

Détection de concepts pertinents pour le résumé automatique de conversations par recombinaison de patrons

Jérémy Trione Benoit Favre Frédéric Béchet

Aix-Marseille Université, CNRS, LIF UMR 7279, 13000, Marseille, France

prenom.nom@lif.univ-mrs.fr

RÉSUMÉ

Ce papier décrit une approche pour créer des résumés de conversations parlées par remplissage de patrons. Les patrons sont générés automatiquement à partir de fragments généralisés depuis un corpus de résumés d'apprentissage. Les informations nécessaires pour remplir les patrons sont détectées dans les transcriptions des conversations et utilisées pour sélectionner les fragments candidats. L'approche obtient un score ROUGE-2 de 0.116 sur le corpus RATP-DECODA. Les résultats obtenus montrent que cette approche abstractive est plus performante que les approches extractives utilisées habituellement dans le domaine du résumé automatique.

ABSTRACT

Relevant concepts detection for the automatic summary of conversations using patterns recombination

This paper describes a template filling approach for creating conversation summaries. The templates are generated from generalized summary fragments from a training corpus. Necessary pieces of information for filling them are extracted automatically from the conversation transcripts given linguistic features, and drive the fragment selection process. The approach obtains ROUGE-2 scores of 0.116 on the RATP-DECODA corpus, which represent a significant improvement over extractive baselines.

MOTS-CLÉS : Résumé, synopsis, patron, abstratif, ROUGE, icsiboost, variables de patron.

KEYWORDS: Summary, synopsis, template, abstractive, ROUGE, icsiboost, slots.

1 Introduction

Le résumé automatique de document repose généralement sur des méthodes par extraction qui sélectionnent dans le texte des passages pertinents et les juxtaposent pour former un résumé. Ces méthodes sont peu adaptées à la problématique du résumé de conversations orales de part la nature spontanée de celles-ci et l'importance de l'interaction entre les locuteurs. En ne sélectionnant que certains passages, les résumés par extraction ne contiennent qu'un verbatim de ce qui a été dit, et non pas une description synthétique de ce qui s'est passé lors de la conversation.

Par exemple, dans le domaine des centres d'appel, il serait souhaitable que les résumés générés renseignent à propos du problème de l'appelant et de comment ce problème a été pris en charge par l'agent ayant traité l'appel. Il n'est pas rare que l'appelant décrive son problème sur plusieurs tours de parole ponctués par des demandes de confirmation ou de reformulation de la part de l'agent, ce qui

est difficile à caractériser à l'aide de méthodes extractives lorsque la taille des résumés est fortement contrainte.

Les méthodes de résumé par remplissage de patron ont montré leur intérêt dans des domaines spécifiques pour le résumé automatique de texte (White *et al.*, 2001). Dans notre cas, elles permettent de traiter du problème de différence de genre entre les données source (transcriptions de conversations) et la forme des résumés à générer (narration synthétique). Toutefois, elles nécessitent l'écriture manuelle de patrons de résumés et l'annotation manuelle de quantités de données source en concepts à détecter pour remplir ces patrons.

Nous proposons d'explorer des méthodes pour le résumé par remplissage de patrons de conversations qui nécessitent moins de supervision de la part d'un expert humain. Nos contributions sont les suivantes :

- l'extraction directe de concepts pertinents à partir des transcriptions pour remplir les patrons, par opposition à une analyse sémantique complète ;
- le transfert des annotations de ces concepts depuis des résumés manuels aux conversations par alignement sémantique, minimisant ainsi le coût d'annotation ;
- la génération dynamique de patrons à partir d'exemples de résumés de référence et des informations détectés dans une conversation ;
- un ensemble d'expériences validant l'approche sur la tâche de génération de synopsis du corpus DECODA.

Le reste de l'article est organisé de la manière suivante : la section 2 positionne le travail par rapport à l'état de l'art, la section 3 décrit notre méthode d'extraction d'information et de génération de résumés, la section 7 décrit le cadre expérimental et les résultats de notre étude, et enfin la section 8 discute ces résultats et dresse une conclusion.

2 Positionnement

Un état de l'art sur le résumé automatique peut être trouvé dans (Nenkova *et al.*, 2011). Malgré de très bons résultats dans le cadre du résumé de textes journalistiques de par la nature de ce média, les méthodes par extraction trouvent leurs limites dans les autres domaines où il existe une différence de genre entre les résumés et les données source (Mehdad *et al.*, 2014). En effet, ces méthodes qui sélectionnent des phrases des documents et les juxtaposent sont contraintes de réutiliser le style de la source, direct et informel pour les transcriptions de centres d'appels, pour générer celui des résumés, indirect et narratif dans notre cas.

Il y a déjà eu des tentatives de résumé automatique de conversations dans des centres d'appel. Par exemple, dans (Byrd *et al.*, 2008), des textes à destination des journaux d'appel sont générés par remplissage de patrons définis manuellement à l'aide de modèles d'extraction, et par extraction de contenu non structuré. (Higashinaka *et al.*, 2010) modélisent les tours de parole à l'aide de HMM sur les thèmes, et génère la sélection à l'aide de l'algorithme de Viterbi. Ces méthodes requièrent soit une forte ingénierie du domaine pour écrire les patrons, soit font de l'extraction.

(Mehdad *et al.*, 2014) proposent une approche de résumé par abstraction en utilisant le remplissage de patron dont la spécificité est de générer automatiquement ces derniers. Les phrases des documents sont d'abord regroupées en communautés, puis fusionnées sous la forme d'un graphe de mots dans

lesquels des mots sont considérés comme équivalents s'ils ont le même hyperonyme. Chaque phrase du résumé est générée comme un chemin dans ce graphe. L'approche semble prometteuse mais les performances rapportées sur un corpus d'enregistrements de réunions sont moins bonnes que celles d'une méthode par extraction tirant parti d'une classification thématique des phrases candidates (Garg *et al.*, 2009).

Nous poursuivons dans nos travaux cette idée de générer automatiquement des patrons. Cependant, contrairement à (Mehdad *et al.*, 2014) nous allons générer ces patrons à partir d'un corpus de résumés plutôt qu'à partir d'un corpus de documents sources.

3 Méthode générale

Exemples de résumés de conversations
- <i>Écharpe</i> oubliée dans <i>bus 140</i> , mais rappeler à <i>onze heures</i> quand le service sera ouvert
- La cliente a oublié <i>son écharpe</i> dans le <i>bus 140</i> à <i>Colombes le soir précédent</i> . L'agent répond que le service n'est pas ouvert avant <i>11h</i> .
- Demande de renseignement sur la perte d'une <i>écharpe</i> dans le <i>bus 140</i> . Attendre l'ouverture du service et rappeler plus tard.
Patron possible
Le client à oublié \$OBJET dans le \$TRANSPORT. Objet non retrouvé, rappeler plus tard.

TABLE 1 – Exemple de patron pour le thème objet perdu. Les variables de patron sont \$OBJET et \$TRANSPORT.

Notre approche pour le résumé de conversations est basé sur le remplissage de patrons de texte avec des parties variables remplies lors de l'analyse des différents tours de parole des conversations. L'originalité de notre approche repose sur le mode de production de ces patrons : alors qu'ils sont habituellement créés manuellement une fois pour toute lors de la phase de mise au point du système, nous proposons de les générer automatiquement de manière dynamique, durant le traitement d'une conversation, à partir d'un corpus d'entraînement constitué de paires de document *transcription de conversations / résumés*.

Cette étape de production de patrons utilisent deux types de séquences de mots extraits du corpus de résumé : des *variables de patron* représentant les concepts spécifiques à la conversation traitée, et des séquences de mots génériques qui peuvent s'appliquer à différentes situations. La figure 1 donne un exemple de patron pour des conversations issues d'un corpus enregistré dans un centre d'appel.

Notre approche consiste en plusieurs étapes :

- tout d'abord la production d'une cohorte de patrons de phrases obtenus à partir de l'annotation d'un corpus de paires de conversations et résumés ;
- puis l'alignement entre les parties variables des patrons et les concepts correspondants dans les conversations traitées afin d'entraîner un classifieur permettant de détecter ces concepts dans une nouvelles conversation ;
- enfin la génération dynamique d'un patron de résumé et son remplissage suite à l'étape

d'analyse permettant d'associer les concepts dans les conversations et les variables dans les patrons de phrases.

Ces étapes sont décrites dans la figure 1 et détaillées dans les sections suivantes.

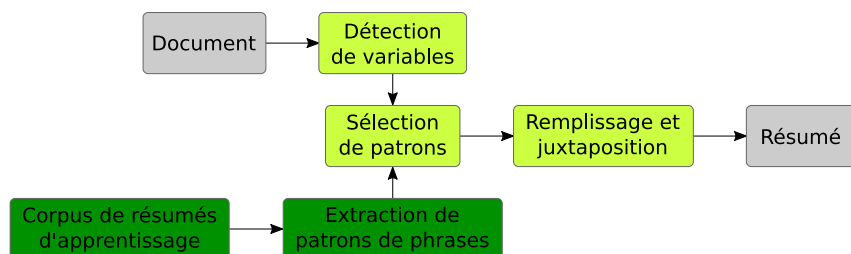


FIGURE 1 – Schéma de l'approche pour le résumé par génération de patrons. Une cohorte de patrons de phrases est générée depuis un corpus d'apprentissage, puis des patrons sont sélectionnés en fonction des variables qui ont été détectées dans le document traité. Le résumé est généré par remplissage de ces patrons et leur juxtaposition.

4 Cadre expérimental

Nous utilisons dans cette étude le corpus RATP-DECODA contenant des transcriptions de conversations enregistrées dans le centre d'appel de la RATP à Paris sur deux jours pendant une période de grève en 2009. Les sujets couverts par les conversations incluent les informations sur le trafic (l'état de fonctionnement des lignes de transport pendant la grève), la recherche d'itinéraire, la perte d'objets sur le réseau RATP, la demande d'informations sur les tarifs, etc.

Étant donné une conversation il est souvent assez facile pour un humain de trouver les informations importantes et de créer un patron pour rassembler ces informations et ainsi créer un résumé de l'appel. On appelle alors synopsis un tel résumé.

Un synopsis correspond à une succession d'évènements entre l'appelant (le client) et l'agent (ou plusieurs agents), il doit contenir une description des besoins ou problèmes du client et comment ces problèmes ont été résolus. Le tableau 2 montre quelques exemples de synopsis issus du corpus DECODA (Bechet *et al.*, 2012).

Si on s'attarde un peu sur les synopsis comme ceux présentés en exemple, il apparaît naturel qu'une approche abstractive est nécessaire. Mais on peut aussi s'apercevoir que ces synopsis suivent une certaine structure (Type de problème, description et résolution). Comme dit plus haut, l'humain sélectionne les informations importantes et les insère dans un patron, il s'agit là du cœur de notre approche : détecter ces informations dans une conversation et créer des patrons de synopsis capables d'accueillir ces informations.

Il est évident qu'il est impossible de créer un patron pour chaque cas rencontré, nous considérons qu'il est possible de trouver une version de patron capable de généraliser le plus grand nombre de synopsis (ce qui correspond à la dernière étape de notre création de patron). Afin de rendre les patrons à la fois génériques et modifiables, nous avons décidé de les définir comme des expressions régulières. Ces expressions régulières comprennent des regroupements, des quantifieurs (point d'interrogation

Conversation	Ann.	Résumé
20091112-0042	1	Est-ce que les bus 172 et 186 circulent ? Non, trafic interrompu à cause de la grève du dépôt de Vitry.
20091112-0042	2	Demande d'information sur la circulation des bus 172 et 186. Grosse perturbation sur ces lignes à cause d'une grève. Plainte de l'appelant.
20091112-0604	1	Itinéraire banlieue, ayant essayé par le site ratp mais pas convaincue du trajet proposé. Récapitulatif du trajet par l'appelant.
20091112-0604	2	Demande d'itinéraire pour aller à la gare de Chilly-Mazarin en partant de la gare de Fontenay-sous-bois. Prendre le RER A en direction de Saint Germain en Laye jusqu'à la gare de Lyon, puis prendre le métro 14 direction Olympiade, descendre à bibliothèque, enfin prendre le RER C jusqu'à Chilly Mazarin. Communication de la durée du voyage et de la fréquence de passage.

TABLE 2 – Exemple de résumés issus du corpus DECODA. Pour chaque conversation, les résumés de deux annotateurs (Ann.) sont donnés.

pour zéro ou une fois, étoile pour l'étoile de Kleene), des alternatives, et des variables (dollars suivi du nom en majuscule). Chaque patron peut alors générer plusieurs synopsis différents en fonction des informations disponibles dans la conversation, ce qui permet d'augmenter la couverture de conversation couverte par un même patron. Le tableau 3 donne des exemples de patrons sous forme d'expressions régulières.

Thème	Patron
Planification	Demande d'horaire (en \$TRANSPORT) ? de \$FROM à \$TO.
Itinéraire	Demande d'itinéraire (en \$TRANSPORT) ? de \$FROM à \$TO (sans prendre \$NOT_TRANSPORT) ?. (Prendre la \$LINE (en direction de \$TOWARDS) ? de \$START_STOP à \$END_STOP)*.
Carte Navigo	Demande de (justificatif remboursement reçu d'informations) pour \$CARD_TYPE. Le client doit se rendre à \$ADRESSE.
Objets perdus	\$ITEM perdu dans \$TRANSPORT (à \$LOCATION) ? (vers \$TIME) ?. (Objet retrouvé, et à récupéré à \$RETRIEVE_LOCATION Objet non retrouvé).
Tarif	Demande de tarif pour se rendre de \$FROM à \$TO. Le coût est de \$BUY.
Trafic	Demande d'information sur l'état de \$TRANSPORT. (La fréquence de passage est de \$FREQUENCY Trafic interrompu à cause de \$ISSUE Impossible de donner d'information à cause de \$ISSUE)

TABLE 3 – Exemples de patrons créés manuellement en utilisant un formalisme de langage régulier.

Nous avons annoté 175 synopsis en variables de patron, en utilisant 17 types de variables différents (cf tableau 4).

Les variables de patron jouent un rôle capital dans les patrons, ce sont elles qui portent l'information de la conversation. Le tableau 4 montre les variables utilisées et leur couverture sur les synopsis annotés.

Comme on pouvait s'y attendre certaines variables sont très répandues comme \$TRANSPORT qui apparaît dans presque la moitié des synopsis. On notera que \$FROM, \$TO, \$LOCATION,

Variable	% Synopsis	Variable	% Synopsis
\$TRANSPORT	42.29	\$TO	27.43
\$FROM	25.14	\$CARD_TYPE	25.14
\$INFO_TARGET	22.29	\$ITEM	20.0
\$ISSUE	18.29	\$LINE	8.0
\$LOCATION	5.14	\$BUY	4.57
\$TOWARDS	2.86	\$END_STOP	2.86
\$TIME	2.29	\$NOT_TRANSPORT	2.29
\$START_STOP	0.57	\$FREQUENCY	0.57
\$RETRIEVE_LOCATION	1.14		

TABLE 4 – Répartition des variables de patrons dans les synopsis.

\$TOWARDS, \$RETRIEVE_LOCATION, \$START_STOP et \$END_STOP sont toutes des variables de type entité nommée, ce qui complique la tâche de différenciation lors de l’annotation.

Une fois définis les patrons à partir des synopsis, il est nécessaire de les aligner avec les conversations qu’ils représentent afin de faire la correspondance entre les concepts exprimés dans les transcriptions de conversation et les variables des patrons. Cette étape est décrite dans la section suivante.

5 Détection des variables de patron

Cette étape consiste à trouver un candidat dans une conversation capable de prendre la place d’une variable de patron. La première hypothèse que l’on pourrait faire serait que les variables sont des entités nommées, car dans de nombreux domaines d’application, elles sont très largement présentes et jouent un rôle pour lier un résumé à la réalité qu’il décrit. Mais une détection simple n’est pas suffisante car, par exemple dans le cas d’un patron relatant un trajet, il faut pouvoir déterminer parmi deux entités nommées de lieu, laquelle est l’adresse de départ et laquelle est celle d’arrivée. Les variables à détecter doivent donc être *pertinentes* pour un patron donné.

D’autre part, certaines informations nécessaires pour remplir les patrons ne sont pas des entités nommées, mais, par exemple, des objets génériques, des actions ou des situations. C’est le cas par exemple de scénarios de perte d’objets, dans lequel le type d’objet perdu, et le fait que l’objet a été retrouvé ne sont pas des entités nommées. Pour couvrir ces types de variables, une solution serait d’avoir recours à d’autres type d’annotations comme par exemple l’annotation en cadre sémantique (Baker *et al.*, 1998), ou encore d’avoir recours à un système d’apprentissage automatique capable de retrouver directement les variables dans la conversation sans passer par une représentation intermédiaire. C’est cette dernière méthode que nous avons choisi d’explorer.

Notre approche consiste à aligner les variables de patron des synopsis avec les transcriptions de conversations afin d’entraîner un classifieur permettant de prédire ces variables dans de nouvelles conversations.

Contrairement à une tâche de recherche de concept dans une conversation, nous ne cherchons pas ici à classifier l’ensemble des variables potentielles de la conversation. Nous cherchons à identifier seulement les variables qui seront présentes dans le synopsis final. Il s’agit d’une double classification qui répond à deux problèmes : s’agit-il d’une variable ? et cette variable sera-t-elle présente dans le

synopsis ?

5.1 Annotation par propagation depuis les résumés

Le principe de la phase de propagation est de retrouver dans les transcriptions d'une conversation les séquences de mots qui correspondent à des variables annotées dans les résumés. Pour effectuer ces alignements, nous effectuons un alignement du contexte de chaque variable avec les séquences de mots des documents sources, annotées syntaxiquement pour obtenir un contexte de correspondance plus riche. Les annotations syntaxiques sont générées par la chaîne de traitement de Macaon adapté au traitement de l'oral (Bazillon *et al.*, 2012).

L'alignement synopsis-conversation suit l'algorithme suivant :

- Les transcriptions de conversations sont tout d'abord analysées syntaxiquement avec Macaon et tous les groupes nominaux sont extraits
- Chaque variable d'un résumé annoté est comparée à l'ensemble des groupes nominaux de la transcription de la conversation correspondante grâce à une fonction de coût utilisant une distance de Levenstein.
- Le groupe nominal de plus faible coût est associé à la variable correspondante.

Pour ce qui est de l'approximation de la correspondance entre les mots, nous avons choisi de normaliser les accents et les majuscules, ce qui permet d'appliquer une distance de Levenstein basée sur les caractères. Cette méthode nous permet d'aligner 316 variables sur les 380 annotées dans les synopsis soit un taux de 83.16% d'alignement.

Les variables de patron non alignées sont dans la majorité des cas dues à des erreurs d'annotation, à des variables trop générales qui ne peuvent être détectées au niveau des mots ou encore à une absence de correspondance entre les variables et la transcription (c'est le cas lorsque l'annotateur a trop généralisé les événements lors de la rédaction du synopsis).

Une fois les conversations annotées en variables, nous voulons désormais être capable d'étendre cette annotation sur de nouvelles conversations pour un coût raisonnable.

5.2 Prédiction dans de nouvelles conversations

L'étape précédente permet de créer un corpus associant variables de patrons et expressions de concepts (sous forme de groupe nominaux) dans les transcriptions de conversations. Le corpus peut alors être utilisé pour prédire directement le type de variable de chaque groupe nominal. Nous utilisons pour cela une nouvelle fois la structure syntaxique des conversations donnée par la chaîne de traitement Macaon (Bazillon *et al.*, 2012) afin d'obtenir un certain nombre de traits lexicaux et syntaxiques sur lesquels repose un système d'apprentissage.

Nous utilisons un classifieur icsiboost (Favre *et al.*, 2007) avec comme données d'entraînement les conversations précédemment annotées en variables de patron. Pour chaque groupe nominal de la conversation, le classifieur est entraîné à prédire un type de variable parmi les 17 disponibles présentés dans la table 4 plus 1 type NULL pour les groupes nominaux qui ne sont pas des variables. Le classifieur repose sur les traits suivants :

1. Tête syntaxique du groupe nominal :

- 1.1. forme fléchie ;
- 1.2. partie de discours ;
- 1.3. lemme ;
- 1.4. type d'entité nommée.
2. Gouverneur syntaxique de la tête du groupe :
 - 2.1. lemme ;
 - 2.2. partie de discours ;
 - 2.3. étiquette de dépendance.
3. Groupe nominal :
 - 3.1. sac de n-grammes de mot ($n \leq 3$) ;
 - 3.2. sac de n-grammes de parties de discours ($n \leq 3$) ;
 - 3.3. longueur en mots.
4. Conversation et discours :
 - 4.1. nombre d'occurrences du type d'entité nommée depuis le début de la conversation ;
 - 4.2. nombre d'occurrences du lemme nommée depuis le début de la conversation ;
 - 4.3. thème de la conversation ;
 - 4.4. position relative du groupe nominal dans la conversation ;
 - 4.5. rôle du locuteur (appelant ou agent).

Il est bon de noter que les traits liés à la conversation et au discours ne sont pas des traits conventionnels pour l'étiquetage en entités nommées ou en concepts. Ils permettent par exemple de discriminer les lieux de départ et d'arrivée. Les détails expérimentaux liés à l'entraînement du classifieur sont donnés dans la section 7.

6 Génération de synopsis

Cette section s'intéresse aux différentes façons d'obtenir un patron dynamiquement pour générer un synopsis. L'approche que nous proposons repose sur l'utilisation des synopsis annotés en variables de patron. L'idée est que parmi tous les synopsis déjà annotés il existe une combinaison de synopsis ou de fragment de synopsis (par exemple des phrases issues d'un tel résumé) capable de décrire les informations détectées dans une nouvelle conversation.

Pour obtenir un tel corpus de synopsis nous avons procédé de la manière suivante :

- pour chaque synopsis annoté du corpus d'apprentissage, remplacer les valeurs des variables par leur type ;
- découper les synopsis en fragments indépendants selon les frontières de phrases.

Cette approche nous permet d'avoir un large corpus de fragment de patrons, que l'on peut combiner pour en créer de nouveaux. Le remplissage et la sélection des patrons est alors effectué à partir des variables détectées dans une conversation : déterminer un ensemble de fragments de patrons qui maximisent la couverture en variables détectées, sous la contrainte qu'un fragment ne peut être exploité que s'il est saturé, et une variable ne peut être utilisée que par un seul fragment. Notre

implémentation repose sur une sélection gloutonne des fragments qui itère sur la population de fragments satisfaisant les contraintes jusqu'à couverture de toutes les variables détectées, ou réduction à l'ensemble vide des candidats acceptables.

L'avantage d'un tel remplissage est que ce sont les variables qui définissent directement le patron, permettant de mieux coller aux diverses situations se déroulant dans les conversations. Mais cela peut aussi être un inconvénient car il se peut que le thème de la conversation ne soit pas retranscrit dans le patron à cause de la mauvaise détection de variables importantes.

À titre comparatif, nous explorons aussi une méthode plus classique de résumé par patrons créés manuellement. Cette méthode consiste à écrire des patrons fixes, et aller chercher dans la conversation les valeurs correspondantes aux variables décrites dans le patron. Les patrons sont alors remplis en fonction des variables détectées dans la conversation. Si une variable n'est pas détectée, elle est laissée vide. Contrairement à la méthode que nous proposons, ici c'est le patron qui détermine le choix des variables et non l'inverse.

Cette méthode présente deux inconvénients majeurs, le premier est que les patrons sont figés et ne peuvent donc pas s'adapter à la particularité d'une conversation. Le second inconvénient est lié à la qualité du détecteur de variable qui peut en laisser vide, ce qui impacte la qualité linguistique du résumé généré.

7 Évaluation

La partie expérimentale de notre travail compare plusieurs variantes du système proposé, ainsi qu'un certain nombre de baselines, sur le corpus DECODA.

7.1 Cadre expérimental

Pour nos expériences nous disposons d'un corpus de 100 conversations provenant du corpus DECODA annotées manuellement en synopsis. Chaque conversation possède entre 1 et 3 synopsis différents pour un total de 175 synopsis au total. Les 175 synopsis sont annotés manuellement en variables de patron. Dans les expériences, nous utilisons les transcriptions de référence, ainsi que les annotations linguistiques de référence fournies avec le corpus.

Le corpus est découpé en un ensemble d'entraînement de 81 conversations, et un ensemble de test de 19 conversations. Le découpage est fait de façon à respecter la distribution des thèmes des conversations sur ce corpus.

Le classifieur est entraîné avec 100 tours de boosting sur l'ensemble d'entraînement, en ignorant les traits apparaissant moins de 5 fois. Il obtient un rappel de 62% et une précision de 43% soit une F-mesure de 51% sur la tâche de classification en type de variable. Les traits les plus discriminants, selon une analyse des poids du modèle, sont dans l'ordre le thème, les occurrences des entités nommées et lemmes, le type d'entité nommée, les sacs de mots et de lemmes.

On considère une variable comme valide si elle a reçu un score de probabilité supérieur à un certain seuil (nous avons fixé ce seuil arbitrairement à 0,1 pour nos expériences). Dans le cas où, pour la même variable, plusieurs valeurs dépassent ce seuil, seule la valeur avec le plus haut score est conservée.

Pour évaluer notre système nous utilisons la métrique ROUGE-2 (Lin, 2004) qui est largement utilisée dans la communauté du résumé automatique. Elle est calculée comme le rappel entre les bigrammes de mots des résumés produits et ceux de l'ensemble des résumés de référence pour une conversation donnée. Bien que cette métrique ait des limites, elle permet d'obtenir des résultats indicatifs de la ressemblance lexicale entre les résumés produits et ceux de référence. On notera qu'elle ne favorise pas les méthodes par abstraction car ne traite pas les paraphrases.

7.2 Expériences

Nous comparons le système proposé dans lequel les variables sont détectées automatiquement et les patrons sont générés à partir de fragments de résumés, à des variantes fondées sur l'écriture manuelle de patrons, ainsi qu'un des méthodes de résumé par extraction.

Les différents systèmes comparés sont les suivants :

- **Variables Icsiboost et patrons générés (IG)** : l'approche proposée ;
- **Variables Icsiboost et patrons manuels (GM)** : la détection des variables est effectuée par le classifieur, mais les patrons sont écrits manuellement ;
- **Variables de référence et patrons manuels (OM)** : un système oracle contenant des valeurs de référence pour les variables, cette variante représente les meilleures performances qu'un système par remplissage de patrons pourrait obtenir ;
- **Détection à base de règles et patrons manuels (RM)** : l'identification des variables repose sur des règles s'appuyant sur les annotations syntaxiques et sémantiques des conversations ;

Ces systèmes permettent d'avoir une bonne idée de l'impact des différents composants (détection de variables à base de règles, d'un classifieur ou de référence ; patrons manuels ou générés). En plus de ces systèmes, nous donnons des résultats comparatifs pour quelques baselines :

- **Entités nommées concaténées (EN)** : les entités nommées jouant un rôle important dans DECODA, les sorties de ce système sont générées en concaténant les entités nommées de la conversation jusqu'à saturer la limite de longueur des synopsis ;
- **Plus proche voisin (PPV)** : On attribue ici à une conversation donnée le synopsis de la conversation la plus proche en terme de distance cosinus entre leurs représentations en sacs de mots, sans y apporter les valeurs des variables de la conversation courante ;
- **Maximal Marginal Relevance (MMR)** : système de résumé par extraction basé sur l'algorithme de maximal marginal relevance (système par extraction) ;
- **Plus long tour de parole @25 (LB)** : tour de parole le plus long dans le premier quart de la conversation (système par extraction) ;
- **Plus long tour de parole (LA)** : tour de parole le plus long sur l'ensemble de la conversation (système par extraction).
- **Variables vides (VV)** : patrons manuels sélectionnés à partir du thème, et dont les variables ne sont pas remplies, pour vérifier l'apport lexical des mots du patron.

Les résultats sont résumés dans le tableau 5. Les systèmes sont définis et entraînés comme décrit plus haut.

Les résultats montrent que les approches "abstractives" produisent de meilleurs résultats que les approches extractives et que systèmes de référence, et ce même si la métrique ROUGE n'est pas

Id.	Système	Patrons	ROUGE-2
OM	Variables de référence	Manuels	0.18067
IG	Variables Icsiboost	Générés	0.11631
IM	Variables Icsiboost	Manuels	0.10675
RM	Règles	Manuels	0.10084
EN	Entités nommées concaténées	-	0.09337
LB	Plus long tour de parole @25 (ext.)	-	0.07816
PPV	Résumé du plus proche voisin	-	0.07196
MMR	Maximal Marginal Relevance (ext.)	-	0.07007
LA	Plus long tour de parole (ext.)	-	0.06419
VV	Variables vides	Manuels	0.02228

TABLE 5 – Résultats ROUGE-2 sur la génération de synopsis sur DECODA. Les systèmes extractifs sont signalés par (ext.)

OM	IG	IM	RM	EN	LB	PPV	MMR	LA	VV	
-	s	s	s	s	s	s	s	s	s	OM
	-	n	n	n	s	s	s	s	s	IG
		-	n	n	s	s	s	s	s	IM
			-	n	n	n	s	s	s	RM
				-	n	n	n	n	s	EN
					-	n	n	n	s	LB
						-	n	n	s	PPV
							-	n	s	MMR
								-	s	LA
									-	VV

FIGURE 2 – Significativité des améliorations à 95% du score ROUGE-2 entre les paires de systèmes. Les paires en "s" sont significativement différentes, alors que les paires en "n" ne le sont pas. Les identifiants de systèmes se réfèrent à la Table 5.

particulièrement adaptée pour les approches abstractives (étant donné qu'elle ne gère pas la synonymie par exemple). La concaténation des entités nommées obtient un meilleur score que les systèmes par extraction, mais le rendu de cette méthode est très peu lisible et n'a pas vraiment de sens pour comprendre ce qui s'est passé dans une conversation. Les systèmes basés sur la prédiction des variables dans les conversations semblent prometteurs dans le sens où ils donnent de meilleurs résultats que le système à base de règles. En revanche le système oracle est loin devant en terme de ROUGE-2, et montre qu'on peut améliorer le système de prédiction des variables. Le résumé par plus proche voisin est au niveau des méthodes extractives, toutefois il faudrait explorer comment cette approche pourrait être combinée avec la génération de patrons. Le système fondé sur des patrons vides obtient de très mauvais résultats montrant que sur le corpus DECODA, la prédiction des variables est nécessaire pour obtenir de bonnes performances.

Le tableau 2 montre la significativité des différences entre les systèmes à 95%. On y observe que l'oracle est significativement meilleur que tous les autres systèmes, que les systèmes par remplissage de patrons ne sont pas significativement différents entre eux, mais différents des baselines, notamment extractives.

7.3 Evaluation subjective des résumés

La mesure ROUGE, même si elle a le mérite de pouvoir comparer facilement de nombreux systèmes entre eux, sous réserve de disposer de résumés références, ne permet pas de rendre compte de l'utilité des résumés produits. Même si le score ROUGE est meilleur, les résumés produits sont-ils lisibles ? compréhensibles ? utiles ? Pour répondre à ces questions il convient d'analyser subjectivement les résumés produits afin d'en déterminer les forces et les faiblesses.

A cet effet, nous donnons, dans cette section, des exemples de synopsis, en commençant par ceux pour lesquels le système a produit un résultat considéré comme acceptable par les juges humains, puis en poursuivant par des synopsis problématiques illustrant les limites du système développé. Dans tous ces exemples les variables détectées par le système sont écrites en gras.

Synopsis acceptables

- Le client a perdu son **sac** dans le **RER** entre **Cergy** et **Poissy**.
- L'appelant voudrait se rendre de **trois rue d'Alger** à **Massy**. Le conseiller lui dit de prendre le **RER B** jusqu'à **Massy**.
- Demande d'horraire pour se rendre **Croix de Berny**. Prochain passage **treize heure trente**.

Synopsis problématiques

- L'appelant voudrait se rendre de **Gare de Drancy** à **Drancy**. (Aucun itinéraire trouvé)
- L'appelant voudrait se rendre de **rue Taclet** à **rue Taclet**. Le conseiller lui dit de prendre le **cent deux** jusqu'à **rue Taclet**. (Répétition de la même entité nommée)
- L'appelant voudrait se rendre de **Denfert-Rochereau** à **trente-six rue d'Assas**. Le conseiller lui dit de prendre le **ligne** jusqu'à **Paris**. (Requête correctement trouvée, mais résolution incorrecte)

Une rapide analyse des synopsis incorrects montre que l'approche proposée peut mener à des erreurs de réutilisation de la même valeur pour différentes variables. Ce problème peut être corrigé en tirant avantage de l'utilisation des coréférences.

Un autre problème rencontré est le fait de ne pas trouver de variable et de laisser une valeur vide. Ce problème est en partie corrigé par la sélection et la construction des patrons de façon automatique ou par fragment de patron.

Cette approche ne gère pas les situations inattendues, c'est-à-dire les conversations qui ne sont couvertes par aucun patron. Ces cas-là méritent plus d'attention afin de pouvoir être traités.

8 Conclusion

Dans le cadre du résumé de conversation parlée provenant de centre d'appel, nous avons présenté une méthode basée sur la génération à partir de fragments de résumés d'un corpus d'apprentissage, et le remplissage de ces patrons par des valeurs détectées dans les conversations grâce à des traits linguistiques et structurels. Notre approche montre la prévalence des résumés par abstraction par rapport aux résumés par extraction et montre que la quantité de supervision pour effectuer le remplissage de patrons peut être réduite.

Une autre conclusion serait que la métrique ROUGE n'est pas une métrique adaptée pour ce type d'évaluation de résumés abstraits du fait de la forte variabilité des termes utilisés. Une approche qui

pourrait être explorée est de se servir des words embedding afin de permettre de se servir de cette variabilité dans l'évaluation et ainsi définir une évaluation plus significative.

Références

- BAKER C. F., FILLMORE C. J. & LOWE J. B. (1998). The berkeley framenet project. In *Proceedings of the 17th international conference on Computational linguistics-Volume 1*, p. 86–90 : Association for Computational Linguistics.
- BAZILLON T., DEPLANO M., BECHET F., NASR A. & FAVRE B. (2012). Syntactic annotation of spontaneous speech : application to call-center conversation data. In *LREC*, p. 1338–1342.
- BECHET F., MAZA B., BIGOUROUX N., BAZILLON T., EL-BEZE M., DE MORI R. & ARBILLOT E. (2012). Decoda : a call-centre human-human spoken conversation corpus. In *LREC*, p. 1343–1347.
- BYRD R. J., NEFF M. S., TEIKEN W., PARK Y., CHENG K.-S. F., GATES S. C. & VISWESWARIAH K. (2008). Semi-automated logging of contact center telephone calls. In *Proceedings of the 17th ACM conference on Information and knowledge management*, p. 133–142 : ACM.
- FAVRE B., HAKKANI-TÜR D. & CUENDET S. (2007). Icsiboost. *Online* : <http://code.google.com/p/icsiboost>.
- GARG N., FAVRE B., REIDHAMMER K. & HAKKANI TÜR D. (2009). Clusterrank : a graph based method for meeting summarization. In *Interspeech*.
- HIGASHINAKA R., MINAMI Y., NISHIKAWA H., DOHSAKA K., MEGURO T., TAKAHASHI S. & KIKUI G. (2010). Learning to model domain-specific utterance sequences for extractive summarization of contact center dialogues. In *Proceedings of the 23rd International Conference on Computational Linguistics : Posters*, p. 400–408 : Association for Computational Linguistics.
- LIN C.-Y. (2004). Rouge : A package for automatic evaluation of summaries. In *Text summarization branches out : Proceedings of the ACL-04 workshop*, volume 8.
- MEHDAD Y., CARENINI G. & NG R. T. (2014). Abstractive summarization of spoken and written conversations based on phrasal queries. In *ACL (1)*, p. 1220–1230.
- NENKOVA A., MASKEY S. & LIU Y. (2011). Automatic summarization. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics : Tutorial Abstracts of ACL 2011*, p. 3 : Association for Computational Linguistics.
- WHITE M., KORELSKY T., CARDIE C., NG V., PIERCE D. & WAGSTAFF K. (2001). Multidocument summarization via information extraction. In *Proceedings of the first international conference on Human language technology research*, p. 1–7 : Association for Computational Linguistics.