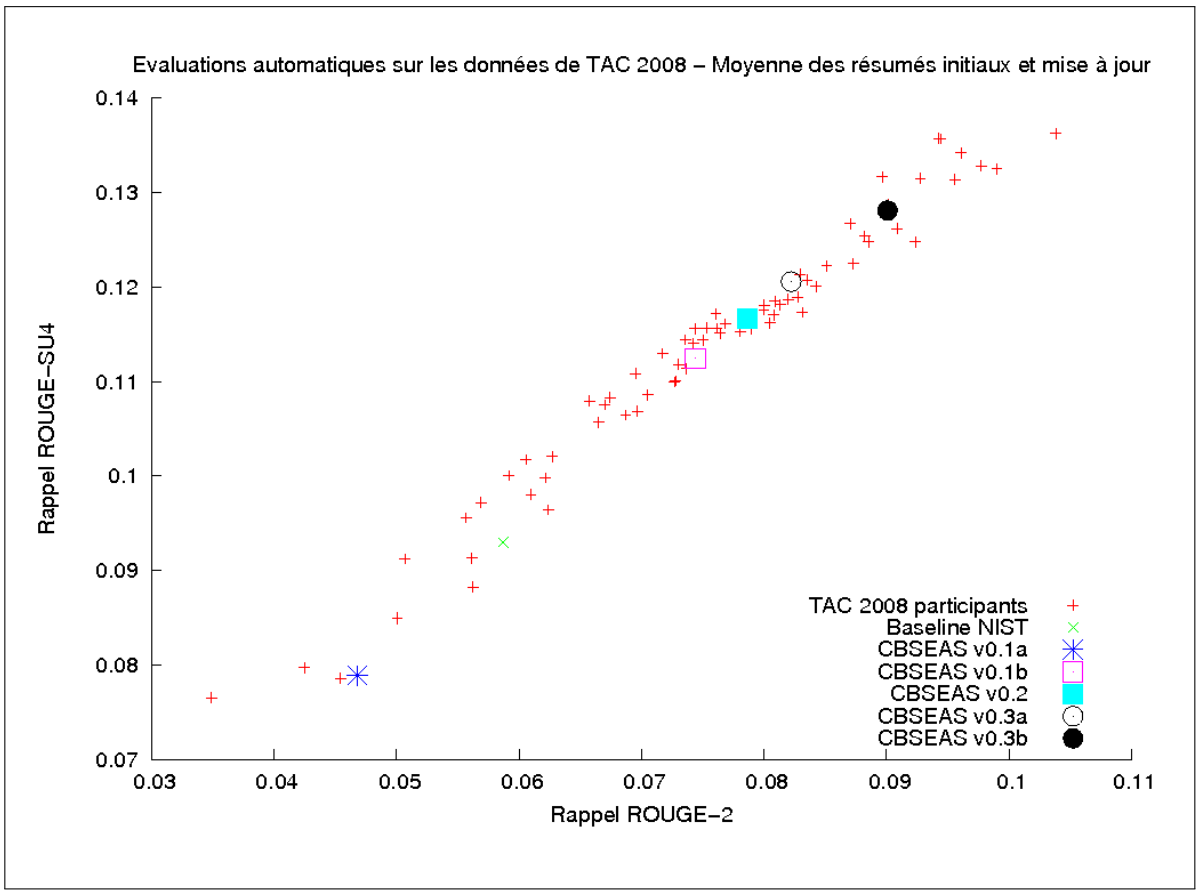


FIGURE 5.5.: Evaluations automatiques des versions de CBSEAS sur les données de TAC 2008 *a posteriori*

évaluations manuelles de TAC 2009, mais les résultats des résumés initiaux et mise à jour y figurent cette fois-ci séparément. On constate que CBSEAS v0.4a figure parmi le premier quart des participants pour le score pyramide, et parmi le premier tiers pour le score de performance générale. Par ailleurs, on notera que notre système génère des résumés contenant, d'après le score pyramide, 83% des informations pouvant être extraites par un système purement extractif –nous considérons que la baseline 3 fournit une mesure précise du maximum d'informations pouvant être véhiculées par un tel système. Les scores du système pour les résumés de mise à jour sont par ailleurs plus élevés que pour les résumés initiaux. Cela montre l'efficacité de notre stratégie de détection de la mise à jour et donc de la classification en classes sémantiques.

La figure B.2 en annexe et la figure 5.7 présentent, sous la même forme que les deux précédentes, les résultats en qualité linguistique et en performance générale des différents systèmes qui ont participé à TAC 2009. Le tableau 5.4 présente les résultats et le classement des systèmes CBSEAS v0.4a et v0.5 dans toutes les évaluations fournies par la campagne TAC 2009. CBSEAS se place dans la deuxième moitié du classement

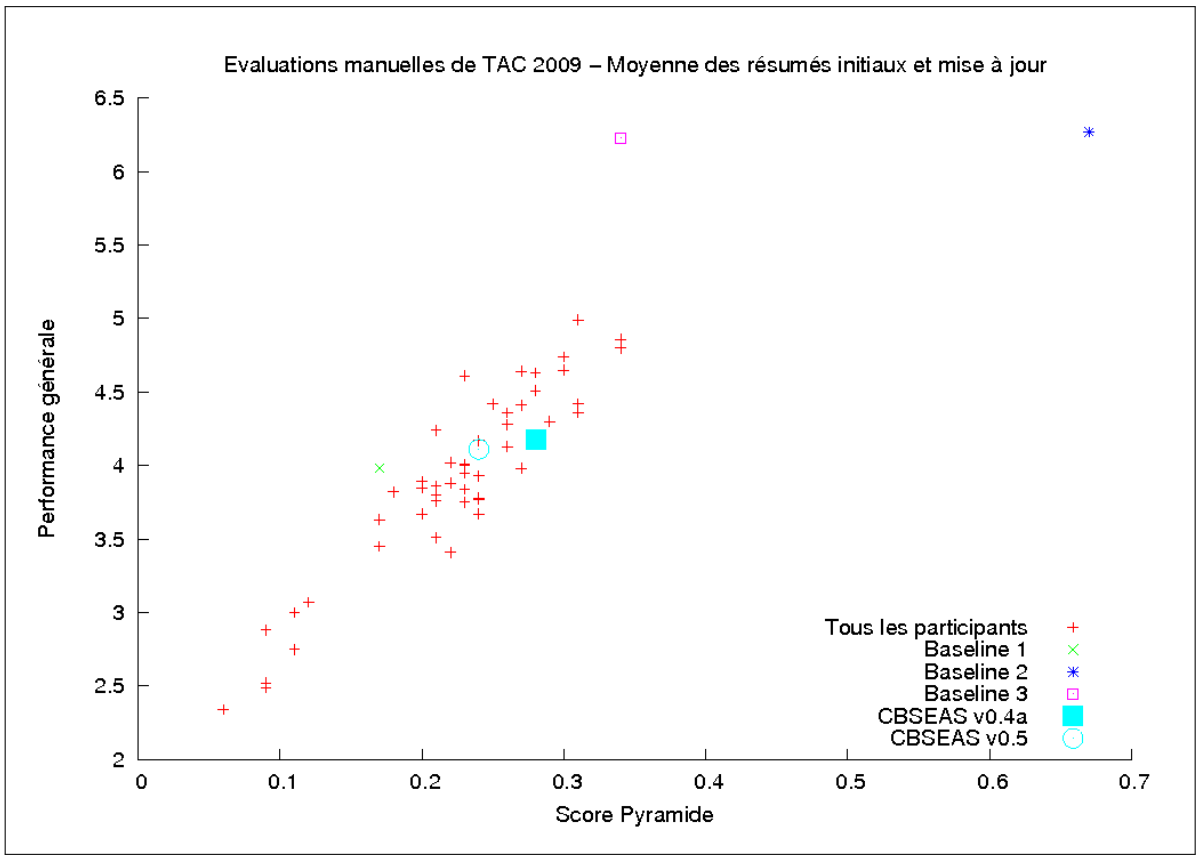


FIGURE 5.6.: Évaluations manuelles de TAC 2009 « Update Task » : moyenne des résumés initiaux et mise à jour

pour ce qui est des scores linguistiques, ce qui explique que notre système obtienne un classement plus faible en « performance générale » que pour le score pyramide, censé refléter la qualité informative des résumés générés.

5.5.4. Discussion

Le premier résultat obtenu par CBSEAS lors de la campagne TAC 2008 a été très décevant. Cependant, nous avons pu en tirer des enseignements afin d'améliorer la qualité des résumés générés par notre système. Premièrement, nous avions fait le choix de respecter strictement la limite de 100 mots imposée par le NIST. Seules les phrases qui se terminaient avant cette limite étaient conservées dans les résumés. Il en a résulté des résumés d'une taille moyenne de 65 mots, plus courts de 28% que les trois systèmes les mieux classés. Or générer un résumé par extraction plus long augmente la probabilité d'intégrer des éléments d'information pertinents. Le NIST ne pénalise pas les systèmes

Moyenne des résumés initiaux et de mise à jour

RunID	ROUGE-2	ROUGE-SU4	Pyramid	Linguistique	Perf. générale
Classement (v0.4a)	9/53	10/53	11/53	31/53	18/53
CBSEAS (v0.4a)	0.0919	0.1305	0.28	4.67	4.18
Classement (v0.5)	11/53	24/53	20/53	33/53	21/53
CBSEAS (v0.5)	0.0823	0.1230	0.24	4.64	4.11
Moins bon score	0.0273	0.0583	0.06	3.40	2.34
Meilleur score	0.1127	0.1452	0.34	5.78	4.99
Moyenne	0.0786	0.1168	0.226	4.751	3.922

Résumés initiaux

RunID	ROUGE-2	ROUGE-SU4	Pyramid	Linguistique	Perf. générale
Classement (v0.4a)	8/53	8/53	15/53	35/53	19/53
CBSEAS (v0.4a)	0.1027	0.1338	0.32	4.91	4.3
Classement (v0.5)	29/53	28/53	24/53	34/53	23/53
CBSEAS (v0.5)	0.0888	0.1263	0.27	4.61	4.25
Moins bon score	0.0282	0.0591	0.06	3.43	2.46
Meilleur score	0.1216	0.1510	0.38	5.93	5.16
Moyenne	0.0853	0.1214	0.252	4.762	4.075

Résumés de mise à jour

RunID	ROUGE-2	ROUGE-SU4	Pyramid	Linguistique	Perf. générale
Classement (v0.4a)	8/53	15/53	10/53	24/53	22/53
CBSEAS (v0.4a)	0.0811	0.1223	0.26	4.75	3.98
Classement (v0.5)	29/53	18/53	21/53	27/53	23/53
CBSEAS (v0.5)	0.0767	0.1198	0.21	4.66	4.00
Moins bon score	0.0264	0.0576	0.05	3.36	2.23
Meilleur score	0.1039	0.1395	0.31	5.89	5.02
Moyenne	0.0719	0.1122	0.198	4.742	3.769

TABLE 5.4.: Résultats de la tâche « Update » : classement des systèmes CBSEAS

qui génèrent des résumés supérieurs à 100 mots, mais coupe tous les résumés à 100 mots. Laisser entiers des résumés dépassant les 100 mots aurait donc statistiquement permis à notre système d'intégrer plus d'éléments pertinents. Cela aurait permis à CBSEAS de se placer au niveau de la majorité des systèmes de résumé de l'état de l'art. C'est la raison pour laquelle nous avons choisi de persévérer dans notre approche, à savoir l'identification de la redondance pour la sélection des phrases à extraire.

Cela s'est révélé pertinent : les différents résultats obtenus depuis TAC 2008 montrent que notre système se comporte bien car la prise en compte d'éléments linguistiques sup-

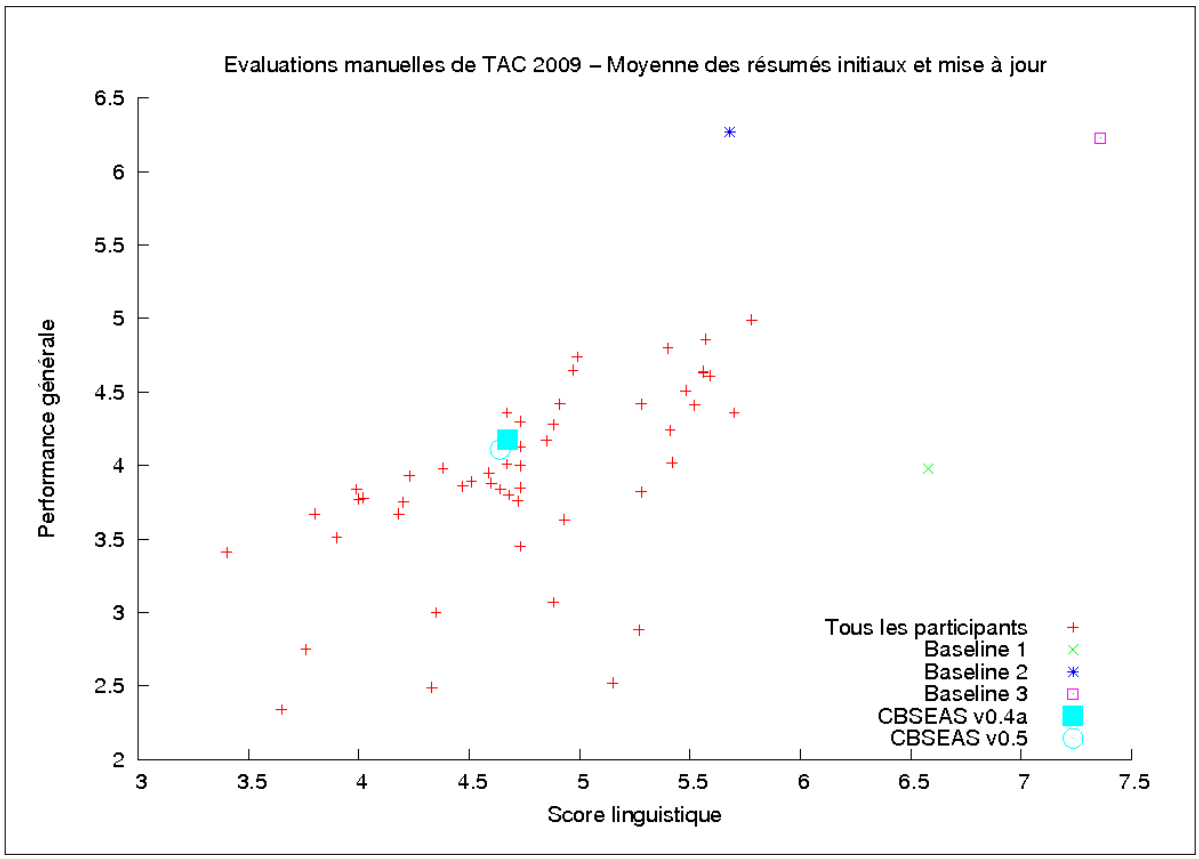


FIGURE 5.7.: Evaluations manuelles de TAC 2009 « Update Task » : moyenne des résumés initiaux et mise à jour

plémentaires améliore la qualité des résumés. Cependant, l'intégration à CBSEAS de la résolution d'anaphores a fait chuter légèrement la qualité des résumés. Cela est dû au mauvais comportement du système de résolution d'anaphores et de co-références utilisé, qui a une précision trop faible pour être intégré au calcul de similarité entre phrases sans en infléchir la justesse. A titre d'exemple, nous avons calculé que sur le jeu de données D0801 de TAC 2008, cet outil a fourni une annotation des entités nommées d'une précision de 83%, et une résolution d'anaphores d'une précision de 72%. Il conviendra de tester ultérieurement l'intégration d'outils d'annotation en entités nommées et de résolution d'anaphore plus performants et d'évaluer leur apport à CBSEAS. La quantité d'informations pertinentes véhiculée par les résumés générés par CBSEAS pour TAC 2009 est satisfaisante (83% des informations pouvant être extraites par un système purement extractif), ce qui démontre la pertinence de notre approche, qui est de surcroît pénalisée par l'absence de module poussé de mesure de la pertinence d'une phrase vis-à-vis de la requête. La mesure de centralité locale est essentielle à notre approche, comme le prouve le changement du *tf.idf* par le *tf.icf*, qui a conduit à une amélioration des résultats des évaluations automatiques de TAC 2009.

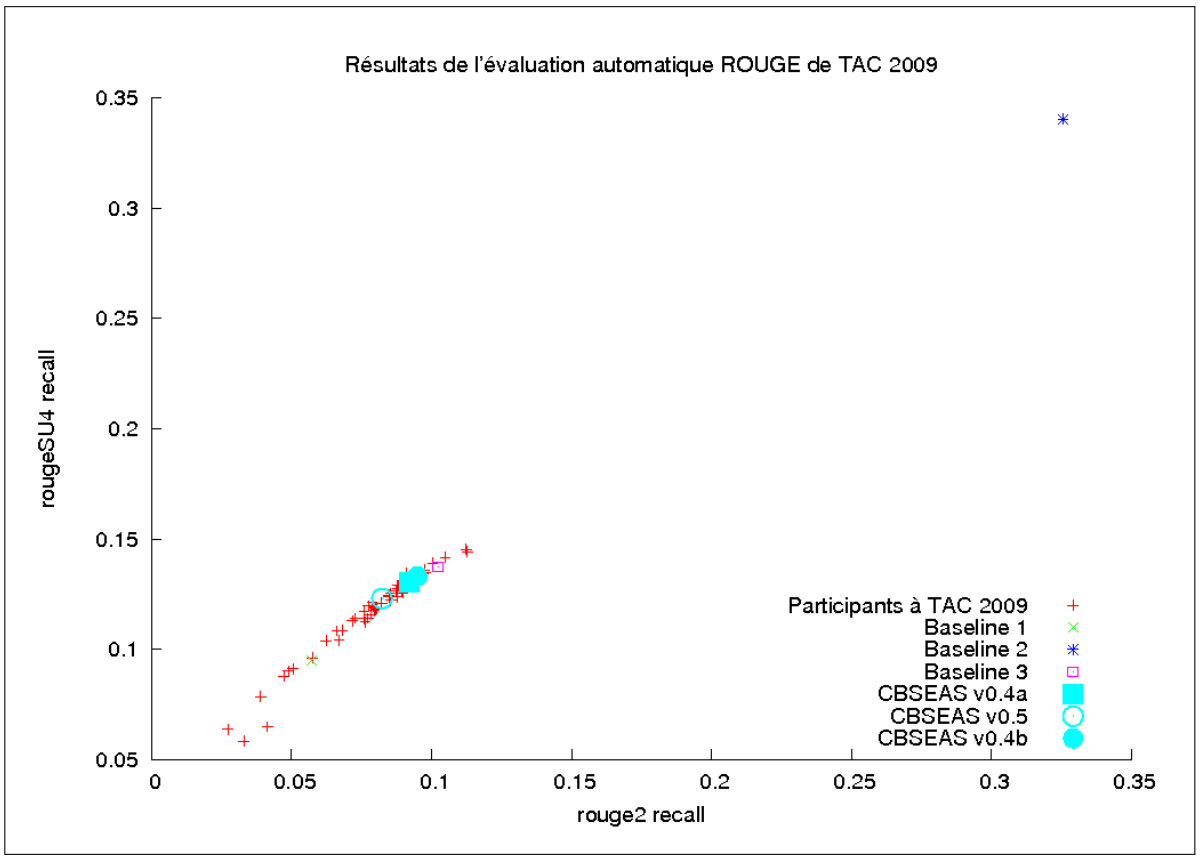


FIGURE 5.8.: Evaluations automatiques de TAC 2009 « Update Task » : moyenne des résumés initiaux et mise à jour. On peut constater la nette évolution de la qualité des résultats de CBSEAS depuis TAC 2008, qui se rapproche désormais de la baseline 3. Les scores de la version 0.4b ont été calculés *a posteriori*.

CBSEAS obtient toutefois de mauvais scores en qualité linguistique lors de la campagne TAC 2009. La note en qualité linguistique est difficile à analyser car elle est dépendante de plusieurs paramètres différents (*cf* §5.5.1) dont les détails ne sont pas fournis dans les résultats. Nous avons cependant identifié deux causes probables à ces mauvais scores : premièrement, les post-traitements ne sont pas assez poussés et beaucoup d'éléments qui défavorisent une lecture rapide sont conservés dans les résumés post-traités(*cf fig.* 5.5). Ensuite, nous avons laissé NIST couper les résumés à 100 mots. Nous n'avons pas coupé nous-même les résumés à la fin de la dernière phrase n'excédant pas 100 mots car la dernière phrase, même incomplète, peut contenir des éléments d'information utiles au résumé. Cependant, il apparaît dans les résultats de NIST que les résumés dont la dernière phrase est incomplète obtiennent des notes linguistiques inférieures à la moyenne, tandis que les résumés dont la dernière phrase est complète obtiennent dans une majorité (62%) des notes supérieures à la moyenne. Ceci peut être

vérifié sur la figure 5.9. Parmi les systèmes qui en constituent les 4 contre-exemples les plus flagrants et dont des résumés issus des résultats de TAC 2009 sont présentés dans les tableaux B.1, B.2, B.3 et B.4, deux sont les systèmes les plus mal classés tous scores confondus (systèmes 5 et 28), et les deux autres (systèmes 9 et 26) semblent utiliser des techniques de compression de phrases qui en dégradent la grammaticalité.

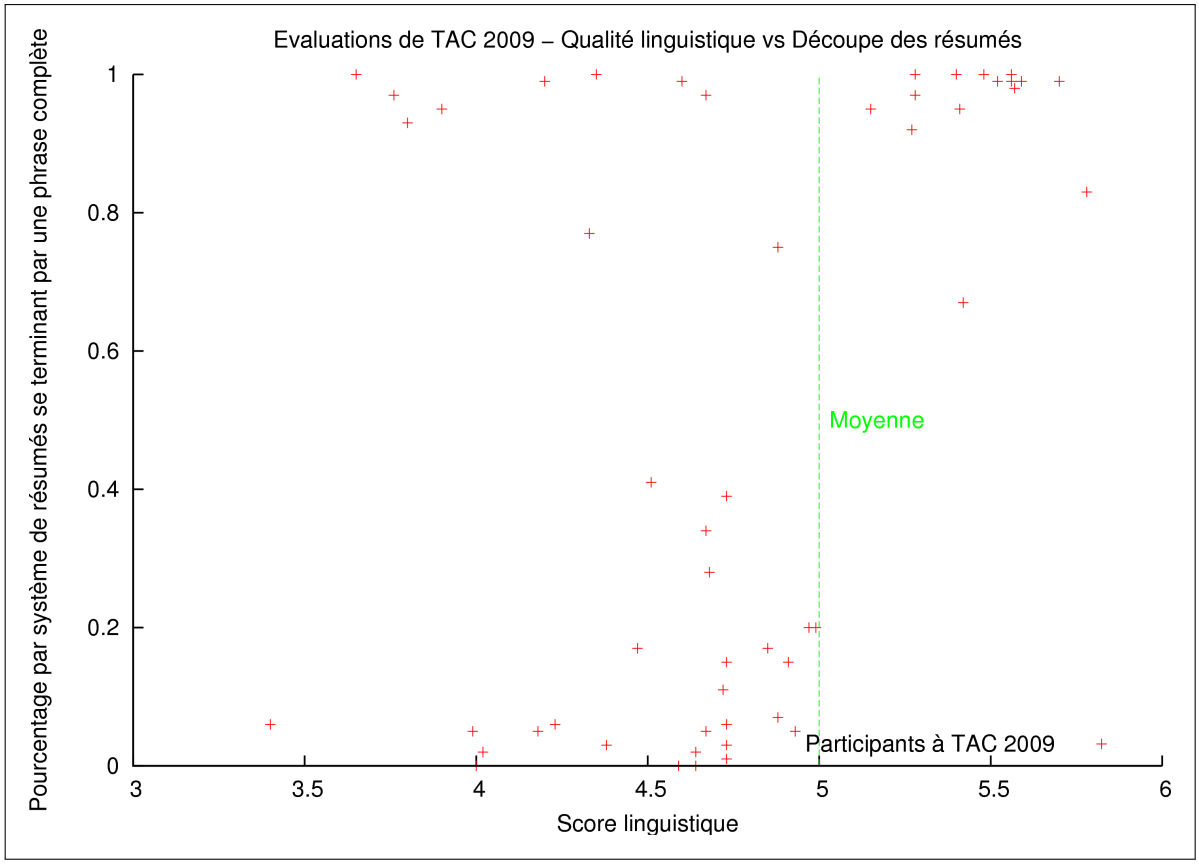


FIGURE 5.9.: Évaluations manuelles de TAC 2009 « Update Task » : Influence des stratégies de découpe des résumés sur la note de qualité linguistique. Tous les systèmes (mis à part un) avec un faible taux de dernière phrase complète ont une note linguistique en dessous de la moyenne.

Les mauvaises notes de notre système en qualité linguistique ont légèrement affecté les notes en performance générale. En effet, la note en performance générale semble être liée à la note de qualité linguistique (*cf fig. 5.7*). Notre stratégie consistant à ne pas supprimer la dernière phrase des résumés a donc certainement influencé la note de performance générale. Cependant, il est difficile d'analyser précisément quel en a été l'impact. En effet, contrairement aux résultats de la tâche « Résumés d'Opinion », les résultats de la tâche « Résumé Mise à jour » en qualité linguistique ne sont pas détaillés. La qualité linguistique dépend de plusieurs paramètres : la structure/cohérence, la grammaticalité, la clarté référentielle, l'absence de redondance, et le *focus*. Or ce dernier paramètre

D0902	In Texas, the debate has formed around a bill under whose original language pharmacists could refuse to fill prescriptions for emergency contraception – **the so-called morning-after pill**. Also on Wednesday, a nurse in Illinois filed a lawsuit against Eastern Illinois University, where Ottenhoff used to work, claiming Ottenhoff was denied a promotion because Ottenhoff objected to dispensing the morning-after pill. Some pharmacists around the country are refusing to fill prescriptions for birth control and morning-after pills, saying that dispensing the medications violates their personal moral or religious beliefs. The FDA's consideration of the morning-after pill's application was delayed because of
D0922	The Patriot Act, passed by Congress a few weeks after the Sept. 11, 2001, terror attacks, gave federal law enforcement officials broader powers of surveillance and prosecution against suspected terrorists, their financiers and their sympathizers. **Washington** – An unusual left-right coalition opened a campaign Tuesday to sharply curtail controversial provisions of the USA Patriot Act, showing that Congress and President Bush face a pointed debate over renewing the law enacted just 45 days after the Sept. 11, 2001, terrorist attacks. The Patriot Act, passed in the wake of the Sept. 11 terrorist attacks, bolstered FBI surveillance and law-enforcement powers in
D0907	John Yoo, a senior Justice Department lawyer who wrote much of the memorandum, exchanged draft language with lawyers at the White House. **The New York Times said in an editorial for Saturday, Feb. 19 :** Of all the claims of an electoral mandate made by President Bush's supporters, none were as bizarre as the one offered by John Yoo, a former Justice Department lawyer who helped draft the cynical justifications for the illegal detention and torture of unlawful combatants.

TABLE 5.5.: Exemples d'éléments de texte (en gras) repérés dans les résumés générés par CBSEAS et qui en perturbent la lecture

est déjà évalué par l'évaluation Pyramide et par l'évaluation de performance générale. Un résumé dont la réponse à la question de l'utilisateur n'est pas assez précise sera donc pénalisé dans trois mesures différentes. Le fait que certains critères d'évaluation interviennent dans plusieurs mesures qui devraient être décorrélées n'est pas nouveau, puisque cela avait déjà été noté par Nenkova *et al.* (2003). De plus, le fait que la précision du résumé entre en jeu dans la note de qualité linguistique rend cette évaluation plus opaque, d'autant que le détail des notes qui la composent n'est pas intégré aux résultats fournis par le NIST. Il est donc difficile de juger grâce aux évaluations manuelles de TAC 2008 et 2009 la lisibilité des résumés générés par notre système.

Les deux versions de CBSEAS soumises officiellement à TAC 2009 (v0.4a et v0.5) obtiennent de meilleurs résultats avec les évaluations ROUGE (*cf fig. 5.8*) qu'avec les évaluations manuelles (*cf fig. 5.6*). L'algorithme génétique utilisé pour paramétrer au mieux le système, et qui utilise les évaluations ROUGE comme fonction de *fitness*, a permis d'améliorer considérablement les scores des résumés. L'amélioration est plus sensible sur des types de mesures d'évaluation similaires aux mesures utilisées pour l'optimisation que sur les évaluations manuelles, celles-ci témoignent tout de même que la qualité des résumé, et non pas seulement les scores automatiques a également bénéficié

de l'optimisation.

Les résultats de la baseline 3, composée uniquement de phrases extraites des documents à résumer, montrent que les meilleurs systèmes de résumé par extraction atteignent aujourd'hui de très bons résultats du point de vue des informations qu'ils véhiculent. Cependant, ceux-ci ne sont pas encore assez performants du point de vue de la facilité de lecture. Malgré les bons résultats obtenus par CBSEAS, le système est encore loin d'égaler les performances linguistiques de cette baseline. Une des façons d'améliorer ses résultats serait d'exploiter pleinement notre système d'étiquetage des dépêches et d'analyse de la structure. Les requêtes de TAC 2009 sont de types multiples (*cf fig. 5.6*) et appellent des réponses de types différents. Par exemple, la requête D0942 de le tableau 5.6 appelle un résumé chronologique de la mise en place d'une loi anti-sécession et des réactions à son sujet. Les dépêches de type chronologie et revue d'opinions sont des candidates idéales pour l'extraction de phrases qui répondent à cette requête. Paramétrer notre système de résumé de manière à sélectionner prioritairement les meilleures sources pour chaque type de requête pourrait permettre de nettes améliorations à la qualité des réponses apportées à l'utilisateur par CBSEAS. Des techniques linguistiques permettant de déterminer la pertinence de l'intégration d'une phrase à un résumé pourrait permettre de combler en partie le fossé qui sépare CBSEAS (et les systèmes de résumé automatique en général) des rédacteurs humains.

D0907	Résumé descriptif	Describe the views and activities of John C. Yoo.
D0920	Résumé informatif	Provide information about Cat Stevens, aka Yusuf Islam.
D0935	Résumé chronologique	Describe the establishment of the MLB team in Washington and its initial impact.
D0939	Résumé événementiel (causes, conséquences)	Describe the Glendale train crash, the cause of the crash, and the arrest and trial of the man accused of causing it.
D0942	Résumé chronologique et résumé d'opinions	Report the developments in the PRC's enactment of an anti-secession law and reactions to that law.

TABLE 5.6.: Illustration des différents types de requêtes de la tâche « Update de TAC 2009

5.6. Conclusion

Dans ce chapitre, nous avons présenté notre approche pour résoudre une tâche de résumé et de mise à jour de résumé. Cette approche se fonde sur la flexibilité de l'algorithme de regroupement des phrases en classes sémantiques afin de résoudre la problématique de la détection de nouveauté et de la génération de la mise à jour d'un résumé selon le même principe que la génération du résumé initial.

Afin de valider notre approche, nous l'avons soumise aux campagnes d'évaluation TAC 2008 et 2009. Les mauvais résultats obtenus par notre premier système lors de TAC 2008 sont le fait d'erreurs de paramétrage, qui ont été résolus lors d'un test *a posteriori* sur les mêmes données. Les résultats obtenus ont alors placé CBSEAS dans la première moitié du classement. Lors de la participation à la campagne TAC 2009, nous avons testé un nouveau calcul de score des phrases fondé sur leur position dans la structure des documents et leur appartenance à certains types de dépêches, identifiés automatiquement ainsi que l'impact d'un algorithme d'optimisation des paramètres de CBSEAS. Le système s'est alors classé dans le premier tiers ou le premier cinquième des systèmes qui ont participé à TAC 2009, selon l'évaluation observée.

Cependant, l'implémentation de certaines techniques de TAL a montré qu'il est nécessaire que de telles techniques soient également soumises à des évaluations applicatives dans lesquelles sont testées non pas les performances pures d'une technique en termes de précision/rappel, mais leur impact sur une application de TAL qui l'intègre dans ses traitements.

6. Application de CBSEAS au résumé d'opinions

Sommaire

6.1. Introduction : de la veille d'information à la veille d'opinions

Les premières applications du résumé automatique ont concerné les dépêches de presse et les articles scientifiques(Mani et Maybury, 1999). Ces applications ont vite trouvé leur place dans l'industrie pour aider aux tâches de veille : veille d'information, veille scientifique – toutes composantes de la veille stratégique.

Rapidement, les industriels ont pris conscience des avantages considérables qu'ils pouvaient tirer des nouvelles ressources électroniques telles que les blogs ou les forums. De telles ressources ouvertes, correctement exploitées, peuvent aider à comprendre les besoins et les attentes de toute une population, et également à analyser les opinions concernant des produits, des personnalités ou même des propositions politiques. L'analyse d'opinion a ainsi connu un essor scientifique considérable ces dernières années.

Cependant, réaliser une analyse d'opinion à partir d'une masse documentaire impor-

tante pose le problème de la lisibilité des résultats. Il est donc nécessaire de les condenser afin de les rendre accessibles. C'est pourquoi la campagne d'évaluation TAC 2008 s'est intéressée au résumé d'opinions au travers de la tâche « pilote » [1] *Opinion Summarization.*

Participer à une campagne d'évaluation automatique telle que TAC 2008, sur une tâche novatrice adaptée à des nouveaux besoins, a permis de situer notre système par rapport aux autres systèmes existants, et surtout de le confronter à des données réelles.

6.2. La tâche « Résumé d'opinions » de TAC 2008

La tâche « Résumé d'opinions » de TAC 2008 est inspirée d'un scenario applicatif réaliste, dans lequel un utilisateur souhaite connaître les opinions exprimées à propos d'une entité clairement identifiée (une personne, un produit, une organisation, un fait de société...). Les dépêches de presse contiennent parfois ce type d'informations. Cependant, celles-ci sont des sources secondaires formulées par des journalistes. Les opinions contenues dans des dépêches de presse sont donc des informations secondaires. Des ressources telles que les blogs permettent à l'utilisateur de disposer d'opinions non déformées. C'est pourquoi TAC 2008 a proposé de réaliser des résumés d'opinion à partir de blogs.

Dans ce scenario, l'utilisateur cherche des réponses à plusieurs questions spécifiques sur une entité donnée ou sur ses caractéristiques. Nous appellerons un jeu de questions à propos d'une même entité un « sujet ». L'utilisateur souhaite comme résultat final un résumé des réponses trouvées dans les blogs à ces questions, où les informations redondantes ne sont pas répétées, mais éliminées . Ce résumé doit être bien organisé et sa lecture aisée. Le nombre de caractères en sortie ne doit pas dépasser 7000 fois le nombre de questions. Par exemple, pour un sujet qui comporte comporte trois questions, le résumé peut comprendre jusqu'à 21000 caractères [2].

L'objectif de cette tâche est de fournir un résumé des réponses à des questions « orientées » vers une polarité d'opinion. Par exemple, un jeu de questions issu de TAC 2008 était :

- Entité : *Rudy Giuliani presidential chances*
    1. *What did American voters admire about Rudy Giuliani ? (+)*
    2. *What qualities did not endear Rudy Giuliani ? (-)*

1. Les tâches pilote visent à explorer de nouvelles pistes applicatives qui sont reconduites si elles satisfont la communauté.

2. Cela paraît beaucoup, surtout en comparaison du nombre limite de mots (100) pour la tâche « Résumé et mise à jour » ; NIST n'a pas donné d'explication pour avoir fixé une limite aussi haute.

Seule une petite fraction des questions issues de la tâche « Résumé d'opinions » de TAC 2008 n'était pas orientée vers une opinion positive ou négative, ou difficilement identifiable comme telle, comme ces deux questions en pourquoi (*why*) :

– Criminalizing flag burning

1. *Why do supporters want to make flag burning a crime ?*
2. *Why do opposers not want to make flag burning a crime ?*

La tâche de résumé d'opinions consiste à coupler une analyse d'opinion avec un système de questions/réponses. Les sujets et les questions proviennent d'une tâche de question/-réponse qui lui est antérieure. Lors de TAC 2008, les sujets étaient au nombre de 25 et étaient constitués d'une à trois questions. Les systèmes doivent, pour chaque sujet, produire en sortie un texte qui résume les réponses à toutes les questions de ce sujet. Les documents associés à chaque sujet sont ceux dans lesquels au moins une réponse à une question a été trouvée par les assesseurs de NIST et les participants à la tâche de question/réponse qui a précédé la tâche de résumé d'opinions.

Afin de permettre aux participants qui ne maîtrisent pas les tâches de question/réponse, les éléments de réponse apportés par les systèmes de question/réponse étaient fournis par NIST. Ces éléments, appelés *snippets*, pouvaient être utilisés par les systèmes de résumé automatique ; cependant, leur utilisation devait être précisée lors de l'envoi des résultats à TAC, car elle permet d'améliorer de façon significative la qualité des résultats.

6.3. Adaptation de CBSEAS pour le résumé d'opinions à partir de blogs

Afin de participer à la tâche la plus réaliste possible, nous avons décidé de ne pas utiliser les *snippets* fournis en option par NIST. En effet, il paraît assez impropable qu'en condition réelle, un système puisse utiliser des requêtes utilisateur, des documents pertinents et des *snippets*. L'idée de coupler un moteur de recherche à un système de résumé automatique semble plus vraisemblable que de mettre à disposition d'un système de résumé automatique, un système de question/réponse.

Les seules adaptations à notre système CBSEAS, présenté en chapitre 3, concernent le pré-traitement des données en entrée –des blogs, donc des contenus de qualité linguistique très variable et souvent bruités par la publicité– et la gestion de l'opinion.

6.3.1. Pré-traitements

Afin de résumer au mieux les blogs, il faut tenir compte de leurs spécificités. Les contenus de blogs sont généralement « pollués » par la publicité (*cf fig.*??), mais aussi par les messages d'individus communément appelés *trolls*[3](*cf fig.* ??).

Afin de remédier à ces deux problèmes, nous avons choisi d'adopter une solution fondée sur le vocabulaire des phrases. Nous éliminons toute phrase dont le ratio :

nombre de mots fréquents / nombre total de mots

est inférieur à un seuil (0.35). Les mots fréquents sont les 100 mots les plus fréquents en anglais, qui consistuent approximativement la moitié des textes écrits(Fry *et al.*, 2000). Le seuil de 0.35, fixé empiriquement, permet une certaine souplesse quant à la sélection de phrases qui n'emploient pas un vocabulaire courant, tout en éliminant celles qui sont écrites dans un anglais plus qu'approximatif.

Une seconde phase de pré-sélection des phrases est alors appliquée, afin d'éviter au système d'avoir à traiter un trop grand nombre de phrases, et d'inclure des phrases non pertinentes dans les résumés. Après avoir éliminé les mots vides des questions concernant le sujet du résumé[4], le système pré-sélectionne les phrases qui contiennent au moins un mot issu des questions posées par l'utilisateur.

6.3.2. Analyse de la polarité des phrases

Nous avons tenté de donner à chaque phrase une polarité positive ou négative en utilisant une approche supervisée qui s'appuie sur un jeu de documents étiquetés, développée par Michel Généreux (Généreux, 2009b). Deux classifieurs SVM ont été entrainés sur ce jeu de documents, l'un pour catégoriser les requêtes, et l'autre pour catégoriser les phrases issues des posts des blogs à résumer.

L'idée sous-jacente à la détection de la polarité d'opinion des requêtes est de pouvoir orienter au maximum le résumé vers les attentes de l'utilisateur. Ainsi, si un utilisateur pose la question « *What did American voters admire about Rudy Giuliani ?* », il faudra détecter qu'il ne demande qu'un résumé des côtés positifs de *Rudy Giuliani*,

3. **troll** : désigne un internaute dont la conversation, généralement envahissante, n'est ni très pertinente ni très polie et engendre de nombreuses discussions inutiles. Il existe également un parallèle fort avec l'emploi à outrance du langage SMS, si ce n'est simplement d'une orthographe déplorable. *source : Wikipedia*

4. Nous avons utilisé la liste des mots vides pour l'anglais issue de http://www.textfixer.com/resources/common-english-words.txt.

et exclure du résumé toute phrase véhiculant une opinion négative sur *Rudy Giuliani*. Les requêtes adoptent des structures assez régulières et comportent souvent des motifs communs. Apprendre à catégoriser les requêtes peut donc se faire avec un nombre limité d'exemples. Les requêtes des campagnes précédentes constituent donc un bon échantillon d'apprentissage. Le classifieur, entraîné sur ces données, a été utilisé afin de catégoriser les requêtes de la tâche « résumé d'opinions » de TAC 2008.

La classification des phrases issues des posts de blogs constitue un tout autre défi. En effet, la variabilité de la longueur des phrases, de leur structure ainsi que du vocabulaire utilisé sont autant d'obstacles à leur classification. De plus, de nombreuses phrases n'ont pas de polarité car elles sont purement objectives, ou n'utilisent pas de vocabulaire spécifique. Pour cette raison, nous avons adopté une stratégie plus simple mais potentiellement moins précise, qui consiste à classifier non pas les phrases seules, mais les posts de blogs. Ainsi, chaque phrase recevra la polarité du post dont elle est issue. Cette approche a été adoptée après avoir observé que les posts de blogs étaient rarement nuancés. Ils véhiculent généralement des opinions quasi-uniformément positives ou négatives, ce qui nous a poussé à adopter cette stratégie : répercuter la tendance générale d'un post sur chaque phrase particulière.

Le classifieur de documents a été entraîné sur un corpus de critiques de films [5]. Ce corpus est certes éloigné de notre tâche — il est dédié aux critiques de film tandis que notre tâche comprend des sujets plus variés—, mais il présente la particularité d'être entièrement annoté en « opinions » : seules les critiques pour lesquelles les auteurs avaient fourni une notation explicite du film (sous forme d'étoiles) ont été sélectionnées. Ces notations ont été converties automatiquement par les auteurs du corpus en « polarité d'opinion » : positives, négatives ou neutres (Pang *et al.*, 2002).

6.3.3. Résumé automatique des blogs

Nous avons utilisé notre système CBSEAS, présenté en chapitre 3, afin de réaliser les résumés de blogs. Ce système comporte deux étapes : la première consiste à regrouper dans des mêmes classes les phrases qui partagent de fortes similarités afin de produire un modèle des documents qui rend compte de la diversité informationnelle que ces documents contiennent. Dans une seconde étape, une phrase est extraite de chaque classe selon le principe de la centralité locale, défini en 3.2.5.

Cette approche se justifie d'autant plus ici que les blogs comportent une caractéristique intéressante. Les posts de blogs, tout comme les posts de forums, comportent des citations (*quote*), très souvent utilisées par les bloggers afin d'approuver ou de réfuter tout ou partie d'un post antérieur. L'étude de telles citations permet d'identifier quelles

5. http://www.cs.cornell.edu/People/pabo/movie-review-data

sont les idées majeures, ou qui portent le plus à controverse. Le fait qu'une même classe –identifiée dans la première étape de CBSEAS– contienne plusieurs phrases identiques issues de posts différents favorise ces dernières lors de la mesure de la centralité locale. La centralité locale reflète alors le lien entre l'importance d'une phrase et le fait qu'elle soit citée dans plusieurs posts différents.

Tout résumé de la tâche « Résumé d'opinion » est limité à un maximum de 7000 caractères multipliés par le nombre de questions de son sujet. Contrairement à la grande majorité des systèmes de résumés, qui permettent de sélectionner les phrases de manière itérative jusqu'à avoir atteint la taille maximum (Kupiec *et al.*, 1995; Radev *et al.*, 2001a; Boudin et Torres-Moreno, 2007; Carbonell et Goldstein, 1998), notre algorithme ne peut pas sélectionner itérativement des phrases jusqu'à ce que les résumés qu'il génère dépassent le nombre maximum de caractères. En effet, les classes sont entièrement recalculées entre la i^{ème} et la i+1ème itération. Nous n'avons donc aucune garantie que les phrases qui auraient été sélectionnées à la i^{ème} itération seront conservées à l'itération suivante. La stratégie la plus sûre consiste alors à fixer le nombre de phrases à extraire avant le début du calcul du résumé. Nous avons donc choisi une stratégie fondée sur la longueur moyenne des phrases. Le système calcule la longueur moyenne des phrases des blogs à résumer, puis extrait un nombre de phrases égal à :

nombre maximum de caractères / nombre moyen de caractères par phrase (nous l'appellerons nombre de phrases idéal).

Une autre approche aurait pu être considérée : sélectionner comme résumé le résultat issu de l'itération de l'algorithme qui contient le nombre de caractères le plus proche de la limite imposée. Cependant, pour être efficace, notre système doit disposer d'assez de phrases en entrée pour pouvoir les regrouper correctement en n classes, et chaque classe doit être assez peuplée pour mesurer correctement la centralité locale. Le nombre de phrases à extraire ne peut donc pas être laissé au hasard et doit être fonction du nombre de phrases pré-sélectionnées. Nous avons décidé empiriquement que la création de n classes pour n^2 phrases permettrait de conserver une efficacité élevée. Cependant, en partant sur l'hypothèse optimiste que la longueur moyenne des phrases est de 80 caractères, il faudrait plus de 7500 phrases en entrée pour réaliser un résumé de 7000 caractères. Nous avons donc baissé ce seuil de qualité à n classes pour $n^{1.4}$ phrases, et le système doit choisir entre ces deux options :

- Les phrases pré-sélectionnées sont au moins au nombre de $(nombre\_de\_phrases\_ideal)^{1,4}$: le résumé est lancé pour un nombre de phrases en sortie égal au nombre de phrases idéal ;
- Le nombre de phrases pré-sélectionnées est inférieur à $(nombre\_de\_phrases\_ideal)^{1,4}$: le résumé est alors lancé pour un nombre de phrases en sortie égal à $\sqrt[1,4]{nombre\_de\_phrases\_pre\_selectionnees}$.

A la date à laquelle nous avons participé à TAC 2008, le système ne disposait pas de

calcul de similarité à la requête, mais appliquait un filtre sur les phrases en entrée (*cf §6.3.1*). Il en résulte une différence majeure par rapport au système CBSEAS présenté en chapitre 3, puisque le calcul final qui permettait de sélectionner les phrases n'incluait pas de score de centralité globale, mais seulement la longueur des phrases et la centralité locale (*cf §3.2.5*).

6.3.4. Réordonnancement

L'algorithme de réordonnancement présenté en §3.2.6 n'est pas ici d'une grande utilité. En effet, les blogs présentés sont rarement composés de discours argumentés. De ce fait, il n'est pas possible de tirer de l'ordre des phrases dans les blogs, des informations concernant l'ordonnancement logique des phrases dans les résumés générés. Nous avons privilégié un réordonnancement par rapport à la requête : les phrases qui comportent le plus d'éléments de la requête sont placées en premier. Ce type de réordonnancement ne suffit pas à organiser les résultats sous la forme d'un résumé d'opinions ; nous lui avons donc adjoint un module de réordonnancement dépendant de la « polarité » des phrases.

Après avoir calculé l'ordre des phrases dans une première phase, les phrases sont séparées en trois groupes distincts : phrases négatives, positives et neutres. Les trois groupes conservent l'ordre des phrases qui a été calculé dans la première phase. Les trois groupes sont ensuite présentés à l'utilisateur selon l'ordre de ses questions : le premier groupe présenté est celui qui partage la même polarité sentimentale que la première question, le second groupe celui qui partage la polarité opposée, et enfin le dernier groupe est celui qui comprend les phrases neutres. Dans le cas où l'utilisateur a posé des questions impliquant des réponses de même polarité, seul le groupe des phrases qui partagent cette polarité est sélectionné.

Etant donné que notre système fonctionne avec un nombre de phrases à extraire, et que les sorties ne doivent pas excéder un certain nombre de caractères, nous extrayons autant de phrases que possible (ceci dépend du nombre de phrases en entrée, *cf §6.3.3*). Nous avons pris cette décision suite à la publication des résultats de TAC 2008, pour lesquels nous avons été pénalisés par la production de résumés trop courts. Présenter les phrases neutres en dernier permet alors de ne pas éliminer de phrases véhiculant des opinions fortes lors du tronquage par NIST des résumés à la taille désirée, tout en évitant de supprimer du résumé des phrases qui pourraient avoir été classées neutres par erreur.

6.4. Évaluation

6.4.1. Protocole

Evaluer de résumés d'opinion est une tâche plus compliquée que d'évaluer des résumés « standard ». En effet, le juge, qu'il soit humain ou artificiel, doit noter non seulement la pertinence d'un résumé vis-à-vis d'une requête, mais également si le résumé est bien constitué de phrases qui véhiculent une opinion conforme à ce qu'a demandé l'utilisateur. Si l'évaluation automatique de la pertinence d'un résumé commence à être maîtrisée et fournit des résultats de plus en plus corrélés aux résultats manuels (*cf §2*), repérer automatiquement si les opinions véhiculées par un résumé sont conformes n'est pas réalisable en l'état actuel des travaux.

NIST a donc choisi de ne pas procéder à une évaluation automatique mais uniquement à une évaluation manuelle des résumés d'opinion. Pour ce faire, des évaluations manuelles guidées par une grille d'évaluation ainsi que la méthode *Pyramide*(Nenkova *et al.*, 2007), présentée en §2.4, ont été utilisées.

Les résumés générés devaient répondre à toutes les questions d'un sujet donné. Par conséquent, les pyramides (constituées des éléments de réponses pertinents) ont été établies en tenant compte de la totalité des questions de chaque sujet. Cependant, pour chaque phrase d'un résumé, le lecteur doit être capable de déterminer soit d'après la phrase elle-même soit d'après son contexte, à quelle question elle répond. Les éléments du résumé n'ont donc été associés à un élément de réponse de la pyramide (*nugget*) que si l'assesseur a pu déterminer à quelle(s) question(s) ces éléments répondaient.

L'instantiation des *nuggets* pour chaque pyramide a été réalisée de la manière suivante : leur poids a été calculé à partir des jugements de 10 assesseurs, qui les ont notés en tant qu'« essentiels » (*vital*) ou « acceptables » (*okay*) vis-à-vis de la question posée. Le poids d'un *nugget* a alors été calculé comme le nombre d'assesseurs qui l'ont marqué comme « essentiel » normalisé par le nombre de jugements « essentiels » reçu par le meilleur *nugget* (le plus essentiel) du sujet à résumer.

Une autre évaluation correspondant à une grille de lecture et à une notation plus subjective a été réalisée. Les assesseurs ont noté de 1 à 10 les résumés selon les 5 critères suivants :

- Grammaticalité : Le résumé ne doit pas comporter de phrases non-grammaticales (fragments de phrases, composants manquants) qui rendent le texte difficile à lire.
- Non-redondance : les résumés ne doivent pas inclure de répétitions inutiles. Cela inclut les phrases entières qui sont répétées aussi bien que des faits relatés plusieurs fois.

– Structure et cohérence : les résumés doivent être bien structurés et organisés. Un résumé ne doit pas être seulement un amoncellement d'informations, mais doit être construit afin de constituer un corps d'informations cohérent. Si une personne ou une entité est mentionnée, leur relation au reste du résumé doit être claire. Il doit être aisé d'identifier à qui ou à quoi les pronoms et les autres éléments non auto-référentiels font référence.
– Score total de lisibilité.
– *Overall responsiveness* : les assesseurs ont attribué un score à chaque résumé, fonction de la réponse du résumé au besoin d'information exprimé par la liste de questions à propos du sujet. Ce score mesure la qualité et l'utilité globales du résumé, considérant le contenu informatif et la lisibilité. De façon plus concrète, les assesseurs devaient se demander : « Si je me posais ces questions à propos de ce sujet, combien est-ce que je serais prêt à payer pour ce résumé des réponses à ces questions ? »(Dang, 2008)

6.4.2. Résultats

Dans cette section, nous présentons nos résultats pour la tâche « Opinion Summarization » de TAC 2008. Notre système n'utilisant pas les « snippets », ressource optionnelle fournie par NIST, nous nous concentrons ici sur les systèmes qui n'ont pas utilisé cette ressource (nous situerons tout de même nos résultats par rapport à ceux obtenus par les systèmes qui ont utilisé cette ressource).

	Pyramide	Grammat.	Non-Red.	Structure	Lisibilité	Overall Resp.
Classement	5/19	3/19	3/19	2/19	2/19	10/19
Score de CBSEAS	0.169	5.955	6.636	3.500	4.455	2.636
Moins bon score	0.101	3.545	4.364	2.045	2.636	1.682
Meilleur score	0.251	7.545	7.909	3.591	5.318	3.909

TABLE 6.1.: Résultats de TAC 2008 « Opinion Summarization » : Classement du système CBSEAS (systèmes sans *snippets*)

RunID	Pyramide	Grammat.	Non-Red.	Structure	Lisibilité	Overall Resp.
Classement	15/34	5/34	6/34	2/34	4/34	22/34
CBSEAS	0.169	5.955	6.636	3.500	4.455	2.636
Moins bon score	0.101	3.545	4.364	2.045	2.636	1.682
Meilleur score	0.489	7.545	8.045	3.591	5.318	5.773

TABLE 6.2.: Résultats de TAC 2008 « Opinion Summarization » : Classement du système CBSEAS (tous systèmes confondus)

Le tableau 6.4 donne des exemples des résumés que notre système produit, en montrant les deux meilleurs résumés ainsi que l'antépénultième d'après l'évaluation *Pyramid* (les résumés classés dernier et avant-dernier sont trop illisibles pour figurer ici). Les résultats

	Pyramide	Grammat.	Non-Red.	Structure	Lisibilité	Overall Resp.
Moy. sans snippets	0.151	5.14	5.88	2.68	3.43	2.61
CBSEAS	0.169	5.955	6.636	3.500	4.455	2.636
Moy. avec snippets	0.312	5.45	5.91	2.87	3.88	4.04

TABLE 6.3.: Résultats de TAC 2008 « Opinion Summarization » : Comparaison des systèmes avec et sans *snippets*

présentés dans les tableaux 6.1 et A.1 montrent le bon comportement de notre système de résumé vis-à-vis de cette tâche de résumé d'opinion. Le système se classe en effet cinquième sur 19 systèmes d'après l'évaluation *pyramide*, et entre la deuxième et la troisième place sur les évaluations de lisibilité. Il est à noter le comportement compétitif du système CBSEAS en ce qui concerne la structuration du résumé : le système est classé second, à seulement 2.5% du meilleur système. Cela montre l'efficacité de la méthode de réordonnancement couplée au regroupement des phrases suivant leur polarité d'opinion. De plus, le système reste très bien classé face aux système qui utilisent les *snippets* (*cf tab.6.2*), malgré l'avantage certain que cela leur a octroyé (*cf tab.6.3*).

Le score *pyramide* des résumés est inversement proportionnel au nombre de caractères (*cf fig. 6.4.2*). En effet, plus un résumé est long, plus la probabilité est grande qu'il intègre beaucoup d'informations pertinentes. Les résumés les plus longs sont ceux qui sont issus des jeux de documents dans lesquels un grand nombre de phrases ont été pré-sélectionnées pour le calcul du résumé. Ces résumés, d'une part sont trop longs pour être acceptables du point de vue de l'utilisateur, d'autre part contiennent beaucoup de phrases potentiellement non pertinentes (le filtre présenté en 6.3.1 est trop laxe). Cela se ressent de manière flagrante sur les résultats de l'évaluation *pyramide*, et a sûrement influencé négativement les évaluateurs lors de l'évaluation subjective. Donner un score fonction des requêtes à chaque phrase des documents en entrée et filtrer uniquement celles qui sont au-dessus d'un certain seuil pourrait pallier ce problème.

Il est difficile d'analyser et de comprendre les résultats de notre CBSEAS pour ce qui est de l'évaluation subjective « *overall responsiveness* » (performance générale). En effet, notre système se classe parmi les cinq meilleurs des systèmes qui n'utilisent pas les *snippets*, et ce quelle que soit l'évaluation considérée (score pyramide, grammaticalité, non-redondance, structure et lisibilité). Cependant, l'évaluation subjective place notre système au milieu du classement. Les différentes évaluations représentant bien ce que l'on attend d'un résumé (quantité d'information véhiculée, clarté et articulation du discours), nous nous attendions à figurer également parmi les meilleurs systèmes dans l'évaluation subjective. Cette dernière évaluation manque de clarté. Elle correspond, selon l'équipe du NIST chargée de TAC, à la réponse à la question « Combien serais-je prêt à payer pour ce résumé ? ». Il aurait été utile de demander aux évaluateurs des précisions sur les motivations de leurs notes. Ainsi, nous aurions tiré des conclusions claires concernant les faiblesses de notre système, voire de reconsidérer notre position sur ce que doit comporter un résumé automatique et la forme qu'il doit prendre. En l'absence de telles annotations,

1. What motivated negative opinions regarding purchasing a car from CARMAX ? 2. What motivated positive opinions of CARMAX from car buyers ?
With Carmax you will generally always pay more than from going to a good used car dealer. Carmax did split the bill which made me happy. We bought it at Carmax, and I continue to have nothing bad to say about that company. Not sure if you have a Carmax near you, but I ve had 2 good buying experiences from them. At Carmax, the price is the price and when you want a car you go get one. Have to say that carmax rocks.

1. Why do people like Mythbusters ?
I watched a couple of episodes of MythBusters which was really entertaining. Mom spent a lot of money on getting the first season of the mythbusters. So the mythbusters did this with the dummy and, man it was so funny. My wife and I are huge Mythbusters fans. amhorach : I love Mythbusters. Mythbusters takes urban myths and proves or disproves them by recreating them in painstaking detail, like Alton, and like him, they use lots of zippy diagrams and helpful interviews with anthropologists and whenever an anthropologist gets a working gig, a tribeman gets a loincloth.

1. What reasons were given for liking Amita ? 2. Why did people like Megan ?
very solid story that felt real and i say that because i know i 've seen stuff like this on shows like primetime and 20/ 20, and it finally gave navi rawat she plays amita a reason to be in the credited cast. don and david stumbled on to this scene in the basement of an la hotel. at the fbi headquarters, amita is the only person that the victim will speak to. turns out the language the girl was speaking in was a dialect common to the area where amita's parents were from. i really liked how this episode forced amita to look at who she was and where she came from. originally, i think the supposition on the battery was as a method of torture. come on, that's nuts. does anyone know when/ if numb3rs is coming out on dvd. how many shows can drag on and on. same scenario ... case opens, call in charlie, larry sneaks up on charlie either in the garage or at school, small mention/ scene with how many shows can drag on and on. she had been gone on a trip around several asian countries for the latter part of october, so we have n't really been e-mailing or text messaging. so she updated me on a lot of shows and we talked about our recent trips. even if i have to sail on a boat to get there. hopefully did n't do too bad on any of them. don comforting terry ... charlie and amita, passing pens michelle and i were probably the only ones who noticed that, as it was n't the focus of the scene and they were just off on the side ... according to my sister, the junior reshuffle is n't so bad. i know the education system there is alot tougher it would seem then ours from example but really it will be to your benefit later on.

Table 6.4.: Exemples de résumés générés par CBSEAS pour la tâche « Opinion Summarization » de TAC 2008

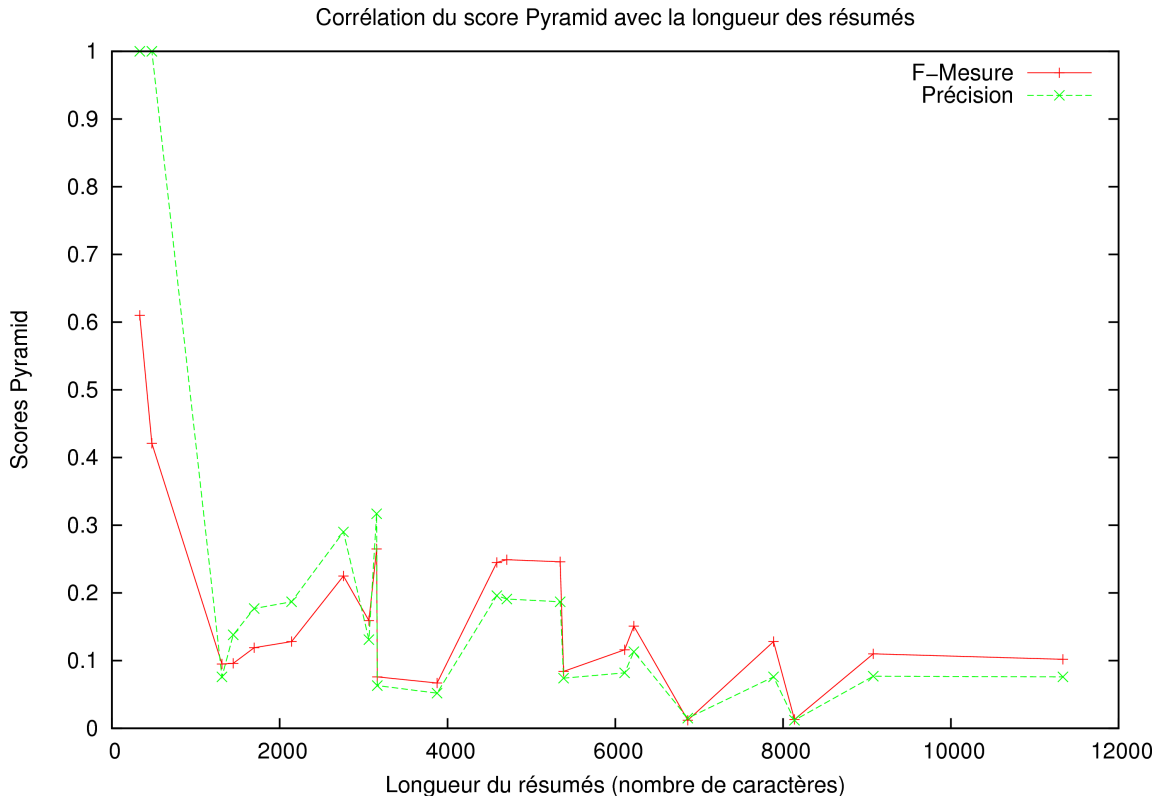


FIGURE 6.1.: Résultats de CBSEAS à TAC 2008 « Opinion Summarization » : Pyramid vs longueur des résumés. Le faible nombre de résumés évalués ne permet pas de calculer de corrélation, mais seulement de dégager une tendance : les résumés les plus longs sont les mieux classés en score pyramide.

il est difficile de considérer sérieusement cette mesure de qualité des résumés, d'autant plus qu'elle apparaît décorrélée de la mesure de qualité linguistique et fortement corrélée au score pyramide, notre système faisant exception à la règle (*cf tab 6.5*).

	Qualité linguistique	Score pyramide
Coefficient de Pearson	0.0322	0.8542

TABLE 6.5.: Corrélation de la mesure d'évaluation subjective aux autres mesures utilisées lors de la tâche « Résumé d'opinions » de la campagne TAC 2008. Elle n'est pas corrélée à la mesure de qualité linguistique, mais est fortement corrélée à la mesure pyramide.

6.5. Conclusion

Dans cette section, nous avons montré comment nous avons adapté notre système, CBSEAS, afin de participer à une campagne d'évaluation internationale sur le résumé automatique d'opinions. Nous avons couplé CBSEAS avec un système de détection d'opinion, ce qui permet de structurer le résumé de manière à exprimer au mieux les opinions contenues dans les textes.

Notre système a obtenu des résultats compétitifs à TAC 2008 en se classant dans les 5 premiers de sa catégorie dans l'évaluation manuelle de type Pyramide et dans les évaluations de lisibilité. Cependant, les performances de l'évaluation subjective ne sont pas aussi bonnes, mais nous ont permis d'identifier les faiblesses de notre système : l'absence de calcul de similarité à la requête et la trop grande taille des résumés générés.

Cependant, les résultats de TAC 2008 mettent aussi en avant la qualité de notre approche de réordonnancement, qui offre à l'utilisateur un texte plus organisé et une lecture plus fluide que ce que proposent la grande majorité des systèmes concurrents.

7. Application au résumé de documents hétérogènes

Sommaire

Réussir à s'approprier un corpus et en tirer les informations essentielles est une tâche impossible à réaliser pour un humain dès lors que les données à traiter sont trop importantes. Les « veilleurs » ont besoin d'aides automatiques afin d'explorer au mieux ces données. Dans cet esprit, de nouvelles perspectives de recherche ont vu le jour afin de faciliter l'accès à un contenu noyé dans un flot d'informations trop important. C'est notamment le cas des tâches de détection et de suivi d'événement (en anglais *topic detection and tracking – TDT*).

La détection et le suivi d'événement consistent à regrouper dans une même classe les documents qui traitent d'un même événement. A titre d'exemple, deux dépêches ayant pour titre « Arrivée en France de Laurent Gbagbo en vue d'une table ronde à Marcoussis » et « Ouverture des négociations entre rebelles et gouvernement ivoirien à Marcoussis » se rapportent à un même événement : « La Table ronde de Marcoussis ». La notion d'événement est cependant une notion vague, qu'il nous appartiendra de préciser dans ce chapitre.

Nous nous sommes inspirés des tâches de TDT afin de réaliser un système d'aide à l'analyse de corpus qui permet à un analyste de repérer les différents événements traités dans un corpus et d'obtenir un résumé de chacun d'eux. Ce besoin d'analyse de corpus

et de suivi d'événement a été exprimé dans le cadre du projet collaboratif Infomagic, et la partie de notre travail consistant à regrouper les documents en classes événementielles a été publiée dans « TALN 2008 » (*cf* C). Après avoir présenté un état de l'art des techniques de détection hors ligne d'événements – e.g. la détection d'événements dans des corpus dont les documents n'arrivent pas en temps réel, éludant le problème de l'incrémentalité –, nous essayons de mieux caractériser la notion d'événement avant de présenter notre système de détection automatique. Nous détaillons ensuite la façon dont les documents sont caractérisés, puis l'algorithme de classification. Enfin, nous présentons l'évaluation de notre système sur un corpus de dépêches AFP en langue française ainsi que des exemples d'utilisation.

7.1. État de l'art : regroupement de documents hors ligne

La détection d'événements permet de suivre en direct des flux de dépêches et de les classer en fonction du thème traité. Nous nous intéressons ici à la détection d'événements hors ligne. Ce thème a été moins traité que la détection en ligne –e.g. incrémentale– mais il est important, au moins dans deux cas de figure bien identifiés :

1. les analystes sont souvent confrontés à des masses de documents traitant de plusieurs thèmes. Avant d'accéder aux documents pertinents, une structuration du fonds documentaire est nécessaire.
2. les systèmes automatiques d'extraction d'information nécessitent des masses de documents homogènes en entrée. Il faut donc les structurer par thème ou par événement avant de passer à la phase d'extraction proprement dite.

Nous visons ces deux buts à la fois, l'objectif de notre application étant *in fine* de produire des synthèses sommaires à partir de masses de documents non structurés car la tâche s'apparente à du résumé multi-documents à partir d'un fonds documentaire non homogène en entrée. La visualisation des données permet en outre à l'analyste de contrôler le processus de regroupement de documents en ensembles pertinents. Nous ne nous intéressons ici qu'à l'étape de regroupement des documents.

Plusieurs auteurs ont décrit des systèmes liés à la détection d'événements hors ligne. Il s'agit de (Yang *et al.*, 1999; Hatzivassiloglou *et al.*, 2000; Zhiwei Li et Ma, 2005). Les premiers et les seconds utilisent des algorithmes de classification hiérarchique, tandis que les troisièmes utilisent des modèles probabilistes.

(Hatzivassiloglou *et al.*, 2000) se sont posé la question des données à utiliser pour la détection d'événements : vaut-il mieux utiliser la totalité des mots/phrases, exclure

des mots qui ne sont pas catégorisables comme termes uniques (e.g. le Palais, peut être du Luxembourg, de l'Elysée...), ou ne tenir compte que des noms propres ? Les auteurs arrivent à la conclusion que les jeux de données avec lesquels ils obtiennent les meilleurs résultats sont ceux prenant en compte tous les mots sans exception. Ils attribuent cela au fait que les outils d'extraction de termes ou de noms propres qu'ils utilisent ne sont pas assez robustes pour ce type de tâche.

(Yang *et al.*, 1999) proposent une méthode de regroupement de documents où chacun des documents est représenté par une liste de termes pondérés par leur *tf.idf* (*cf. fig. 7.2*). Sur le programme TDT, en utilisant un algorithme k-means multi-passes, les auteurs arrivent à des résultats de 61 % et 69 % en précision/rappel.

(Zhiwei Li et Ma, 2005) proposent quant à eux une approche probabiliste pour le regroupement de documents en utilisant comme représentation d'un document une matrice composée de quatre vecteurs : les noms de personnes, de lieux, les dates et des mots-clés. Leur modèle probabiliste appliqué à un extrait du corpus du programme TDT4 produit des résultats de l'ordre de 85 % de précision et 67 % de rappel, en fixant à la main le nombre de classes dans lesquelles ranger les documents. Sur des jeux de données ne séparant pas les entités nommées des mots-clefs, les résultats sont inférieurs de 10 %. Les auteurs l'expliquent par le fait que lorqu'elles ne sont pas distinguées des mots-clefs, les entités nommées se retrouvent noyées dans les données, alors que ce sont les éléments clés pour la construction d'un modèle d'événement.

Les expériences visant à regrouper dynamiquement un flux de documents *en ligne* utilisent globalement les mêmes méthodes, à l'instar de (Binsztok *et al.*, 2004) : il s'agit généralement d'approches probabilistes combinant des sacs de mots et une fenêtre temporelle associée à chaque groupe de documents.

Toutes les approches présentées ici, particulièrement (Hatzivassiloglou *et al.*, 2000), utilisent pour caractériser un document des vocabulaires assez étendus. La taille des données induite par ce type de caractérisation fait chuter les performances et la vitesse des systèmes de classification. Par ailleurs, il a été montré dans (Zhiwei Li et Ma, 2005) que la prise en compte de tout le vocabulaire est moins pertinente que la focalisation sur les seuls éléments clés, notamment les entités nommées. Celles-ci ont par ailleurs un rôle déterminant puisque les fondre dans la masse de données fait chuter les performances.

Enfin, par rapport à des approches comme (Zhiwei Li et Ma, 2005), nous souhaitons élaborer une méthode qui évite d'avoir à préciser à la main *a priori* le nombre de classes visées. Plusieurs solutions existent pour cela, dont l'utilisation de l'indice de Davies-Bouldin (Davies et Bouldin, 1979), une mesure permettant de quantifier la validité d'un clustering. Cet indice correspond au rapport des inerties inter et intra-classes. Parmi les autres méthodes de sélection du nombre de classes, on peut citer l'approche de (Hamerly et Feng, 2006), qui permet d'obtenir une meilleure mesure de la validité du clustering dans des cas difficiles, comme les grandes dimensions. Dans notre cas, l'indice de Davies-

Bouldin semble approprié : nous travaillons sur des données de taille modeste. Nous projetons d'examiner ces autres mesures sur des données plus volumineuses.

7.2. Comment caractériser la notion d'événement ?

Si tous les travaux présentés dans l'état de l'art obtiennent des résultats satisfaisants, aucun n'a tenté de décrire formellement ce qu'est un événement. Donner une définition d'un tel concept est certes difficile, mais nous avons cependant considéré qu'il était nécessaire de caractériser un événement, ne serait-ce que pour rendre plus objective l'évaluation.

Nous avons exploré différentes définitions d'un événement, notamment d'un point de vue sociologique et d'un point de vue linguistique. D'un point de vue sociologique tout d'abord, l'événement prend autant de définitions que de champs disciplinaires dans lesquels il est considéré (Prestini-Christophe, 2006). Il existe cependant des points communs à toutes ces définitions :

- un événement est un fait inattendu et correspond à une rupture,
- un fait devient événement en fonction du monde dans lequel il advient : l'événement est subjectif.

D'un point de vue linguistique, Pustejosky a élaboré une théorie (Pustejovsky, 2000) dans laquelle il fonde sa définition de l'événement sur le repérage de structures prédicat/arguments et sur la notion de changement. L'événement est identifiable grâce à un *prédicat d'événement* (i.e. un verbe qui implique un changement) et une structure argumentale équivalente (c'est-à-dire une identité référentielle des arguments du prédicat, même si les formes de surface employées sont différentes). Une implémentation de cette théorie nécessiterait des connaissances sémantiques très fines sur les verbes, mais également des connaissances sur les différents arguments possibles.

Etant donné le nombre de définitions et le peu de formalisation du concept d'événement, il est sûrement plus judicieux de partir de nos besoins afin de définir le concept d'événement dans le cadre de notre travail. Notre tâche consiste à regrouper les documents d'actualité (des dépêches) qui traitent du même événement, afin de réaliser une synthèse automatique des groupes créés. On part du principe qu'une dépêche traite d'un événement unique (ceci est généralement vérifié, sauf pour certaines dépêches qui retracent tous les faits marquants d'une journée, ou le déroulement d'une succession de faits plus ou moins liés entre eux). Les dépêches sont en outre majoritairement rédigées en forme de pyramide inversée : le premier paragraphe doit normalement contenir l'information sur l'événement brut, les détails, commentaires et opinions étant normalement détaillés dans la suite du document.

Il s'agit alors, afin de réaliser notre tâche, de trouver les traits communs entre les dépêches qui traitent d'un même événement. Un événement consiste en une action réalisée par une ou plusieurs personnes, à une certaine date, dans un certain lieu. Il est donc assez intuitif d'utiliser ces marqueurs afin de reconnaître que deux dépêches traitent du même sujet. Nous retrouvons ici les éléments mis en évidence par (Pustejovsky, 2000) concernant les arguments du prédicat. On peut par ailleurs utiliser des mots clefs, tels que « procès », « élections », « bombardement », afin de départager des dépêches partageant des entités nommées mais dont le lecteur a l'intuition qu'ils se rapportent malgré tout à des événements différents.

Reste à définir la portée d'un événement. En d'autres termes, quelle fenêtre temporelle doit-on adopter pour exclure d'un groupe un document qui traite du même événement qu'un document précédent ? Contrairement à (Binsztok *et al.*, 2004), nous considérons la notion de portée indépendante de toute notion temporelle. Un fait directement lié à un événement initial peut se produire longtemps après le fait initial, comme la reconnaissance de l'innocence d'un homme vingt ans après sa condamnation.

Afin d'éviter toute ambiguïté, nous appellerons dorénavant l'ensemble des dépêches regroupées ensemble, parlant de faits directement liés entre eux, le « sujet », et le fait initial, celui qui aura conduit à cette succession de faits, l'« événement ». Les regroupements que nous chercherons à effectuer seront donc des regroupements par sujet, et non des regroupements par événement. Cette notion de sujet se rapproche de la notion d'« activité » telle que définie dans TDT4, par opposition à la notion d'« événement », toujours dans TDT4.

Malgré cette tentative d'« objectivisation » des notions d'événement et de sujet, ces notions restent largement subjectives. A titre d'exemple, prenons deux documents : l'un parlant des élections anticipées en Côte d'Ivoire consécutives aux accords de Marcoussis, l'autre parlant de la table ronde de Marcoussis. Faut-il créer deux sujets pour ces deux dépêches – l'un concernant la réunion de Marcoussis, l'autre l'application des accords décidés à Marcoussis, voire les élection anticipées – ou les réunir au sein d'un même sujet concernant les accords de Marcoussis, de leur négociation à leur application ? Il y a là une part de subjectivité qui ne peut être complètement évitée.

7.3. Modèle de description des documents

Afin de regrouper les documents par sujet, mais également afin d'obtenir une représentation graphique qui permette à un utilisateur de prendre rapidement connaissance du contenu d'un fonds documentaire, celui-ci est représenté à la manière d'un graphe pondéré (fig 7.1). Chaque document a des liens avec les autres documents du corpus. Un événement se caractérise principalement par la réponse aux questions : qui ? Quoi ? Où ?

Quand ? Nous avons volontairement exclu le quoi de la description des documents. En effet, cela nous permet de nous concentrer uniquement sur les entités nommées présentes dans les documents. Beaucoup de langues disposent aujourd'hui d'outils d'annnotation d'entités nommées. Des annotateurs multilingues existent même, qui normalisent les entités nommées de documents écrites dans différentes langues (Brun *et al.*, 2009). Fonder notre approche uniquement sur les entités nommées permet donc d'imaginer des applications d'exploration de corpus multilingues. Ainsi, le poids des liens entre deux documents est calculé selon les entités nommées qu'ils partagent. Notre méthode de regroupement de documents est similaire à celle que nous utilisons pour regrouper les phrases pour la génération de résumés, à ceci près que nous n'utilisons ici que les entités nommées.

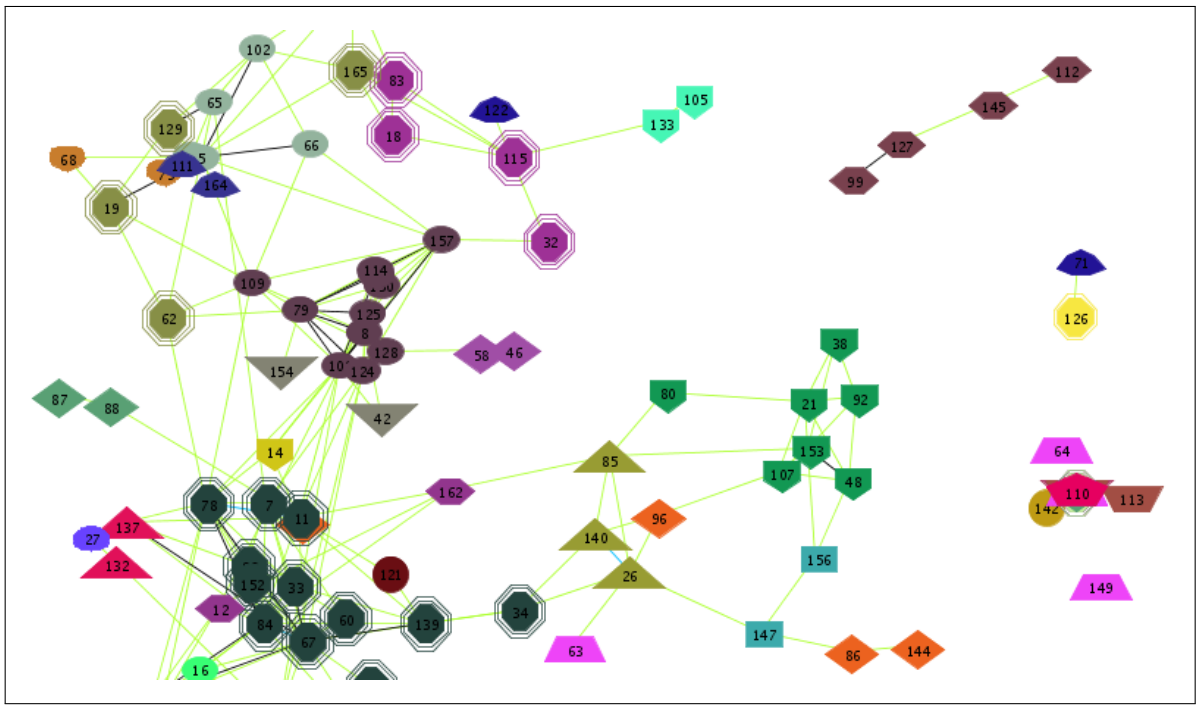


FIGURE 7.1.: Un corpus vu comme un graphe : les noeuds sont les documents, leurs forme et couleur dénotent leur appartenance à une classe, et la couleur des liens est fonction de leur poids : plus le lien est foncé, plus il est important.

Nous avons utilisé, pour calculer la similarité de contenu entre deux documents, un indice Jaccard pondéré. L'indice Jaccard est une métrique utilisée en statistique pour comparer la similarité de deux ensembles, fondée sur le rapport entre la cardinalité de l'intersection et la cardinalité de l'union des ensembles.

Chaque ensemble est formé des entités nommées contenues dans chaque document. Afin d'affiner l'importance relative des différentes entités, il est nécessaire de les pondérer. En effet, les corpus privilégient souvent certains thèmes : certaines entités centrales apparaissent alors de manière relativement uniforme et ont de ce fait un pouvoir discri-

minant très faible. Leur participation au calcul de similarité doit être en conséquence.

Pour cela, nous avons testé deux mesures : la mesure `idf` et la mesure du `tf.idf` (Salton et Buckley, 1988) (nous utilisons le mode de calcul classique de ces deux mesures, cf. figure 2). Cette pondération permet en outre de prendre en compte la fréquence d'un terme. Il y a en effet une très forte probabilité que deux documents qui contiennent les mêmes entités nommées avec des fréquences proches fassent partie d'un même sujet, et la mesure jaccard pondérée par le `tf.idf` permet de rendre compte de ce fait.

Calcul du tf d'un terme t_i dans le document d_j :
$\mathrm{tf}_{t_i,d_j} = \frac{|t_i|_j}{\sum_{k=0}^{n} |t_k|_{d_j}}$

Calcul de l'`idf` d'un terme t_i au sein d'un corpus D de documents d_j :
$\mathrm{idf}_{t_i} = \log \frac{(|D|)}{\{|d_j : t_i \in d_j\}|}$

Calcul du `tf.idf` du terme t_i pour un document d_j :
$\mathrm{tf.idf}_{t_i,d_j} = tf_{t_i,d_j} \times idf_{t_i}$

FIGURE 7.2.: Calcul du `tf.idf`

Cette pondération selon le `tf.idf` comme réalisée dans (Yang *et al.*, 1999) ne suffit cependant pas : certains types d'entités nommées ont un rôle plus important que d'autres dans la description d'un événement. En effet, les types d'entités les plus instables au sein d'un même sujet sont les types « Personnes » et « Organisations ». Si les références aux dates et aux lieux restent globalement inchangées dans le suivi d'un sujet, les intervenants sont en revanche multiples et variables. Il faut donc distinguer les entités suivant leur type afin d'éviter la création de sujets distincts correspondant à chaque intervenant. Nous avons alors affecté un poids à chaque type d'entités nommées, en privilégiant les lieux et les dates par rapport aux noms de personnes et d'organisations. Le calcul de la similarité entre documents est celui présenté en figure 7.3

7.4. Classification

Nous réalisons un regroupement des documents en utilisant une variante de l'algorithme k-means (Forgy, 1965b), *fast global k-means*. Cet algorithme réalise une classification en k classes, en minimisant la variance intra-classe. *Fast global k-means* est itératif et permet facilement d'étendre la portée de notre système : comme pour la mise à jour de résumé exposée en §5.3.3, le système peut prendre en charge l'ajout de documents à un corpus déjà traité. *Fast global k-means* est décrit en figure 3.4.

- $S(i,j) = \frac{N_{11}(i,j)}{(N_{11}(i,j)+N_{10}(i,j)+N_{01}(i,j))}$
- $N_{11}(i,j) = \sum_{k=0}^{n} poids(e_k) \times tf.idf(e_k)$, où les e_k sont les entités nommées présentes dans i et j.
- $N_{10}(i,j) = \sum_{k=0}^{n} poids(e_k) \times tf.idf(e_k)$, où les e_k sont les EN présentes seulement dans i.
- $N_{01}(i,j) = \sum_{k=0}^{n} poids(e_k) \times tf.idf(e_k)$, où les e_k sont les EN présentes seulement dans j.

FIGURE 7.3.: Calcul de la mesure de similarité entre documents

L'algorithme fast global k-means prend comme paramètre le nombre de classes (k). Dans le cadre du regroupement non-supervisé de documents, il est nécessaire de laisser ce paramètre libre, et de trouver le meilleur k possible. Pour chaque itération de l'algorithme, l'indice de Davies-Bouldin est calculé (cet indice permet de calculer la validité de la classification, en mesurant la cohérence des regroupements ; les regroupements qui minimisent la distance entre objets de la même classe et maximisent la distance entre objets de classes différentes ont le meilleur indice). Finalement, la classification établie à l'itération qui minimise l'indice de Davies-Bouldin Davies et Bouldin (1979) est retenue.

7.5. Evaluation

Nous avons choisi d'évaluer le travail de classification par un ensemble de mesures quantitatives. Cette section décrit le cadre applicatif et le corpus utilisé puis les évaluations menées.

7.5.1. Description du cadre applicatif

Cette recherche s'inscrit dans le cadre du projet Infom@gic, du pôle de compétitivité Cap Digital [1]. Ce cadre nous permet d'avoir accès à des besoins opérationnels précis et à des corpus variés.

Un de ces corpus largement mis à contribution dans le cadre d'Infomagic est un en-

1. Cap Digital porte sur l'indexation multimedia. Cette recherche s'insère dans le cadre de l'« axe texte » du projet, qui regroupe des entreprises et des laboratoires de recherche en traitement des langues, ainsi que des industriels ayant des besoins spécifiques qui servent de cadres d'application communs.

semble de dépêches AFP portant sur la Côte d'Ivoire. Le corpus compte 15000 documents. L'annotation des documents avec les entités nommées nous a été fourni par la société Arisem (partenaire du projet Infomagic). Nous avons travaillé sur un extrait du corpus comptant 200 dépêches, afin de pouvoir évaluer au mieux la classification. Ainsi, nous avons fait le choix de réaliser une double annotation, afin de pouvoir comparer notre approche à deux résultats obtenus manuellement par des personnes différentes. Le corpus compte 1030 entités nommées différentes. Nous avons établi un corpus de référence en classant les dépêches par sujet, et obtenu 44 classes avec une répartition très inégale.

Nous avons voulu évaluer toutes les étapes de la classification, afin d'identifier les faiblesses de la méthode utilisée. Dans un premier temps, nous avons évalué la qualité de la classification en examinant les résultats de l'algorithme *fast global k-means* pour k de 2 à 100, ce qui correspond à diviser le corpus en un nombre de classes allant de 2 à 100. Dans un second temps, nous avons évalué le résultat « optimal » de l'apprentissage, qui correspond à la classification qui minimise l'indice de Davies-Bouldin en le comparant au meilleur résultat obtenu sur toutes les itérations de *fast global k-means*, qualitativement parlant.

7.5.2. Evaluation globale

Nous avons évalué la pertinence de la classification par deux méthodes :

- les micro-moyennes ;
- les macro-moyennes.

Ces deux mesures donnent des résultats très différents : le résultat de la macro-moyenne tient plus compte des catégories ayant peu de documents pertinents, tandis que le résultat des micro-moyennes fait plus ressortir les résultats sur les plus grosses classes. Ces deux mesures sont donc nécessaires pour la bonne interprétation des résultats. La méthode des macro-moyennes consiste à comparer chaque classe C_i du corpus étiqueté automatiquement à la classe du corpus de référence majoritaire dans C_i. Les deux classifications comparées doivent avoir le même nombre de classes. Nous avons donc fixé k au nombre de classes du corpus de référence. Une moyenne est effectuée sur les mesures de chaque classe, avec un poids égal pour chaque classe. La méthode des micro-moyennes consiste à fusionner les tables de contingence de toutes les classes et à calculer les mesures sur la table fusionnée. Nous avons utilisé les mesures classiques de précision, rappel et F-mesure. La F-Mesure est la moyenne harmonique de la précision et du rappel, et favorise les systèmes qui ont des mesures de rappel et de précision proches.

Il est intéressant de noter que deux annotateurs humains ont obtenu sur ce même corpus, des résultats dont la F-Mesure est à 44%. Les deux annotations différentes se

	Précision	Rappel	F-Mesure
Macro-Moyenne	61.4%	65.3%	63.2%
Micro-Moyennes	52.3%	55.6%	53.9%

FIGURE 7.4.: Résultats de l'évaluation avec le nombre de classes fixé

défendent, les résultats de celles-ci dépendant fortement du choix de granularité utilisé par les annotateurs dans les sujets, et du choix de raccorder certains événements à un sujet ou de créer un nouveau sujet pour ceux-ci. Ceci pose le problème de la comparaison de classifications dont le nombre de classes diffère et de l'évaluation d'une classification par des méthodes quantitatives.

7.5.3. Sélection du k par l'indice de Davies-Bouldin

Nous avons évalué la sélection du k selon les deux points suivants :
- la différence entre le nombre de classes choisi automatiquement et le nombre de classes du corpus de référence ;
- le rapport entre l'indice de Davies-Bouldin et une mesure d'évaluation de la classification.

En moyenne sur 15 lancements de l'algorithme (donc 15 configurations de départ différentes), nous avons trouvé 37,4 classes contre 44 classes dans le corpus de référence. Le même corpus annoté par un deuxième annotateur contient quant à lui 65 classes, l'annotateur ayant fait des choix différents concernant les sujets existants. On constate donc une erreur dans le choix automatique du nombre de classes pouvant aller de 16% à 57%. Cette mesure n'est donc pas asez significative pour évaluer la qualité du choix du nombre de classes. Nous avons évalué la pertinence de l'utilisation de l'indice de Davies-Bouldin ; la figure 7.5 montre l'évolution de l'indice pour tous les k ainsi qu'une mesure d'évaluation du clustering, qui consiste à considérer tous les documents liés au sein d'un même cluster, et à effectuer les calculs de précision/rappel sur la présence/absence des liens du corpus de référence dans le corpus classifié automatiquement.

On constate la non-corrélation de l'indice de Davies-Bouldin et le choix d'un nombre de classes qui maximise la F-mesure. Ceci est en grande partie dû au fait que le découpage du corpus de référence n'est pas homogène : une des classes contient à elle seule 1/6 des documents du corpus. Or le choix d'un k par Davies-Bouldin et la classification des documents par k-means seront optimaux dans le cas où les classes ont toutes approximativement le même nombre d'objets.

Une évaluation qualitative de notre outil d'analyse de corpus est nécessaire afin de valider son utilisabilité, son utilité ayant déjà été validée au sein du projet Infomagic. Une telle évaluation nécessite une collaboration étroite avec des industriels afin d'établir un

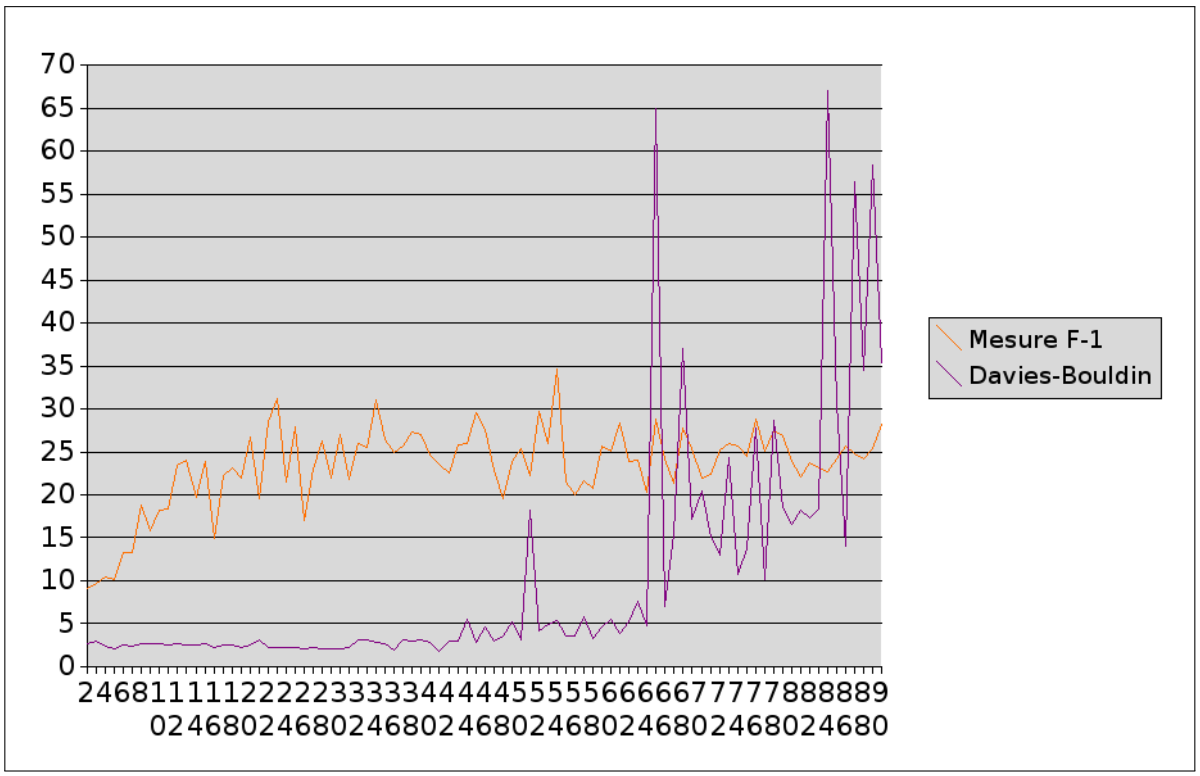


FIGURE 7.5.: Evaluation de l'indice de Davies-Bouldin et du choix du k

protocole cohérent, en adéquation avec leurs besoins. De plus, notre système de résumé automatique a été évalué sur des tâches de résumé automatique de documents en anglais guidé par une requête tandis que notre outil d'analyse de corpus réalise des résumés en français non guidés ; il faudra également évaluer notre système de résumé sur une telle tâche afin de valider son apport à l'analyse de corpus.

7.6. Visualisation du corpus

Une fois les documents regroupés en classes événementielles, il faut proposer à l'utilisateur une visualisation efficace qui lui permette d'appréhender le corpus dans sa globalité, mais également le contenu de chacune des classes. Nous présentons ici la manière dont nous adressons à l'utilisateur une indication du contenu de chaque classe ainsi qu'un résumé à la demande.

Calcul du tf d'un terme t_i dans la classe de documents c_j :
$\mathrm{tf}_{t_i,c_j} = \frac{|t_i|_j}{\sum_{k=0}^{n} |t_k|_{c_j}}$

Calcul de l'icf d'un terme t_i au sein d'un corpus C composé de documents répartis dans les classes c_j :

$\mathrm{icf}_{t_i} = \log \frac{(|C|)}{\{|c_j : t_i \in c_j\}|}$

Calcul du tf.icf du terme t_i pour une classe c_j :
$\mathrm{tf.icf}_{t_i,c_j} = tf_{t_i,c_j} \times icf_{t_i}$

FIGURE 7.6.: Calcul du tf.icf

7.6.1. Titrage des classes événementielles

Les différentes classes sont établies en fonction des entités nommées qu'elles comportent. En étudiant les entités nommées repérées dans une classe et leur distribution dans le corpus, il est possible de trouver lesquelles ont été discriminantes dans le regroupement des documents au sein de cette classe. Ces entités nommées sont censées être les plus représentatives de l'événement narré par les documents de la classe. Nous caractérisons la discriminance des entités nommées par leur *tf.icf*, une variante du *tf.idf* utilisant les classes de documents à la place des documents. Sa formule est définie en figure 7.6.

Les trois entités nommées les plus discriminantes de chaque catégorie d'entité sont proposées à l'utilisateur comme titre de la classe. Afin de lui permettre de se rendre compte visuellement rapidement de l'importance des entités du titre, nous les surlignons dans des nuances de vert fonction de leur tf.icf. Nous avons choisi de ne pas normaliser les *tf.icf* pour chaque classe en fonction du *tf.icf* maximum. En effet, certains regroupements sont anarchiques car les documents marginaux sont souvent placés dans des classes « fourre-tout ». Ne pas normaliser les *tf.icf* au sein de chaque classe permet donc à l'analyste d'identifier rapidement ces classes grâce aux couleurs sombres des entités nommées qui en composent le titre. La palette de couleurs varie en RGB de [0.1,0.1,0.1] à [0.1,1,0.1], ce qui permet d'obtenir des couleurs allant du gris sombre au vert clair. Les *tf.icf* sont donc normalisés entre 0.1 et 1 en fonction du maximum et du minimum du corpus, ce qui permet d'obtenir la palette de couleurs la plus étendue possible. La formule du *tf.icf* normalisé devient donc la formule de la figure 7.7.

$$tf.icfnorm(e_i, C_j) = \frac{tf.icfnorm(t_i,C_j) - argmin_{C_k \in C, e_n \in C_k}(tf.icf(e_n,C_k))}{argmax_{C_k \in C, e_n \in C_k}(tf.icf(e_n,C_k)) - argmin_{C_k \in C, e_n \in C_k}(tf.icf(e_n,C_k))} \times (F - D) + D$$

où C est le corpus étudié,
C_n les classes du corpus,
$e_n \in C_n$ les entités nommées de la classe C_n,
D la borne inférieure de l'intervalle dans lequel on veut contenir les valeur du *tf.icf* normalisé, et F la borne supérieure de cet intervalle.

FIGURE 7.7.: Formule du *tf.icf* normalisé

Accords de Marcoussis

Acteurs	Lieux	Evénements nommés Organisations	Dates	Divers
Gbagbo	Marcoussis	Accords de Marcoussis	vendredi	quatre mois
Seydou Diarra	France	MPCI	samedi	14 ans
Guillaume Soro	Paris	Accord de marcoussis	samedi après-midi	plus de quatre mois de crise

Combats pour l'aéroport de Man

Acteurs	Lieux	Evénements nommés Organisations	Dates	Divers
Guillaume Gbatto	Man	MPIGO	28 novembre	26 ans
Ange-Antoine	Aéoroport de Man	MJP	5 décembre dernier	14h15
Réné Dessonh	Danané	Mouvement pour la Justice et la Paix	de noël 1999 à octobre 2000	30 km

« Inclassables »

Acteurs	Lieux	Evénements nommés Organisations	Dates	Divers
Gbagbo	France	PDCI	2000	deux mois
Bédié	Paris	MPCI	décembre 1999	40%
Apollinaire	Bouaké	SNBA	septembre 2002	cent litres

FIGURE 7.8.: Illustration des résumés indicatifs des classes événementielles pour trois classes établies à la main et issues du corpus « Côte d'Ivoire ». Plus les couleurs sont claires, plus l'entité est discriminante.

La figure 7.8 présente l'exemple du titrage de trois classes différentes issues d'un des corpus de référence décrit en §7.5.1 classifié à la main par un annotateur. La première classe correspond au suivi des accords de Marcoussis qui ont tenté de mettre fin à la crise ivoirienne de 2002, la seconde a trait aux combats pour la prise de contrôle de l'aéroport de Man (Côte d'Ivoire), et la troisième est constituée de documents hétérogènes inclassables ailleurs ; elle illustre bien les classes « fourre-tout » dont il est question ci-dessus.

La technique utilisée pour donner un « titre » aux classes événementielles s'apparente au résumé indicatif tel qu'il est défini en §1.1.1. Elle se démarque des méthodes existantes par la coloration des entités nommées du titre qui permet à l'utilisateur d'une part de visualiser la cohérence du titre vis-à-vis des documents, d'autre part de voir si les

regroupements effectués au préalable par l'algorithme de classification sont cohérents.

7.6.2. Résumé des classes événementielles

L'utilisateur de notre outil d'analyse de corpus a accès à une représentation du corpus et à une visualisation de la classification des documents en classes événementielles. Il peut alors demander un résumé de chacune des classes établies.

Nous avons relié notre système de résumé automatique –CBSEAS, présenté en chapitre 3– à l'outil d'analyse de corpus. Un résumé automatique peut être guidé ou non par une requête. Dans le cas de résumés guidés par une requête, un des scores participant au calcul de la pertinence d'une phrase est le score de pertinence vis-à-vis de la requête. Ici, les résumés ne sont pas guidés par une requête, car nous cherchons à fournir à l'utilisateur une indication sur le contenu de chaque classe événementielle. Le score de pertinence à la requête est donc remplacé par un score « centroïde », qui évalue l'importance d'une phrase vis-à-vis du contenu global de sa classe. Plus une phrase contient des mots centraux dans sa classe, plus son score centroïde est élevé. La section 3.2.2 détaille le calcul de ce score.

Nous avons vu en §4.1 que l'utilisation d'une annotation en entités nommées dans CBSEAS améliore la qualité des résumés produits. Par conséquent, nous utilisons dans cette expérience l'annotation fournie par la société Arisem pour le corpus Côte d'Ivoire. Celle-ci ne contient pas de résolution de co-référence. Une annotation simple des groupes nominaux constituant les entités nommées permet tout de même d'améliorer nos calculs de similarité entre phrases, et par là-même la qualité des résumés générés.

7.7. Conclusions

Nous avons présenté un système d'aide à l'analyse et à la prise de connaissance de corpus. Celui-ci regroupe automatiquement les documents en classes événementielles qu'il titre grâce à un résumé indicatif sommaire et propose un résumé informatif des classes à la demande de l'utilisateur. Ce résumé informatif est obtenu grâce à CBSEAS, un système de résumé automatique qui a obtenu récemment de bons résultats sur la tâche de « résumé et mise-à-jour de résumé de dépêches » en langue anglaise de la campagne d'évaluation internationale TAC 2009.

Le système de regroupement de dépêches en classes événementielles est quant à lui fondé sur les entités nommées partagées entre documents. Les résultats obtenus suite à la classification automatique sont comparables à ceux obtenus par des annotateurs

humains. Dans nos expériences, la classification automatique est même plus corrélée aux différentes classifications manuelles que les différentes classifications manuelles entre elles.

L'évaluation quantitative devrait à l'avenir être complétée par une évaluation qualitative. En effet, les systèmes d'évaluation quantitative ne permettent pas de valider l'utilisabilité des résultats : la séparation d'une classe en deux par un algorithme peut faire chuter le rappel de 50%. Les résultats obtenus automatiquement sont toutefois intéressants et peuvent faire ressortir d'autres regroupements que ceux choisis par les auteurs du corpus de référence, le regroupement de documents étant très subjectif. De plus, la classification étant intégrée à un système plus large, il convient d'en évaluer l'utilisabilité pour notre tâche d'analyse de corpus.

L'algorithme utilisé pour la classification, k-means, n'est pas exempt de défauts : moins efficace sur des données à regrouper dans des classes non-homogènes, il devient également moins performant sur des corpus de grande dimension. Pour passer à l'échelle supérieure, il nous faudra donc explorer d'autres méthodes de classification, comme les SVM et les cartes de Kohonen non supervisées.

Conclusions et perspectives

Dans cette thèse, nous avons proposé une approche générique pour le résumé automatique, CBSEAS, qui établit un modèle des documents à résumer fondé sur le regroupement des phrases en classes sémantiques préalablement à la sélection des phrases à extraire. Le modèle en question nous sert non seulement à l'extraction des phrases, mais également au réordonnancement des phrases du résumé généré et à la gestion de la mise à jour de résumés.

Après avoir utilisé un algorithme génétique pour paramétrer notre système de résumé automatique, nous avons cerné les limites liées à la sélection de phrases par des méthodes statistiques, et identifié le besoin d'attaquer la problématique du résumé automatique par d'autres types de méthodes. Nous nous sommes alors porté sur un domaine spécialisé, le résumé de dépêches de presse, et avons proposé une classification des dépêches et un étiquetage automatique des zones qu'elles comportent afin de mieux apréhender la sélection des phrases. Pour finir, nous avons montré l'utilité du résumé automatique pour des tâches d'analyse de corpus.

Résultats

Nos approches ont été évaluées sur deux tâches différentes des campagnes d'évaluation « *Text Analysis Conference* » (TAC). La première tâche, le résumé d'opinions, a été gérée en couplant notre système, CBSEAS, à un outil automatique d'analyse d'opinions. Seules les phrases détectées comme véhiculant l'opinion demandée par la requête utilisateur sont sélectionnées dans le résumé. Notre système s'est révélé compétitif sur cette tâche, se classant dans le premier quart des participants.

CBSEAS a également été évalué sur une tâche de résumé de dépêches et de mise à jour de résumé. Si les résultats témoignent de la performance de CBSEAS en ce qui concerne l'informativité des résumés produits, ils révèlent les lacunes de notre système en confort de lecture pour l'utilisateur. Le fait que les résumés soient coupés automatiquement à 100 mots par les organisateurs de TAC en est certes une raison, cela ne doit cependant pas nous faire oublier que nous n'avons développé aucune technique de post-traitement linguistique permettant de couper des fragments de phrases du résumé qui peuvent être

jugés inutiles.

Cette tâche nous a permis de montrer l'utilité de notre modèle des documents, qui représente la diversité informationnelle qu'ils contiennent. En effet, notre approche de détection de la nouveauté pour la génération de résumés de mise à jour, qui se fonde sur ce modèle, s'est révélée plus efficace que la moyenne. Les résultats de l'évaluation ont montré la complexité de cette tâche, puisque les scores Pyramide [2] des systèmes qui ont participé à TAC 2009 chutent en moyenne de 23%, tandis que notre système perd 11% seulement sur cette même évaluation.

Il est intéressant de noter que les systèmes de résumé purement extractifs ont pratiquement atteint leurs limites. En effet, la *baseline* 3 de la tâche résumé et mise à jour de TAC 2009, composée uniquement de phrases extraites des documents sources obtient des scores d'informativité à peine meilleurs que ceux des meilleurs systèmes de résumé automatique actuels. En revanche, la satisfaction de l'utilisateur vis-à-vis du résumé produit n'est pas encore à la hauteur des résumés manuels. Il est donc nécessaire de se concentrer sur les techniques de réordonnancement et sur des méthodes de sélection de phrases plus adaptées aux tâches de résumé automatique.

Notre système se comporte bien sur les tâches « Résumé et mise à jour » de TAC 2009 et « Résumé d'opinions » de TAC 2008. Cependant, il n'échappe pas aux conclusions générales tirées ci-dessus, et doit être amélioré afin de mieux répondre aux attentes des utilisateurs. La section suivante présente nos propositions dans ce sens.

Perspectives

Les évaluations auxquelles nous avons participé ont révélé certaines lacunes de notre système de résumé automatique. Premièrement, les résumés générés par CBSEAS obtiennent des notes en linguistique en dessous des scores fondés sur l'inclusion d'éléments d'information essentiels. Nous concentrer sur la définition de critères linguistiques permettant de déterminer la pertinence de l'extraction d'une phrase dans un résumé pourrait résoudre en partie le problème. Utiliser des techniques de compression de phrases apparaît également comme essentiel afin d'éliminer toute information superflue et donc d'améliorer le rapport du lecteur au résumé.

La résolution d'anaphores mériterait également d'être poussée plus avant. Une technique privilégiant la précision au rappel serait plus adaptée à une tâche de résumé automatique. Etudier et mettre en place un remplacement sélectif des pronoms par leur référent dans les résumés améliorerait le rendu des résumé par rapport à la méthode

2. Pyramide : méthode manuelle d'évaluation de la qualité informative des résumés, *cf* §2.4.

naïve que nous avons employée, à savoir le remplacement systématique des pronoms par leur référent.

Enfin, la classification et l'analyse de la structure des dépêches de presse est incomplète. Nous voulons identifier les différentes zones qui composent les dépêches en nous fondant notamment sur les indices définis dans Lucas (2004). Cela nous permettra notamment d'établir des critères d'extraction fondés sur la position d'une phrase dans la structure d'une dépêches plus adaptés aux besoins des utilisateurs.

Des post-traitements tels que la compression de phrases permettraient également d'améliorer la qualité des résumés générés. Premièrement, cela permettrait de supprimer les informations non essentielles, et donc d'améliorer le ressenti de l'utilisateur vis-à-vis du résumé automatique. De plus, dans le cas où les résumés sont limités à un certain nombre de mots, la compression de phrases permet de libérer de la place afin d'inclure d'autres phrases qui véhiculent leur lot d'informations, et participent donc à la quantité d'informations portées par les résumés, un des critères de qualité les plus important pour un résumé automatique. Le but des techniques de compression de phrases étant d'éliminer les informations non pertinentes tout en préservant la grammaticalité des phrases, nous pourrions apporter aux techniques existantes un autre critère de pertinence en nous servant de notre modèle de représentation du corpus et de la classification des phrases en classes sémantiques.

La diversité des axes de recherche futurs que nous avons identifiés témoigne de la complexité des tâches de résumé automatique et des domaines variés qu'elles font intervenir. Le domaine du résumé automatique, à la croisée des méthodes statistiques et linguistiques, reste un problème ouvert auquel il nous faudra apporter des solutions plus complètes afin de rapprocher les résumés automatiques des résumés produits par des humains.

A. Annexes relatives à TAC2008

A.1. Tâche Résumé d'opinions

RunID	Pyramid	Grammat.	Non-Red.	Structure	Lisibilité	Overall Resp.
CLASSY2	0.123	7.545	7.909	3.591	5.318	1.773
DLSIUAES2	0.155	3.545	4.364	3.091	2.636	2.227
ICTCAS1	0.113	5.045	6.455	2.182	3.682	2.864
ICTCAS2	0.110	6.000	5.818	2.409	3.455	2.955
IIITSum081	0.101	5.273	6.045	2.045	3.545	2.364
IIITSum082	0.102	5.045	6.318	2.045	3.545	2.500
LIPN2	**0.169**	**5.955**	**6.636**	**3.500**	**4.455**	**2.636**
Miracl1	0.163	5.727	5.500	2.636	3.591	3.909
Miracl2	0.125	5.455	5.909	2.273	3.500	2.591
PolyU1	0.251	5.864	6.909	3.045	3.864	2.864
TOC1	0.176	5.091	4.500	3.409	3.591	1.909
TOC2	0.150	4.864	4.727	3.409	3.545	1.955
uavua2	0.133	4.545	4.636	2.364	3.182	3.455
UBC1	0.161	3.864	6.091	2.091	2.909	2.818
UBC2	0.223	5.455	5.864	2.591	4.045	3.318
UIUC1	0.165	4.500	5.000	3.091	3.636	2.818
UIUC2	0.181	5.091	5.955	2.318	3.227	3.318
UMD1	0.130	4.318	6.364	2.409	3.318	1.682
UMD2	0.136	4.409	6.682	2.364	3.545	1.727

TABLE A.1.: Résultats de TAC 2008 « Opinion Summarization » : systèmes automatiques sans utilisation des *snippets*

A.2. Tâche Résumé et Mise à Jour

A.2.1. Sujet D0805

Nous présentons ici pour exemple les deux jeux de documents (documents initiaux et documents de mise à jour) du sujet D0805.

Jeu de documents initial

TIPS ON CHOOSING RIGHT PLANS

People eligible for the Medicare prescription drug benefit that takes effect Jan. 1 can choose from a variety of plans that include the new coverage by itself or as part of a broader health insurance package. Here are some tips on narrowing the options :
IF you have no prescription drug coverage now,
THEN you should purchase a stand-alone prescription drug plan, or a broader Medicare Advantage Plan that includes drug coverage along with coverage for traditional Medicare services such as visits to doctors and hospitalization.
IF you now have drug coverage through a health plan offered by your former employer or union,
THEN you may not need to do anything. Your former employer or union will let you know if your current coverage is at least as good as the new Medicare benefit. Signing up for a Medicare Part D drug plan could jeopardize your existing coverage.
IF you now receive benefits from both Medicare and Medicaid,
THEN your drug coverage will be moved from Medicaid to Medicare. You will automatically be assigned to a Medicare drug plan, but with the option of switching if you find another plan. If you do not sign up for a drug plan, you will automatically be assigned to one.
IF you have a so-called Medigap health plan, which provides coverage for benefits not provided by traditional Medicare (including prescription drug coverage),
THEN you should study the cost of the Medicare prescription drug plans and decide if one of them is better than the drug coverage provided under your existing Medigap plan. If so, you can cancel the drug coverage in your existing Medigap plan.
IF you now receive health coverage through a Medicare Advantage Plan,
THEN your current health plan will offer you a new option that incorporates prescription drug benefits with traditional Medicare benefits.
IF you now receive drug benefits through the Massachusetts Prescription Advantage Plan, and are eligible for Medicare,
THEN you will automatically be assigned to a randomly selected Medicare drug plan. Low-income beneficiaries will be required to apply for Medicare subsidies.

- JEFFREY KRASNER

FIGURE A.1.: NYT_ENG_20051113.0163

Making Sense of the New Medicare Drug Program

Which one should I pick ?
Millions of people are asking that question as they consider Medicare's new prescription drug plan, which rolls out Jan. 1. The voluntary program – also known as Medicare Part D – will allow seniors and the disabled who are covered under the Medicare program to get subsidized drug coverage.
Medicare beneficiaries have been bombarded with mail, phone calls and advertisements over the last several weeks from plans eager to sign them up. Enrollment starts Nov 15.
By now, they've probably realized it's more confusing than anything else.
Medicare beneficiaries have many plans to choose from – each with its own premium, lists of covered drugs and participating pharmacies.
The Bush administration estimates the coverage (which will cost the federal government at least $ 720 billion during the first 10 years) will save beneficiaries an average of $ 1,300 a year. But in fact, the savings can vary wildly depending on factors such as what kind of drug coverage someone currently has or to how much they spend on prescription drugs.
A person who spends $ 1,500 a year on prescriptions, for example, can save about $ 550 on the new plan. But another person who spends just $ 500 may actually pay $ 198 more out of pocket because of premiums and drug co-payments.
How to make a choice among the blizzard of plans ? "You don't have to understand every detail and every option," said Dr. Mark McClellan, administrator of the federal Medicare agency. "People just need to focus on what they want."
Following are some highlights and things to consider about the new drug plan.
ELIGIBILITY
One of the biggest misconceptions about the new drug benefit is that it's only for low-income seniors. In fact, all Medicare beneficiaries are eligible. People won't be automatically enrolled, however. Enrollees have to choose a plan and sign up through a private insurer or with Medicare, either on the agency's phone or its Web site.
DECIDING IF IT'S WORTH IT
The plans make most sense for people who are already spending a lot of money on prescription drugs, and for low-income seniors, who may be eligible for financial assistance with premiums and drug co-payments.
The costs for each plan can vary considerably depending on what's offered. This year, monthly premiums vary from nothing to as high as $ 60 a month, although cheaper plans usually have much leaner coverage, with higher deductibles and fewer covered drugs.
WHICH PLAN TO PICK
This essentially boils down to what prescriptions you use and how you like to pay for them, and the kind of medical coverage you have.
Plans can look very different, because the government is allowing insurers to structure them as they want. They merely have to be "actuarially equivalent" to each other – they must, in other words, provide the same benefit in the end. For most people, the first $ 250 is not covered, but the next $ 2,000 is – except for any co-payments you may have to make. Following this, you will have to pay the next $ 1,350 : Experts refer to this as a "doughnut hole" in coverage. But after that, 95 percent of drugs are covered, which is meant to help people with catastrophically high drug costs in any given year.
There are two classes of plans among the dozens of offerings. The one you'll pick will depend on whether you're buying just a drug plan or adding it to your current managed care plan.
Seniors who choose to stay in Medicare's traditional fee-for-service plan can only buy what's called a stand-alone prescription drug plan. Those who are in Medicare Advantage HMO plans – about one-fourth of all beneficiaries – will now get the drug benefit added to their current planWhether the premium will rise, and by how much, depends on the plan.
"People need to sit down, take a breath and figure out what is best for them," said Bonnie Burns, Medicare policy and training specialist for California Health Advocates, a nonprofit Medicare education and advocacy group.
EMPLOYER DRUG COVERAGE : KEEP IT OR NOT
About 30 percent of Medicare recipients have drug coverage through a former employer, which is often better than the coverage the government is offering. Think hard before dumping that coverage, experts say. "Retirees who cancel their employer coverage and take the new federal benefit may not be able to go back if they later realize they've made a mistake," said Tricia Neuman, director of the Medicare Policy Project at the Henry J. Kaiser Family Foundation.
WHICH DRUGS ARE COVERED, WHICH AREN'T
Don't automatically assume that the medications you currently take will be covered by a new plan. Each insurer will have its own list of covered drugs, known as a formulary. If you take lots of medications – especially if they're expensive ones – you'll want to carefully go over each plan and make sure those drugs are on the list. Check if your pharmacy is covered too. Something else to consider : Several plans offer discounts for people who use generic brands.
If a drug plan decides to remove a drug you use from its formulary in the future, they must notify you 60 days in advance. You will be able to switch your plan, but you may have to wait till the next enrollment period.
WHEN TO SIGN UP
You can sign up any time from Nov. 15 through May 15, 2006. If you sign up after that, you may face a penalty of 1 percent a month until you enroll. That also applies to people who turn 65 after May and delay signing up. For example, the premium for someone who is 65 or older who waits four years may be 48 percent higher for the rest of his or her life. Remember : This is an insurance plan, and it may make sense to sign up now. Even someone who doesn't use a lot of prescriptions today may come to use more as they get older. You can switch your plan once a year.
IF THE NEW DRUG PLAN IS TOO COSTLY
About one-third of enrollees will qualify for some type of financial assistance. If your annual income is less than about $ 14,000 (or $ 19,000 per couple) and you have less than $ 11,500 in assets ($ 23,000 per couple), not including your home or car, you are eligible for additional assistance. The amount will vary depending on your exact situation. If you think you qualify, you should contact your local Social Security office.
WEIGHING THE PLANS AGAINST MEDIGAP
Medigap is private insurance that seniors can purchase to pay for costs Medicare doesn't cover, which up until now has included prescription drugs. But because the government is heavily subsidizing the new Medicare drug plan and doesn't subsidize Medigap plans, Medigap will prove more costly for most people, and they should switch.
DEMISE OF THE DRUG DISCOUNT CARD
In 2003, after Congress passed legislation for the new drug benefit, seniors were given temporary discount drug cards to carry them over until the plan began. You can continue to use your Medicare-approved drug discount card until you enroll in a Part D plan. But whether you sign up or not, on May 15, 2006, the card will stop working.

FIGURE A.2.: LTW_ENG_20051102.0069

Retirement Journal : Medicare's Part D as Plan B

After a two-year wait, on Tuesday more than 40 million Americans eligible for Medicare can start to sign up for Medicare Part D, the program's voluntary prescription drug benefit. And by Tuesday, a quarter of those people, some 10 million retired American workers, will find out something more specific : whether their former employers will continue to pay for their prescription drug benefits in 2006. Tuesday is the deadline for employers to say "yes," "no" or, perhaps, "maybe."

This group of retirees includes my wife, Sara, and me. In our case, we were waiting for a letter from General Electric Co., where Sara worked for 23 years. When she retired, she signed up for a GE retiree prescription drug plan, which pays 70 to 85 percent of our drug costs. We get our medications by mail and pay $ 25 for a 90-day supply. It's a benefit we did not want to lose.

We began to worry about whether GE would continue its drug benefit when Congress voted in 2003 to let Medicare pay for prescription drugs, which opened the door for some companies to drop their retiree coverage.

The new voluntary drug benefit begins Jan. 1 and is expected to be of financial help to many seniors who have no drug coverage.

Part D allows people on Medicare to get their drugs in one of two ways : by buying an individual drug policy or by signing up with a Medicare Advantage managed care plan. The new drug plans are being offered by private health insurance companies.

Retirees who have approved drug coverage from former employers will not need to buy a Part D plan. So Sara and I were relieved when we received a "yes" letter from GE, telling us that our drug coverage would continue without change. One of the reasons for our concern was that our out-of-pocket drug expenses under the GE plan last year were $ 1,685. But if we had been in a standard Medicare Part D plan, our expenses would have been about $ 5,000. We did not want to go there. The new Medicare drug benefit, expected to cost more than $ 724 billion over 10 years, is being regulated by the Centers for Medicare and Medicaid Services (CMS).

To encourage employers to retain their retiree prescription coverage, Congress offered them a 28 percent tax-free subsidy of the cost of that coverage. But the question was : Would employers keep their drug plans and take the subsidy, or would they dump their plans to save money, knowing that retirees could now get prescription coverage from Medicare Part D ?

The speculation seemed reasonable. The number of corporate employers offering retiree health benefits has declined steadily for 15 years. Part D could have provided an escape hatch for employers who wanted out.

In addition, employers had to jump through some difficult regulatory hoops to prove to CMS that their plans were as good as the standard Part D drug plan. Developing the proof turned out to be a costly, time-consuming process that required employers to hire a small army of actuaries and consultants.

Employers who passed the CMS tests were able to claim they had "creditable coverage," a requirement for getting the 28 percent subsidy.

When Sara and I got our letter from GE, it said :

"GE has determined that the prescription drug coverage offered by the GE Pensioners' Prescription Drug Plan is, on average for all plan participants, expected to pay out as much or more than the standard Medicare prescription drug coverage will pay in 2006."

In Medicare-speak, that meant that GE had passed the CMS tests, would take the subsidy from the government and would continue to provide our prescription coverage. Enclosed in the GE packet was a "Certificate of Creditable Coverage," suitable for framing. GE said we should keep it in a safe place.

The reason for the caution is that people who do not sign up for a Part D prescription plan by May 15 will face a 1 percent per month premium penalty if they sign up later. The penalty will be waived for those who can show they already had "creditable coverage" when the deadline passed.

GE, of course, was not the only employer to face a decision on whether to continue its retiree health coverage. Thousands of corporations and state and local governments had to make similar choices.

While most retirees are getting "yes" letters, others will not be so fortunate. If you receive a "no" letter – meaning your employer's plan is not as good as the standard Medicare plan – and you want to buy better coverage, you will have to sift through dozens of Medicare drug plans offered by health care companies.

If you receive what is, in essence, a "maybe" letter, it may say that your former employer will continue to pay for some of your retiree drug benefits, at least for a while, but that you may have to move into a Part D plan.

How many employers are staying, how many going ?

The CMS says it is too soon to provide a scorecard. But Mark McClellan, CMS administrator, in an interview predicted : "The vast majority of employers will take the retiree subsidy." And that means, he said, that millions of retirees will continue to get drug coverage from former employers in 2006.

Meanwhile, several retired friends told me they, too, had received "yes" letters from their former employers. One, a former accounting manager at GMAC, received a "yes" letter from General Motors Corp., which provides his health and drug benefits. The letter said retirees who want to keep their coverage should not join a Medicare Part D plan. He said he likes his GM health benefits and was happy not to have to join a Part D plan.

I also talked with people at Lockheed Martin Corp. in Bethesda, Md. The company has about 50,000 retirees eligible for prescription drug benefits, according to spokesman Thomas Greer.

Lockheed Martin, he said, will maintain its benefits, take the government subsidy and send out notices to retirees by the Tuesday deadline. "We feel it is important to keep our plans unchanged to ensure stability for our retirees during introduction of Medicare Part D," Greer said.

In some cases, the subsidy will allow employers to lower retiree premiums in 2006. For example, Dow Chemical Co. said it would share its Medicare subsidy money with retirees on a 50-50 basis. As a result, said Janet Boyd, director of government relations, "We expect that our eligible Medicare retirees' 2006 premium will be reduced by about $ 20 a month." The decision to share rather than keep the subsidy money, Boyd said, was not only good for the retirees but also helped increase the company's contribution to the retiree drug plan. By doing that, she said, it made it easier for the company to qualify for "creditable coverage."

Meanwhile, IBM said it would share one-third of its subsidy with retirees. The action will help reduce scheduled premium increases by $ 13 a month for individuals, $ 26 for couples.

The Medicare employer subsidy applies to retiree drug expenses between $ 250 and $ 5,000 a year in 2006. The average subsidy is estimated to be $ 668 a year. By comparison, Medicare will contribute an estimated $ 1,200 a year per person for people who sign up for Part D stand-alone drug plans and Medicare managed care plans.

Thus Medicare will save money if employers can be persuaded to take the subsidy – which by itself will cost $ 41 billion over 10 years. I couldn't immediately find employers who sent "no" or "maybe" letters. But the Fresno Bee reported that retailer JC Penney Co. had decided to discontinue all health benefits for 9,500 retirees who are 65 and older. "The fact that Medicare was implementing prescription drug coverage in 2006 was the impetus for the change," said spokesman Tim Lyon. JC Penney will help retirees enroll in an AARP plan and will pay 55 percent of the premium for one year, he said.

However, dropping out is not going to be widespread, said Jon Gabel, vice president of the Washington-based Center for Studying Health System Change. Dropping a retiree health benefit, he said, is "the very last thing employers want to do." They know, he added, "that it will lead to all kinds of public relations problems."

But employer enthusiasm for the subsidy may be temporary, predicted John Gorman, president of Gorman Health Group, a Washington consulting firm. "In 2007," he said, "you're going to start to see an exodus of employers and unions from their own plans in order to move their retirees into Part D options." Many employers, Gorman added, view the subsidy as "putting a Band-Aid on a shotgun wound. ... It will slow down the bleeding ... but it is not going to stop it."

FIGURE A.3.: LTW_ENG_20051114.0029

SENIORS CONFUSED ABOUT PART D MEDICARE DRUG BENEFIT

SAN BERNARDINO, Calif.
Donald Anglin, 68, takes five different medications, none of which he can recall the names of even though he swallows them on a daily basis.
His left arm is bruised from a recent surgery that introduced pins to his elbow and a plate to his forearm. His ability to speak is often hampered by a stroke that happened several years ago. And his clunky walker appears more like a hurdle than an escort he can depend on for mobility.
"I think the drugs I'm taking are making me an addict rather than actually helping the problem," said Anglin, a retired construction worker who lives in subsidized housing next to the San Bernardino Senior Center. "And with all of these new drug benefits, they're just gonna make things worse. I don't know how these new benefits are going to help me."
Anglin is not alone. Scores of seniors throughout the nation are confused about the new Medicare prescription-drug benefit known as Part D. It's called that because the program is next in line after Part C, another plan that essentially provides HMO benefits to seniors.
The issue with Part D is that enrollment begins Nov. 15, coverage starts Jan. 1, and many seniors have little knowledge about their own prescription drug coverage, let alone a new program that's only months away. The prognosis : Some seniors could be paying more for their drugs without even realizing the scope of Part D.
Only 33 percent of seniors recently surveyed by the Henry J. Kaiser Family Foundation felt they had enough information about the new Medicare benefit. And only 37 percent of those interviewed felt the drug benefit would be "very" or "somewhat" helpful.
At its core, Part D is supposed to help seniors gain control of spiraling drug costs. But the cost of drugs don't seem to bother a group of lunching retirees at the San Bernardino Senior Center. Some of the seniors receive benefits through Medicare-sponsored HMO plans.
"So I'm not going to worry about it," said Kenneth Erwin, 76, whose mailbox has been bombarded by health plans seeking Part D business. "And why should I worry ? This stuff is so confusing and my ability to be analytical is not the same as it used to be."
A cluster of California-based health care companies offer Part D plans including Health Net, Kaiser Permanente, PacifiCare and others. Most of these health plans say their members need not worry about Part D. In fact, many seniors won't even notice a change in their health benefits.
But for those shopping for a Part D plan, there are several variables worth considering. All Medicare beneficiaries, no matter what their income or marital status, are eligible for Part D prescription drug coverage. Medicare patients can obtain the drug benefit through two types of private plans : a prescription drug plan (PDP), or a Medicare Advantage plan (MA), which includes HMOs and regional PPOs.
Those who sign up for a PDP will only get drug coverage and consequently must receive their Medicare benefits via a fee-for-service program. The MA route will cover drugs and all other Medicare benefits.
As for the cost of a Medicare drug plan, seniors will have to pay a monthly premium that is set to cover about 25 percent of a standard monthly premium. In other words, patients may be billed $ 30 a month for plans that cover brand-name and generic drugs. In addition to that, participants will be charged a co-pay of about 25 percent of the drug's price. There is also a standard $ 250 annual deductible, but that only covers $ 2,250 in drug costs. After that's used up, beneficiaries then enter the "doughnut hole."
The hole is a $ 2,850 gap in which the beneficiary has to pay 100 percent of the price of drugs. Once $ 2,850 is surpassed, Medicare kicks back in and pays 95 percent of the cost of drugs.
As confusing as the formula is, health plans are in favor of the program. Jacklyn Kosecoff, executive vice president of specialty companies at Cypress-based PacifiCare, deemed the Medicare overhaul as "the greatest single change in her career."
WellPoint Health Networks also looks favorably upon Part D. The Indianapolis-based company believes that "supporting national solutions for seniors and our aging population is imperative to helping meet the overall immediate and future health care needs of our country," Susan Rawlings, president of senior services of WellPoint, said in a statement.
Of course, many health plans are happy with Part D because it allows them to market a new product and the government has agreed to foot a significant portion of the bill.
But for lower-income individuals who are dual-eligible, meaning they qualify for Medicaid or Medi-Cal and Medicare, Part D deserves more scrutiny than accolades. That's because under the new benefit, lower-income seniors could be forced to switch plans in order to continue receiving the same prescriptions they're used to taking.
"The drug coverage they (lower-income seniors) have will not stop, although the source will change," said Juliette Cubanski, senior policy analyst with the Kaiser Family Foundation, a nonprofit organization that focuses on health care issues. "This is why a lot of people are on alert for what the experience might be for the dual-eligible population."
Anglin is dual-eligible. But he has no idea what the future holds for his health needs. A bulletin board is filled with information about Part D at the senior center he frequents. Incidentally, the information is sponsored by a drug company.
Of the more than 42 million elderly and disabled Americans enrolled in the Medicare program, an estimated 14.4 million will be eligible for low-income subsidies. To qualify for those subsidies, incomes must be below $ 14,355 for individuals and $ 20,000 for couples. Medicare is expected to pay $ 4,189 of drug costs for beneficiaries receiving low-income assistance. Low-income beneficiaries are also expected to spend 83 percent less for their drugs once Part D goes into effect.
But Jo Wales, coordinator of information assistance at the Long Beach Senior Center, isn't convinced seniors will be the ones saving money.
"They're going to be left in the dark. They're going to be up a creek and left without a paddle," said Wales, 71, who is enrolled in an HMO plan. "A lot of people, especially lower-income, are used to having all of their medication paid for. And the fact that they might have to start paying or carrying around a formulary of all of the plans only adds to the confusion."
Jim Anderson, a Pasadena-based spokesman for Kaiser Permanente, said his concern is that seniors will purchase a Part D program, even if they already have the benefit through their Medicare plan. If a senior does purchase a Part D plan and already has the benefit, it will immediately void their initial coverage.
"So we're trying to let people know that we will automatically enroll them in Part D and there is no premium at all," Anderson said.
Despite educational efforts on behalf of health plans and nonprofit organizations, many seniors remain apathetic. Bobbie Foster, 71, doesn't care about Part D.
"I've heard about it. But when I get this stuff in the mail, I just throw it in the trash," said Foster, who carts around an oxygen tank on most days at the San Bernardino Senior Center. "My problem is that I just switched from UHP to Secure Horizons (another HMO) and now I have to pay $ 57 a month to get my tank filled."

FIGURE A.4.: NYT_ENG_20051022.0103

A PLAN FOR YOUR FUTURE

November 15 will never rival July 4 or December 25, but it is definitely a date to remember.
That's when Medicare finally enters the age of wellness and prevention.
When it started in 1965, Medicare was built around episodes of illness. If you developed cancer and needed surgery, Medicare would defray hospital bills. Flu ? Broken hip ? Medicare paid for doctors and nursing home rehab.
But prescription drugs never entered the picture - an omission that grew increasingly glaring as modern science proved more and more effective at keeping people well. It makes no sense to forgo a $ 30-a-month generic statin while shelling out $ 50,000 when a heart attack leads to hospitalization, surgery and rehabilitation.
Politically, drug coverage became a potent campaign issue among the coveted senior vote. The oft-repeated image of a poor widow choosing between food and Fosamax always energized a crowd.
On the other hand, Medicare already gobbled up a huge chunk of the federal budget. With 40-million younger Americans uninsured, could we really afford to beef up insurance benefits to everyone older than 65 ? Even for people who are healthy and wealthy ?
Two years ago, the Bush Administration and Congress crafted an uneasy compromise with advocates like AARP :
Medicare would underwrite a portion of drug costs, but average beneficiaries would still have substantial out-of-pocket bills.
Low-income people would get a better deal than people of means.
The government would not dictate prices or even jawbone discounts out of the manufacturers, as it does with doctors, hospitals and nursing homes.
Instead, insurance companies would act as private drug brokers, setting up plans that would seek discounts and compete for customers.
All of which brings us to Nov. 15.
That's when people on Medicare can start choosing how to arrange their health care for 2006. The new drug benefit - Medicare's biggest transformation ever - complicates those decisions.
The drug program, called Part D, is not just new, it is all over the block. Nineteen companies are authorized to offer Medicare drug plans in Florida, many with two or three options.
Premiums range from $ 10.35 a month to $ 104.89. Formularies cover some drugs, but not others. Then there is the so-called "coverage gap," or more informally, the "doughnut hole," or "Krispy Kreme," a financial Bermuda Triangle that can swallow coverage when total drug costs are more than $ 2,250 and less than $ 5,100 a year.
How complicated is Part D ? Medicare's Web site (www.medicare.gov) contains a list of "frequently asked" questions about the new drug benefit. Earlier this month, it listed 983 "frequently asked" questions.
Such complications aside, there is much to like about these drug provisions - at least from a beneficiary's standpoint.
Medicare is going to pay the insurance companies $ 700 a year for everyone who signs up for a drug plan. That underwrites a lot of drug-buying and still leaves room for profit. Unlike most insurance - where premiums and copayments cover the risk - the bulk of this benefit is financed by taxpayers as a whole, not by the contributions from individual Medicare beneficiaries.
Humana's Standard plan carries a premium of just $ 10.35 a month. Coventry Advantra, PacificCare and WellCare all have plans under $ 20. At those prices, even beneficiaries who take no drugs will benefit from signing up. If they suddenly have a illness, they will quickly earn back the premium. And if their health declines later in life, they can get cheap drugs without paying a stiff premium penalty.
Many low-income people who blanch at a $ 20 premium can qualify for what Medicare calls "special help." This subsidy reduces premiums, deductibles and copayments. Yearly out-of-pocket drug costs will average $ 130 for those people.
By adding hundreds of thousands of Medicare clients to their rolls, the drug plans can buy in bulk and can negotiate discounts from manufacturers. Discounts will come in handy when beneficiaries enter the "coverage gap," and have to pay 100 percent of costs.
The new Medicare also improves HMOs and PPOs.
Their clients cannot enroll in separate Part D drug plans, which are designed to supplement Medicare's traditional Parts A and B. However, Medicare is also beefing up payments to HMOs, to help them stay competitive. In return, the HMOs must enhance their drug coverage or other benefits.
That suits Pauline Policastri, 72, just fine. The Hudson resident and her husband, Salvatore, belong to a Humana HMO. She loves its Silver Sneakers wellness program, which pays for a YMCA membership so she can take dance and exercise classes three days a week.
Like other HMOs, the Policastris' plan charges no monthly premium and provides drugs with minimal copayments.
If the Policastris were to return to traditional Medicare, they would face copayments for doctors and hospitals and an extra Medicare premium if they wanted the new drug coverage.
"It's frightening," she says. "What little pension we have doesn't go up while everything else everything goes up around us. If you get sick and don't have something to protect you, you're in trouble."
Lastly, the Bush administration has a lot of political capital riding on Medicare Part D and has budgeted $ 100-million to inform people and guide them through the process. Among other things, administration officials swear that beneficiaries can get all the help they need by calling a toll-free 1-800-633-4227 hotline. A live person is supposed to answer, around the clock.
Advocates and state agencies figure the drug benefit is a good deal and are devoting a lot of energy to help as well. For example, a two-day senior expo scheduled for Nov. 16-17 at Tropicana Field will feature seminars and enrollment help for Medicare beneficiaries to assess plans, sign up for coverage or apply for low-income subsidies.
So pull up a chair and browse through this issue of Seniority. It is full of useful information about the new Medicare.
And mark Nov. 15 on the calendar. You can't sign up for anything until then. This is just the study phase.

FIGURE A.5.: NYT_ENG_20051028.0370

RETIREE DRUG PLANS URGE CLIENTS TO SHUN MEDICARE

Louis A. deBottari, a 78-year-old retired aerospace worker with good drug insurance, recently opened an envelope from Boeing, hoping to be reassured.
The notice, sent to about 100,000 Boeing retirees and their dependents, discussed the new Medicare drug insurance program that begins in January and what it means for retirees already covered by the company's health plan. "In general," deBottari read, "your Boeing prescription drug coverage is more generous than standard Medicare prescription drug coverage." So far, so good.
But a bit farther down came a warning, if deBottari was nonetheless tempted to give the new Medicare program, called Part D, a try : "Your Boeing prescription drug coverage is part of your Boeing retiree medical plan. If you cancel your Boeing prescription drug coverage, your Boeing medical coverage also will be canceled."
Many other retirees fortunate enough to already have drug coverage are receiving similar warnings from their former employers as the new Medicare drug benefit approaches.
This was not necessarily what Congress intended when it passed the Medicare law two years ago, extending drug insurance to more than 40 million older and disabled Americans. The aim then was to provide benefits similar to those already enjoyed by a more fortunate minority of retirees – about 10 million, currently – who have drug coverage from former employers.
In practice, though, that fortunate minority is worrying about the future of their own benefits. And the message many are receiving from their former employers is a surprisingly stern one : Stay put – or else.
The attitude may be understandable. Many employers generally offer drug benefits within a comprehensive package that also covers doctors' services and hospital care. They typically do not charge a separate premium for drug coverage and do not administer it as a separate benefit. And so they say they have no convenient way to split up the benefits package they now offer.
There is another factor. Like other employers who continue offering drug insurance as good or better than Medicare's, Boeing will receive a federal subsidy for the drug purchases of each retiree it retains. Many of those companies have shown little interest in continuing to provide other, increasingly expensive health coverage to a retiree if they lose the drug subsidy.
And so deBottari is staying put and trying to ignore that big employers like General Motors are trimming their retiree health benefits and that Wall Street analysts are predicting many other employers will eventually scale back or even end their own drug programs, relegating retirees to the less generous Medicare plans they are now being advised to avoid.
By Nov. 15, all employers who offer drug coverage to retirees and employees 65 and older are required to send word to those beneficiaries, indicating how the company's drug plan compares with Part D. Employers must also inform retirees of any changes in their existing health benefits.
In a survey early this year of 458 large employers that had retiree health plans, 82 percent said they planned to continue drug coverage for retirees under those plans, according to the benefits consulting firm Watson Wyatt.
During the current advisory period, many large employers that offer retiree drug coverage – including GM, Caterpillar, Verizon Communications, SBC Communications and the Southern Co. – are, like Boeing, telling their people that the company's plan is as good or better than the Medicare drug program.
And "more so than not, plan sponsors are saying if you join Medicare Part D, you will be out of our plan," said Edward Kaplan, a senior health care consultant at the Segal Co., a benefits consulting firm. A notable exception is GM, which is letting retirees keep their other company health benefits, even if they sign up for Part D.
(BEGIN OPTIONAL TRIM.)
Dow Chemical, for example, is telling its people that they have a choice to make. Low-income retirees may want the new, special Medicare subsidies. But if they decide to go with a Medicare plan, they cannot stay in the Dow retiree Medicare plan, which does include drug coverage. "They cannot be in two drug plans, under CMS rules," said Teri Ferguson, benefits manager for Dow. Retirees at Dow can return to the company plan in 2007, an option that isn't available at some companies.
JP Morgan Chase, in its letter, warns retirees who choose Medicare and lose Morgan benefits : "You may not be eligible to reinstate this coverage following a two-year transition period."
(END OPTIONAL TRIM.)
Staying in a company's plan allows retirees to fend off an often bewildering array of choices touted by insurers and other private sponsors of Part D plans, which are overseen by Medicare but actually provided by private industry. In some parts of the country, older Americans may have up to 45 Medicare plans to choose among, with premiums that average $ 32 a month but whose coverage and co-payments vary widely.
"I don't have to deal with Medicare Part D and I'm delighted," said Marlene Barnes, 69, a retired federal office worker in Memphis, Tenn., whose health benefits are provided by the Federal Employees Health Plan.
People who do want to sign up for a Medicare drug plan can do so penalty free from Nov. 15 to May 15 ; there could be increasingly stiff late-fee penalties for those who delay. But the penalties will not apply to retirees who currently have solid employer-sponsored coverage but who shift later to a Medicare plan – if, say, their current employer-sponsored plans are curtailed or ended.
The number of companies providing retiree health benefits has been shrinking for decades, as rising costs for drugs and other health needs have cut sharply into profits. And benefit consultants say that many employers, after seeing how things work out with the Medicare subsidies in 2006, will be weighing possible changes in their retiree benefits for 2007 and beyond.

FIGURE A.6.: NYT_ENG_20051103.0182

MEDICARE PLAN OFFERS A LOT TO FIGURE OUT

<DATELINE> ATLANTA </DATELINE>
Reuben and Claudia Beck understand complex things.
Reuben, 84, who was a sailor in World War II and an engineer for 40 years, designed conveyor systems for automobiles before retiring at age 70. Claudia, 66, a registered nurse, is a retired insurance claims adjuster who knows how to slash through red tape.
Even so, the Becks fear they have met their match in Medicare's new prescription drug benefit plan, designed to help seniors pay for medicine.
"It's just so confusing," Claudia Beck said. "And it seems the more I call, the more runaround I get."
The couple, along with 42 million other seniors and disabled citizens eligible for Medicare, will have access to a benefit that seniors have wanted for a long time.
And there's no doubt the prescription drug program will help millions.
First, however, they must survive the jungle of jargon and confusing array of options that are tripping up thousands already – and the benefit hasn't even started yet.
Medicare is the government program that provides health insurance for Americans 65 and older. It has not previously covered outpatient prescription drugs.
But in 2003, Congress passed a law that will make prescription drug coverage available to those on Medicare starting Jan. 1.
Private insurers are providing the coverage. Those eligible for the benefit can pick an insurance plan from dozens being marketed to them. Seniors generally will pay premiums and co-payments as with most insurance policies.
But the confusion begins right at the beginning – when seniors try to pick a plan. In Georgia, there are 82 plans offered by 18 companies.
Deductibles, co-payments, out-of-pocket expenses and covered drugs vary among the plans. Imagine 82 plans with, say, five variables – premium, deductible, co-payments, drug list and pharmacy choice. The variables become mathematically if not emotionally daunting.
"The new Medicare drug benefit is 10 times more complicated than the Medicare drug discount card," said Robert Hayes, president of the Medicare Rights Center, a New York-based consumer advocacy organization, "and 10 times more important."
Even after comparing premiums, co-payments, deductibles and rosters of drugs, the work will still not be complete. Seniors will also need to determine how they'll stomach what's termed "the doughnut hole" – a gap where there's no coverage after seniors rack up drug costs of $ 2,250.
Under most drug plans, patients will have coverage up to $ 2,250 in spending. But between $ 2,250 and $ 5,100, they'll generally be on their own. Medicare will pay 95 percent of covered drug costs above $ 5,100.
Worth looking into
Despite the difficulties, it's worthwhile for seniors to explore the new $ 700 billion benefit, especially those with limited incomes, those who lack drug coverage now and those with high prescription costs, experts said.
The best way for seniors to compare plans is by using Medicare's Web site or calling Medicare or a local insurance counseling service to use the Internet tool for them.
Gary Karr, a spokesman for the federal agency that runs Medicare, acknowledged "a little bit of anxiety" among consumers over the benefit. But the Centers for Medicare and Medicaid Services also emphasized there's no immediate rush to sign up, that Nov. 15 is just the starting point.
"People do have time to take a deep breath," said Dr. Mark McClellan, the Medicare agency's administrator. "They don't need to rush into a decision."
Enrollment is open between Nov. 15 and May 15. Coverage for those who enroll before Dec. 31 will begin Jan. 1.
If you wait until after May 15 to enroll, you might have to pay extra to join a plan.
The Medicare agency said seniors now can narrow down their choices based on what's important to them – premium, deductible, available drugs, or other factors.
McClellan said CMS had added thousands of customer service representatives, and that during a recent week handled more than 700,000 calls "with no or minimal delay."
The competition among drug plans has helped consumers, he said. "What we're seeing is that drug plans are coming in at a lower cost and with better benefits than were expected." Some plans have no doughnut hole or no deductible, he said.
"This benefit is really worth it," McClellan said.
Can't call everybody
The Becks have struggled to find information about the 82 drug plans and analyze which one might work best for them.
For example, the Medicare pamphlet about the program could not list all the provisions of the 82 plans listed in Georgia.
"Am I going to have to call all 82 companies ?" Claudia Beck asked one day last week.
And the Becks wonder : If figuring out a prescription drug plan is this hard for them, how hard will it be for millions of others not as savvy or healthy ?
Such insecurity among seniors appears to be widespread, according to Lisa Federico, coordinator of GeorgiaCares, a nonprofit insurance counseling service. Her office is receiving about 300 calls a day from seniors frantic for help. The staff can process about 80 of them, she said, creating a pileup of 220 calls a day and putting them at least a week behind.
"I've never seen anything like this, just the level of confusion this is generating," Federico said.
After some delays, the Medicare Web site (www.medicare.gov) for the drug benefit has finally opened, but still isn't fully functional, with a key piece of the "Prescription Drug Plan Finder" tool not working. It frustrates counselors trying to help seniors. The Medicare agency said the site would be completely functioning "in the next few days."
Even then, for many reasons, seniors will need insurance counseling, Hayes said.
"There will be insurers who will promote a low premium and disguise the extremely high co-payments," Hayes said. A senior can find out what drugs are covered on a plan's formulary, he said, "but people aren't being told how much they are going to pay" for the drugs.
State government retiree Robert Hauss, 75, said he feels a time crunch on his drug plan decision.
State retirees were told they must make a decision about keeping their current state drug plan by Tuesday, but Hauss said he didn't get his booklet from the state detailing the changes until Oct. 27. Other state retirees have also complained to GeorgiaCares about a tight deadline for the state plan.
On Friday, Gov. Sonny Perdue announced that he was extending the state's enrollment deadline 10 days, to Nov. 18.
Hauss, meanwhile, said his retiree health plan will be about twice as expensive as its current cost if he doesn't join a Medicare drug plan. "The state is forcing us to go into Medicare Part D," he said. "I'm not happy with the changes."
But a spokeswoman for the state agency overseeing the retiree plan denied the state was pushing people into Medicare drug coverage.
"They have choices," said Julie Kerlin, spokeswoman for the Georgia Department of Community Health. "Employees make choices every time there is an open enrollment."
Hauss takes seven medications for chronic conditions, and worries that not all his drugs will be covered as they are now. "Medicare Part D plans may not have the same drug formulary as the state health plan," he said.
Hauss isn't hooked into the Internet, so he's waiting for insurance counselors to call back.
Frustration is high
Jim and Marlys Antonoff have spent hours compiling information on Medicare's drug benefit. They've neatly filed brochures and information from seven drug plans in a plastic container. They've called the insurance companies – even politicians – seeking answers.
Despite the painstaking work they have done to understand Part D, they're still frustrated by its complexity.
And they find themselves chuckling about the string of roadblocks they run into.
One roadblock : Three drug plans' phone numbers listed in the Medicare handbook are wrong, said Jim Antonoff.
Virginia Anderson and Andy Miller write for The Atlanta Journal-Constitution. E-mail : landerson@ajc.com ; jamiller@ajc.com

FIGURE A.7.: NYT_ENG_20051105.0111

SIGN UP BEGINS FOR MEDICARE PRESCRIPTION DRUG COVERAGE FOR MONDAY, NOV.14

Cox News Service
WASHINGTON - Nearly two years after President Bush signed into law the most sweeping change in Medicare's history, beneficiaries can begin signing up Tuesday for coverage of their prescription drugs.
Despite recent polls showing widespread confusion about how the benefit will work and concern that the coverage choices are too complicated, administration officials say they are confident the drug program will catch on.
By the time the open enrollment period ends next May 15, beneficiaries will have a better understanding of the program and "those polls will change," said Health and Human Services Secretary Michael Leavitt, in a conference call with reporters last week.
"This is the beginning of a new era in Medicare," Leavitt said. "We will slowly build understanding."
Administration officials stress there is no need for beneficiaries to hastily enroll. The drug benefit will not take effect until Jan. 1, 2006, and beneficiaries can continue signing up until May 15, 2006, without paying a penalty. But those who are eligible and do not sign up by then will pay a 1 percent a month penalty when they do enroll. The next enrollment period after May 15 will be in November 2006.
With the enrollment period about to begin and beneficiaries being flooded with information from Medicare and dozens of prescription drug providers, interest in the program is quickly growing.
"More and more people are starting to pay attention." said Lisa Nora, coordinator and volunteer specialist at the Medicare Answers Prescription Savings Center in West Palm Beach, Fla. "They know that they have to make a decision."
Nationwide, the federal government is working with about 10,000 local organizations to provide assistance, Leavitt said. That help will be needed, experts say, because the program is so complicated.
Advocates of a drug benefit for the elderly had lobbied for years to expand the program within Medicare, but when Congress finally acted, it tossed the benefit to the private sector with a hefty government subsidy. The result is that beneficiaries must enroll in private plans - not in Medicare directly - for the prescription drug benefit known as Medicare Part D.
In virtually every state, beneficiaries are confronted with more than 40 competing drug plans, many of them different options offered by the same company. It's like deciding which cell phone service to buy from competing companies and then selecting a particular calling plan within the chosen company.
A survey of Medicare beneficiaries late last month found that more than 60 percent said they do not understand the new drug law.
The survey, conducted by the non-profit Kaiser Family Foundation and the Harvard School of Public Health, also found that 37 percent of beneficiaries said they do not plan to enroll in a plan ; 43 percent were undecided and only 20 percent said they planned to enroll for the drug benefit.
Most elderly who have not followed the issue are unaware that they have so many plans to choose from. When told they have so many choices, nearly three-fourths said that "makes it confusing and difficult to pick the best plan" while 22 percent said it "provides an opportunity to choose the best plan."
Critics of the drug law have argued for months that it is too complex. They also have argued that if Medicare purchased drugs directly, the way the Department of Veterans Affairs does for veterans, beneficiaries would pay much less.
Ron Pollack, executive director of Families USA, which opposed the drug bill when it was in Congress, said its flaws "are producing bewilderment and confusion among seniors" and will also result in a far more costly program.
But Dr. Mark McClellan, head of the federal Centers for Medicare and Medicaid Services, said competition among the private plans has resulted in lower premiums, lower deductibles and lower copayments than had been predicted.
A study of the prices scheduled to be charged under the drug law for the top 25 most commonly prescribed medications used by the elderly found that the drug benefit will result in average savings of 20 percent on brand name drugs bought at a retail pharmacy and 27 percent through a mail order pharmacy.
For generic drugs, the savings are 44 percent for retail and 64 percent for mail order. The study, based on filings with the Medicare program, was conducted by the Pharmaceutical Care Management Association, the national trade organization representing pharmacy benefit managers.
McClellan said he was unconcerned that more than a third of Medicare beneficiaries say they don't intend to enroll in a drug plan because most of them have more generous coverage through a former employer. The Medicare drug law provides substantial incentives for employers to continue to offer retiree drug coverage.
Of the more than 40 million elderly and disabled Medicare beneficiaries nationwide, administration officials expect about 28 million to 30 million to enroll in a drug plan or remain in a federally subsidized retiree plan.
That includes about 10 million retirees with employer-sponsored coverage, 12 million very low income people whose drug coverage automatically will be switched from Medicaid to Medicare, and 6 million in Medicare managed care plans that already provide drug coverage.
To help beneficiaries wade through the various plans, Medicare has created a series of tools that can compare costs and benefits. Those tools - available on the Internet at www.medicare.gov or by calling the Medicare hotline at 1-800-633-4227 - allow beneficiaries to list the specific drugs they take and find out which plan will give them the best deal in terms of cost and convenience.
The tool enables beneficiaries to compare premiums, deductibles and copayments for each plan based on the drugs listed. It also tells them which local pharmacies will accept the various plans, or whether drugs can be purchased through a bulk mail-order or online supplier.
"It's a wonderful tool," Nora said. "It really gets in depth on how much you are going to save. You plug in (the information) and it's the best thing since sliced bread."
Drew. E. Altman, president of the Kaiser Family Foundation, said it's unclear how many beneficiaries will enroll in a plan, but he noted that "in the long run the bigger question is whether seniors believe they are getting ?enough help with their drug costs from the plans they are in."

Larry Lipman's byline is larryl@coxnews.com

FIGURE A.8.: NYT_ENG_20051111.0050

CONTACTING MEDICARE : A FEW TIPS

Medicare has a toll-free hotline and a Web site to help people sort through the new Part D drug coverage. Both options can be frustrating as well as helpful. If you want Medicare to help you compare drug plans, here are a few suggestions.
BY TELEPHONE
Medicare has hired 8,000 people to answer questions about the drug benefit and they are still learning their jobs. So when you call, try these tips.
1) Have your Medicare card and a list of all your medications handy. You will be asked for this information.
2) Call toll-free (800) 633-4227. A recording will offer you some phone-tree options. Choose "drug coverage." 3) After listing some pertinent dates, you will be asked to select either "enrollment" "information" or "publication." The fastest way to talk to a live person is to select "enrollment." After giving some of your personal information, select "agent" the next time you have an option. That should connect you to a live person.
4) If you pick the "information" option, you may find yourself in a loop of various recordings. When you are ready to talk to a live person, either return to the main menu and start over, or select the "status" option and follow that with the "agent" option.
5) Tell the agent you want a list of stand-alone drug plans and their costs to you, based on your particular medications. Make sure the agent types in your specific medications before comparing plan costs.
6) If the agent asks if you want to limit your premium and deductible cost, say "no." It may sound cheaper to pick plans with zero deductibles or a low premiums, but those plans may cost you more in the long run because they cover fewer drugs. Tell the agent to compare all plans without restricting the premiums or deductibles.
7) When the agent lists three or four plans that seem to be the cheapest, find out what each plan will cost. Ask if those plans place restrictions on your drugs. If they all do, ask the agent to look for a plan that places fewer restrictions, even if it costs you a little more.
8) Write down your top three or four plans and contact those companies for more detailed information.
9) Ask the agent to mail you a written brochure covering the information you discussed.
ONLINE
Be prepared to spend at least 30 minutes on this search.
1) Go to www.medicare.gov and click on "Compare Prescription Drug Plans."
2) About two-thirds of the way down the next page, click on the orange button next to "Find a Medicare Prescription Drug Plan."
3) Next page : Fill in your personal information and click on "search plans."
4) Next page : Fill in parts A, B and C and click on "continue."
5) Next page : Click on "choose a drug plan type."
6) Next page : Click on C "search for Medicare prescription drug plans."
7) Next page : Click on B "Enter my medications." (Do not click on C "limit your drug plans." If you do, it may exclude the best and cheapest plan for you.)
8) Next page : List your medications into A "find your drugs by name." When you are done, click on B "continue with selected drugs." Then click on C "choose my drug dosage" and use the drop down list next to each drug to enter your dosage. Then click on B "continue with selected drugs" then B "continue with plan list." (Skip the pharmacy option for now.)
9) The next page lists the cheapest five plans in order of cost. Go to a drop-down box at bottom right and select the "All" function. That will let you peruse all the plans more quickly.
10) To the right of each plan is a drop-down box. Go to the first plan and select "view cost details" in the drop-down box. Note that it tells you how to save money by using a 90-day mail-order pharmacy. Also, see if your drugs have asterisks next to them. That means that the plan somehow restricts their usage (more about that later).
11) Go back to the plan list and click on "lower my cost share" in the drop-down box. This lists possible substitute drugs and how they will lower your cost. If you click on the red link to each substitute drug, the computer will identify the substitutes so you can ask your doctor if you can safely switch to them.
12) Go back to the plan list and click on ". of pharmacies" next to the first plan. This shows the pharmacies that will accept this plan. The list is preset to show pharmacies within 2 miles of your ZIP code. A drop-down box lets you widen that distance and add more pharmacy options.
13) Repeat steps 10 through 12 with each plan that interests you. Then contact the plans for full details of costs and any restrictions on your drugs. You can find plan phone numbers by going to the plan list and clicking on the plan name.
14) For more information on drug restrictions, go back to Medicare's home page and click on "Formulary finder." Follow that thread to the "plan list." In the "Plans per page" box at bottom right, change the "5" to "All" and scroll down to the plans that interest you. (All stand-alone drug plans will have "PDP" in the third column. Click on those plans for more information on how they restrict access to your drugs.)

FIGURE A.9.: NYT_ENG_20051112.0134

A LOT OF INFO, IF YOU CAN GET TO IT

TARPON SPRINGS - Robert Loos, 83, can Google with the best of them.
He searches travel deals on the Internet, pays bills online and stores crucial phone numbers in a digital address book.
So with enrollment in Medicare's new Part D drug benefit to begin on Tuesday, Loos powered up his Compaq 7550 last week and headed for the government's Web page.
In Florida, 19 companies are marketing 43 Medicare drug plans, and that doesn't count the various HMOs and PPOs that people can join.
Medicare has never been so complicated, so full of choices. It's a classic chance for computer-savvy people to bypass commercial advertising and free-lunch seminars by searching www.medicare.gov on their own.
Medicare literature advises people to do just that.
But Loos fiddled at it for an hour Thursday with only mixed success. The Web site was often ponderous. It answered some questions, but rarely quickly.
Things got worse when Loos called Medicare's toll-free hotline, 1-800-633-4227. He meandered around in an irritating phone-tree loop for about five minutes until he finally hung up and called back. The second time, he found his way to one of Medicare's 8,000 new "customer service representatives," a polite, patient man who turned out to be a fount of misinformation.
Medicare beneficiaries need not panic. They have six more weeks to choose their 2006 coverage before it begins on Jan. 1. That also gives Medicare's designers time to tweak the Web site and hotline employees time to learn their jobs.
Still, Loos worries about his fellow retirees, especially those who don't use computers.
"I can't see how older people are ever going to determine what the best plans are," he says. "They are going to be very confused."
Loos, who owned a trucking equipment business before retirement, does not want to join a Medicare HMO, which usually carries drug coverage. He and his wife, Eileen, don't want to change doctors if an HMO shuffles its provider network.
The seven drugs Loos takes would cost more than $ 8,000 a year if he bought them at his local Publix pharmacy. Instead, he buys one from Canada, one from New Zealand and gets two free from Merck's patient assistance program. That winnows his bill to about $ 2,700 a year.
He would like to join a new Medicare Part D drug plan because he thinks the government is going to crack down on foreign drug importation.
When he checked www.medicare.gov, his high-speed Internet connection moved him quickly through preliminary questions until he was asked to enter his seven medications. Then it slowed to a crawl by forcing him to register his medications one at a time instead of all at once.
At one point, he tried to register Flomax and the computer said "service unavailable." He back-tracked, tried again, and it worked the second time.
"It's very, very time-consuming," he muttered.
After about 15 minutes, the computer finally listed how much each plan would cost, from cheapest to most expensive, ranging from about $ 3,600 to more than $ 8,000. When he clicked on other buttons, Medicare's Web site showed him how he could knock about $ 500 off the cheapest plan - Humana Complete - by using its mail-order pharmacy. He could knock off $ 770 more by swapping his drugs for similar ones preferred by Humana.
That brought the cost to about $ 2,400 - less than he pays now.
"Boy, there's a lot of stuff here," he said. "So far, it's working pretty good."
After an hour, Loos was ready to quit, even though he still lacked price-reduction information that might make other plans more attractive. The Web site would only let him work through one drug plan at a time, which slowed him down.
He also wanted to check which plans restrict how his doctor can prescribe his drugs. The Humana Complete plan, for example, required either prior-authorization, quantity limitations or "step therapy" on every one of his seven drugs.
Maybe cheapest wasn't best.
Medicare's Web site can provide details about these restrictions. But Loos had previously played around with that section of the Web site and found it confusing and time-consuming.
Perhaps Medicare's toll-free hotline could help. Medicare literature urges people to call, particularly if they don't have access to a computer.
Medicare has 43-million beneficiaries, so it was no surprise that Loos was greeted by a recorded female voice guiding him through a phone tree.
When she asked if he wanted "information," that seemed like a good choice, but it sent him into a loop of various messages and didn't seem to tell him how to reach a live person.
After five minutes, he hung up and tried again. This time, he tried other options and a customer service representative came on the line. Loos asked what various drug plans would cost.
Would Loos like to keep his premiums and deductibles low, the man asked. Deductibles range from zero to $ 250 a year.
"I'd like to pay nothing if I can," Loos responded. As for premiums, maybe something like $ 25 to $ 30 a month.
It was a mistake on both their parts. Given Loos' seven medications, the cheapest plan is Humana Complete, with a $ 61-a-month premium. Its total cost is less because it offers broader coverage that more than makes up for the higher premium.
The Medicare worker was following a script, as trained. But when he limited Loos' premiums to $ 30 or less, the computer spit out a different Humana plan, an AARP plan and a United HealthCare plan as the three cheapest.
Then the worker inaccurately said each of the three plans carried co-payments of 25 percent. That's the amount for a hypothetical boilerplate plan that Medicare uses as a guideline in its literature. But it bears no relationship to actual co-payments.
"Who determines the cost of prescriptions?" Loos asked.
"I guess the pharmacist," said the worker, when in fact the companies that offer the plans determine the cost of prescriptions.
Medicare spokesman Peter Ashkenaz said hotline workers are trained for at least a week before answering calls and receive at least two more hours a week of ongoing training. Supervisors monitor about four calls a month for each worker. The part of Medicare's computer program that lists the cheapest plan was installed last week, so the worker who helped Loos had only a few days of experience with it, at best.
Medicare hopes hotline calls will average about 10 minutes, Ashkenaz said.

Loos' call took 30 minutes.

FIGURE A.10.: NYT_ENG_20051112.0174

Jeu de documents de mise à jour

COVERAGE ENROLLMENT FOR 2006 BEGINS TODAY WHAT IS SPECIAL ABOUT TODAY ?

This is the first day people on Medicare can determine their health coverage for 2006 - whether they receive care through Medicare's traditional Parts A and B or through a Medicare HMO or PPO. For the first time, people on Parts A and B also add a Part D drug benefit to help pay their prescription bills. To get Part D coverage, you must sign up with one of 19 private insurance companies offering 43 different plans in Florida. Today is the first day you can enroll in a Part D plan or in a Medicare HMO or PPO.
When does Part D drug coverage begin ?
Coverage starts Jan. 1 if you have signed up by the end of this year. You can also sign up between Jan. 1 and May 15, but the drug plan will not defray your costs until you have signed up.
How do Part D drug plans work ?
The company negotiates discounts from drug manufacturers and resells to you, usually at a below-market price. You will pay a monthly premium to the company and receive a drug plan card to show your pharmacist when you fill prescriptions. The company will charge you an annual deductible, if any, and copayments. Medicare will reduce these costs by subsidizing the company for serving you.
How much do plans cost ?
Your costs will vary widely from plan to plan. Premiums in Florida range from $10.95 to $104.89 a month, deductibles from $0 to $250 a year. Often coverage ceases when total drug costs hit $2,250, then begins again when total costs hit $5,100. Copayments can range from $0 for a 30-day generic to $75 for a brand-name drug. There is no free lunch. Plans with broader coverage charge higher premiums.
How do HMOs fit in all this ?
If you want to get your hospital and doctor care through a Medicare HMO or PPO, it will also provide your drug coverage, and usually more cheaply than traditional Medicare. Part D drug plans only accompany Medicare's traditional Parts A and B, where you pick your own doctors and hospitals. Warning : You can't have both a Medicare HMO and a Part D drug plan. If you sign up for both, Medicare will disenroll you from the HMO and put you on traditional Medicare Parts A and B. Some companies, like Humana, Blue Cross-Blue Shield and WellCare offer both HMOs and Part D plans so be careful what you sign up for.
I'm healthy. Must I get drug coverage ?
No. Just like Part B doctor coverage, the new Part D drug coverage is optional. However, if you reject coverage now, Medicare will charge a penalty if you change your mind later. Monthly premiums will rise 1 percent for each month you reject coverage. If you have no drug coverage now, it's probably wise to enroll in Part D. You can join a plan for as little as $10.95 a month. For that, you get coverage if you suddenly become sick and you avoid premium penalties later in life, when you are taking drugs.
What if I can't afford the premiums ?
If you qualify for Medicare and already receive drugs through Medicaid or state assistance programs, you will automatically be enrolled in a Part D drug plan without having to pay premiums or large copayments. As of Jan. 1, Medicaid will no longer provide drugs to people who also qualify for Medicare. Even if you are not on Medicaid, you might qualify for reductions on premiums, deductibles and copayments if your income falls below $14,356 ($19,246 for a couple) and you don't own a lot of liquid assets such as savings, CDs and stocks. Some income is not counted, so if you are anywhere close to qualifying, check with Social Security toll-free at 1-800-772-1213.
How do I sign up ?
If you want to join an HMO or PPO, contact them directly. If you want to add a Part D drug plan to traditional Medicare Parts A and B, you can contact the private drug plan directly or sign up online at www.medicare.gov or by calling toll-free 1-800-MEDICARE (633-4227).
How do I pick a plan that is right for me ?
The Medicare & You 2006 booklet that Medicare recently sent you lists information about drug plans and company phone numbers. The people who answer 1-800-MEDICARE can give give a list of the three cheapest plans based on your medications. Or you can research that directly at www.medicare.gov. Florida's SHINE program can give one-on-one advice if you call toll-free 1-800-963-5337. Expect to leave a message, but a SHINE volunteer will call you back. Information on companies is also available on www.sptimes.com/medicare.
What are the pitfalls ?
If you take very many medications, be wary of picking the plan with the lowest premium. You often can save money by paying more for broader coverage. Make sure the company includes your drugs on its formulary and will let you use the pharmacy of your choice. Many plans require prior authorization or restrict prescription quantities for some drugs. Ask each company if they restrict your drugs and how. Consider paying a slightly higher premium for a plan with fewer restrictions on your specific drugs.
What if I get drug benefits from my ex-employer's or union's retiree health plan ?
Retiree plans offered by employers and unions are often cheaper than Part D plans. You probably should keep these retiree benefits as long as Medicare deems them "creditable coverage." If so, you can always switch to a Medicare Part D plan later without paying any penalty. Ask your ex-employer or union whether your retiree plan qualifies as "creditable." If it isn't, you probably should switch to a Medicare HMO or a Part D drug plan to avoid paying a premium penalty later on.
Will manufacturers still offer free drugs ?
Many "patient assistance" programs for low-income people will end Jan. 1. If you rely on one of them, contact the manufacturer for details.

FIGURE A.11.: NYT_ENG_20051115.0033

Résumés générés par les différentes versions de CBSEAS

Leap in the Dark

The following editorial appeared in Tuesday's Washington Post :
Over the past few days, the Medicare.gov Web site has been jammed. Tuesday, it could grind to a halt. That's because it's the first day of open enrollment for Medicare Part D, the Medicare drug benefit for seniors. Already, anecdotal and polling evidence shows that many seniors are baffled by the benefit, which requires beneficiaries to choose from several dozen private plans. More than six in 10 told a Kaiser Family Foundation pollster that they understand the new drug benefit "not too well" or "not at all." Only one in five said that they plan to enroll. Worse, more than three-quarters said they have never used the Internet, although the Medicare.gov Web site provides the most efficient means of comparing the different plans.
All of that bodes ill for a benefit that its creators have promoted as the wave of the future : the first example of 21st-century, public-private health cooperation. But as seniors start signing up – and they have until May 15, when a late fee kicks in – it's important not to judge the plan by the inevitable initial glitches and anticipatory fears. For one, it is possible that seniors, by gravitating toward the simplest plans, will force the system to become simpler.
In the long term, the benefit's success will be measured in at least two ways. The first is by cost : to seniors and to the government. At the moment, the average senior will pay about $32 in monthly premiums, some $5 less than predicted. The average government subsidy per person is running at $94 monthly, rather than $109. These savings are the result of better-than-expected competition among private insurance companies, many of whom got better-than-expected deals from drug manufacturers. The real test, though, is whether those are lasting savings or whether the insurers will find it necessary to raise costs over time.
The other (related) measure of success is how many seniors enroll. Former Louisiana Sen. John Breaux, D, one of the benefit's congressional authors and now chairman of the Medicare Rx Education Network, reckons that the benefit will be a success if half of eligible seniors enroll. But that's just an estimate : The real test is whether enough healthy people stay in the program to keep insurance companies interested in offering it and to keep premiums low.

FIGURE A.12.: LTW_ENG_20051115.0016

Seniors Find Medicare Drug Plan Options Bewildering

<DATELINE> WASHINGTON </DATELINE>
They came equipped with notepads of questions and lists of medications, and with the thick red-white-and-blue handbook "Medicare & You." Nearly 200 strong, they came looking for help in figuring out which among dozens of bewildering plans would buy them the most medicines for the least price.
Several hours later, many left a little less confused – but not necessarily any closer to a decision.
"It's wearing me out," Betsy Curtin confessed as she and her husband departed the James Lee Community Center near Falls Church, Va. She sounded almost indignant about the complexity of the new Medicare Part D Prescription Drug Coverage, the federal government's effort to offer millions of older Americans a way to better afford their pills.
"I'm a college graduate. I worked about 20 years for the Air Force," Curtin said. And she has a computer and some savvy using it. And still, papers in hand, she grimaced. "It's challenging."
Just days into the launch of Medicare's biggest expansion since its inception 40 years ago – a program that could run taxpayers $720billion in its first decade – seniors and the people trying to advise them are incredulous about the multiplicity of options, deductibles, premiums and exceptions, which make comparisons seem overwhelming, if not impossible.
They say the same thing, whether in information sessions large and small or in one-on-one appointments. "It's been driving me crazy," said Rita Reitman Sobel of Alexandria, Va., her face betraying her doubt and anxiety after Thursday's presentation.
What she and the rest of the audience heard was hardly reassuring. Although Part D will greatly benefit certain people, its savings are not universal. The annual costs of its various plans, which are being marketed by private companies, can vary by thousands of dollars, and a penalty could multiply many beneficiaries' bills by thousands more if they sign up after the deadline.
And the best way to sort out the competing information ? Via the Internet – and searches that most seniors say are beyond their ken.
Enrollment started Tuesday and continues, for those 65 and older, until May 15. Actual coverage begins Jan.1.
The program's arcane intricacies and implementation, preceded by a barrage of advertising, have been so criticized that some lawmakers are pushing for a two-year delay.
"This is a bad piece of legislation, and it never should have been enacted," Rep. James P. Moran Jr., D-Va., told the community center crowd, which responded with applause.
The program is too complicated, he said, too expensive and "too much of a giveaway to the pharmaceutical companies."
Yet, Moran added, "I want to emphasize, I don't think you have a choice. You've got to sign up for it."
Questions and concerns took on greater urgency after enrollment officially commenced this week. Calls to the Medicare Rights Center, an independent, New York-based organization, quadrupled over previous weeks' volume.
"People know the marketing materials arrived in the mail ... that there's a benefit, but whether it's right for them is a complete mystery to them," spokeswoman Deane Beebe said Friday afternoon. "We're still trying to figure out what to tell them."
In northern Virginia, the Fairfax Area Agency on Aging has trained volunteers to assist local residents who walk through the door, sometimes carrying a shoebox of pill bottles. The seniors listen carefully but, invariably toward the end of the conversation, lean in close and lower their voices.
"They say, 'What would you do ?"' program coordinator Howard Houghton said. When the counselors explain that they can't cross that line, their clients try a different approach : "'What would you tell your mother to do ?"'
Around computers in an Alexandria apartment complex's community center, several seniors paired Friday with workers from the city's Office of Aging and Adult Services and tried to make sense of the future.
Social worker Irene Ilachinski sat with Maria Tarzi and looked over her list of medicines. "I don't know if there's a plan that takes all of them," she said, "but we'll see what's available at least."
Ilachinski called up the online Medicare Prescription Drug Plan Finder and began clicking on specific brand names : Zocor, Plavix, Nexium and a half-dozen more. Thirty-five plans popped up.
A few feet away, MaryAnn Griffin typed in the same details and appropriate dosages for Tarzi's cousin, Mahboub Rafik, and then hit enter. The results : 40 possibilities, from a Humana plan with annual costs of $3,880.66 to a plan from PacifiCare that might run $10,257.12 a year.
"Well," Griffin said to Rafik, "we're not going to take that one, are we ?"

FIGURE A.13.: LTW_ENG_20051119.0012

SENIORS LOOKING FOR ADVICE ON MEDICARE DRUG COVERAGE

<DATELINE> SACRAMENTO, Calif. </DATELINE>
Enrollment in Medicare's new drug coverage started Tuesday, and seniors are clamoring for advice on how to navigate a crucial program that offers more than 50 options.
Joan Parks, director of Sacramento's Health Insurance Counseling and Advocacy Program, or HICAP, said her office is getting more than 200 phone messages a day – far more than the staff of eight can quickly return.
"We're just deluged," said Parks, whose program provides free, unbiased consumer advice. "But I refuse to sink."
Medicare's new drug coverage, called Part D, goes into effect Jan. 1. During this initial enrollment, the state's 4.3 million beneficiaries have until May 15 to make a choice.
A HICAP-sponsored information session Tuesday at the Elk Grove Senior Center drew an audience of 85. Questions, and concerns, were plentiful.
"I take about 20 medications, and half of them are extremely expensive, so I am really worried about this," Helen Casillas, 64, of Elk Grove told the counselors.
Later, the retired Realtor said, "I am terrified by it. One of my drugs alone is $200, and others are nearly that high."
In addition to presentations to senior groups across the region, HICAP's staff and 40 trained volunteers provide one-on-one counseling. But seniors and the disabled may have to be patient.
Because of the number of calls, Parks said, it may take her staff five days to return calls. Some are calling more than once when they don't get a return call the same day, adding to the volume.
"We're doing our best to meet that need," Parks said. "They may get a voicemail."
She said people should call once and "know we will call them back."
Parks said HICAP isn't taking on new volunteers at this time because she doesn't have the staff or time needed to train them on the complicated new coverage.
She encouraged Medicare beneficiaries to attend a community presentation to learn the basics.
"There's a lot of resources at those events and they're constantly bringing in the new information," Parks said.
Casillas said she attended Tuesday's Elk Grove session because she is having trouble getting her questions answered.
"The worst part is that the people in charge are just learning it," she said. "They know very little. I've been calling them for a while."
Medicare officials are encouraging people to call the (800)MEDICARE telephone assistance line or go online to get enrollment information. The challenge is for beneficiaries to figure out which of the 47 stand-alone drug plans or seven Medicare managed-care health plans available in Sacramento is right for them.
Jack Cheevers, spokesman for the federal Centers for Medicare and Medicaid Services in San Francisco, said there has been a "robust" use of Medicare's Web site at www.medicare.gov, which offers details of the plans.
But Parks said many of those seeking advice are older people without computer access or skills to use Internet search tools to compare plans. They may not have family or friends to turn to for help.
HICAP receptionist Denise King said some seniors have complained to her they are put on hold or transferred when they call Medicare, and do not get their questions answered.
There's a wait for one-on-one HICAP counseling appointments. While some feel rushed to make a choice, others may wait until the last minute.
Michael Negrete, an official with the California Pharmacists Association, spoke at a community event Monday night to residents of a gated mobile-home park. Only 20 of the 200 residents invited actually attended, he said.
Negrete said pharmacists are learning about the coverage to be prepared to help customers.
His best advice to consumers is get a good understanding of the coverage you have now, make a list of all medications now needed and decide which pharmacy you'd like to use.
Beneficiaries who get traditional fee-for-service Medicare may want to consider one of the 47 stand-alone prescription drug plans offered statewide.
Others who now are enrolled in a Medicare managed-care plan may consider signing up for an enhanced package with that insurer that will include prescription drug coverage, he said.
Many retirees may want to stay with their current prescription drug coverage through a former employer's health plan if it is comparable to what's offered through Medicare.

FIGURE A.14.: NYT_ENG_20051115.0354

EDITORIAL : MEDICARE DRUG PLAN GIVES WITH ONE HAND, TAKES WITH THE OTHER

The Bush administration is about as likely to admit the Republican Medicare drug assistance plan that was rushed through Congress has flaws as it is to say, "Uh, you know there were some questions about the intelligence on Iraq."
Senior citizens who are trying to sort through their drug coverage options face many problems, a main one being that those options are complicated and difficult to compare. Another, it now turns out, is that many drug companies that have been giving drugs (or selling them at low cost) to seniors in order to promote their products while giving elderly people a break aren't going to offer such programs anymore.
Why ? The companies fear they may find themselves wrestling with Uncle Sam if they try to maintain what they're doing for seniors while participating in government programs at the same time.
The government's logic evidently is that if the companies waive drug costs, then they might be considered to be making illegal kickbacks. The thinking is that the companies might be setting aside their fees to induce patients to take drugs that were more expensive than ones offered by competitors. If the patients then signed up for the Medicare program, the government – and taxpayers – would have to pay the difference.
Confusing ? Oh, it's not as bad as it will be for a lot of seniors. Consider, among many facts in an N&O report from staff writer Jean P. Fisher, that the Medicare Part D plan is going to be run by dozens of private insurance companies across the country. In North Carolina, 16 companies are offering 38 different plans. Individuals who have less than $14,355 in annual income get free or reduced premiums. That income figure is maddeningly low, of course, and seniors who are just a little above it, or even a good bit above it, may be in dire straits indeed. Drug companies that offer patient assistance programs typically have a higher income cutoff, about $19,000 a year.
The concept of covering prescription costs through Medicare is worthwhile. But the plan now going into effect smacks of a political taint : The president and his party vowed to do something about drug costs for seniors but wanted at the same time to protect drug manufacturers and insurance companies no matter what. It's a mess. It's hopelessly confusing. It's a plan that won't help nearly as many people as it should help, and will leave some worse off than they are now.
Instead of turning the tide against a movement toward a national health care system – which is what this administration and the drug and insurance industries are desperately trying to do – this program will only encourage it.
The next edition of Congress, which unfortunately won't be in until after the 2006 elections, will have a repair job on its hands. And its members will have plenty of reason to get busy. It's the right thing to do for senior citizens for whom a reliable supply of affordable medicine can be a cost-effective life-saver. It's also politically smart, as anyone can tell who's paying any attention to the considerable voting clout elderly Americans have.

Raleigh News & Observer editorial

FIGURE A.15.: NYT_ENG_20051122.0177

HELP IS ON THE WAY FOR MEDICARE DRUG COVERAGE

<DATELINE> RALEIGH, N.C. </DATELINE>
Confused about the new Medicare prescription drug coverage ? A new source of help is on the way.
A national coalition of senior and consumer advocacy groups said Tuesday it will dispatch a fleet of vans equipped with wireless computers and trained staff to North Carolina and 26 other states. The goal : to help Medicare enrollees understand the optional prescription drug coverage, which takes effect Jan. 1. Vans should start rolling into the Triangle and Charlotte areas by mid-December.
The education campaign – "My Medicare Matters" – comes a week after people enrolled in Medicare began signing up for the new prescription benefits. Many older adults have complained that the program is overly complex, and some advocates fear that people who would be helped by the benefit will throw their hands up and fail to sign up.
"Doing nothing ... is not the right response for America," said James L. Firman, president of the National Council on the Aging and chairman of the Access to Benefits Coalition, a collection of more than 100 groups that is coordinating the campaign.
My Medicare Matters, which is supported by a $10 million grant from pharmaceutical company AstraZeneca, will aim to help people enrolled in Medicare get the information they need to make an informed choice about whether to sign up for Medicare Part D, as the new prescription coverage is known.
Its counselors will do that largely by helping people tap existing online resources, such as the Medicare drug plan finder that is available at www.medicare.gov. People can punch in the medicines they take and compare prescription plans available in their area by cost and coverage, greatly simplifying comparison shopping.
However, few Medicare enrollees are using the online tool or any of the other information Medicare makes available at its Internet site. A national survey of Medicare members conducted recently by the Kaiser Family Foundation and Harvard School of Public Health found that just 6 percent of respondents had even visited the Web site.
My Medicare Matters will supplement existing education and outreach efforts in North Carolina. The largest single resource now operating in the state is the N.C. Seniors' Health Insurance Information Program, which has seen the number of calls to its information line, at (800) 443-9354, nearly triple in recent weeks, largely because of questions about Medicare Part D.
"Anything that is going to result in [North Carolina] having additional counselors who are educated and can help people is wonderful," said Carla Obiol, the SHIIP program's director. "The magnitude of this program change is so great we almost can't have enough help."

FIGURE A.16.: NYT_ENG_20051122.0321

MAKING THE ROUNDS FOR MEDICARE'S DRUG PLAN

<DATELINE> MINNEAPOLIS, Minn. </DATELINE>
Six counselors toting laptop computers will begin scouring the Twin Cities in a van in search of Medicare beneficiaries who need help signing up for the new drug benefit, officials said Tuesday.
"We'll be a lot more effective helping people who need a hand figuring out this new benefit – people whose health and finances may improve significantly because of help paying for their drugs," said Janine Stiles of the Minnesota Senior Federation.
The new cadre of counselors also may help ease pressure on the statewide Minnesota Linkage Line, which is wrestling with a backlog of more than 1,000 calls for help enrolling in the Medicare drug program.
Counselors will work with the Minnesota Access to Benefits Coalition, which Stiles coordinates.
The coalition has been in the field for a year helping educate older and disabled beneficiaries about the Medicare drug benefit.
In about two weeks, the counselors will start showing up at already scheduled informational sessions about the drug benefit and will visit other sites. They will help beneficiaries sort through more than 60 private plans offering drug insurance. The benefit will start Jan. 1 for those who enroll by Dec. 31.
The local effort is part of a national campaign and funded by a $10 million grant from British drugmaker AstraZeneca to the National Council on the Aging.
The council organized the national Access to Benefits Coalition last year. This year it is financed largely by a $2 million grant from the drug manufacturers trade association.
NEW ONLINE HELP
In its announcement Tuesday, the council also unveiled a streamlined Web site – www.MyMedicareMatters.org – to help Medicare beneficiaries and their families learn about the drug benefit and plug into the enrollment process.
The new computer site, financed in part by the federal government, supplements the Medicare Web site that some older people have complained is cumbersome and at times is slow because of heavy Internet traffic.
Demand on the Medicare Web site was so heavy Tuesday that its enrollment tool was unavailable for much of the day, officials at the Minnesota Linkage Line said.
THANKSGIVING CONVERSATIONS
Medicare officials have not released any information about how many people have joined a Medicare drug-insurance plan since enrollment started Nov. 15.
"My hunch is we'll know more – maybe even get a big push in enrollments – after families talk about the drug benefit over their Thanksgiving dinner," said Peter Ashkenaz , a spokesman for the federal Centers for Medicare and Medicaid Services.
Medicare program administrator Mark McClellan urged families Tuesday to have those conversations.
"We'll do as much outreach as we can," he said. "This new initiative with counselors and computers will help. But the best outreach will come from families, talking it over with their parents and aunts and uncles."

FIGURE A.17.: NYT_ENG_20051123.0004

WHICH PLAN ? THE ANSWERS ARE OUT THERE

Signing up my father for a Medicare drug plan should be simple, especially because the Medicare.gov Web site is well designed and easy to use. If only the real world would cooperate.
My father is 92 years old, lives in a nursing home and takes 10 prescription drugs, paid for from his savings and Social Security income. His medicines cost him $636.88 a month, or $7,642.56 a year.
He now gets a discount on his drugs because he has a Medigap policy, which he buys for $93.23 a month from AARP to cover some health care costs not picked up by Medicare.
Without the AARP Medigap discount, his drugs would cost nearly 40 percent more : $1,044.72 a month, or $12,536.64 a year.
So how does he – or actually, how do my sister and I, who manage his affairs – slash through the dozens of Medicare drug plans available in the part of suburban Washington where he lives ? How do we decide whether he should sign up for any of them and, if so, figure out the best plan for him ?
My sister, Martha Bari, to avoid the onerous research she was fearing, considered simply choosing an insurer with a trusted name. A likely candidate was UnitedHealthcare, the company used by the AARP Medigap program.
But when I looked at the Medicare Web site, the research tool there seemed easy to use. The site lets the user compare drug plans, based on the information one enters. I called Martha and told her I wanted to give it a try.
It turns out that you really do need a computer to sort through drug plans because there are so many variables. What drugs are you taking ? Where do you live ? What plans are in your area ? Are you willing to switch to generics if they are an option ? Do you care which pharmacy supplies your drugs ? Do you want to order them by mail ? Are you in a nursing home that only uses one pharmacy ? Those considerations can affect the price you will be charged. And, depending on how big a discount a plan offers, you may be better off with one that does not cover all of them but that deeply discounts the drugs it does cover.
In terms of your out-of-pocket expenses, such a plan might work better than a plan that covers every drug but offers smaller discounts.
I set to work, blithely thinking I would could sort things out in half an hour.
I discovered, instead, that it took two days to get the information I needed before even starting to use the Medicare Web site. First I had to find out what drugs my father was actually taking and how much he paid for them.
Initially, Martha and I understood that my father was taking eight drugs. Only after checking with the nursing home did she get the complete list of 10.
Then I called AARP to find out exactly what type of Medigap plan my father has and what it covers.
Next came a call to the pharmacy to find out what the drugs would cost without my father's AARP discount, which involved providing a list of the medications and then waiting for a call with the prices.
And how would my father's Medigap policy be affected if he signed up for the Medicare drug plan ? Would the Medigap plan pick up some of his leftover drug costs ? And if it no longer covered his drugs, would his Medigap premium decrease ?
I eventually learned in a phone conversation with Jim Pogue, who manages AARP insurance for the UnitedHealth Group, that my father's Medigap plan would provide discounts on any drug not covered by a Part D plan. But the Medigap premium would remain the same whether or not he needed the supplemental drug coverage.
That meant that unless a Medicare plan was cheaper than the $7,642.56 a year my father pays for his drugs with the AARP discount, he might as well not sign up for a new plan.
Finally, after two days and many phone calls, I had the information I needed. And the actual process of comparing drug plans was fun – a breeze, in fact.
Last Friday, I plugged my father's drugs into the Medicare Web site. I specified their doses and the preferred pharmacy, which is 35 miles from his nursing home but the only one the nursing home will use.
Voila ! I got a list of 42 plans. They were ranked in order, based on my father's total annual out-of-pocket expense for his drugs would be, including the monthly drug premium and the annual deductible for each plan. I had the ability to compare any three plans side by side, in as many combinations as I chose.
The variations were astonishing, with the estimated yearly cost varying by more than 100 percent.
Under the least expensive plan for my father's situation, one offered by First Health Premier, he would pay an estimated $4,008.72 a year, or $3,633.84 less than he pays now. His monthly co-payments would be $142.70, lower than with any other plan, and his monthly premium would be $22.91, among the lowest of all the plans.
Under the most expensive plan, AmeriHealth Advantage Rx Option, he would pay $8,454.33 for the same drugs from the same pharmacy – or $811.77 more than he pays now for those drugs from that pharmacy. His monthly co-payments would be $673.42, and his monthly drug premium would be $23.16.
It took only 10 minutes to get the list of the drug plans and their prices for my father from the Medicare Web site. But the most amazing thing was what happened next.
I went to the Web site again last Tuesday, four days after first entering my father's data. This time the results were different.
I thought there must be some mistake. But, no, I had a printout from the first session, and there had been a change.
There were six additional plans that were not there before, for a total of 48 instead of 42. And the prices were lower, especially for plans at the high end.
First Health Premier was still the cheapest – but $3,977.01 now, compared with $4,008.72 four days earlier. AmeriHealth Advantage was still the most expensive, but its price had dropped to $6,978.94, from $8,454.33.
What was going on ?
Peter Ashkenaz, a Medicare spokesman, explained that the agency updated the plans and their prices every Monday. He said that the insurers were starting to assess one another's plans and revise them to be more competitive.
I asked Ashkenaz : What happens if you choose a plan and then the next week it becomes cheaper, or another plan becomes cheaper ?
As it turns out – and Ashkenaz allowed that Medicare was not particularly publicizing this wrinkle – you can change plans every day from now until the end of the year if you want to.
Whichever plan you are enrolled in on Dec. 31 is the plan you will be in when the program starts on Jan. 1. And you will be paying the rate that was quoted when you enrolled.
Even after that, you can change plans one more time before May 15, the last time to sign up for the Medicare drug benefit with no penalty, Ashkenaz said.
For those who do not want to use the Web site, or cannot use it, there is another option. Call Medicare's toll-free number, 1-800-MEDICARE, specify "enrollment" on the voice-mail prompt, and tell the customer service representative the drugs you take and where you live.
Medicare says it has 7,500 such service representatives handling calls 24 hours a day, seven days a week.
The service agents are assigned to walk the caller "through everything that's up there and will send them a printout – we call it a personalized brochure," Ashkenaz said.
I called the number and got right through to a service agent in the middle of the day last Thursday, with no wait. I told her that I was a reporter and that I was just testing the system ; I did not ask her to go through the whole drug plan search process for me. And then I tried again the next day with the same result.
As for Martha and me, we have not yet signed my father up for a plan. I told her that I now think the best strategy is to sign up for the cheapest plan and then keep checking back every week on the Medicare site to see if a better option comes along.
Who knows – maybe the prices will rise in the next month, and we would want to be signed up before that happens.
Martha says it is up to me to do all that checking if I want to, but it is not something she cares to take on. And she wonders how many people will.

FIGURE A.18.: NYT_ENG_20051123.0247

MILLIONS FACE DEADLINE FOR CHOOSING NEW MEDICARE PLAN

The deadline is approaching for elderly and disabled people who must decide whether and when to enroll in one of the new Medicare Part D drug plans that start Jan. 1, 2006.
The so-called Part D program for the first time brings drugs under the Medicare tent that covers 42 million people.
The program is set up to be federally subsidized but run by the commercial insurance industry, a combination that some say generates confusion.
Companies are clamoring to compete, and some are adding to the confusion by offering dozens of plans with a welter of wrinkles and widely varying prices. Further, some insurers keep adjusting their offerings.
But there are some basic steps to help consumers decide which plan, if any, is right for an individual. Hari Peterson, 69, a retired nurse in Warwick, R.I., has survived the process.
"I found it to be confusing, but we really studied it," said Peterson, whose husband, Ray, 72, has drug coverage from the Department of Veterans Affairs that he plans to keep for now.
She collected information on computer printouts from senior centers and talked with state insurance advisers at a Medicare fair.
"I gave them my prescriptions and they pinpointed it down to two or three plans," she said. Peterson also plans to consult her longtime pharmacist before making a final selection.
Consumers who want coverage to start in January can make a selection by Dec. 31, but they can also wait until May 15 to select a plan, without penalty.
The Web site medicare.gov can answer many questions and help with plan comparisons.
Medicare is also operating a telephone help line around the clock, 1-800-Medicare, which has an automated voice system that also allows callers to reach a real person.
Those who decide now can still check for a better offer next week. Through Dec. 31, consumers can change their minds. And it may pay to keep checking, because some insurers are continually adjusting their offerings.

FIGURE A.19.: NYT_ENG_20051123.0294

EDITORIAL : A ROCKY ROLLOUT FOR MEDICARE PART D

<DATELINE> MINNEAPOLIS </DATELINE>
The government's celebrated new prescription drug program for senior citizens got off to a rocky start this month, with elderly consumers fuming over its complexity and federal websites crashing under heavy traffic. But perhaps the worst news came last Tuesday, when Rep. Henry Waxman released a report showing that a typical senior citizen could get a better deal at Costco or Drugstore.com than in the elaborate new Medicare marketplace.
It's too late now to change the law, which passed Congress in 2003, and too soon to rewrite it. But it's not premature to study whether it's really serving the interests of Medicare beneficiaries and the taxpayers who will pay its massive cost.
Waxman, a California Democrat, asked congressional investigators to study the 10 top-selling drugs among senior citizens, and then compare prices available at popular discounters against prices in the new Medicare Part D, which relies on private insurance companies to deliver coverage. They found that, on average, drug prices in the new Medicare were about 3 percent higher than prices available to any customer at Costco or Drugstore.com and 60 percent higher than prices in Canada.
Given the enormous sums that Congress poured into the Medicare subsidy, that's bizarre.
To be sure, this is not the last word on Medicare Part D. Some elderly consumers, buying some medications, probably will find better prices through Medicare plans. And it's worth remembering that congressional skeptics were wrong before : A year ago they predicted that insurance companies would balk at entering the Medicare market, but today there are dozens with Medicare drug offerings. It's only fair to give this big, complicated new system a year or so to prove itself.
That said, congressional Republicans who promised that competition among private insurance companies would improve Medicare, delivering low prices for consumers and low costs for taxpayers, now bear a heavy burden of proof. The Veterans Administration, for example, doesn't operate that way. It simply negotiates bulk discounts with drugmakers and pays far less than retail. So does Medicaid, the federal-state insurance program for the poor. And yet the 2003 law that created Part D specifically forbids Medicare to negotiate the same sort of bulk discounts.
The theory of market competition is powerful, but it shouldn't be taken as gospel. As economist Paul Krugman has argued, private markets might be the best way to produce cars and computers, but are a horrible way to deliver health care. The United States relies on private markets to deliver health care far more than any other developed nation, and yet Americans pay far more for health care than consumers in other advanced nations.
If the new Medicare Part D turns out to be expensive but wonderful, it might be worth it. If it turns out to be inexpensive but complicated, that also might be acceptable. But if it turns out to be complicated and wasteful too, Americans will find it's time to go back to the drawing board on health care.

FIGURE A.20.: NYT_ENG_20051126.0135

D0805 Medicare Part D : résumé initial Describe the developments in the Medicare Part D implementation.
But if we had been in a standard Medicare Part D plan, our expenses would have been about $5,000. "The state is forcing us to go into Medicare Part D". Part D allows people on Medicare to get their drugs in one of two ways : by buying an individual drug policy or by signing up with a Medicare Advantage managed care plan. Their clients cannot enroll in separate Part D drug plans, which are designed to supplement Medicare's traditional Parts A and B. However, Medicare is also beefing up payments to HMOs, to help them stay competitive. You can
Score ROUGE-SU4 : 0.09448

D0805 Medicare Part D : résumé de mise à jour Describe the developments in the Medicare Part D implementation.
Warning : You ca n't have both a Medicare HMO and a Part D drug plan. com than in the elaborate new Medicare marketplace. If the new Medicare Part D turns out to be expensive but wonderful, it might be worth it. Although Part D will greatly benefit certain people, its savings are not universal. The deadline is approaching for elderly and disabled people who must decide whether and when to enroll in one of the new Medicare Part D drug plans that start Jan. 1, 2006. The so-called Part D program for the first time brings drugs under the Medicare
Score ROUGE-SU4 : 0.09018

TABLE A.2.: Résumé généré par CBSEAS sans gestion de la structure

D0805 Medicare Part D : résumé initial Describe the developments in the Medicare Part D implementation.
Medicare has a toll-free hotline and a Web site to help people sort through the new Part D drug coverage. After a two-year wait, on Tuesday more than 40 million Americans eligible for Medicare can start to sign up for Medicare Part D, the program's voluntary prescription drug benefit. "I do n't have to deal with Medicare Part D and I 'm delighted," said Marlene Barnes, 69, a retired federal office worker in Memphis, Tenn., whose health benefits are provided by the Federal Employees Health Plan. The question was : Would employers keep their drug plans and take the
Score ROUGE-SU4 : 0.13115

D0805 Medicare Part D : résumé de mise à jour Describe the developments in the Medicare Part D implementation.
Medicare has a toll-free hotline and a Web site to help people sort through the new Part D drug coverage. After a two-year wait, on Tuesday more than 40 million Americans eligible for Medicare can start to sign up for Medicare Part D, the program's voluntary prescription drug benefit. "I do n't have to deal with Medicare Part D and I 'm delighted," said Marlene Barnes, 69, a retired federal office worker in Memphis, Tenn., whose health benefits are provided by the Federal Employees Health Plan. The question was : Would employers keep their drug plans and take the
Score ROUGE-SU4 : 0.09773

TABLE A.3.: Résumé généré par CBSEAS avec gestion de la structure et des entités nommées

D0805 Medicare Part D : résumé initial Describe the developments in the Medicare Part D implementation.
Medicare has a toll-free hotline and a Web site to help people sort through the new Part D drug coverage. After a two-year wait, on Tuesday more than 40 million Americans eligible for Medicare can start to sign up for Medicare Part D, the program's voluntary prescription drug benefit. The question was : Would employers keep their drug plans and take the subsidy, or would they dump their plans to save money, knowing that retirees could now get prescription coverage from Medicare Part D. "I do n't have to deal with Medicare Part D and I 'm delighted," said Marlene Barnes,
Score ROUGE-SU4 : 0.13115

D0805 Medicare Part D : résumé de mise à jour Describe the developments in the Medicare Part D implementation.
The so-called Part D program for the first time brings drugs under the Medicare tent that covers 42 million people. Waxman, a California Democrat, asked congressional investigators to study the 10 top-selling drugs among senior citizens, and then compare prices available at popular discounters against prices in the new Medicare Part D, which relies on private insurance companies to deliver coverage. If it is n't, you probably should switch to a Medicare HMO or a Part D drug plan to avoid paying a premium penalty later on. Today is the first day you can enroll in a Part D plan
Score ROUGE-SU4 : 0.08641

TABLE A.4.: Résumé généré par CBSEAS avec gestion de la structure et des anaphores

B. Annexes relatives à TAC2009

D0941 Huygens probe Track the progress of the Huygen space probe.
Huygens was designed for only a brief mission. It is scheduled to arrive at its destination on January 14. The mission involved 260 scientists from 17 nations. *Some kind of flow on the surface ?** The European Space Agency built and managed the development of Huygens and directs the probes operations. *Perhaps, said Jean-Pierre Lebreton, the Huygens project manager for the European Space Agency.* The Cassini mission is a joint project of NASA, ESA and the Italian space agency.* **15, 1997, and arrived at Saturn in June**. *And then, it was on its way.**
Score pyramide : 0,14 Note linguistique : 1

D0925 Apple Computer switch to Intel chips Trace plans for and progress of the switch to Intel chips by Apple Computer.
The original Mac, released in 1984, used a processor made by Motorola Inc. **, firm plans to switch to Intel Corp.** A decade ago, Apple switched to PowerPC, a chip designed by Motorola and IBM. **A stormy, decade-long relationship between Apple Computer Inc.** Intel makes the Pentium M, a cool-running chip designed especially for laptops. **announced plans Monday to switch its Macintosh computers to the same Intel Corp.** **Officials from Apple, Intel Corp.** *Where does this lead us ?** So Mac OS X is cross-platform by design right from the very beginning. Intel expressed delight at Apples switch.
Score pyramide : 0,34 Note linguistique : 2

TABLE B.1.: Illustration des systèmes mauvais en linguistique (1/4). Système de résumé numéro 9 (soumission anonyme). En gras : phrases grammaticalement incorrectes ; en italique : phrases dont les références sont opaques ; * : phrases sans rapport à la requête

D0942 PRC's anti-secession law Report the developments in the PRC's enactment of an anti-secession law and reactions to that law.
The top agency in charge of Taiwan's policies toward the mainland.* **Stressing that Taiwan is not part of the People's Republic of China as the Beijing authorities claim.** **Meanwhile, Michel Lu, spokesman of the Ministry of Foreign Affairs.*** **The government will also continue to express its opposition to the proposed anti-secession law, Lu went on.** **Shu Yung-ming, an assistant researcher with Academia Sinica, Taiwan's highest research institute.*** **Japan and EU countries last year to brief officials.*** **Raymond F. Greene, chief of the political section at the AIT's Taipei office***
Score pyramide : 0 Note linguistique : 1

D0940 opening of Disneyland in Hong Kong Follow the developments connected with the opening of Disneyland in Hong Kong.
Feng shui experts also designated "no fire zones" in the kitchens, water, wood, fire and earth in balance.* The park is a landscaping achievement, reproducing a slice of America with Asian accents. The opposition may seem odd, shark fin soup is hugely popular, and stray dogs are summarily dealt with as health hazards.* **From August to Sept. 11, including public roads, the Park Promenade** The rehearsal period at HK Disneyland will last for approximately four weeks prior to the grand opening. The ticket sale of the Hong Kong Disneyland has been proceeding well.
Score pyramide : 0,21 Note linguistique : 1

TABLE B.2.: Illustration des systèmes mauvais en linguistique (2/4). Système de résumé numéro 5 (soumission anonyme). En gras : phrases grammaticalement incorrectes ; en italique : phrases dont les références sont opaques ; * : phrases sans rapport à la requête

D0941 Huygens probe Track the progress of the Huygen space probe.
Huygens headed Saturday for a historic up-close encounter with Titan, after successfully separating from the US spacecraft Cassini. On Monday, Cassini will perform a course change to avoid following the probe into Titan. **Cassini, with the Huygens probe riding piggy-back.** **After it detaches from Cassini, the 9-foot Huygens probe will fall toward Titan and.** Huygens separated from its NASA mothership late Friday to begin a descent to Saturn's largest moon Titan. The Huygens probe, designed and built by the European Space Agency, had been carried by NASA's Saturn probe Cassini.
Score pyramide : 0,46 Note linguistique : 1

D0942 PRC's anti-secession law Report the developments in the PRC's enactment of an anti-secession law and reactions to that law.
The top agency in charge of Taiwan's policies toward the mainland.* **Stressing that Taiwan is not part of the People's Republic of China as the Beijing authorities claim.** **Meanwhile, Michel Lu, spokesman of the Ministry of Foreign Affairs.*** The government will also continue to express its opposition to the proposed anti-secession law, if passed, Lu went on.* Beijing will diminish cross-Taiwan Strait exchanges and thus will be harmful to the development of bilateral ties.* **Shu Yung-ming, Taiwan's highest research institute.*** **Japan and EU countries last year to brief officials, which targets Taiwan.***
Score pyramide : 0 Note linguistique : 1

TABLE B.3.: Illustration des systèmes mauvais en linguistique (3/4). Système de résumé numéro 26 (soumission anonyme). En gras : phrases grammaticalement incorrectes ; en italique : phrases dont les références sont opaques ; * : phrases sans rapport à la requête

D0907 John Yoo Describe the views and activities of John C. Yoo.
780 words, by Eric A. Posner and John C. Yoo (Times special).* **835 words, by Harold Meyerson (Post special) (k).*** **900 words, by Reuel Marc Gerecht (Post special) (k).*** **935 words, by Robert J. Delahunty and John C. Yoo (Times special) (k).*** **670 words, by Lisa Simeone (Sun special) (k).*** **780 words, by Vivan Gornick (Times special) (k).*** **1,430 words, by William H. Hooke (Post special) (k).*** **890 words, by Michael W. Kauffman (Sun special) (k).*** **490 words, by Mike Berland and Doug Shoen (Times special) (k).*** **TIMES EDITORIAL, upcoming ; POST EDITORIAL, moved (both k).***
Score pyramide : 0,00 Note linguistique : 1

D0928 Trials of Tyco's Dennis Kozlowski Follow the legal proceedings in the criminal trials of Tyco Chief Dennis Kozlowski.
*According to some estimates, investors lost $60 billion (euro48.33 billion).** *This will be an important test of the idiot CEO defense (see Richard Scrushy and Ken Lay, below).** *This authorized compensation – as huge as it was – was just not enough for these two men, Heimer said.** *Heimer repeatedly spoke to the jury of a process he described as instruct, deceive and conceal.** Stillman said the minutes cannot be relied on to prove anything because they are incomplete.* On Wednesday, Jordan sat in the back row as Heimer finished his presentation.*
Score pyramide : 0,00 Note linguistique : 1

Table B.4.: Illustration des systèmes mauvais en linguistique (4/4). Système de résumé numéro 28 (soumission anonyme). En gras : phrases grammaticalement incorrectes ; en italique : phrases dont les références sont opaques ; * : phrases sans rapport à la requête

B.1. Résultats de la tâche « Résumé et mise à jour »

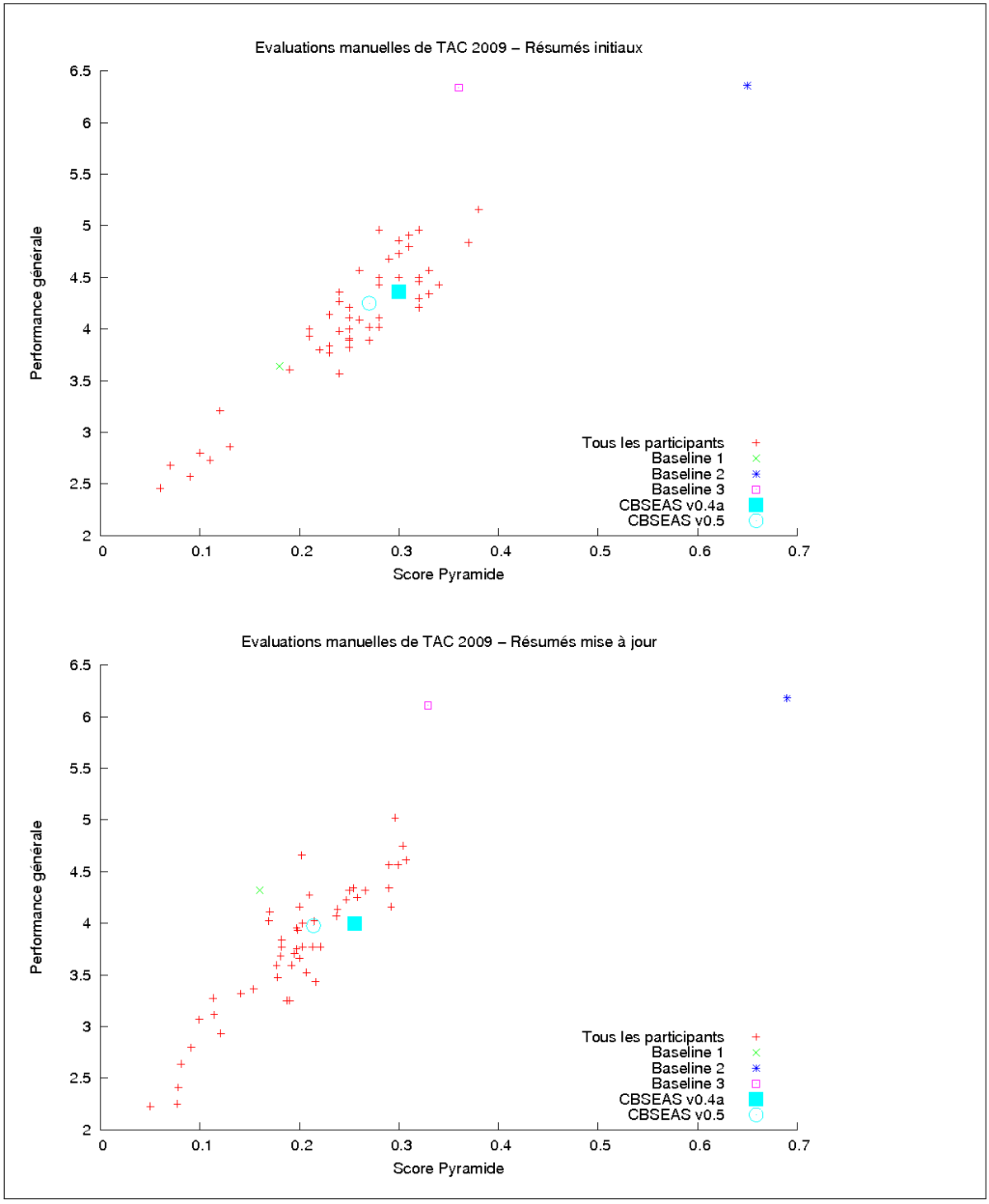


FIGURE B.1.: Évaluations manuelles de TAC 2009 « Update Task » : score pyramide et performance générale

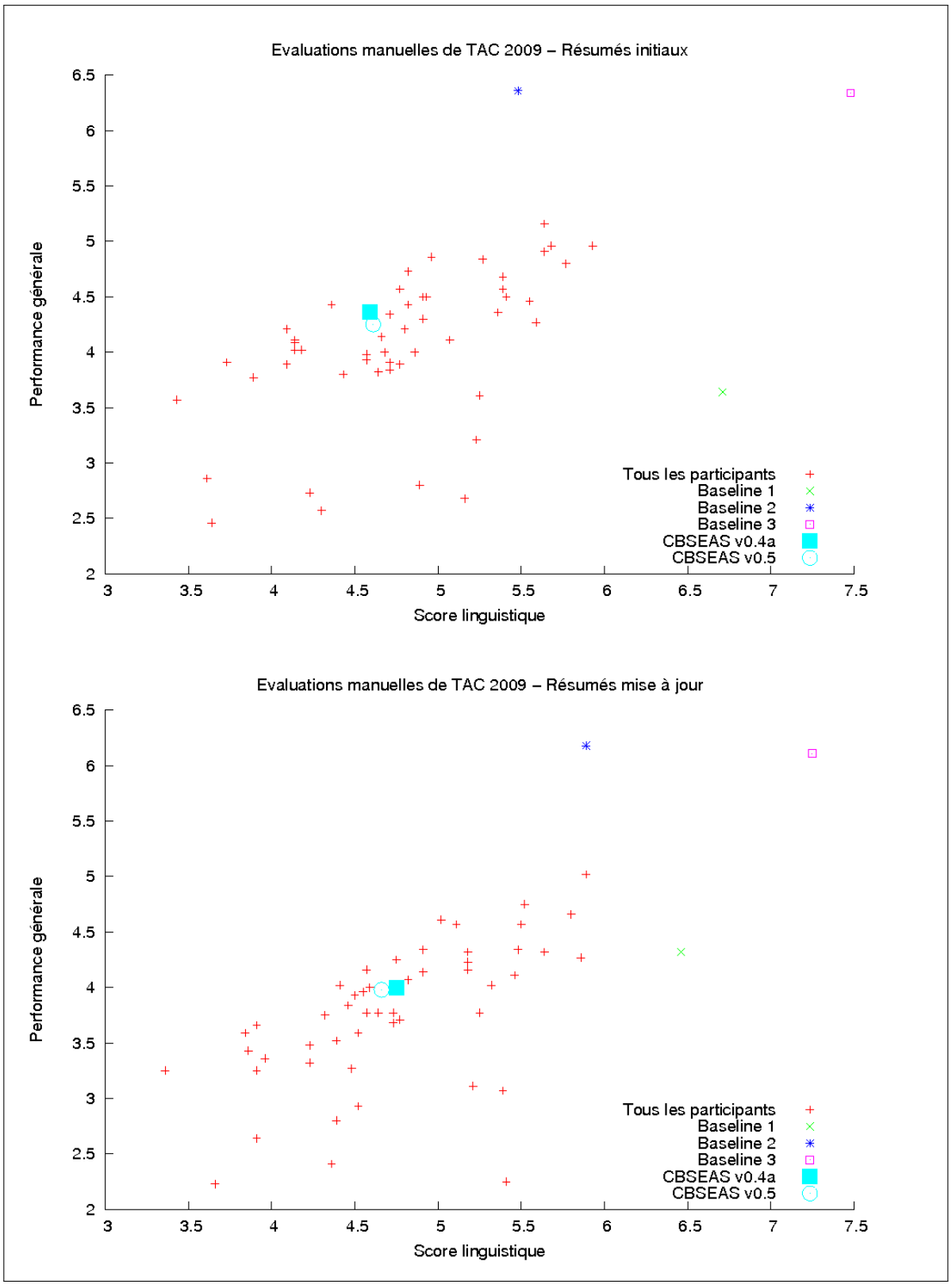


FIGURE B.2.: Evaluations manuelles de TAC 2009 « Update Task » : qualité linguistique et performance générale

C. Liste des Publications liées à la thèse

Communications internationales avec actes

Aurélien Bossard et Thierry Poibeau
Integrating Document Structure to an Automatic Summarizer
RANLP 2009, Septembre 2009 - *Poster*

Aurélien Bossard
CBSEAS, a New Approach to Automatic Summarization
SIGIR 2009 - Doctoral Consortium, juillet 2009

Communications nationales avec actes

Aurélien Bossard
Une approche mixte-statistique et structurelle - pour le résumé automatique de dépêches
TALN 2009, Juin 2009

Aurélien Bossard et Michel Généreux
Résumé automatique de textes d'opinion
TALN 2009, Juin 2009

Aurélien Bossard et Thierry Poibeau
Regroupement Automatique de Documents en Classes Evénementielles
TALN 2008, juin 2008 - *Poster*

Workshops internationaux avec comité de sélection

Aurélien Bossard, Michel Généreux and Thierry Poibeau
CBSEAS, a Summarization System Integration of Opinion Mining Techniques to Summarize Blogs
EACL 2009, System Demonstration, avril 2009

Workshops internationaux

Aurélien Bossard
Description of the LIPN Systems at TAC2009
Text Analysis Conference 2009, Workshop on Summarization Tracks, novembre 2009

Aurélien Bossard, Michel Généreux et Thierry Poibeau
Description of the LIPN Systems at TAC2008 : Summarizing Information and Opinions
Text Analysis Conference 2008, Workshop on Summarization Tracks, novembre 2008

D. Liste des Acronymes

AFP	:	Agence France Presse
ANNIE	:	A Nearly-New Information Extraction system
AQUAINT	:	Advanced Question Answering for Intelligence
BE-HM	:	Basic Elements, Head Modifier pairs
CBSEAS	:	Clustering-Based Sentence Extractor for Automatic Summarization
CNA	:	Central News Agency
CSIS	:	Cross-sentence informational subsumption
DUC	:	Document Understanding Conference
EN	:	Entité Nommée
GATE	:	General Architecture for Text Engineering
ICF	:	Inverse Cluster Frequency
IDF	:	Inverse Document Frequency
LCS	:	Longest Common Subsequence
MMR	:	Maximal Marginal Relevance
NIST	:	National Institute of Science and Technology
NYT	:	New York Times
NLP	:	Natural Language Processing
ROUGE	:	Recall-Oriented Understudy for Gisty Evaluation
RST	:	Rhetorical Structure Theory
SCU	:	Single Content Unit
SVM	:	Support Vector Machine
TAC	:	Text Analysis Conference
TAL	:	Traitement Automatique de la Langue
TDT	:	Topic Detection and Tracking
TF	:	Term Frequency
TLF	:	Trésor de la Langue Française

Bibliographie

Yves Agnès : *Manuel de journalisme : écrire pour le journal*. Paris : la Découverte, 2008. ISBN 978-2-7071-4393-8.

Chinatsu Aone, Mary Ellen Okurowski et James Gorlinsky : Trainable, scalable summarization using robust nlp and machine learning. *In Proceedings of the 17th international conference on Computational linguistics*, pages 62–66, Morristown, NJ, USA, 1998. Association for Computational Linguistics.

T Aït El Mekki et A Nazarenko : L'index de fin de livre, une forme de résumé indicatif ? *Revue Traitement Automatique des Langues*, 1(45):121–150, 2004.

Regina Barzilay, Noemie Elhadad et Kathleen McKeown : Inferring strategies for sentence ordering in multidocument news summarization. *J. Artif. Intell. Res. (JAIR)*, 17:35–55, 2002.

Regina Barzilay et Kathleen McKeown : Sentence fusion for multidocument news summarization. *Computational Linguistics*, 31:297–328, September 2005.

P. B. Baxendale : Machine-made index for technical literature : an experiment. *IBM Journal for Research and Development*, 2:354–361, 1958.

P. Beineke, T. Hastie, C. Manning et S. Vaithyanathan : An exploration of sentiment summarization. AAAI, 2003.

Sylvain Bellemare, Sabine Bergler et René Witte : ERSS at TAC 2008. *In Proceedings of the TAC 2008 Workshop*, volume Notebook Papers and Results, 2008.

Henri Binsztok, Thierry Artières et Patrick Gallinari : Un modèle probabiliste de détection en ligne de nouvel evénement. *In Reconnaissance des Formes et Intelligence Artificielle (RFIA 2004)*, Toulouse, France, 2004.

Aurélien Bossard, Michel Généreux et Thierry Poibeau : Description of the LIPN System at TAC 2008 : Summarizing Information and Opinions. *In Proceedings of the 2008 Text Analysis Conference (TAC 2008) TAC 2008*, pages 282–291, Gaithersburg, United States, 11 2008. URL http://hal.archives-ouvertes.fr/hal-00397010/en/.

Florian Boudin : *Exploration d'approches statistiques pour le résumé automatique de texte*. Thèse de doctorat, 2008.

Florian Boudin, Frederic Bechet, Marc El-Beze, Benoit Favre, Laurent Gillard et Juan-Manuel Torres-Moreno : The lia summarization system at duc-2007. *In Proceedings of DUC 2007 Document Understanding Conference*, volume Document Understanding Conference Workshop, April 2007.

Florian Boudin, Marc El-Bèze et Juan-Manuel Torres-Moreno : The LIA update summarization systems at TAC-2008. *In Proceedings of the TAC 2008 Workshop*, volume Notebook Papers and Results, 2008a.

Florian Boudin et Juan-Manuel Torres-Moreno : Neo-cortex : A performant user-oriented multi-document summarization system. *In* Alexander F. Gelbukh, éditeur : *CICLing*, volume 4394 de *Lecture Notes in Computer Science*, pages 551–562. Springer, 2007. ISBN 3-540-70938-X.

Florian Boudin, Juan-Manuel Torres-Moreno et Marc El-Bèze : A scalable mmr approach to sentence scoring for multi-document update summarization. *In COLING Conference*, pages 21–24, Manchester,UK, August 2008b.

Florian Boudin, Juan-Manuel Torres-Moreno et Patricia Velázquez-Morales : An efficient statistical approach for automatic organic chemistry summarization. 5221/2008:89–99, 2008c.

Richard Braddock : The frequency and placement of topic sentences in expository prose. *Research in The Teaching of English*, 8:287–302, 1974.

Ronald Brandow, Karl Mitze et Lisa F. Rau : Automatic condensation of electronic publications by sentence selection. *Information Processing Management*, 31(5):675–685, 1995.

Caroline Brun, Nicolas Dessaigne, Maud Ehrmann, Baptiste Gaillard, Sylvie Guillemin-Lanne, Guillaume Jacquet, Aaron Kaplan, Marianna Kucharski, Claude Martineau, Aurélie Migeotte, Takuya Nakamura et Stavroula Voyatzi : Une expérience de fusion pour l annotation d'entités nommées. *In Actes de TALN 2009 (Traitement Automatique du Langage Naturel)*, juin 2009.

Jaime Carbonell et Jade Goldstein : The use of mmr, diversity-based reranking for reordering documents and producing summaries. *In SIGIR 1998 : Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Melbourne, Australie, August 1998. ACM.

J. Chauché : Un outil multidimensionnel de l'analyse du discours. *In ACL-22 : Proceedings of the 10th International Conference on Computational Linguistics and 22nd annual meeting on Association for Computational Linguistics*, pages 11–15, Morristown, NJ, USA, 1984. Association for Computational Linguistics.

Massimiliano Ciaramita et Yasemin Altun : Broad-coverage sense disambiguation and information extraction with a supersense sequence tagger. *In EMNLP '06 : Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pages 594–602, Morristown, NJ, USA, 2006. Association for Computational Linguistics. ISBN 1-932432-73-6.

H. Cunningham, D. Maynard, K. Bontcheva et V. Tablan : GATE : A Framework and Graphical Development Environment for Robust NLP Tools and Applications. *In Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics (ACL'02)*, 2002.

Hoa Trang Dang : Overview of the tac 2008 opinion question answering and summarization tasks. *In Proceedings of TAC 2008 Workshop*, 2008.

Hoa Trang Dang et Karolina Owczarzak : Overview of the tac 2008 opinion question answering and summarization tasks. *In Proceedings of the TAC 2008 Workshop*, volume Notebook Papers and Results, 2008a.

Hoa Trang Dang et Karolina Owczarzak : Overview of the tac 2008 update summarization task (draft). *In Proceedings of TAC 2008 Workshop*, 2008b.

D. L Davies et D. W. Bouldin : A cluster separation measure. *In IEEE Trans. on Pattern Analysis and Machine Intelligence*, pages 224–227, 1979.

Dan Dolan : Locating main ideas in history textbooks. *Journal of Reading*, pages 135–140, 1980.

H. P. Edmundson : New methods in automatic extracting. *J. ACM*, 16(2):264–285, 1969. ISSN 0004-5411.

Maud Ehrmann et Guillaume Jacquet : Vers une double annotation des entités nommées. *TAL*, 47:63–88, 2006.

Christian Girardi Emanuele Pianta et Roberto Zanoli : The textpro tool suite. *In Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco, may 2008. European Language Resources Association (ELRA). ISBN 2-9517408-4-0. http ://www.lrec-conf.org/proceedings/lrec2008/.

Güneş Erkan et Dragomir R. Radev : Lexrank : Graph-based centrality as salience in text summarization. *Journal of Artificial Intelligence Research (JAIR)*, 2004.

Atefeh Farzindar et Guy Lapalme : Production automatique du résumé de textes juridiques : évaluation de qualité et d'acceptabilité. *In TALN 2005*, volume 1, pages 183–192, Dourdan, France, jun 2005.

Benoit Favre, Frédéric Béchet, Patrice Bellot, Florian Boudin, Marc El-Bèze, Laurent Gillard, Guy Lapalme et Juan-Manuel Torres-Moreno : The LIA-Thales summarization system at DUC-2006. *In Document Understanding Conference Workshop, HLT-NAACL'06, New York (USA)*, 2006.

Olga Feiguina et Guy Lapalme : Query-based summarization of customer reviews. *In Proceedings of the 20th Canadian Conference on Artificial Intelligence*, 4509, pages 452–463, Montréal, may 2007. Springer-Verlag.

Christiane Fellbaum, éditeur. *WordNet An Electronic Lexical Database*. The MIT Press, Cambridge, MA ; London, May 1998. ISBN 978-0-262-06197-1. URL http://mitpress.mit.edu/catalog/item/default.asp?ttype=2&tid=8106.

Silvia Fernández Sabido et Juan-Manuel Torres-Moreno : Une approche exploratoire de compression automatique de phrases basée sur des critères thermodynamiques. *In Actes de la Conférence sur le Traitement Automatique du Langage Naturel 2009, Volume Posters et Démonstrations*, 2009.

E. Forgy : Cluster analysis of multivariate data : Efficiency vs. interpretability of classifications. *Biometrics*, pages 21–768, 1965a.

E. Forgy : Cluster analysis of multivariate data : Efficiency vs. interpretability of classifications. *Biometrics*, pages 21–768, 1965b.

Edward B. Fry, Jacqueline E. Kress et Dona Lee Fountoukidis : *The Reading Teachers Book of Lists*. 2000.

Dimitrios Galanis et Prodromos Malakasiotis : AUEB at TAC 2008. *In Proceedings of the TAC 2008 Workshop*, volume Notebook Papers and Results, 2008.

William Gale, Kenneth Church et David Yarowsky : One sense per discourse. *In Proceedings of the workshop on Speech and Natural Language*, pages 224–227, New York, 1992. Harriman.

Pierre-Étienne Genest, Guy Lapalme et Mehdi Yousfi-Monod : Hextac : the creation of a manual extractive run. *In Proceedings of the TAC 2008 Workshop*, volume Notebook Papers and Results, Gaithersburg, Maryland, USA, nov 2009.

Christopher Buckley Gerard Salton : Term-weighting approaches in automatic text retrieval. *Information Processing and Management : an International Journal*, 24:513–523, 1988.

Daniel Gillick, Benoit Favre, Dilek Hakkani-Tur, Berndt Bohnet, Yang Liu et Shasha Xie : The ICSI/UTD Summarization System at TAC 2009. *In Proc. of the Text Analysis Conference workshop, Gaithersburg, MD (USA)*, 2009.

Michel Généreux : Summarizing a blog search engine hits. *In WWW 2009, Workshop on Web Search Result Summarization and Presentation*, April 2009a.

Michel Généreux : Summarizing a blog search engine hits. *In Proceedings of the WWW 2009 Conference, Workshop on Web Search Result Summarization and Presentation*, 2009b.

Jade Goldstein, Vibhu Mittal, Jaime Carbonell et Mark Kantrowitz : Multi-document summarization by sentence extraction. *In NAACL-ANLP 2000 Workshop on Automatic summarization*, pages 40–48, Morristown, NJ, USA, 2000. Association for Computational Linguistics.

Iryna Gurevych et Michael Strube : Semantic similarity applied to spoken dialogue summarization. *In COLING '04 : Proceedings of the 20th international conference on Computational Linguistics*, page 764, Morristown, NJ, USA, 2004. Association for Computational Linguistics.

Greg Hamerly et Yu Feng : Pg-means : learning the number of clusters in data. *In The Twentieth Annual Conference on Neural Information processing systems*, Vancouver, Canada, 2006.

Vasileios Hatzivassiloglou, Luis Gravano et Ankineedu Maganti : An investigation of linguistic features and clustering algorithms for topical document clustering. *In SIGIR 2000 : Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 224–231, Athens, Greece, 2000. ACM.

Tingting He, Jinguang Chen, Zhuoming Gui et Fang Li : CCNU at TAC 2008 : Proceeding on using semantic method for automated summarization yield. *In Proceedings of the TAC 2008 Workshop*, volume Notebook Papers and Results, 2008.

John H. Holland : *Adaptation in natural and artificial systems : An introductory analysis with applications to biology, control, and artificial intelligence*. University of Michigan Press, 1975. ISBN 0472084607.

Eduard Hovy : Automated discourse generation using discourse structure relations. *Artificial Intelligence*, 63:341–385, 1993.

Eduard Hovy, Chin-Yew Lin et L Zhou : A be-based multi-document summarizer with sentence compression, 2005.

Eduard Hovy, Chin-Yew Lin, Liang Zhou et Junichi Fukumoto : Automated summarization evaluation with basic elements. *In In Proceedings of the Fifth Conference on Language Resources and Evaluation (LREC)*, 2006.

Meishan Hu, Aixin Sun et Ee P. Lim : Comments-oriented blog summarization by sentence extraction. *In CIKM '07 : Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*, pages 901–904, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-803-9. URL http://dx.doi.org/10.1145/1321440.1321571.

Minqing Hu et Bing Liu : Mining and summarizing customer reviews. *In KDD '04 : Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 168–177, New York, NY, USA, 2004. ACM. ISBN 1-58113-888-1.

Paul Jaccard : *Bulletin de la Société Vaudoise des Sciences Naturelles*, 37:241–272, 1901.

J. J. Jiang et D. W. Conrath : Semantic similarity based on corpus statistics and lexical taxonomy. *In International Conference Research on Computational Linguistics (ROCLING X)*, pages 9008+, September 1997. URL http://adsabs.harvard.edu/cgi-bin/nph-bib_query?bibcode=1997cmp.lg....9008J.

Hongyan Jing : Sentence reduction for automatic text summarization. *In Proceedings of the 6th Applied Natural Language Processing Conference*, pages 310–315, 2000.

Karen Sparck Jones : What might be in a summary ? *In Information Retrieval*, pages 9–26, 1993.

David E. Kieras : Thematic processes in the comprehension of technical prose. *In* B. K. Britton et J. B. Black, éditeurs : *Understanding Expository Text*, pages 89–107. Hillsdale, NJ : Lawrence Erlbaum, 1985.

Kevin Knight et Daniel Marcu : Summarization beyond sentence extraction : A probabilistic approach to sentence compression. *Artif. Intell.*, 139(1):91–107, 2002.

Julian Kupiec, Jan Pedersen et Francine Chen : A trainable document summarizer. *In SIGIR '95 : Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 68–73, New York, NY, USA, 1995. ACM. ISBN 0-89791-714-6.

Claudia Leacock et Martin Chodorow : Combining local context and wordnet similarity for word sense identification. *An Electronic Lexical Database*, pages 265–283, 1998.

A. Likas, N. Vlassis, Aristidis Likas, Nikos Vlassis et J.J. Verbeek : The global k-means clustering algorithm. *Pattern Recognition*, 36:451–461, 2001.

Chin-Yew Lin : Rouge : a package for automatic evaluation of summaries. *In Proceedings of the Workshop on Text Summarization Branches Out (WAS 2004)*, Barcelona, Spain, 2004.

Chin-Yew Lin, Guihong Cao, Jianfeng Gao et Jian-Yun Nie : An information-theoretic approach to automatic evaluation of summaries. *In Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics*, pages 463–470, Morristown, NJ, USA, 2006. Association for Computational Linguistics.

Chin-Yew Lin et Eduard H. Hovy : Identifying topics by position. *In ANLP*, pages 283–290, 1997.

Nadine Lucas : La rhétorique des dépêches de presse à travers les marques énonciatives du temps, du lieu et de la personne. *In Semaine du Document Numérique (SDN 2004)*, 06 2004.

H. P. Luhn : The Automatic Creation of Literature Abstracts. *IBM Journal of Research and Development*, 2(2):159–165, 1958.

J. MacQueen : Some methods for classification and analysis of multivariate observations. *In* Lucien M. Le Cam et Jerzy Neyman, éditeurs : *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, Statistics. University of California Press, 1967.

Inderjeet Mani et Mark T. Maybury : *Advances in Automatic Text Summarization*. 1999. URL http://www.mitpressjournals.org/doi/abs/10.1162/coli.2000.26.2.280?cookieSet=1&journalCode=coli.

Inderjeet Mani et D. George Wilson : Robust temporal processing of news. *In Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics*, October 2000.

William C. Mann et Sandra A. Thompson : Rhetorical structure theory : Toward a functional theory of text organization. *Text*, 8(3):243–281, 1988.

Daniel Marcu : The rhetorical parsing, summarization, and generation of natural language texts. Rapport technique, 1997.

Mónica Marrero, Sonia Sánchez-Cuadrado, Jorge Morato Lara et George Andreadakis : Evaluation of named entity extraction systems. *In* Alexander F. Gelbukh, éditeur : *CICLing*, volume 5449 de *Research in Computing Science*, pages 1–12. Springer, 2009.

Ryan McDonald : A study of global inference algorithms in multi-document summarization. *In* Giambattista Amati, Claudio Carpineto et Giovanni Romano, éditeurs : *Proceedings of the 29th European Conference on Information Retrieval Research*, volume 4425 de *Lecture Notes in Computer Science*, pages 557–564. Springer, 2007. ISBN 978-3-540-71494-1.

Kathleen R. McKeown, Regina Barzilay, David Evans, Vasileios Hatzivassiloglou, Judith L. Klavans, Ani Nenkova, Carl Sable, Barry Schiffman et Sergey Sigelman : Tracking and summarizing news on a daily basis with columbia's newsblaster. *In Proceedings of the second international conference on Human Language Technology Research*, pages 280–285, San Francisco, CA, USA, 2002. Morgan Kaufmann Publishers Inc.

Kathleen R. McKeown, Judith L. Klavans, Vasileios Hatzivassiloglou, Regina Barzilay et Eleazar Eskin : Towards multidocument summarization by reformulation : progress and prospects. *In AAAI '99/IAAI '99 : Proceedings of the sixteenth*

national conference on Artificial intelligence and the eleventh Innovative applications of artificial intelligence conference innovative applications of artificial intelligence, pages 453–460, Menlo Park, CA, USA, 1999. American Association for Artificial Intelligence. ISBN 0-262-51106-1.

R. Mihalcea et P. Tarau : TextRank : Bringing order into texts. *In Proceedings of EMNLP-04 and the 2004 Conference on Empirical Methods in Natural Language Processing*, July 2004.

Jean-Luc Minel : *Filtrage sémantique. Du résumé à la fouille de textes.* Hermès, 2003. URL http://hal.archives-ouvertes.fr/hal-00022142/en/.

Johanna D. Moore et Cecile L. Paris : Planning text for advisory dialogues : Capturing intentional and rhetorical information. *COMPUTATIONAL LINGUISTICS*, 19:651–694, 1993.

G. Murray, S. Renals et J. Carletta : Extractive summarization of meeting recordings. *In Proc. Interspeech*, septembre 2005.

Ani Nenkova, Rebecca Passonneau et Kathleen McKeown : The pyramid method : Incorporating human content selection variation in summarization evaluation. *ACM Trans. Speech Lang. Process.*, 4(2):4, 2007. ISSN 1550-4875.

Ani Nenkova, Barry Schiffman, Andrew Schlaiker, Sasha Blair-Goldensohn, Regina Barzilay, Sergey Sigelman, Vasileios Hatzivassiloglou et Kathleen McKeown : Columbia at the document understanding conference 2003. *In In Proceedings of the Document Understanding Workshop DUC 2003*, 2003.

Chikashi Nobata, Satoshi Sekine et Hitoshi Isahara : Evaluation of features for sentence extraction on different types of corpora. *In Proceedings of the ACL 2003 workshop on Multilingual summarization and question answering*, pages 29–36, Morristown, NJ, USA, 2003. Association for Computational Linguistics.

Normes françaises : Recommandations aux auteurs des articles scientifiques et techniques pour la rédaction des résumés. décembre 1984.

Naoaki Okazaki, Yutaka Matsuo et Mitsuru Ishizuka : Improving chronological ordering of sentences extracted from multiple newspaper articles. *ACM Transactions on Asian Language Information Processing (TALIP)*, 4(3):321–339, 2005. ISSN 1530-0226.

Kenji Ono, Kazuo Sumita et Seiji Miike : Abstract generation based on rhetorical structure extraction. *In Proceedings of the 15th conference on Computational linguistics*, pages 344–348, Morristown, NJ, USA, 1994. Association for Computational Linguistics.

Miles Osborne : Using maximum entropy for sentence extraction. *In Proceedings of the ACL-02 Workshop on Automatic Summarization*, pages 1–8, Morristown, NJ, USA, 2002. Association for Computational Linguistics.

Bo Pang, Lillian Lee et Shivakumar Vaithyanathan : Thumbs up ? Sentiment classification using machine learning techniques. *In Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 79–86, 2002.

Randall L. Popken : A study of topic sentence use in academic writing. *Written Communication*, 4(2):209, April 1987.

Mireille Prestini-Christophe : La notion d'evénement dans différents champs disciplinaires. *Pensée Plurielle*, 13:21–29, 2006.

James Pustejovsky : Events and the semantics of opposition. *In* Carol Tenny et James Pustejovsky, éditeurs : *Events as Grammatical Objects*, chapitre 13, pages 445–482. CSLI Publications, 2000.

J. Ross Quinlan : *C4.5 : programs for machine learning*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1993. ISBN 1-55860-238-0.

Dragomir Radev, Timothy Allison, Sasha Blair-Goldensohn, John Blitzer, Arda Çelebi, Stanko Dimitrov, Elliott Drabek, Ali Hakim, Wai Lam, Danyu Liu, Jahna Otterbacher, Hong Qi, Horacio Saggion, Simone Teufel, Michael Topper, Adam Winkel et Zhang Zhu : MEAD - a platform for multidocument multilingual text summarization. *In Proceedings of LREC 2004*, Lisbon, Portugal, May 2004.

Dragomir R. Radev, Sasha Blair-goldensohn et Zhu Zhang : Experiments in single and multidocument summarization using mead. *In In Proceedings of the First Document Understanding Conference*, 2001a.

Dragomir R. Radev, Sasha Blair-Goldensohn, Zhu Zhang et Revathi Sundara Raghavan : Interactive, domain-independent identification and summarization of topically related news articles. *In ECDL '01 : Proceedings of the 5th European Conference on Research and Advanced Technology for Digital Libraries*, pages 225–238, London, UK, 2001b. Springer-Verlag. ISBN 3-540-42537-3.

Owen Rambow, Lokesh Shrestha, John Chen et Christy Laurdisen : Summarizing email threads. *In* Daniel Marcu Susan Dumais et Salim Roukos, éditeurs : *HLT-NAACL 2004 : Short Papers*, pages 105–108, Boston, Massachusetts, USA, May 2 - May 7 2004. Association for Computational Linguistics.

Philip Resnik : Using information content to evaluate semantic similarity in a taxonomy. *In In Proceedings of the 14th International Joint Conference on Artificial Intelligence*, pages 448–453, 1995.

Horacio Saggion : Topic-Based Summarization at DUC 2005. *In Proceedings of DUC 2005 Document Understanding Conference*, volume Document Understanding Workshop, October 2005.

Thiago Alexandre Salgueiro Pardo, Lucia Helena Machado Rino et Maria das Graças Volpe Nunes : Extractive summarization : how to identify the GIST of a text. *In Proceedings of I2TS'2002 - International Information Technology Symposium*, Florianopolis, SC, Brasil, 2002.

Gerard Salton et Christopher Buckley : Term-weighting approaches in automatic text retrieval. *Information Processing and Management : an International Journal*, 24:513–523, 1988.

Gerard Salton et M.J. McGill : *Introduction to Modern Information Retrieval*. 1983.

Helmut Schmid : Probabilistic part-of-speech tagging using decision trees. *In Proceedings of the International Conference on New Methods in Language Processing*, pages 44–49, 1994.

Craig G. Smith : Braddock revisited : The frequency and placement of topic sentences in academic writing. *The Reading Matrix*, 8(1), April 2008.

Karen Sparck Jones : Automatic summarising : factors and directions. *CoRR*, cmp-lg/9805011, 1998.

C. Stoyanov : Toward opinion summarization : Linking the sources. 2005.

Simone Teufel et Marc Moens : Summarizing scientific articles - experiments with relevance and rhetorical status. *Computational Linguistics*, 28:2002, 2002.

Kristina Toutanova, Chris Brockett, Michael Gamon, Jagadeesh Jagarlamudi, Suzuki Hisami et Lucy Vanderwende : The pythy summarization system : Microsoft research at DUC 2007. *In Proceedings of the HLT-NAACL Workshop on the Document Understanding Conference (DUC-2007)*, Rochester, USA, 2007.

Vasudeva Varma, Prasad Pingali, Rahul Katragadda, Sai Krishna, Surya Ganesh, Kiran Sarvabhotla, Harish Garapati, Hareen Gopisetty, Vijay Bharath Reddy, Kranthi Reddy, Rohit Bharadwaj et Praveen Bysani : Iiit hyderabad at tac 2008. *In Proceedings of the TAC 2008 Workshop*, volume Notebook Papers and Results, 2008.

Juha Vesanto, Esa Alhoniemi et Student Member : Clustering of the self-organizing map. *IEEE Transactions on Neural Networks*, 11:586–600, 2000.

Stephen Wan et Kathy McKeown : Generating overview summaries of ongoing email thread discussions. *In COLING '04 : Proceedings of the 20th international conference on Computational Linguistics*, page 549, Morristown, NJ, USA, 2004. Association for Computational Linguistics.

WIKIPEDIA : Micro-trottoir. URL http://fr.wikipedia.org/wiki/Micro-trottoir.

WIKIPEDIA : Wordnet. URL http://fr.wikipedia.org/wiki/WordNet.

René WITTE et Sabine BERGLER : Fuzzy Clustering for Topic Analysis and Summarization of Document Collections. *In* Z. KOBTI et D. WU, éditeurs : *Proc. of the 20th Canadian Conference on Artificial Intelligence (Canadian A.I. 2007)*, LNAI 4509, pages 476–488, Montréal, Québec, Canada, May 28–30 2007. Springer.

Zhibiao WU et Zhibiao WU : Verb semantics and lexical selection, 1994.

Yiming YANG, Jaime G. CARBONEL, Ralf D. BROWN, Thomas PIERCE et Xin Liu BRIAN T. ARCHIBALD : Learning approaches for detecting and tracking news events. *In IEEE Intelligent Systems*, pages 32–43, Cambridge, Massachusetts, 1999.

Jen-Yuan YEH, Hao-Ren KE et Wei-Pang YANG : Chinese text summarization using a trainable summarizer and latent semantic analysis. *In ICADL '02 : Proceedings of the 5th International Conference on Asian Digital Libraries*, pages 76–87, London, UK, 2002. Springer-Verlag. ISBN 3-540-00261-8.

Chin yew LIN et Eduard HOVY : Neats : A multidocument summarizer. *In Proceedings of the Document Understanding Conference (DUC) Workshop on Automatic Summarization*, 2001.

Mehdi YOUSFI-MONOD et Violaine PRINCE : Utilisation de la structure morpho-syntaxique des phrases dans le résumé automatique. *In Actes de la Conférence sur le Traitement Automatique du Langage Naturel 2005*, pages 193–202, 2005.

David M. ZAJIC, Bonnie J. DORR et Jimmy LIN : Single-document and multi-document summarization techniques for email threads using sentence compression. *Information Processing Management*, 44(4):1600–1610, 2008. ISSN 0306-4573.

Klaus ZECHNER : Automatic summarization of open-domain multiparty dialogues in diverse genres. *Comput. Linguist.*, 28(4):447–485, 2002. ISSN 0891-2017.

Mingjing Li ZHIWEI LI, Bin Wang et Wei-Ying MA : A probabilistic model for retrospective news event detection. *In SIGIR 2005 : Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Salvador, Brazil, 2005. ACM.

L. ZHOU et Eduard HOVY : Fine-grained clustering for summarizing chat logs, 2005a.

Liang ZHOU et Eduard HOVY : Digesting virtual "geek" culture : the summarization of technical internet relay chats. *In ACL '05 : Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, pages 298–305, Morristown, NJ, USA, 2005b. Association for Computational Linguistics.

Liang Zhou et Eduard Hovy : On the summarization of dynamically introduced information : Online discussions and blogs. *In Proceedings, The Twenty-First National Conference on Artificial Intelligence and the Eighteenth Innovative Applications of Artificial Intelligence Conference*, pages 237–245. AAAI Press, july 2006.

Résumé

Que ce soit pour des professionnels qui doivent prendre connaissance du contenu de documents en un temps limité ou pour un particulier désireux de se renseigner sur un sujet donné sans disposer du temps nécessaire pour lire l'intégralité des textes qui en traitent, le résumé est une aide contextuelle importante. Avec l'augmentation de la masse documentaire disponible électroniquement, résumer des textes automatiquement est devenu un axe de recherche important dans le domaine du traitement automatique de la langue.

La présente thèse propose une méthode de résumé automatique multi-documents fondée sur une classification des phrases à résumer en classes sémantiques. Cette classification nous permet d'identifier les phrases qui présentent des éléments d'informations similaires, et ainsi de supprimer efficacement toute redondance du résumé généré. Cette méthode a été évaluée sur la tâche « résumé d'opinions issues de blogs » de la campagne d'évaluation TAC 2008 et la tâche « résumé incrémental de dépêches » des campagnes TAC 2008 et TAC 2009. Les résultats obtenus sont satisfaisants, classant notre méthode dans le premier quart des participants.

Nous avons également proposé d'intégrer la structure des dépêches à notre système de résumé automatique afin d'améliorer la qualité des résumés qu'il génère. Pour finir, notre méthode de résumé a fait l'objet d'une intégration à un système applicatif visant à aider un possesseur de corpus à visualiser les axes essentiels et à en retirer automatiquement les informations importantes.

Mots clés : Résumé automatique, Extraction d'information, Traitement automatique de la langue.

Abstract

Professionals who have to peruse documents in a limited amount of time or private individuals who want to be informed about a specific topic without having the time to read all the texts about it both need summaries. The increase in electronic documents available have made the research in automatic summarization an important domain in the field of natural language processing.

We propose a method based on a sentence classification in semantic clusers, using similarity calculation between sentences. This step allows us to identify the sentences which convey the same information and to remove redundancy from the automatically generated summaries. This method has been evaluated on the « opinion summarization » task of TAC 2008 evaluation campagin, and on the « news summarization » task of TAC 2008 and TAC 2009 campaigns. Our system ranks itself among the the first quarter of the participating systems.

We also propose to integrate newswire articles structure to our summarization system in order to improve the quality of the summaries it generates. Our summarization method has also been integrated to a larger application which aims to help the user to visualize the main topics of a corpus and to automatically extract the essential information.

Keywords : Automatic summarization, Information extraction, Natural language processing.

Discipline : Informatique

Laboratoire d'informatique de Paris-Nord – UMR CNRS 7030