

# Résumé Automatique Multi-Document Dynamique : État de l'Art

Maâli Mnasri<sup>1,2</sup>

(1) CEA,LIST, Laboratoire Vision et Ingénierie des Contenus, Gif-sur-Yvette, F-91191 France.

(2) Univ. Paris Sud, Orsay, France

maali.mnasri@cea.fr

**Résumé.** Les travaux menés dans le cadre du résumé automatique de texte ont montré des résultats à la fois très encourageants mais qui sont toujours à améliorer. La problématique du résumé automatique ne cesse d'évoluer avec les nouveaux champs d'application qui s'imposent, ce qui augmente les contraintes liées à cette tâche. Nous nous intéressons au résumé extractif multi-document dynamique. Pour cela, nous examinons les différentes approches existantes en mettant l'accent sur les travaux les plus récents. Nous montrons ensuite que la performance des systèmes de résumé multi-document et dynamique est encore modeste. Trois contraintes supplémentaires sont ajoutées : la redondance inter-document, la redondance à travers le temps et la grande taille des données à traiter. Nous essayons de déceler les insuffisances des systèmes existants afin de bien définir notre problématique et guider ainsi nos prochains travaux.

## Abstract.

### Automatic multi-document update summarization : State of the Art

The field of automatic text summarization is characterized both by some interesting achievements and a lot of issues to address, especially with the introduction of new tasks brought by applications. In this article, we focus more particularly on the multi-document update summarization task and review the existing work about it with a special emphasis on recent work. We show that the results for this task are still low because of the necessity to take into account three important constraints : information redundancy through documents and time and the size of data. We analyze the strengths and weaknesses of existing systems according to these constraints to propose subsequently new solutions.

**Mots-clés :** Résumé multi-document, résumé dynamique, redondance, évaluation.

**Keywords:** Multi-document summarization, update summarization, redundancy, evaluation.

## 1 Introduction

Le résumé multi-document dynamique (dit aussi évolutif ou mis à jour) est l'une des thématiques de recherche récentes à laquelle les chercheurs en Traitement Automatique des Langues (TAL) se sont intéressés, en particulier à l'occasion des campagnes d'évaluation TAC (Text Analysis Conference) et précisément à travers les tâches de résumé dynamique. Étant donné un flux de documents (par exemple des dépêches de presse), l'objectif est de générer un résumé se concentrant sur les nouveautés apportées par les documents les plus récents par rapport aux documents plus anciens. Cet article dresse un état de l'art des travaux sur le résumé automatique. Il présente une revue des développements récents dans ce domaine qui constitueront le point de départ dans nos travaux à venir. Plusieurs revues sur le résumé automatique ont déjà été publiées (Spärck Jones, 2007; Das & Martins, 2007; Gupta & Lehal, 2010; Nenkova & McKeown, 2012; Torres-Moreno, 2014). Nous essayons, à travers cet article, de les compléter. À cet effet, nous présentons l'évolution du domaine ainsi que les dernières réalisations remarquables. De plus, nous structurons notre analyse autour de deux axes principaux : le résumé multi-document et le résumé dynamique. Il existe d'autres axes intéressants, par exemple les résumés multilingues et crosslingues, mais qui ne font pas partie des objectifs de cet article.

Les recherches que nous considérons ici sont motivées par le besoin croissant d'applications variées de synthèse d'information en général et de résumé automatique de texte en particulier. Dans le cas des moteurs de recherche par exemple, un système de résumé permet de présenter un contenu bref et pertinent par rapport à une requête de l'utilisateur (Nenkova & McKeown, 2012). Le résumé d'articles scientifiques fait aussi partie des principaux cas d'utilisation (Jaidka *et al.*, 2013) compte tenu du nombre croissant de documents scientifiques disponibles sous une forme numérique dont la lecture et la compilation deviennent de plus en plus difficiles. Le domaine de la presse est aussi un champ d'application majeur. La consultation des dépêches en ligne à partir des terminaux mobiles est pénible si les articles sont longs et nombreux. Le résumé automatique résout ce problème en présentant les informations les plus importantes sous une forme réduite (Plaza

*et al.*, 2010).

Dans la section suivante, nous analysons le problème du résumé automatique en caractérisant les différents types de résumés. Nous présentons dans la troisième section le principe et les réalisations traitant la problématique du résumé abstraitif. L'explication détaillée des techniques utilisées dans le cadre du résumé extractif fera l'objet de la quatrième section. Nous nous focalisons dans la cinquième section sur le résumé multi-document dynamique, point central de notre attention dans cet article. La sixième section est consacrée aux méthodes d'évaluation actuelles des systèmes de résumé automatique. Nous clôturons cet article par un travail de synthèse présenté dans la dernière section.

## 2 Résumé automatique : Analyse du problème

Les résumés automatiques et leurs méthodes peuvent être catégorisés selon différents critères (Nenkova & McKeown, 2012). Nous citons les plus importants et les plus utilisés dans la littérature.

**Mode de production du résumé.** Nous distinguons les méthodes *extractives* (Dalal & Malik, 2013) des méthodes *abstractives* (Genest & Lapalme, 2012). Le résumé extractif est formé de phrases extraites du texte source. Les premiers travaux en résumé automatique se sont appuyés sur cette approche (Luhn, 1958) en exploitant la fréquence des mots. Les critères de sélection ont ensuite été enrichis en tenant compte du contenu et de la structure du texte (Edmundson, 1969) (cf. section 4.1). Jusqu'à aujourd'hui, ces méthodes sont les plus exploitées parce qu'elles évitent le problème de la génération de texte, toujours considéré comme une tâche complexe. Les méthodes abstractives sont inspirées des travaux en psycholinguistique cognitive et en intelligence artificielle et notamment du modèle théorique de la compréhension de van Dijk et Kintsch (Kintsch & van Dijk, 1978). Ce dernier considère le résumé d'un texte comme le produit de sa compréhension. Celle-ci est modélisée par la mise en relation sémantique des composants du texte dans une structure adaptée (par exemple un graphe de cohérence). Un résumé abstraitif est le produit de la synthèse de la représentation sémantique du texte source avec des phrases générées automatiquement.

**Portée du résumé.** Les systèmes de résumé automatique peuvent être mono-document ou multi-document. Les premiers produisent des résumés pour un seul document et peuvent être plus ou moins adaptés à des tailles différentes de documents : résumer un article ne pose pas tout à fait le même problème que résumer un rapport scientifique. Le système CHORAL (García Flores *et al.*, 2009) fondé sur l'analyseur lexical LIMA (de Chalendar, 2014) se distingue ainsi par son efficacité sur les documents longs. Il produit des résumés de 1 à 5 pages pour un rapport de thèse. Les systèmes de résumé multi-document, plus récents, génèrent des résumés de taille ajustable d'un ensemble de documents.

**Généricité du résumé.** Un résumé de texte est soit générique, soit orienté. Le résumé générique est produit en se référant uniquement au contenu du texte source et indépendamment de son contexte. En revanche, le résumé orienté est guidé par une tâche ou une requête. Dans ce cas, seule l'information en relation avec la tâche ou la requête est sélectionnée. Ce type de résumé dépend donc fortement du contexte. Ce dernier peut être défini comme un ensemble de facteurs d'entrée du système de résumé automatique (Spärck Jones, 2007). Il couvre l'audience, l'usage, le cadre spatio-temporel, etc.

**Style du résumé.** Un résumé est soit informatif, soit indicatif. Le résumé informatif est un modèle rétréci du texte d'origine relatant le plus largement possible les informations du document. En revanche, un résumé indicatif rend compte des sujets les plus importants évoqués par le texte. Certains systèmes de résumés guidés (Saggion & Lapalme, 2002) génèrent un résumé indicatif du texte comme étape initiale. L'utilisateur choisit parmi les sujets proposés par le résumé ceux qui l'intéressent. Le système produit alors un résumé informatif du texte guidé par la requête de l'utilisateur. La requête dans ce cas est l'ensemble des sujets sélectionnés à partir du résumé indicatif.

## 3 Résumé abstraitif

Bien que nous nous intéressions surtout aux systèmes de résumé extractifs, les systèmes abstraitifs partagent avec le résumé dynamique une certaine forme de modélisation du contenu des documents, même si les critères d'extraction dans le cas dynamique sont généralement sémantiquement moins profonds. Les méthodes de résumé abstractives imitent, jusqu'à un certain degré, le processus naturel accompli par l'homme pour résumer un document. Par conséquent, elles produisent des résumés plus similaires aux résumés manuels. Ce processus peut être décrit par deux étapes majeures : la compréhension du texte source et la génération du résumé (Khan & Salim, 2014). Ces deux tâches sont assez complexes. C'est pourquoi elles ont été simplifiées. La première étape vise à analyser sémantiquement le contenu du texte et à identifier les parties à exprimer dans le résumé. Elle a parfois pris la forme d'une tâche d'extraction d'information liée au domaine abordé (Genest & Lapalme, 2011, 2012) ou de regroupement des phrases du texte source (Filippova, 2010). La génération de texte est un domaine en soi. Une des approches simplifiées consiste à appliquer des techniques de génération *text-to-text* : utilisation de paraphrases (Madnani & Dorr, 2010) ou fusion et compression de phrases (Filippova, 2010). Une alternative consiste à induire un modèle textuel du domaine (patron) et de l'instancier lors de la génération (Cheung

*et al.*, 2013). Ces méthodes ne sont pas très largement exploitées. Ceci peut être dû à la rareté des outils de génération du texte. Le domaine de la génération est toujours en cours de développement mais pas encore à maturité, ce qui freine parfois l'implémentation des systèmes abstraits. Les chercheurs préfèrent alors se tourner vers les méthodes extractives, qui ne dépendent pas de ce prérequis. La majorité des travaux s'étant intéressés aux méthodes extractives, ces dernières ont connu un développement plus important. La majorité des évaluations menées sur le résumé ont aussi été conçues pour des résumés plutôt extractifs. Ceci explique en partie les résultats moins encourageants des systèmes abstraits. Néanmoins, certains chercheurs pensent que ceux-ci pourraient susciter un certain regain d'intérêt. Cette prédiction est justifiée d'une part, par le besoin de résumés plus proches des résumés manuels (Nenkova & McKeown, 2012) et d'autre part, par le plafonnement des performances des techniques extractives.

## 4 Résumé extractif

Le point fort du résumé par extraction est qu'il évite la génération de texte. Ceci permet d'une part, de se concentrer sur la sélection du contenu pertinent et d'autre part, d'obtenir un résumé lisible et linguistiquement correct. La cohérence n'est en revanche pas garantie. Par exemple, si le système de résumé sélectionne des phrases contenant des références (acronyme, pronom personnel, etc.) et ne sélectionne pas les phrases contenant leurs antécédents, il est fort probable que le résumé produit soit incompréhensible. Pour pallier ce problème, certains travaux considèrent le paragraphe comme unité d'extraction au lieu de la phrase (Salton *et al.*, 1996). Ceci permet de garder la cohérence du texte source mais ne peut pas être applicable dans le cas de résumés courts. De plus, il est évident que cette méthode réduit la précision du résumé en y incluant des phrases peu importantes juste pour améliorer la cohérence. D'autres chercheurs procèdent à des étapes de pré/post-traitement du texte qui améliorent partiellement la cohérence globale du résumé, comme par exemple la résolution des références anaphoriques dans le texte source (Trandabâţ, 2011). Le processus principal dans le résumé extractif est la sélection des segments de textes (généralement les phrases) pertinents et non redondants sans dépasser une taille limite de résumé. Ce principe limite la couverture des informations apportées par le texte source. Les résumés abstraits souffrent moins de ce problème puisque l'information peut y être reformulée.

### 4.1 Critères de sélection

Dans cette partie nous détaillons les critères de sélection des unités textuelles utilisés par les systèmes de résumé. Les unités textuelles dépendent du modèle de langue choisi. Elles peuvent être des phrases, des N-grammes ou n'importe quel segment du texte. Ces critères ne sont pas exclusifs d'une méthode bien déterminée mais sont applicables à tous les types de résumés extractifs qu'ils soient mono-document, multi-document ou dynamiques.

#### 4.1.1 Critères liés au contenu du texte

Cet ensemble de critères s'intéressent au contenu du texte et aux informations qu'il apporte. Le contenu est analysé soit par des approches de surface, comme le calcul des fréquences d'occurrence des mots, soit par des approches sémantiques qui exploitent les sens des mots et leurs relations sémantiques, comme avec l'annotation en rôles sémantiques. Nous citons, dans ce qui suit, les critères les plus utilisés.

**Fréquence d'occurrence des mots.** Ce critère a été introduit initialement par Luhn (Luhn, 1958). L'idée est que les mots les plus fréquents sont les plus liés au sujet du texte. La fréquence d'occurrence des mots est largement exploitée, même dans des systèmes récents. La différence est qu'elle est combinée à d'autres critères. Même les méthodes reposant sur l'analyse sémantique des mots utilisent la fréquence d'occurrence comme première étape pour déterminer les thèmes principaux abordés par le texte. Le point fort de ce critère est qu'il est totalement indépendant de la langue.

**Similarité entre les phrases.** La similarité des textes est une notion très importante en TAL. Plusieurs mesures de similarité textuelle ont été établies (Bär *et al.*, 2015). Dans le domaine du résumé automatique, elle est d'abord exploitée pour l'élimination de la redondance mais aussi plus indirectement pour la sélection de phrases pertinentes, sans oublier la comparaison avec des résumés modèles lors de l'évaluation. Certaines méthodes de résumé s'appuient uniquement sur ce critère. Tel est le cas de l'algorithme de résumé mono-document *TextRank* (Mihalcea, 2004). Ce critère est par ailleurs particulièrement important dans le cas multi-document. Dans ce contexte, les documents sont généralement représentés par des vecteurs de mots pondérés avec une mesure comme TF\*IDF (Term Frequency \* Inverse Document Frequency) (Sammut & Webb, 2010) et regroupés selon la similarité de leurs vecteurs. Plus une phrase est similaire au barycentre du regroupement, plus elle décrit les informations caractéristiques du groupe de documents considéré (Radev *et al.*, 2004; Neto *et al.*, 2003) et peut être alors considérée comme représentative de ce groupe, ce qui est un critère de sélection important.

**Reconnaissance d'entités nommées / Annotation en rôles sémantiques.** La reconnaissance des entités nommées dans un texte améliore le filtrage des informations pertinentes (Hassel, 2003). Elle sert aussi à répondre à des requêtes standards

(OÙ, QUI, QUAND, etc.) dans le résumé guidé (Ng *et al.*, 2011). Certains vont au delà de cette étape et déterminent les rôles sémantiques des entités reconnues (Trandabâ, 2011). L'entité la plus fréquente est identifiée, c'est l'entité principale. Par la suite, les phrases contenant cette entité sont sélectionnées. Enfin, seules les phrases où l'entité principale possède un rôle sémantique fondamental (non auxiliaire) sont gardées pour le résumé. Les rôles sémantiques peuvent aussi être utilisés pour simplifier les phrases complexes, c'est à dire les phrases contenant deux prédicats ou plus. Le prédicat est généralement un verbe. Dans ce cas, les prédicats pour lesquels l'entité principale a un rôle auxiliaire sont éliminés.

Ces critères mettent l'accent sur le contenu du texte et le message qu'il communique. Il existe d'autres critères qui ne s'intéressent pas au contenu du texte, mais qui renferment des informations très importantes et décisives dans l'étape de sélection. Elles font l'objet du paragraphe suivant.

#### 4.1.2 Critères liés à la forme et à la structure du texte

La structure du texte est très importante dans le jugement de la pertinence d'une phrase. En effet, lors de la rédaction d'un texte, l'ordre des phrases n'est pas arbitraires. De plus, les styles de rédaction diffèrent d'un domaine à l'autre. Par exemple, dans le domaine journalistique, les informations les plus importantes sont souvent mentionnées au début du texte. Ceci n'est pas toujours le cas dans un article scientifique ou un roman. Ce facteur a été exploité par les chercheurs en TAL pour déterminer l'importance des segments textuels. Nous expliquons dans cette partie les critères les plus importants.

**Position de la phrase.** Ce critère dépend de la nature du document et de son genre. Les phrases se trouvant au début sont généralement plus informatives et décrivent le sujet principal du document. De plus, les phrases situées au début de chaque paragraphe tendent à apporter plus d'informations pertinentes (Lin & Hovy, 1997; McKeown *et al.*, 1999). Dans le résumé des articles scientifiques, certains travaux se sont appuyés principalement sur la structure des articles (Jaidka *et al.*, 2013) pour générer des revues scientifiques. Les revues descriptives (résumé informatif) sont formées par les phrases des parties *Résumé* et *Introduction* sont extraites. En revanche, dans le cas des revues intégratives (critique et comparaison des études), les phrases les mieux notées sont celles des parties *Résultats et discussion* et *Conclusion*. Cette approche est déduite de l'analyse d'un corpus de 20 revues scientifiques et de 349 références pointées par ces revues. Il a été constaté que plus que 25% des informations contenues dans les revues ont été extraites de la partie *Résumé* des articles sources.

**Similarité avec le titre.** Plus une phrase est similaire au titre, plus elle est liée au sujet principal du texte (Edmundson, 1969; Perret, 2005) étant donné que dans la majorité des cas le titre informe de façon très brève sur le contenu principal du texte. La similarité avec les sous-titres est aussi considérée comme indicateur de pertinence.

**Longueur de la phrase.** La longueur moyenne d'une phrase dans un texte dépend de son genre. Généralement, les phrases très courtes sont considérées comme peu informatives alors que les phrases très longues sont soupçonnées de détailler des informations déjà exprimées dans l'ensemble des documents par des phrases plus courtes et donc de favoriser la redondance. Cette caractéristique est exploitée en fixant un intervalle de longueur (entre 15 et 30 mots). Une phrase ayant une longueur en dehors de cet intervalle est pénalisée (Schiffman *et al.*, 2002).

**Les mots indices (cue words).** Ce critère prend la forme d'une liste de mots activant ou inhibant la sélection d'une phrase, généralement en fonction du rôle qu'ils permettent d'attribuer à la phrase dans laquelle ils apparaissent (exemple, conclusion, etc.) (Edmundson, 1969). Ces listes sont constituées manuellement ou définies par apprentissage à partir d'un corpus de documents représentatifs (Mani, 2001). Elles peuvent inclure des noms propres (Neto *et al.*, 2003) et des dates.

**Analyse du discours.** La Rhetorical Structure Theory (RST) (Mann & Thompson, 1988) est une méthode d'analyse de discours. Elle s'appuie sur une segmentation des textes en EDU, unités discursives élémentaires (phrases ou parties de phrase). Chaque EDU est soit noyau soit satellite selon son importance dans le texte. Un noyau apporte à lui seul une information jugée pertinente. Un satellite est en revanche complémentaire. La RST représente la cohérence et la structure du texte par un ensemble de relations rhétoriques entre les noyaux et les satellites : exemplification, preuve, justification, etc. Contrairement aux autres critères de sélection, ceux exploitant la RST tiennent ainsi compte de la structure du texte (Marcu, 1997, 1998). Les analyseurs RST (Joty *et al.*, 2013; Feng & Hirst, 2014) produisent un arbre binaire dont les feuilles représentent les EDU. Les autres nœuds représentent les relations rhétoriques entre les EDU et portent le numéro de leurs fils noyaux. De cette façon, les nœuds les plus proches de la racine sont les plus importants. L'ordre d'importance des EDU peut alors être déduit en parcourant l'arbre de la racine vers les feuilles. L'architecture d'arbre imposée par la RST n'est pas souvent représentative du texte, notamment lorsque deux EDU présentent plus qu'une relation rhétorique ou quand un nœud doit avoir deux parents. Pour résoudre ce problème, un nouveau modèle sous forme de graphe dont les arcs représentent les relations rhétoriques a été proposé (Wolf & Gibson, 2005). La comparaison de ces deux représentations (graphe et arbre) pour le résumé automatique a fait l'objet d'un certain nombre d'études (Louis *et al.*, 2010).

Nous avons cité les critères de sélection les plus utilisés pour le résumé automatique. Le choix de bons critères n'est pas suffisant pour obtenir un bon résumé. Il faut savoir comment les utiliser et quel degré d'importance accorder à chacun pour produire un résumé satisfaisant. La section suivante s'intéresse à cette problématique.

## 4.2 Exploitation et intégration des critères

Il est très rare qu'un système de résumé automatique utilise un seul critère pour sélectionner les phrases du texte source. Plusieurs critères sont combinés. Les méthodes d'intégration sont assez nombreuses. Nous décrivons dans cette partie les différentes méthodes pour combiner les critères et les utiliser pour sélectionner les phrases du résumé.

### 4.2.1 Méthodes par apprentissage

Du point de vue de l'apprentissage, le résumé automatique a été considéré aussi bien comme un problème de classification que comme un problème de régression. Étant donné un ensemble de textes source et leurs résumés, les méthodes par apprentissage visent à apprendre un modèle de choix des phrases du résumé. Les phrases des textes source sont caractérisées par divers critères de sélection. Dans l'approche par régression, le modèle prédit les scores des phrases (Conroy *et al.*, 2011). La décision est alors quantifiée. L'ordonnement des phrases reste à la charge du système de résumé. Dans l'approche par classification, le modèle choisi distingue les phrases du texte à inclure dans le résumé et celles à ne pas inclure dans le résumé. Le modèle bayésien naïf donne généralement les meilleurs résultats (Neto *et al.*, 2003). Les réseaux de neurones artificiels (RNA) ont aussi été utilisés pour l'apprentissage supervisé (Kaikhah, 2004). Chaque phrase du texte est modélisée par un vecteur de  $n$  composantes, chacune correspondant à un critère de sélection. Le RNA est composé de  $n$  neurones d'entrées (un neurone par critère), une couche cachée de  $p$  neurones et un neurone de sortie. Ce dernier indique si la phrase en entrée doit être incluse dans le résumé. L'apprentissage permet d'adapter le poids des liaisons entre les couches en fonction d'un ensemble d'exemples. À l'issue de cet apprentissage, les liaisons de très faible poids sont éliminées, de même que les neurones isolés.

### 4.2.2 Méthodes fondées sur la programmation linéaire en nombres entiers

McDonald fut le premier à modéliser le problème du résumé automatique multi-document par le biais de la Programmation Linéaire en Nombres Entiers (PLNE) (McDonald, 2007). Depuis, plusieurs chercheurs utilisent cette méthode pour combiner différents critères de sélection (Li *et al.*, 2011; Woodsend & Lapata, 2012; Li *et al.*, 2013). Le principe consiste à maximiser une fonction objectif favorisant les unités textuelles satisfaisant les critères de sélection et pénalisant la redondance entre les unités sélectionnées. Le calcul est effectué sous un ensemble de contraintes à fixer selon le système de résumé. La contrainte triviale est la taille souhaitée du résumé. Le modèle proposé par McDonald attribue des poids aux phrases du texte source. La PLNE permet de choisir les phrases ayant les poids maximaux et une redondance minimale avec les phrases déjà sélectionnées sans dépasser la taille maximale du résumé. D'autres modèles extraient du texte les concepts qu'il évoque et leur associent des poids. La PLNE permet de choisir les phrases couvrant les concepts ayant les poids les plus élevés sans dépasser la longueur du résumé (Gillick & Favre, 2009). Une contrainte d'exclusivité des concepts a été ajoutée pour éviter la redondance. Un concept n'est autorisé à apparaître qu'une seule fois dans le résumé. La difficulté réside dans le choix des concepts pouvant être les sujets les plus importants, des entités nommées ou des relations sémantiques. Cette étape a été simplifiée dans certains travaux en supposant que les concepts sont tous les bigrammes du texte source (Gillick & Favre, 2009). Ce choix a été justifié par la complexité et la fragilité des analyses sémantiques, qui peuvent donner des concepts erronés ou ambiguës et par conséquent un résumé hors-sujet.

### 4.2.3 Méthodes fondées sur les graphes

En représentant un texte sous la forme d'un graphe de phrases, il devient possible d'appliquer un certain nombre d'algorithmes génériques, comme l'algorithme PageRank (Page *et al.*, 1999), pour déterminer l'importance relative de celles-ci. PageRank est un algorithme de classement utilisé par le moteur de recherche de Google. Il représente les pages Web par les sommets d'un graphe et les liens par les arcs. Il attribue récursivement à chaque nœud un score dépendant de la structure de tout le graphe. TextRank est un algorithme pour le résumé automatique mono-document fondé sur les graphes (Mihalcea, 2004). Le texte est représenté par un graphe où les sommets sont tout simplement les phrases du texte. Alors que les arcs des graphes rhétoriques (cf. section 4.1.2) représentent des relations rhétoriques entre les phrases, les arcs dans TextRank représentent leurs similarités. Pour ne pas favoriser les phrases longues au détriment des phrases courtes, la valeur de la similarité entre deux unités textuelles est divisée par la somme de leur longueur. Initialement, à chaque sommet est attribué un score aléatoire. Par la suite, à chaque itération de l'algorithme TextRank, le score de chaque nœud est calculé récursivement en fonction de sa similarité avec ses voisins et des scores de ces derniers. La même approche a été appliquée pour le résumé multi-document en français (Boudin & Torres-Moreno, 2009). Pour éliminer la redondance, un seuil de similarité maximal a été fixé. Au-delà de ce seuil, deux sommets ne sont plus connectés.

## 5 Résumé multi-document dynamique

Depuis quelques années déjà, les recherches se concentrent beaucoup plus sur le résumé multi-document que sur le résumé mono-document. Plus récemment, a émergé le résumé dynamique. Nous décrivons dans ce paragraphe les spécificités de chaque type de résumé et les contraintes qu'il impose.

## 5.1 Résumé multi-document

Un système de résumé multi-document permet de produire un résumé d'une collection de textes en rendant compte de ses idées principales. Les méthodes de résumé citées à la section précédente peuvent être appliquées pour le résumé mono ou multi-document. Cependant, certaines sont plus adaptées que d'autres au résumé multi-document. Par exemple, les méthodes fondées sur la programmation linéaire ont montré plus de succès que les méthodes fondées sur les graphes. En effet, la pluralité des documents impose de nouvelles contraintes que nous détaillons ci-dessous.

### 5.1.1 Redondance inter-document

Le problème de la redondance est davantage présent dans le cadre du multi-document, apparaissant à deux niveaux : entre les phrases du même document et entre les phrases de différents documents. Il se pose de façon plus aiguë encore lorsque les documents à résumer sont thématiquement homogènes, ce qui est souvent le cas. Par exemple, si les textes source sont des articles de presse concernant le même événement, il est très probable que les phrases les plus importantes de chaque texte soient très similaires. L'adoption d'une approche de résumé statistique fondée sur la fréquence d'occurrence des mots conduit à un résumé tendant à sur-représenter la même information. Bien que cette information soit la plus pertinente dans tous les documents, le résumé obtenu est pauvre et ne rappelle pas tout ce dont parle l'ensemble des textes. Ce type d'approches convient plus à l'identification du sujet principal d'une collection de documents. Pour résoudre le problème de redondance inter-document, différentes solutions ont été proposées. La première famille d'approches commence premier lieu par ordonner les phrases par ordre décroissant de pertinence. Dans un second temps, les phrases du résumé sont sélectionnées en débutant par les phrases les mieux notées et en comparant chaque phrase aux phrases déjà choisies pour le résumé. Si leur similarité dépasse un seuil donné, la phrase n'est pas retenue pour le résumé. La deuxième famille d'approches repose dans la plupart des cas sur l'analyse thématique des documents exploitant des facteurs superficiels comme la similarité lexicale ou des facteurs plus profonds comme la similarité sémantique. Cette dernière permet de détecter la redondance des informations au-delà d'une similarité de surface. En effet deux phrases peuvent exprimer exactement la même information sans pour autant avoir des mots en commun.

**Clustering de documents.** Une manière assez répandue d'aborder le résumé multi-document est d'adopter une approche en deux temps (Radev *et al.*, 2004). Les documents similaires sont d'abord regroupés en *clusters*. Chaque cluster est ensuite résumé en extrayant une ou plusieurs phrases des documents qu'il contient. Cette extraction peut le cas échéant être réalisée par des méthodes mono-document en considérant le cluster comme un unique document. Le résumé final est la concaténation des phrases représentant chacun des clusters. Cette façon de faire permet en particulier de limiter la combinatoire de recherche des redondances entre phrases.

**Segmentation thématique.** La segmentation thématique permet de découper les textes en segments contigus thématiquement homogènes en s'appuyant sur la distribution du vocabulaire dans les textes ou sur des marques linguistiques. Dans le cadre du résumé multi-document, elle constitue un outil permettant de travailler avec des unités textuelles homogènes entre le niveau du texte et celui de la phrase et facilite ainsi la détection des similarités thématiques tout en réduisant la combinatoire des comparaisons (Angheluta *et al.*, 2002; Ferret *et al.*, 2004). TextTiling (Hearst, 1997) est un exemple d'algorithme de segmentation thématique très utilisé dans ce cadre (Neto *et al.*, 2000).

**Identification des thèmes.** D'autres chercheurs ont choisi d'identifier d'abord les thèmes ou les événements majeurs mentionnés dans le texte (Arora & Ravindran, 2008; Li *et al.*, 2011). Ensuite, ils classifient les phrases par thème et choisissent une ou plusieurs phrases pour couvrir chaque thème.

### 5.1.2 Problème combinatoire

Nous avons montré précédemment que les approches classiques du résumé automatique mono-document ne sont pas toujours adéquates dans le cas du multi-document parce qu'elles ne prennent pas en compte la redondance inter-document. Une autre raison de l'insuffisance de ces méthodes est qu'elles ont été conçues pour opérer sur un seul document à la fois, c'est-à-dire sur des données de petite taille. Le passage au résumé multi-document signifie le passage à des données plus volumineuses. Certaines approches, (Li *et al.*, 2011) pour ne citer qu'un exemple, organisent les traitements effectués sur les textes en pipeline. Cette architecture oblige à parcourir l'ensemble des documents autant de fois que le nombre de traitements à réaliser. Une telle organisation est très coûteuse et peut réduire considérablement l'utilisabilité du système proposé. La grande taille des données doit être prise en compte en amont de la conception du modèle de façon à réduire les parcours séquentiels des documents et paralléliser les opérations au maximum.

## 5.2 Résumé dynamique : une dimension temporelle

Le résumé dynamique est une variante du résumé automatique multi-document incluant la dimension supplémentaire du temps. Alors que dans le problème du résumé multi-document les données d'entrée sont statiques, le résumé dynamique introduit une difficulté supplémentaire en faisant varier les données d'entrée sur l'axe du temps. Les travaux sur ce type de résumé peuvent être classés en deux catégories. Les systèmes de résumé dynamiques séquentiels produisent un résumé

rendant compte de l'information portée par les documents couvrant une période donnée en prenant comme point de référence les informations connues juste avant cette période, incarnées par un résumé (Yang *et al.*, 2013; Xu *et al.*, 2013). Les systèmes de résumé dynamiques incrémentaux produisent quant à eux des mises à jour d'un résumé initial à chaque fois que des informations nouvelles apparaissent concernant l'objet du résumé initial (Chowdary & Kumar, 2008; McCreadie *et al.*, 2014).

### 5.2.1 Formalisation du problème

La formalisation classique du problème considère deux instants  $t$  et  $t+1$ . Étant donné un ensemble de documents A à l'instant  $t$  et un autre ensemble B à l'instant  $t+1$  plus récent, il s'agit de produire un résumé des textes de l'ensemble B sous l'hypothèse que le lecteur a déjà pris connaissance de toutes les informations apportées par l'ensemble A. Autrement dit, il faut résumer les documents de B sans répéter ce qui a été évoqué dans A. Ceci peut être considéré comme la combinaison des problèmes du résumé automatique et de la détection de nouveauté.

### 5.2.2 Redondance à travers le temps

La contrainte nouvelle imposée par ce type de résumé est la gestion de la redondance à travers le temps entre les deux ensembles A et B, qui s'ajoute aux contraintes de redondance inter-document et intra-document héritées respectivement du résumé multi-document et mono-document. Une première approche adoptée pour répondre à cette question a réduit le problème du résumé dynamique en un problème de résumé multi-document. Un résumé est généré d'abord pour chaque ensemble (A et B). Ensuite, le résumé de B est modifié de façon à éliminer ce qui est redondant avec le résumé de A. Cette méthode n'est pas très performante car seul le contenu du résumé de l'ensemble A n'est pas autorisé à apparaître dans le résumé de B. Rien n'empêche alors que d'autres informations des documents de A soient incluses dans le résumé de B. C'est pourquoi les meilleurs systèmes de résumé dynamiques actuels considèrent la totalité des textes de A pour l'élimination de la redondance. Les solutions peuvent être plus précisément classées en deux catégories : des solutions par élimination et des solutions par évitement. Les solutions par élimination traitent l'ensemble de documents B sans aucune prise en compte de l'ensemble A. Une fois les phrases sélectionnées ou ordonnées, la redondance est éliminée par la suppression des phrases similaires au contenu de l'ensemble A. Les solutions par évitement considèrent au contraire la redondance comme critère lors de l'attribution des scores. Dans certains cas, la redondance peut être justifiée voire bénéfique. En effet, les informations marginales de l'ensemble A peuvent acquérir plus d'importance à travers le temps. Elles apparaissent alors comme des informations principales dans B. Actuellement, les systèmes conçus ne considèrent pas ce cas.

### 5.2.3 Travaux récents

Dans ce qui suit, nous décrivons les méthodes adoptées par les meilleurs systèmes de résumé dynamique dans TAC 2011 en soulignant parallèlement les approches proposées pour le résumé multi-document puisque la majorité des travaux traitent les deux problèmes simultanément. Certaines approches consistent à entraîner un modèle de régression SVR (*Support Vector Regression*) sur les données de TAC 2010 afin de prédire le score de chaque phrase des textes sources (Ng *et al.*, 2011). Pendant la phase d'apprentissage, les scores sont calculés en fonction de la similarité ROUGE-2 (cf. paragraphe 6.1) avec les résumés manuels. Le problème de redondance est traité lors du classement des phrases en utilisant l'algorithme *Maximal Marginal Relevance*. Celui-ci pénalise le score d'une phrase par sa similarité avec les phrases déjà retenues dans le résumé et par sa similarité avec la phrase la plus similaire de l'ensemble de documents A. D'autres chercheurs ont suivi une approche par évitement (Wan, 2012). Ils modélisent toutes les phrases des ensembles A et B par les sommets d'un seul graphe. De cette manière, la redondance est évitée *a priori*. Pour chaque sommet, sont calculés itérativement l'apport de nouveauté et la corrélation avec les sujets principaux du document en déterminant les similarités et dissimilarités entre les composantes de tout le graphe. À l'issue de la convergence de l'algorithme, les sommets ayant l'apport de nouveauté maximal sont sélectionnés pour le résumé final. Le résumé dynamique était proposé comme tâche à DUC 2007 et dans le cadre de TAC Summarization de 2008 à 2011. En 2013, TREC (*Text REtrieval Conference*) propose pour la première fois une tâche de résumé temporel (track Temporal Summarization). Ce dernier est un résumé dynamique avec une forte variabilité des documents à résumer en fonction du temps. Les données d'entrée sont des flux de documents horodatés. L'évaluation TREC 2013 a ainsi fourni, pour la tâche de résumé temporel, un corpus de documents allant d'octobre 2011 jusqu'à janvier 2013. L'objectif était de fournir à l'utilisateur des résumés liés à un événement ou à un sujet donné et mis à jour au fil du temps.

## 6 Évaluation des résumés automatiques

L'évaluation des résumés automatiques est une problématique importante à laquelle les travaux de recherche n'ont répondu que partiellement. Avec le développement du domaine et l'abondance des travaux proposés, des campagnes d'évaluation annuelles (DUC, TAC, TREC) ont été organisées afin de comparer les systèmes de résumé. Les premières évaluations

reposaient sur le jugement des lecteurs concernant la qualité linguistique et le contenu du résumé, soit en estimant la similarité des résumés candidats avec un résumé manuel (évaluation objective), soit en jugeant la qualité du résumé sans se référer à un modèle (évaluation subjective). La dernière variante correspond à la mesure *Responsiveness* utilisée jusqu'à aujourd'hui pour évaluer le résumé de point de vue du contenu et de la qualité linguistique. Ces méthodes nécessitent un fort investissement en temps et en effort, ce qui pose problème pour le développement des systèmes de résumé. C'est pourquoi des métriques standards, avec une mise en œuvre automatique, ont été proposées pour rendre plus facile la comparaison des différentes approches. Les méthodes répondant à cette problématique s'intéressent plus à l'évaluation du contenu sélectionné qu'à la qualité linguistique ou grammaticale. Par ailleurs, l'automatisation n'est que partielle. En effet, pour juger un résumé, celui-ci est comparé à un résumé manuel (idéal, modèle ou de référence). Ces systèmes dépendent donc de la disponibilité des résumés manuels. Trois métriques sont généralement utilisées pour quantifier la comparaison.

**Précision.** Elle traduit à quel point les données sélectionnées sont pertinentes. Concrètement, il s'agit du rapport du nombre d'unités textuelles communes entre le résumé candidat et les résumés de référence sur le nombre de toutes les unités textuelles du résumé candidat.

**Rappel.** Il reflète à quel degré le résumé candidat rappelle (évoque) des données pertinentes qu'il est sensé inclure. Il désigne le rapport des unités textuelles communes aux résumés candidat et de référence sur le nombre de toutes les unités textuelles du résumé de référence.

**F-mesure.** C'est la moyenne harmonique de la précision et du rappel. D'après les résultats d'évaluation des systèmes de résumé, le rappel est généralement plus difficile à obtenir que la précision.

Dans ce qui suit, nous présentons les deux méthodes d'évaluation semi-automatique les plus utilisées : ROUGE et PYRAMID. Leur succès est en particulier lié à leurs fortes corrélations avec les jugements humains. Nous donnons ensuite un aperçu sur les travaux en cours sur l'automatisation complète des systèmes d'évaluation.

## 6.1 ROUGE

ROUGE évalue les résumés en les comparant à des résumés modèles. Cette comparaison est automatique et ne nécessite pas un pré-traitement particulier. Elle est déduite à partir du recouvrement entre les N-grammes des deux textes. Cette méthode a montré une forte corrélation avec les jugements humains (Lin, 2004). La corrélation de Pearson des scores ROUGE-2 avec les jugements humains, pour le résumé multi-document, varie entre 0,85 et 0,94 en utilisant 3 résumés de référence et en éliminant les mots vides. Cette corrélation augmente avec le nombre de résumés modèles. Il existe plusieurs variantes de ROUGE exploitant des modèles autres que les N-grammes, comme la plus longue sous-séquence commune ou les bi-grammes distants. Comme l'indique ses premières lettres, ROUGE est orienté rappel (*Recall Oriented*). La dernière implémentation de ROUGE permet de calculer en plus la précision et la f-mesure. Jusqu'à présent ROUGE est l'outil d'évaluation le plus utilisé.

## 6.2 PYRAMID

Cette méthode permet de comparer un résumé candidat à un ensemble de résumés de référence (Nenkova & Passonneau, 2004). Étant donné qu'un résumé idéal n'existe pas et que les styles de rédaction diffèrent d'une personne à l'autre, l'utilisation d'un seul résumé de référence ne satisfait pas la condition d'équité entre les résumés candidats. Pour relaxer cette contrainte, les campagnes d'évaluation présentent au moins 4 résumés modèles. Le principe de la méthode PYRAMID consiste à annoter les résumés de référence afin d'identifier les unités appelées SCUs (*Summary Content Units*). Un SCU est un ensemble d'unités textuelles des résumés de référence exprimant la même information. Il lui est assigné un poids égal au nombre de résumés de référence qui l'instancient. Ces SCUs peuvent être organisés en pyramide où chaque couche regroupe les SCUs de même poids. Pour évaluer un résumé, ce dernier est annoté afin de repérer les SCUs candidats qu'il contient. Par la suite, chaque SCU candidat hérite du poids du SCU le plus similaire dans la pyramide. Le score PYRAMID du résumé est finalement le rapport de la somme des poids de tous ses SCUs candidats sur la somme des poids d'un résumé idéal ayant le même nombre de SCUs. L'inconvénient de cette méthode est qu'elle nécessite une étape d'annotation des résumés. Le calcul du score PYRAMID a été automatisé en utilisant la sémantique distributionnelle (Passonneau *et al.*, 2013). Malheureusement, l'annotation des résumés modèles reste difficile à automatiser.

## 6.3 Autres méthodes d'évaluation automatique

Les méthodes d'évaluation citées ci-dessus restent assez coûteuses en termes de temps et de ressources humaines à mobiliser. Elles ne permettent pas ainsi de mettre en œuvre des évaluations à large échelle. Ce problème a motivé les chercheurs pour proposer des méthodes d'évaluation entièrement automatiques, c'est-à-dire sans avoir besoin de résumés de référence. Une des solutions proposées consiste à utiliser un moteur de recherche pour ordonner un ensemble de documents et un ensemble de leur résumés par ordre de pertinence par rapport à une requête donnée (Radev *et al.*, 2003). Les meilleurs systèmes de résumé sont ceux qui préservent le plus l'ordre de classement entre les documents sources et leurs résumés.

Ensuite, une méthode d'évaluation automatique fondée sur des métriques de similarité entre les documents sources et les résumés générés a été proposée (Louis & Nenkova, 2008). Elle permet d'atteindre une corrélation de 0,88 avec le score PYRAMID et 0,74 avec le score Responsiveness (Louis & Nenkova, 2013). D'autres travaux ont étendu cette approche dans la perspective du multilinguisme (Saggion *et al.*, 2010). Entre 2009 et 2011, TAC a proposé la tâche AESOP (*Automatically Evaluating Summaries Of Peers*) pour encourager le développement des méthodes automatiques d'évaluation des résumés.

## 7 Synthèse

Depuis les années 2000, les travaux sur le résumé automatique ont connu un fort développement, caractérisé par une grande quantité et une grande diversité de méthodes dont nous n'avons rendu compte que partiellement dans cet article. Les premières approches du résumé multi-document exploitaient des méthodes statistiques, des classifieurs bayésiens ou des techniques de regroupement. Les réseaux de neurones artificiels ont été aussi utilisés pour pondérer les critères de sélection. Les résultats préliminaires obtenus ont motivé les chercheurs à faire évoluer ces systèmes. L'analyse des documents textuels a été approfondie et les recherches ont été poussées vers la compréhension du texte par des analyses sémantiques et thématiques. Pour ce faire, des ressources externes comme Wikipédia et WordNet ont été exploitées et d'autres ont été créées (Ferret, 2004). Toutefois, les performances atteintes restent à améliorer. Les approches sémantiques peuvent ainsi être étendues pour intégrer davantage de critères dans le calcul des similarités entre les mots. C'est le cas dans le système CHORAL (García Flores *et al.*, 2009) sur lequel nous nous appuyons au niveau de son exploitation des sens de mots. L'établissement d'un modèle abstrait soulignant les concepts mis en valeur par les documents a toujours été l'objectif de plusieurs travaux de recherche. Néanmoins, cet objectif n'a pas été réalisé complètement compte tenu de la complexité de l'analyse sémantique et de la rareté des ressources. Il a été souvent remplacé par un modèle des N-grammes les plus fréquents et de mots indices (Schlesinger *et al.*, 2008; Gillick & Favre, 2009). Une autre piste pourrait être constituée par la construction de représentations lexicales distribuées par des réseaux de neurones dans le cadre du *Deep Learning*, approche qui a montré des résultats intéressants en TAL (Yu & Deng, 2011; Luong *et al.*, 2013; Zheng *et al.*, 2013). Très peu de travaux ont exploité ce type de représentations pour le résumé automatique (Denil *et al.*, 2014) jusqu'à présent. En ce qui concerne le résumé dynamique, qui est plus récent, la performance des systèmes actuels est faible. Les travaux existants peuvent être divisés en deux familles. Certains réutilisent des systèmes de résumé multi-document pour le résumé dynamique et se contentent d'éliminer la redondance entre les anciens et les nouveaux documents après avoir classé les segments textuels. Cette solution n'apporte pas de résultats très satisfaisants car elle dépend de la précision de la mesure de similarité utilisée. D'autres travaux considèrent la problématique du résumé dynamique plutôt comme un problème d'extraction d'information. Ces systèmes ciblent la détection de nouveauté dans le document mais n'accordent pas assez d'importance aux analyses sémantique, discursive et thématique des documents. Par ailleurs, le contexte d'usage influence considérablement la formulation du problème du résumé dynamique. Alors que le problème générique est l'élimination de la redondance entre les anciens et les nouveaux documents, certaines situations exigent cette redondance. La variation des données à travers le temps peut changer la pertinence d'une information. De plus, il peut s'avérer nécessaire de reprendre certaines informations afin de rappeler ce qui est important et situer le nouveau résumé dans son cadre. L'élimination de la redondance dans les résumés dynamiques doit certainement prendre en compte ces aspects. Nos prochains travaux seront inspirés des études présentées dans cet article et fondés sur les leçons qui en sont tirées. Nous essaierons aussi d'exploiter de nouvelles pistes dans l'intention de rapprocher les performances des systèmes de résumé multi-document et dynamique vers les attentes de l'utilisateur final.

## Références

- ANGHELUTA R., BUSSE R. D. & MOENS M.-F. (2002). The use of topic segmentation for automatic summarization. In *Proceedings of the ACL-2002 Workshop on Automatic Summarization*, p. 11–12, New Brunswick, NJ.
- ARORA R. & RAVINDRAN B. (2008). Latent dirichlet allocation based multi-document summarization. In *Proceedings of the Second Workshop on Analytics for Noisy Unstructured Text Data*, AND'08, p. 91–97, New York, NY, USA.
- BÄR D., ZESCH T. & GUREVYCH I. (2015). *Composing Measures for Computing Text Similarity*. Rapport interne TUD-CS-2015-0017, TU Darmstadt, Allemagne.
- BOUDIN F. & TORRES-MORENO J.-M. (2009). Résumé automatique multi-document et indépendance de la langue : une première évaluation en français. In *Actes de la 16ème conférence sur le Traitement Automatique des Langues Naturelles*, Senlis, France.
- CHEUNG J. C. K., POON H. & VANDERWENDE L. (2013). Probabilistic Frame Induction. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics : Human Language Technologies*, p. 837–846, Atlanta, Georgia.

- CHOWDARY C. R. & KUMAR P. S. (2008). An incremental summary generation system. In *Proceedings of the 14th International Conference on Management of Data, December 17-19, 2008, IIT Bombay, Mumbai, India*, p. 83–92.
- CONROY J. M., SCHLESINGER J. D. & KUBINA J. (2011). CLASSY 2011 at TAC : Guided and Multi-lingual Summaries and Evaluation Metrics. In *Proceedings of the Fourth Text Analysis Conference (TAC 2011)*.
- DALAL V. & MALIK L. (2013). A Survey of Extractive and Abstractive Text Summarization Techniques. In *6th International Conference on Emerging Trends in Engineering and Technology (ICETET)*, p. 109–110.
- DAS D. & MARTINS A. F. T. (2007). *A Survey on Automatic Text Summarization*. Rapport interne, Literature Survey for the Language and Statistics II course at Carnegie Mellon University.
- DE CHALENDAR G. (2014). The LIMA Multilingual Analyzer Made Free : FLOSS Resources Adaptation and Correction. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC-2014), Reykjavik, Iceland*, p. 2932–2937.
- DENIL M., DEMIRAJ A., KALCHBRENNER N., BLUNSOM P. & DE FREITAS N. (2014). *Modelling, Visualising and Summarising Documents with a Single Convolutional Neural Network*. Rapport interne arXiv :1406.3830, University of Oxford.
- EDMUNDSON H. P. (1969). New Methods in Automatic Extracting. *Journal of the Association for Computing Machinery*, **16**(2), 264–285.
- FENG V. W. & HIRST G. (2014). A linear-time bottom-up discourse parser with constraints and post-editing. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, p. 511–521.
- FERRET O. (2004). Discovering word senses from a network of lexical cooccurrences. In *Proceedings of Coling 2004*, p. 1326–1332, Geneva, Switzerland.
- FERRET O., CHÂAR S. L. & FLUHR C. (2004). Filtrage pour la construction de résumés multi-documents guidée par un profil. *Traitement Automatique des Langues*, **45**(1), 65–93.
- FILIPPOVA K. (2010). *Dependency Graph-Based Sentence Fusion and Compression*. PhD thesis, TU Darmstadt, Allemagne.
- GARCÍA FLORES J., FERRET O. & DE CHALENDAR G. (2009). Summarizing through sense concentration and contextual exploration rules : the CHORAL system at TAC 2009. In *Proceedings of the Second Text Analysis Conference (TAC 2009)*, Gaithersburg, Maryland, USA.
- GENEST P.-E. & LAPALME G. (2011). Framework for Abstractive Summarization using Text-to-Text Generation. In *Proceedings of the Workshop on Monolingual Text-To-Text Generation*, p. 64–73, Portland, Oregon.
- GENEST P.-E. & LAPALME G. (2012). Fully Abstractive Approach to Guided Summarization. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Short Papers)*, p. 354–358.
- GILLICK D. & FAVRE B. (2009). A Scalable Global Model for Summarization. In *Proceedings of the Workshop on Integer Linear Programming for Natural Language Processing, ILP '09*, p. 10–18, Boulder, Colorado.
- GUPTA V. & LEHAL G. S. (2010). A Survey of Text summarization Extractive Techniques. *Emerging Technologies in Web Intelligence*, **2**(3), 258–268.
- HASSEL M. (2003). Exploitation of Named Entities in Automatic Text Summarization for Swedish. In *In Proceedings of NODALIDA 03 - 14 th Nordic Conference on Computational Linguistics*, Reykjavik, Iceland.
- HEARST M. A. (1997). TextTiling : Segmenting Text into Multi-paragraph Subtopic Passages. *Computational Linguistics*, **23**(1), 33–64.
- JAIDKA K., KHOO C. & NA J.-C. (2013). Deconstructing Human Literature Reviews – A Framework for Multi-Document Summarization. In *Proceedings of the 14th European Workshop on Natural Language Generation*, p. 125–135.
- JOTY S., CARENINI G., NG R. & MEHDAD Y. (2013). Combining Intra- and Multi-sentential Rhetorical Parsing for Document-level Discourse Analysis. In *Proceedings of ACL*.
- KAIKHAH K. (2004). Automatic Text Summarization with Neural Networks. In *Proceedings of IEEE International Conference on Intelligent Systems*, volume 1, p. 40–44.
- KHAN A. & SALIM N. (2014). A Review on Abstractive Summarization Methods. *Journal of Theoretical and Applied Information Technology*, **59**(1), 64–72.
- KINTSCH W. & VAN DIJK T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, **85**(5), 363 – 394.

- LI C., QIAN X. & LIU Y. (2013). Using supervised bigram-based ILP for extractive summarization. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, p. 1004–1013, Sofia, Bulgaria.
- LI P., WANG Y., GAO W. & JIANG J. (2011). Generating aspect-oriented multi-document summarization with event-aspect model. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, p. 1137–1146, Edinburgh, Scotland, UK.
- LIN C.-Y. (2004). Rouge : A package for automatic evaluation of summaries. In *Text Summarization Branches Out : Proceedings of the ACL-04 Workshop*, p. 74–81, Barcelona, Spain.
- LIN C.-Y. & HOVY E. (1997). Identifying Topics by Position. In *Proceedings of the Fifth Conference on Applied Natural Language Processing*, p. 283–290, Washington, DC, USA.
- LOUIS A., JOSHI A. & NENKOVA A. (2010). Discourse indicators for content selection in summarization. In *Proceedings of the SIGDIAL 2010 Conference*, p. 147–156, Tokyo, Japan.
- LOUIS A. & NENKOVA A. (2008). Automatic summary evaluation without human models. In *Proceedings of the First Text Analysis Conference (TAC 2008)*.
- LOUIS A. & NENKOVA A. (2013). Automatically assessing machine summary content without a gold standard. *Computational Linguistics*, **39**(2), 267–300.
- LUHN H. P. (1958). The automatic creation of literature abstracts. *IBM Journal of Research Development*, **2**(2), 159–165.
- LUONG M.-T., SOCHER R. & MANNING C. D. (2013). Better Word Representations with Recursive Neural Networks for Morphology. In *CoNLL*, p. 104–113, Sofia, Bulgaria.
- MADNANI N. & DORR B. J. (2010). Generating phrasal and sentential paraphrases : A survey of data-driven methods. *Computational Linguistics*, **36**(3), 341–387.
- MANI I. (2001). *Automatic Summarization*. John Benjamins Publishing.
- MANN W. C. & THOMPSON S. A. (1988). Rhetorical Structure Theory : Toward a functional theory of text organization. *Text*, **8**(3), 243–281.
- MARCU D. (1997). *The Rhetorical Parsing, Summarization, and Generation of Natural Language Texts*. Rapport interne CSRG-371, Computer Systems Research Group, University of Toronto.
- MARCU D. (1998). Improving summarization through rhetorical parsing tuning. In *Proceedings of The Sixth Workshop on Very Large Corpora*, p. 206–215, Montreal, Canada.
- MCCREADIE R., MACDONALD C. & OUNIS I. (2014). Incremental update summarization : Adaptive sentence selection based on prevalence and novelty. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management, CIKM '14*, p. 301–310, New York, NY, USA : ACM.
- MCDONALD R. (2007). A study of global inference algorithms in multi-document summarization. In *Proceedings of the 29th European Conference on IR Research, ECIR'07*, p. 557–564, Berlin, Heidelberg : Springer-Verlag.
- MCKEOWN K. R., KLAVANS J. L., HATZIVASSILOGLOU V., BARZILAY R. & ESKIN E. (1999). Towards Multidocument Summarization by Reformulation : Progress and Prospects. In *Proceedings of the Sixteenth National Conference on Artificial Intelligence and the Eleventh Innovative Applications of Artificial Intelligence Conference Innovative Applications of Artificial Intelligence, AAAI '99/IAAI '99*, p. 453–460, Menlo Park, CA, USA.
- MIHALCEA R. (2004). Graph-based ranking algorithms for sentence extraction, applied to text summarization. In *Proceedings of the ACL Interactive Poster and Demonstration Sessions*.
- NENKOVA A. & MCKEOWN K. (2012). *Mining Text Data*, chapter A Survey of Text Summarization Techniques. Springer.
- NENKOVA A. & PASSONNEAU R. J. (2004). Evaluating content selection in summarization : The pyramid method. In *HLT-NAACL*, p. 145–152.
- NETO J. L., FREITAS A. A. & KAESTNER C. A. A. (2003). Automatic text summarization using a machine learning approach. In *Advances in Artificial Intelligence*, p. 205–215 : Springer Berlin Heidelberg.
- NETO J. L., SANTOS A., KAESTNER C. & FREITAS A. (2000). Generating text summaries through the relative importance of topics. In M. MONARD & J. SICHMAN, Eds., *Advances in Artificial Intelligence*, volume 1952 of *Lecture Notes in Computer Science*, p. 300–309. Springer Berlin Heidelberg.
- NG J. P., BYSANI P., LIN Z., KAN M.-Y. & TAN C.-L. (2011). SWING : Exploiting Category-Specific Information for Guided Summarization. In *Proceedings of the Text Analysis Conference (TAC 2011)*, Gaithersburg, Maryland, USA.

- PAGE L., BRIN S., MOTWANI R. & WINOGRAD T. (1999). *The PageRank Citation Ranking : Bringing Order to the Web*. Technical Report 1999-66, Stanford InfoLab.
- PASSONNEAU R. J., CHEN E., GUO W. & PERIN D. (2013). Automated pyramid scoring of summaries using distributional semantics. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Short Papers)*, p. 143–147, Sofia, Bulgaria.
- PERRET L. (2005). *Extraction automatique d'information : génération de résumé et question-réponse*. PhD thesis, Université de Neuchâtel Faculté des Sciences.
- PLAZA L., DÍAZ A. & GERVÁS P. (2010). Automatic Summarization of News Using WordNet Concept Graphs. *IADIS International Journal on Computer Science and Information Systems*, **5**(1), 45–57.
- RADEV D. R., JING H., STYŠ M. & TAM D. (2004). Centroid-based summarization of multiple documents. *Information Processing and Management*, **40**(6), 919–938.
- RADEV D. R., TEUFEL S., SAGGION H., LAM W., BLITZER J., QI H., ÇELEBI A., LIU D. & DRABEK E. (2003). Evaluation challenges in large-scale document summarization. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, p. 375–382, Sapporo, Japan.
- SAGGION H. & LAPALME G. (2002). Generating Indicative-Informative Summaries with SumUM. *Computational Linguistics*, **28**(4), 497–526.
- SAGGION H., TORRES-MORENO J.-M., DA CUNHA I., SANJUAN E. & VELAZQUEZ MORALES P. (2010). Multilingual summarization evaluation without human models. In *23rd COLING International Conference on Computational Linguistics (Posters)*, p. 1059–1067.
- SALTON G., SINGHAL A., BUCKLEY C. & MITRA M. (1996). Automatic text decomposition using text segments and text themes. In *Proceedings of the the Seventh ACM Conference on Hypertext, HYPERTEXT '96*, p. 53–65, New York, NY, USA : ACM.
- SAMMUT C. & WEBB G. I. (2010). *Encyclopedia of Machine Learning*. Springer US.
- SCHIFFMAN B., NENKOVA A. & MCKEOWN K. (2002). Experiments in multidocument summarization. In *Proceedings of the Second International Conference on Human Language Technology Research, HLT '02*, p. 52–58, San Francisco, CA, USA : Morgan Kaufmann Publishers Inc.
- SCHLESINGER J. D., O'LEARY D. P. & CONROY J. M. (2008). Arabic/English Multi-document Summarization with CLASSY : The Past and The Future. In *Proceedings of the 9th International Conference on Computational Linguistics and Intelligent Text Processing, CICLing'08*, p. 568–581, Berlin, Heidelberg : Springer-Verlag.
- SPÄRCK JONES K. (2007). Automatic summarising : The state of the art. *Inf. Process. Manage.*, **43**(6), 1449–1481.
- TORRES-MORENO J.-M. (2014). *Automatic Text Summarization*. Wiley-ISTE.
- TRANDABĂȚ D. (2011). Using semantic roles to improve summaries. In *Proceedings of the 13th European Workshop on Natural Language Generation*, p. 164–169, Nancy, France.
- WAN X. (2012). Update summarization based on co-ranking with constraints. In *Proceedings of COLING 2012 (Posters)*, p. 1291–1300, Mumbai, India.
- WOLF F. & GIBSON E. (2005). Representing Discourse Coherence : A Corpus-Based Study. *Computational Linguistics*, **31**(2), 249–287.
- WOODSEND K. & LAPATA M. (2012). Multiple aspect summarization using integer linear programming. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, p. 233–243, Jeju Island, Korea.
- XU T., MCNAMEE P. & OARD D. W. (2013). HLTCOE Submission at TREC 2013 : Temporal Summarization. In *The Twenty-Second Text REtrieval Conference Proceedings*.
- YANG Z., YAO F., SUN H., ZHAO Y., LAI Y. & FAN K. (2013). BJUT at TREC 2013 Temporal Summarization Track. In *The Twenty-Second Text REtrieval Conference Proceedings*.
- YU D. & DENG L. (2011). Deep learning and its applications to signal and information processing. *Signal Processing Magazine, IEEE*, **28**(1), 145–154.
- ZHENG X., CHEN H. & XU T. (2013). Deep Learning for Chinese Word Segmentation and POS Tagging. In *Proceedings of the Conference on EMNLP*, p. 647–657, Seattle, USA.