

## Méthodes par extraction pour le résumé automatique de conversations parlées provenant de centres d'appels

Jérémy Trione<sup>1</sup>

(1) Aix-Marseille Université, CNRS, LIF UMR 7279, 13000, Marseille, France

Jeremy.trione@lif.univ-mrs.fr

**Résumé.** Dans ce papier nous traitons des résumés automatiques de conversations parlées spontanées. Pour cela nous utilisons des conversations provenant de cas réels d'appels téléphoniques de centre d'appels issues du corpus DECODA. Nous testons des méthodes extractives classiques utilisées en résumé de texte (MMR) ainsi que des méthodes basées sur des heuristiques du dialogue dans le cadre des centres d'appels. Il s'agit de la sélection du tour de parole le plus long dans le premier quart de la conversation, dans l'ensemble de la conversation et dans le dernier quart de la conversation. L'ensemble est évalué avec la métrique ROUGE. Les résultats obtenus soulignent les limites de ces approches « classiques » et confirment la nécessité d'envisager des méthodes abstraites intégrant des informations de structures sur les conversations. En effet, ces premiers résultats montrent que les méthodes heuristiques basées sur la structure produisent des résultats comparables, voir meilleurs que des méthodes telles que MMR.

**Abstract.** In this paper we speak about automatic spoken conversation summaries. We use conversation from some real cases call from a call center extracted from the DECODA corpus. We test some extractive summary methods used in text summary (MMR) and some dialogue heuristics methods. It's mainly to select the longest speaker turn in different part of the dialogue, the first quarter, the whole dialogue, and the last quarter of the dialogue. All the results are evaluated thanks the ROUGE software. The results show the limits of these classical approaches and suggest that we need some abstractive methods including some structural features of the conversation. In fact, these results show that the structural heuristics based methods are even or better than the classic method like MMR.

**Mots-clés :** Résumé de conversations parlées, résumé par extraction, ROUGE, corpus DECODA, MMR.

**Keywords:** spoken conversation summarization, extractive summary, ROUGE, DECODA corpus, MMR.

## 1 Introduction

Dans les centres d'appels, l'étude des traces (ou « logs ») d'interaction entre conseillers et clients permet d'évaluer le travail des conseillers téléphoniques afin d'optimiser les relations avec les usagers, ou encore faciliter la recherche d'informations sur l'ensemble des appels pour détecter d'éventuels problèmes et extraire des statistiques sur les besoins des usagers du service. Aujourd'hui seule une infime partie des données recueillies dans les centres d'appels est utilisée pour les tâches citées au-dessus (moins de 1%). Le traitement automatique de ces conversations et notamment la génération de résumés pourrait alors permettre de généraliser les traitements et ainsi améliorer les services proposés. Dans la suite de cet article nous nous intéresserons donc à la génération de résumés de conversations téléphoniques provenant de centres d'appels.

Le corpus RATP-DECODA<sup>1</sup> contient des transcriptions de conversations entre des usagers et des conseillers de la RATP. Il contient également des résumés de ces conversations, sous la forme de « synopsis » de quelques lignes, établies par des experts de ce service à des fins d'analyse. L'étude présentée dans cet article consiste à comparer ces synopsis à des résumés produits par des méthodes classiques de résumés automatiques de texte par extraction. En analysant les limites

---

<sup>1</sup> <http://sldr.org/sldr000847/fr>

de ces systèmes, nous justifions la nécessité d'enrichir les systèmes automatiques avec des informations de structure sur les conversations dans la perspective produire des résumés par abstraction.

## 2 Etat de l'art

De façon générale on peut identifier deux types de résumés, les résumés extractifs et les résumés abstraits. Le résumé extractif est aujourd'hui l'approche la plus répandue, elle consiste à pondérer les phrases selon leur représentativité (Li, et al, 2006). Le résumé abstrait consiste à extraire des informations ou des notions du document à résumer afin de rédiger un tout nouveau texte. Ces informations peuvent être par exemple des concepts d'opinions (Ganesan et al, 2010).

La plupart des travaux réalisés sont des résumés extractifs, certains traitent des journaux télévisés (Ribeiro et Martins de Matos, 2007), d'autres concernent des résumés de réunions comme les travaux de G. Murray (2008), en se basant sur le corpus des réunions de l'ICSI<sup>2</sup> (Janin. et al, 2003) et AMI<sup>3</sup> (Carletta et al, 2006). Il base ses premières expériences sur des systèmes de résumés de textes classiques (en utilisant une version modifiée de MMR), ainsi que sur des approches basées sur la structure des réunions. De la même façon des travaux ont été réalisés sur le résumé d'e-mails comme G. Murray (2008) ou encore J. Lin (2007) principalement sur le corpus d'e-mails d'Enron. Ici deux approches sont étudiées, le résumé de l'ensemble des e-mails d'une conversation, et le résumé des e-mails individuellement. J. Lin et al (2008) utilisent un modèle probabiliste basé sur la méthode de *sentence compression* (K. Knight et D. Marcu, 2000). D'autres travaux réalisés par Xiaodan Zhu et Gerald Penn (2006) portent sur le résumé de conversations spontanées, ils utilisent les données de SWITCHBOARD, qui contiennent des conversations annotées manuellement. Ils utilisent des méthodes de résumés classiques (i.e. MMR) ainsi que des heuristiques de localisation, de prosodie ou de disfluence. Enfin des recherches concernent aussi le résumé de cours d'enseignement oraux (Togashi, 2006) en utilisant aussi des transcriptions manuelles et automatiques et en prenant en compte d'autres concepts audio comme la prosodie pour la génération de résumés. Dans cette étude aussi il est question d'utiliser des techniques de résumés de textes classiques, ainsi que des informations structurelles mais aussi des informations sur l'audio (prosodie et disfluences). Nous noterons aussi que l'ensemble de ces études a été évalué grâce à la métrique ROUGE.

## 3 Corpus

Dans le cadre de notre étude notre corpus sera composé de 200 conversations téléphoniques issues d'un centre d'appels de la RATP. Ces appels proviennent du corpus du projet DECODA<sup>4</sup> (Bechet, et al. 2012). Chaque conversation est disponible en version audio et textuelle, les transcriptions utilisées ont été réalisées manuellement.

Ces conversations ont été recueillies dans un centre de la RATP sur une période d'une journée. Etant donné qu'elles ont été enregistrées à partir d'un centre d'appels de transport, elles traitent de tous sujets se rapportant de près ou de loin au transport. Cela va de la demande d'itinéraire, aux oublis d'objets sur le réseau, en passant par des plaintes de fonctionnement. Ci-dessous le tableau regroupe les dix sujets d'appel les plus courants :

Raison de l'appel	%
Info trafic	22.5
Itinéraire	17.2
Objets trouvés/perdus	15.9
Souscription aux forfaits	11.4
Horaires	4.6
Billets	4.5
Appels spécialisés	4.5
Aucun sujet particulier	3.6
Nouvel enregistrement	3.4
Information tarifaire	3.0

TABLE 1 : Top 10 des sujets d'appels sur le corpus DECODA.

2 International Computer Science Institute

3 <https://www.idiap.ch/dataset/ami/>

4 DEPouillement automatique de Conversation provenant de centre D'Appels : <http://decoda.univ-avignon.fr/>

**METHODES PAR EXTRACTION POUR LE RESUME AUTOMATIQUE DE CONVERSATIONS PARLEES PROVENANT DE CENTRES D'APPELS**

Notons que la répartition de ces sujets reste la même sur les 200 conversations choisies au préalable.

La durée de l'appel varie entre 55 secondes pour les plus courts et 16 minutes pour les plus longs. Le tableau ci-dessous regroupe plus de détails en ce qui concerne la taille du corpus en termes de mots.

	En moyenne par conversation	Sur l'ensemble du corpus
Nombre de mots	414.1	82819
Nombre de phrases	66.8	13351

TABLE 2 : Détails structurels sur le corpus de DECODA

Afin d'enrichir ce corpus pour notre étude, chaque conversation s'est vu attribuer un minimum de deux résumés réalisés par deux annotateurs différents. Le premier est un expert du domaine ayant réalisé un premier jeu de synopsis afin de se repérer dans les conversations traitées. Le second annotateur est un étudiant n'ayant aucune notion particulière dans la réalisation de résumé. Dans un premier temps aucune contrainte n'a été donnée aux annotateurs, que ce soit des contraintes de temps, de langue ou autre. Le but était d'observer dans quelle mesure les synopsis de chacun pouvaient différer et si des schémas se retrouvaient entre les individus.

La principale contrainte qui nous concernait était la taille, après l'étude des 400 premiers synopsis recueillis nous avons obtenu une taille moyenne pour les résumés de 6% à 7% de la taille de la conversation originale (cette taille est basée sur le nombre de mots de la conversation et du synopsis). De cette seule contrainte découle plusieurs autres. Par exemple le langage utilisé ne devra pas forcément suivre une syntaxe très poussée, les phrases simples et courtes seront alors privilégiées par les annotateurs, de la même façon la quantité de détails rapportés sera elle aussi limitée par la taille du document, effectivement seule l'information principale devra remonter sous peine de dépasser la limite de taille imposée précédemment.

Ci-dessous est présenté un exemple de conversation que l'on peut trouver dans le corpus RATP-DECODA.

<b><u>Usager</u></b> : allô bonjour monsieur monsieur je m'excuse de vous déranger je vous appelle de la Haute-Loire pourriez-vous m'indiquer s'il vous plaît le bus qui correspond de la Gare de Lyon à la Gare heu Montparnasse ?
<b><u>Conseiller</u></b> : alors vous avez le 91 Madame
<b><u>Usager</u></b> : c'est le 91 ?
<b><u>Conseiller</u></b> : oui
<b><u>Conseiller</u></b> : Gare de Lyon Gare Montparnasse ce sera le 91
<b><u>Usager</u></b> : d'accord Monsieur
<b><u>Conseiller</u></b> : oui
<b><u>Usager</u></b> : et c'est une ligne directe donc ?
<b><u>Conseiller</u></b> : c'est une ligne directe tout à fait
<b><u>Usager</u></b> : vous êtes très gentil monsieur je vous remercie
<b><u>Conseiller</u></b> : je vous en prie
<b><u>Usager</u></b> : bonne journée au revoir
<b><u>Conseiller</u></b> : au revoir Monsieur bonne journée
<b><u>Usager</u></b> : au revoir
<b><u>Conseiller</u></b> : merci au revoir

TABLE 3 : Exemple de conversation du corpus DECODA

## 4 Résumés de conversations

En partant de ce qui existe dans la littérature pour les textes, et en se basant sur quelques heuristiques du dialogue concernant la position de l'information au sein d'une conversation issue d'un centre d'appels, nous appliquons des méthodes classiques utilisées en résumé automatique, l'ensemble de ces méthodes constitue ce que nous appellerons *baseline*. Les conversations utilisées sont issues du corpus du projet DECODA. Ce sont des conversations spontanées, provenant de cas concrets et réels (c'est-à-dire non joués par des acteurs) entre des utilisateurs de la RATP<sup>5</sup> et les conseillers téléphoniques.

Définissons dans un premier temps ce qu'est une conversation dans cette étude ainsi que le résumé qui peut lui être associé.

### 4.1 Conversation

La définition la plus classique d'une conversation, serait : un échange d'informations entre au moins deux individus portant sur un ou plusieurs sujets précis. Dans notre cas la partie qui va nous intéresser concerne l'échange d'informations. En effet dans le cadre de la rédaction d'un résumé, la détection des informations importantes et intéressantes est primordiale.

D'un point de vu structurel, le découpage d'une conversation ne peut pas se faire grâce à des phrases, puisque la notion même de phrase est difficilement définissable au sein d'une conversation, ceci étant dû à l'absence de ponctuation concrète. Pour nous donner alors une unité de découpage, nous utiliserons le tour de parole. On appelle tour de parole le laps de temps pendant lequel un interlocuteur s'exprime. Chaque tour de parole sera alors susceptible de contenir une certaine quantité d'informations relatives à la conversation.

À la différence d'un texte classique (c'est-à-dire rédigé et réfléchi) une conversation est un échange spontané entre individus, de ce fait les informations au sein de celle-ci peuvent être altérées par de nombreux phénomènes directement liés à la spontanéité. Ce caractère spontané de la conversation introduit du bruit dans les données, il s'agit par exemple de nombreuses répétitions, des changements brusques d'idées, des erreurs de langue et autres. À cela nous pouvons aussi ajouter le fait que nous travaillons essentiellement sur des conversations téléphoniques, la compréhension devient encore plus bruitée par des problèmes liés à la qualité sonore. Afin de pallier ces problèmes, nous utilisons essentiellement les transcriptions manuelles des conversations pour la suite de ce papier.

### 4.2 Synopsis

En ce qui concerne le résumé, une définition simple serait : Forme abrégée du contenu d'un texte, d'un document, d'un film. Pour une conversation cette forme abrégée doit contenir l'ensemble des informations clés qui ont été abordées au cours de celle-ci. Nous appellerons cette forme « synopsis » pour la suite de l'étude. Chaque synopsis doit être capable de retranscrire les informations véhiculées dans la conversation en un nombre de phrases (ou mots) réduit. Il est important de préciser que la forme de nos synopsis est textuelle et abstraite, c'est-à-dire que la forme résumée de nos conversations ne sera pas une nouvelle conversation plus courte ou une sélection des tours de parole les plus pertinents, mais un court texte rappelant les idées abordées à l'instar des synopsis de films.

Le principal problème de ces synopsis est directement lié à leur nature. En effet un synopsis est le résultat produit par une personne qui souhaite résumer une conversation, mais cette même conversation pourrait très bien être résumée d'une façon totalement différente par un second individu. Afin de limiter ce caractère subjectif dans notre étude, nous nous basons pour notre évaluation sur plusieurs (au moins deux) résumés de références pour une même conversation. À cela s'ajoute la création d'un guide d'annotation en synopsis pour tout annotateur désireux d'enrichir le corpus, afin que tous les synopsis produits suivent la même orientation.

À ce caractère de subjectivité lié à l'individu, on peut ajouter des variantes dans l'orientation d'un résumé. Dans le cadre des centres d'appels on peut identifier deux catégories de synopsis : des synopsis basés sur le contenu sémantique, c'est-à-dire sur le sujet réel de la conversation, et des synopsis basés sur les interactions entre l'utilisateur et le conseiller, cela correspondrait par exemple à privilégier la durée et l'efficacité du conseiller par rapport au problème même de l'utilisateur. Dans notre étude nous nous intéressons principalement aux synopsis basés sur le contenu sémantique de la conversation.

---

<sup>5</sup> Régie Autonome des Transport Parisien (<http://www.ratp.fr>)

### 4.3 Exemples

Pour illustrer nos propos voici deux exemples concrets de synopsis rédigés par nos annotateurs.

	Annotateur 1	Annotateur 2
Conversation 1	quel bus pour gare de Lyon vers Montparnasse	Demande de renseignement sur une ligne de bus pour aller de de Gare de Lyon à Gare Montparnasse.
Conversation 2	horaires RER E de Meaux à la Gare de l'Est	Demande d'horaires de train de la gare de Maux à la gare de l'Est à une heure donnée

TABLE 4 : Exemples de synopsis

Comme on peut facilement le voir, les synopsis de l'annotateur 2 sont syntaxiquement mieux construits que ceux de l'annotateur 1, mais en termes d'informations, les deux synopsis sont très similaires.

## 5 Méthodes de résumés

Dans cette partie nous présentons les méthodes classiques de résumés extractifs et heuristiques de sélection selon la position que nous utilisons dans notre étude.

### 5.1 Maximal Marginal Relevance (MMR)

MMR (Corbonnell et Goldstein, 1998) est largement utilisé dans le résumé de conversation du fait de sa simplicité et de son efficacité. Il sélectionne les tours de paroles les plus riches de sens d'un texte tout en évitant la redondance des informations. Dans le cadre du résumé extractif le score attribué à un tour de parole  $S$  est calculé comme suit :

$$MMR(S_i) = \lambda \times Sim_1(S_i, D) - (1 - \lambda) \times Sim_2(S_i, Summ)$$

Où  $D$  représente le vecteur document,  $Summ$  représente les tours de paroles qui ont été extraits pour le résumé, et  $\lambda$  est une constante utilisée pour ajuster la relation entre la pertinence et la redondance. Les deux fonctions de similarité ( $Sim_1$  et  $Sim_2$ ) représentent respectivement la similarité entre un tour de parole par rapport à l'ensemble du document et par rapport au résumé courant. Les tours de paroles avec le plus haut score MMR sont sélectionnés itérativement pour générer le résumé, jusqu'à ce que la limite de taille soit atteinte. Ici on sélectionne des tours de paroles jusqu'à ce que le nombre de mots excède 6% de la totalité des mots qui composent le texte de base.

### 5.2 Heuristiques basées sur le dialogue.

Pour établir notre *baseline* nous avons utilisé quelques heuristiques basées sur les conversations. Nos conversations sont essentiellement les appels provenant de centres d'appels. De ce fait on peut facilement penser qu'il existe une certaine structure qui se répète dans ces appels. On notera que pour la sélection des tours de parole selon leur position, tous les tours de parole sont pris en compte, autant les tours de parole de l'utilisateur que ceux du conseiller.

Le premier constat et le plus évident est que l'utilisateur appelle le centre pour obtenir des informations sur un sujet bien précis. Il est donc normal de penser que l'information essentielle de l'appel se trouve en début de conversation dans les tous premiers tours de parole, juste après les échanges de politesse conventionnels (« bonjour »). D'autre part l'opérateur se doit d'écouter l'appelant pour savoir quelle est sa requête. À partir de ces deux constats nous établissons notre première heuristique : Le résumé (que l'on notera LB pour la suite de ce papier) sera constitué de l'unique tour de parole dont la taille en mot est maximale parmi tous les tours de parole du premier quart de la conversation.

Dans la même optique que précédemment nous prenons cette fois comme résumé (que l'on notera LA pour la suite) l'unique tour de parole dont la taille est maximale sur l'ensemble de la conversation. Celui-ci peut symboliser soit une explication détaillée de la requête par l'utilisateur, soit une explication de la réponse à cette requête par l'opérateur. Dans les deux cas nous espérons aussi révéler une prise d'informations importante dans la conversation.

Afin d'observer si la fin de la conversation contient ou non des informations utiles, nous utiliserons comme résumé (que l'on notera LE) l'unique tour de parole dont la taille est maximale parmi les tours de parole contenus dans le dernier quart

## METHODES PAR EXTRACTION POUR LE RESUME AUTOMATIQUE DE CONVERSATIONS PARLEES PROVENANT DE CENTRES D'APPELS

de la conversation. La fin d'une conversation pourrait très bien représenter un rappel global de ce qu'il s'est dit tout au long de l'appel, mais pourrait aussi contenir la solution apportée ou non par l'opérateur à l'utilisateur.

Enfin nous testons un système complètement aléatoire pour se donner une idée de son efficacité dans cette situation. Pour cela nous relevons des tours de parole aléatoirement tant que la contrainte de taille n'a pas été atteinte. Nous noterons ce résumé RS.

## 6 Résultats et interprétations

Dans cette partie nous présentons d'une part la méthode d'évaluation utilisée, et d'autre part les résultats obtenus à partir de cette dernière.

### 6.1 Evaluation

L'évaluation de résumés par des humains est une tâche longue et coûteuse. De ce fait pour évaluer nos systèmes nous nous sommes tournés vers une évaluation automatique. Pour cela nous utiliserons l'évaluation ROUGE. Celle-ci est basée sur les occurrences de mots entre les résumés produits automatiquement et les résumés humains dits « idéaux ». ROUGE compare alors le texte produit par les systèmes avec un ensemble de résumés humains sur le même document original. ROUGE-1 à ROUGE-4 sont de simples mesures d'occurrences de n-grammes, qui détectent les mêmes segments entre les résumés produits et ceux de références. ROUGE-L mesure les séquences communes entre les deux types de résumés.

$$ROUGE - N = \frac{\sum_{S \in \{ReferenceSummaries\}} \sum_{gram_n \in S} Count_{match}(gram_n)}{\sum_{S \in \{ReferenceSummaries\}} \sum_{gram_n \in S} Count(gram_n)}$$

Où  $n$  correspond au nombre de mots contenus dans la séquence de mots à rechercher lors de l'évaluation,  $Count_{match}(gram_n)$  représente le nombre maximum de mots qui interviennent à la fois dans le résumé à évaluer et dans les échantillons de résumés de référence.

Lin (Lin, 2003) a montré que ROUGE-1 et ROUGE-2 constituaient un bon indicateur du jugement humain.

### 6.2 Résultat et interprétation

Nous avons donc lancé nos évaluations avec les résumés générés à partir d'heuristiques basées sur les conversations : le plus long tour de parole en début de conversation (LB), le plus long tour de parole sur toute la conversation (LA) et le plus long tour de parole en fin de conversation (LE), ainsi que la méthode d'extraction de tours de parole aléatoires (RS). Nous utilisons aussi à titre de comparaison un résumé constitué de l'ensemble de la conversation (FT). Les résultats sont regroupés dans le tableau ci-dessous :

	ROUGE-1	ROUGE-2	ROUGE-L
LB	0.183	0.613	0.145
LA	0.175	0.051	0.132
MMR	0.150	0.051	0.119
FT	0.092	0.035	0.070
LE	0.055	0.009	0.043
RS	0.049	0.011	0.041

TABLE 5 : Résultats de l'évaluation

On notera tout d'abord que l'évaluation des synopsis extractifs générés se fait avec des synopsis abstraits. Cela aurait pour effet de réduire le score de chaque méthode, mais étant donné que toutes les méthodes évaluées ici sont basées sur de l'extraction de tours de parole, la variation des scores sera la même pour toutes. Cependant ces résultats pourraient être difficilement comparables avec d'autres résultats impliquant l'utilisation de ces méthodes dans un autre cadre que celui décrit dans ce papier.

Cela mis de côté on s'aperçoit que les meilleurs résultats sont obtenus avec l'extraction du tour de parole le plus long en début de conversation (LB), cela met bien valeur le fait que dans les conversations du corpus de RATP-DECODA

## METHODES PAR EXTRACTION POUR LE RESUME AUTOMATIQUE DE CONVERSATIONS PARLEES PROVENANT DE CENTRES D'APPELS

l'information essentielle est souvent contenue dans les tout premiers échanges entre l'utilisateur et l'opérateur. De la même façon il n'y a que très peu d'informations en fin de conversation comme nous le montre le résultat de l'évaluation LE.

L'extraction du plus long tour de parole en début de conversation et sur toute la conversation donne des résultats assez proches. Cela peut s'expliquer par la nature la conversation. En effet lors d'un appel, le consommateur va souvent expliquer son problème en début d'appel sans que l'opérateur ne l'interrompe. En revanche au cours de la conversation, c'est-à-dire pendant la résolution dudit problème, il y aura souvent de nombreux échanges rapides entre les deux interlocuteurs.

En ce qui concerne les résultats donnés par l'algorithme de MMR, ce-dernier peut facilement être biaisé avec ce genre de données. La nature spontanée et naturelle de la conversation introduit de nombreux bruits, dont la répétition fréquente d'un même terme ou d'une même idée. Au niveau de l'appel cela peut se retranscrire par une période d'incompréhension entre les deux parties, menant alors à de nombreuses répétitions d'informations peu importantes. MMR va alors se concentrer sur ces répétitions et ne renvoyer que des tours de parole de faible importance au niveau information. Cependant les résultats restent tout de même relativement proches des heuristiques basées sur le dialogue les plus efficaces.

### 6.3 Exemples de résumés produits

Ci-dessous est présenté un exemple de résumé produit en utilisant le MMR et les heuristiques simples de la conversation.

	ROUGE-1	Synopsis
Humain 1	/	perdu téléphone mais pas retrouvé
Humain 2	/	Demande de renseignements sur la perte d'un téléphone portable Alcatel Blanc sur la ligne de bus 20. Non retrouvé.
LB	0.205	allô oui bonjour monsieur je téléphonais pour savoir si par hasard vous auriez trouvé un téléphone portable euh hier dans l'autobus euh la ligne 20 en fin en fin d'après midi disons en soirée plutôt
LA	0.205	allô oui bonjour monsieur je téléphonais pour savoir si par hasard vous auriez trouvé un téléphone portable euh hier dans l'autobus euh la ligne 20 en fin en fin d'après midi disons en soirée plutôt
MMR	0.178	allô oui bonjour monsieur je téléphonais pour savoir si par hasard vous auriez trouvé un téléphone portable euh hier dans l'autobus euh la ligne 20 en fin en fin d'après midi disons en soirée plutôt. ligne numéro 20, un téléphone portable.
LE	0	bonne journée au revoir

TABLE 6 : Différents synopsis obtenus sur un exemple concret.

Comme on peut le voir, le MMR et LB sont les deux méthodes qui semblent le plus efficace en terme de récupération d'informations, avec un point supplémentaire pour le MMR, mais ceci est dû au fait qu'il récupère deux tours de parole pour son synopsis alors que LB, LA et LE n'en récupère qu'une seule. De la même façon LE n'est clairement pas significatif.

## 7 Conclusion et perspectives

Dans notre étude nous n'avons testé et évalué que des systèmes basés sur des méthodes extractives classiques des résumés de textes afin de se donner une idée de l'efficacité de ces derniers. Cependant nos résumés de références sont abstraits, ce qui fausse nos résultats dans la mesure où les résultats obtenus ne sont comparables qu'entre eux. Notre système d'évaluation ROUGE atteint aussi ses limites en ne capturant pas la variabilité lexicale, de ce fait il faudra penser à faire une évaluation manuelle.

Cependant on s'aperçoit tout de même d'après les résultats obtenus que les méthodes les plus efficaces sont directement liées à la structure de l'appel. La méthode de résumé par MMR qui fonctionne bien dans les cas des résumés de texte, est ici battue par les méthodes basées sur les heuristiques du dialogue (LB et LA). De ce fait pour la suite de nos études nous nous orienterons plus dans la direction de la structure des appels reçus dans les centres téléphoniques, en essayant de trouver par exemple une segmentation de la conversation en unités logiques.

## Remerciements

Ces travaux de recherche ont été financés en partie par l'Union Européenne à travers le projet SENSEI<sup>6</sup> (FP7/2007-2013 - n° 610916 – SENSEI).

## Références

JANIN, A., BARON, D., EDWARDS, J., ELLIS, D., GELBART, D., MORGAN, N., ... & WOOTERS, C. (2003). The ICSI meeting corpus. *Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on*. Vol. 1. I-364. IEEE.

GODFREY, J. J., HOLLIMAN, E. C., & MCDANIEL, J. (1992). SWITCHBOARD: Telephone speech corpus for research and development. *Acoustics, Speech, and Signal Processing, 1992. ICASSP-92., 1992 IEEE International Conference on* (Vol. 1, pp. 517-520). IEEE.

CARLETTA, J., ASHBY, S., BOURBAN, S., FLYNN, M., GUILLEMOT, M., HAIN, T., ... & WELLNER, P. (2006). The AMI meeting corpus: A pre-announcement. *Machine learning for multimodal interaction*. 28-39. Springer Berlin Heidelberg.

RIBEIRO R., DE MATOS DM. (2007). Extractive Summarization of Broadcast News: Comparing Strategies for European Portuguese. *Text, Speech and Dialogue*, 115-122.

MURRAY G., CARENINI G. (2008). Summarizing Spoken and Written Conversations. *EMNLP*, 773-782.

LIN J., ZAJIC D. M., DORR B. J. (2007). Single-document and multi-document summarization techniques for email threads using sentence compression, *IPM 44*, 1600-1610.

KNIGHT K., DANIEL M. (2000) Statistics-based summarization-step one: Sentence compression, *AAAI/IAAI*, 703-710.

ZHU X., PENN G (2006). Summarization of Spontaneous Conversations, *INTERSPEECH*, 1531-1534.

TOGASHI S., YAMAGUSHI M, NAKAGAWA S. (2006). Summarization of spoken lectures based on linguistic surface and prosodic information, *STL*, 34-37.

BECHET F., MAZA B., BIGOUROUX N., BAZILLON T., EL-BEZE M., DE MORI R., ARBILLOT E. (2012). DECODA: a call-center human-human spoken conversation corpus. *LREC*.

GANESAN K., ZHAI CX, HAN J. (2010). . Opinosis: A Graph-Based Approach to Abstractive Summarization of Highly Redundant Opinions. *International Conference on Computational Linguistics 23*, 340-348.

LI W., WU M., LU Q, XU W., YUAN C (2006). Extractive Summarization using Inter- and Intra- Event Relevance. *ACL 44*, 369-376.

LIN C.-Y., HOVY E. H. (2003). Automatic evaluation of summaries using n-gram co-occurrence statistics, *HLT-NAACL*.

CARBONNELL J., GOLDSTEIN J. (1998), The use of MMR, the diversity-based reranking for reordering documents and producing summaries, *SIGIR*.

---

<sup>6</sup> <http://www.sensei-conversation.eu/>