

Analyse automatique de sondages téléphoniques d'opinion*

Nathalie CAMELIN¹, Frédéric BÉCHET¹
Géraldine DAMNATI², Renato DE MORI¹

¹ LIA/CNRS, University of Avignon,
BP1228, 84911 Avignon cedex 09 France

² France Télécom R&D – TECH/SSTP/RVA, 2 av. Pierre Marzin,
22307 Lannion Cedex 07, France

{nathalie.camelin, frederic.bechet, renato.demori}@
univ-avignon.fr
geraldine.damnati@francetelecom.com

Résumé. Cette étude présente la problématique de l'analyse automatique de sondages téléphoniques d'opinion. Cette analyse se fait en deux étapes : tout d'abord extraire des messages oraux les expressions subjectives relatives aux opinions de utilisateurs sur une dimension particulière (efficacité, accueil, etc.) ; puis sélectionner les messages *fiab*les, selon un ensemble de mesures de confiance, et estimer la distribution des diverses opinions sur le corpus de test. Le but est d'estimer une distribution aussi proche que possible de la distribution de référence. Cette étude est menée sur un corpus de messages provenant de vrais utilisateurs fournis par France Télécom R&D.

Abstract. This paper introduces the context of the automatic analysis of opinion telephone surveys. This analysis is done by means of two stages : firstly the subjective expressions, related to the expression of an opinion on a particular dimension (efficiency, courtesy, ...), are extracted from the audio messages ; secondly the *reliable* messages, according to a set of confidence measures, are selected and the distribution of the positive and negative opinions in these messages is estimated. The goal is to obtain a distribution as close as possible to the reference one. This study is carried on a telephone survey corpus, provided by France Télécom R&D, obtained in real field conditions.

Mots-clés : détection d'opinions, classification automatique, reconnaissance automatique de la parole, champs conditionnels aléatoires.

Keywords: opinion extraction, automatic classification, automatic speech recognition, conditional random fields.

1 Introduction

Face à la quantité grandissante de donnée disponible, l'extraction d'information pertinente est devenue un des défis de ces dernières années. Plus précisément, l'extraction d'opinion a ré-

*Travaux réalisés en collaboration avec France Télécom's R&D, contrat 021B178.

comment fait l'objet d'une grande attention de la part de la communauté TALN (atelier d'ACL 2006 *Sentiment and Subjectivity in Text*, DEFT'07). Ce domaine a donné lieu à de nombreuses publications (Wiebe *et al.*, 2005; Choi *et al.*, 2005) portant principalement sur deux aspects : la détection automatique d'opinions à partir d'avis rédigés par des consommateurs (Popescu & Etzioni, 2005) et d'autre part l'analyse de la subjectivité d'une phrase pour les systèmes de résumé automatique ou de question/réponse (Riloff & Wiebe, 2003).

Les travaux présentés dans cette étude concernent le premier cadre applicatif : la détection automatique d'opinions à partir de sondage d'utilisateurs. Il s'agit ici de sondages téléphoniques effectués par France Télécom auprès d'utilisateurs réels. Une des principales caractéristiques de cette étude est la détection d'opinions à partir de messages vocaux, contenant de la parole complètement spontanée, collectée dans des conditions réelles.

A cause des difficultés intrinsèques à ce type de corpus (bruits, téléphone mobile, disfluences, grande variété d'accents, nombreuses digressions entraînant un nombre important de mots hors-vocabulaire), il est très important de développer des méthodes robustes, peu sensibles aux erreurs de Reconnaissance Automatique de la Parole (RAP). En contrepartie, le nombre d'opinions susceptibles d'être détectées est forcément réduit. Nous avons présenté dans (Camelin *et al.*, 2006) une stratégie de RAP utilisant des modèles de langage spécifiques à la détection d'opinions afin de limiter le bruit généré par les portions de messages hors-sujet (ou digressions). La problématique des sondages d'opinions à partir de corpus de messages vocaux et une stratégie d'analyse automatique de ces mêmes sondages a été présentée dans (Béchet *et al.*, 2006). Les travaux présentés ici proposent de nouvelles stratégies mettant l'accent sur l'introduction de connaissances explicites dans le processus de classification automatique. La définition de critères de calibrage du système de détection, spécifiques à la problématique des sondages, est également abordée.

Cet article est organisé comme suit : le paragraphe 2 formalise le problème de la détection d'opinion et de l'analyse de sondages ; le paragraphe 3 introduit le corpus utilisé dans cette étude ; le paragraphe 4 présente les stratégies de détection d'opinions sur des transcriptions manuelles des messages et sur des sorties du module de RAP ; enfin la section 5 permet de comparer ces 2 stratégies et détaille l'ajout de connaissances explicites dans la stratégie analysant les messages vocaux ; elle expose également comment le système peut être paramétré pour le problème particulier de l'analyse de sondages.

2 Formulation du problème

L'analyse automatique d'opinions à partir de messages collectés dans des conditions réelles est une tâche difficile. Une très grande variété de locuteurs exprime leurs opinions de très nombreuses façons, avec des messages de tailles variables, et un grand nombre de répétitions, corrections et contradictions. À cette variabilité des locuteurs s'ajoute la variabilité acoustique due à des environnements et des canaux de transmissions très variés (téléphones portables, bruits ambiants). Pour toutes ces raisons, les résultats obtenus par les systèmes de RAP sur ce type de corpus sont très variables, avec des taux d'erreurs sur les mots dépassant les 50%.

Ces erreurs de reconnaissance vont grandement affecter les performances de détection d'opinions pour les messages très bruités, cependant le problème de l'analyse de sondages ne nécessite pas le traitement de la totalité des messages enregistrés. En effet, le but recherché est de connaître la distribution des étiquettes d'opinions sur le corpus, pas les étiquettes individuelles

de chaque message. On retrouve ici la problématique traditionnelle des sondages d'opinion : comment collecter un sous-ensemble traitable d'observations qui conserve les mêmes distributions d'opinions que celles du corpus général. Le *sous-ensemble traitable*, dans notre cas, est l'ensemble des messages considérés comme *fiabiles* par les modèles de RAP et de détection d'opinion. Nous verrons au paragraphe 4 comment cette fiabilité est estimée à l'aide de mesures de confiance.

2.1 Analyse de sondage

Le problème est formalisé de la manière suivante :

Soit C un corpus de n messages oraux m_1, m_2, \dots, m_n .

Soit $C' \in C$ un sous-ensemble de n' messages de C sélectionné par la stratégie d'analyse automatique.

Soit $O(m, x) \in \{o_1, \dots, o_k\}$ l'opinion exprimé dans le message m à propos de la dimension $x \in D$. Les dimensions correspondent ici aux différentes analyses effectuées dans le sondage, par exemple sur l'efficacité d'un service, la courtoisie des opérateurs, le coût, etc. Les valeurs d'opinions $o_i \in O$ sont les différentes opinions possibles. Dans cette étude on considère les opinions suivantes :

$O = \{\text{entièrement positives, entièrement négatives, mitigées, sans opinion}\}$.

Les opinions $O_r(m, x)$ sont les opinions de référence, données par des annotateurs humains. Les opinions $O_h(m, x)$ sont les étiquettes d'opinions attribuées automatiquement.

Le but des stratégies proposées dans cette étude est de minimiser la distance entre la distribution des opinions de référence PC pour la dimension x :

$$PC(x) = [p(o_1), \dots, p(o_k)] \text{ et } p(o_i) = \frac{|C_{o_i}|}{n}$$

avec $m \in C_{o_i}$ ssi $m \in C$ et $O_r(m, x) = o_i$.

et la distribution PC' estimée sur le sous-corpus extrait C' :

$$PC'(x) = [p'(o_1), \dots, p'(o_k)] \text{ et } p'(o_i) = \frac{|C'_{o_i}|}{n'}$$

avec $m \in C'_{o_i}$ ssi $m \in C'$ et $O_h(m, x) = o_i$.

Cette distance est évaluée grâce à la divergence de Kullback-Leibler (KLD) entre les deux distributions :

$$D_{KL}(PC(x)||PC'(x)) = \sum_{i=1}^k p(o_i) \cdot \log \frac{p(o_i)}{p'(o_i)} \quad (1)$$

2.2 Détection d'opinions

Le formalisme introduit dans le paragraphe précédent nécessite le calcul de $O(m, x)$ qui représente l'opinion o_i contenue dans le message m à propos de la dimension x . Ces opinions o_i

sont exprimées dans le message sous la forme d'*expressions subjectives* notées $W(o_i, x)$. Par exemple pour la dimension $x = \text{courtoisie}$ et $o_i = \text{entièrement positive}$ on peut trouver dans notre corpus : $W(o_i, x) = [\text{l'accueil était parfait}]$.

Le rôle du module de détection d'opinions est d'extraire des messages vocaux m ces séquences $W(o_i, x)$ afin de calculer $O(m, x)$. Cette opération se fait en deux étapes : tout d'abord extraire du message les segments susceptibles de représenter des expressions subjectives, puis caractériser ces segments en fonctions des différentes étiquettes d'opinions.

Une fois ces étapes effectuées, un message m est décrit par une séquence de segments : $m = W_1(o_1, d_1) \dots W_i(o_i, d_i)$ avec $o_i \in O$ et $d_j \in D$.

L'attribution de l'opinion $O(m, x)$ au message m pour la dimension x est effectuée de la manière suivante :

$$O(m, x) = \begin{cases} \text{sans opinion} & \text{si } \forall W_i(o_i, d_i) \in m \text{ on a } d_i \neq x \\ \text{satisfait} & \text{si } \forall W_i(o_i, x) \in m \text{ on a } o_i = \text{satisfait} \\ \text{insatisfait} & \text{si } \forall W_i(o_i, x) \in m \text{ on a } o_i = \text{insatisfait} \\ \text{mitigé} & \text{sinon} \end{cases}$$

3 Description du corpus de sondage téléphoniques

Le corpus de sondage téléphonique a été collecté auprès de réels clients d'un service de France Télécom. Les personnes contactées sont invitées par un court message à appeler un numéro gratuit qui leur permet d'exprimer leur satisfaction vis à vis du service-client qu'ils ont récemment appelé. En composant ce numéro, le message vocal suivant les invite à laisser un message : [...] *Vous avez récemment contacté notre service clientèle. Nous souhaitons nous assurer que vous avez été satisfait de l'accueil et de la suite donnée à votre appel. N'hésitez pas à me faire part de tous vos commentaires et de vos suggestions sur notre service, ceux-ci nous aideront à nous améliorer. Nous vous remercions de votre aide et nous restons à votre disposition. Laissez votre message après le signal sonore. [...]*

A l'origine ces messages étaient destinés à être traités par des opérateurs. Ainsi aucune consigne de nature à faciliter le traitement automatique n'a été donnée : pas de conseils sur le mode d'élocution, question ouverte et même incitation à laisser des commentaires. Pour cette étude un ensemble de 1779 messages, collectés sur une période de 3 mois, a été transcrit manuellement au niveau mots, expressions subjectives et marqueurs (indication de disfluence et marqueurs discursifs). Ce corpus a été divisé en deux sous-corpus : un corpus d'apprentissage contenant environ 80% des messages et un corpus de test contenant les 20% restant.

L'analyse de la satisfaction des utilisateurs par l'équipe d'analyse des sondages se fait selon trois dimensions : la qualité de l'accueil (notée *accueil*), la rapidité d'accès au service (notée *attente*) et enfin l'efficacité du service (notée *efficacité*). Cette dernière dimension est la plus représentée dans le corpus, elle concerne à la fois l'évaluation des réponses aux attentes des utilisateurs (est ce que le problème a été réglé ?) mais aussi la qualité des informations données. Chaque expression subjective peut recevoir deux polarités : *positive* et *négative*. Nous avons donc un total de 6 étiquettes pour caractériser les expressions subjectives du corpus.

Dans la transcription manuelle, au sein de chaque message, ces expressions sont indiquées par des balises. Nous disposons ainsi d'un corpus de segments, chacun porteur d'une opinion

particulière. Le but du traitement automatique est de retrouver ces segments et de les étiqueter avec l'une des 6 étiquettes. Voici un exemple de message avec les balises manuelles :

oui c'est monsieur NOMS PRENOMS j'avais appelé donc le service client ouais <seg label=accueil,pos> j'ai été très bien accueilli </seg> des <seg label=efficacité,pos> bons renseignements </seg> sauf que <seg label=efficacité,neg> ça ne fonctionne toujours pas </seg> donc je sais pas si j'ai fait une mauvaise manipulation ou y a un problème enfin voilà sinon <seg label=efficacité,pos label=accueil,pos> l'accueil était et les conseils très judicieux </seg> même si <seg label=efficacité,neg> le résultat n'est pas n'est pas là </seg> merci au revoir

4 Détection et classification d'expressions subjectives

Deux stratégies ont été développées pour extraire et classifier les expressions subjectives des messages vocaux : l'une (notée *ref*) s'appuie sur les transcriptions manuelles des messages ; l'autre (notée *asr*) est intégrée dans le processus de décodage de parole. Ces deux stratégies nous permettent de dissocier les erreurs dues à une mauvaise transcription en mots des erreurs de détection d'opinions. Il a été nécessaire de différencier les traitements sur les transcriptions manuelles des traitements sur les transcriptions automatiques à cause de la très mauvaise qualité de ces dernières : les méthodes développées sur le texte *propre* ne sont pas assez robustes pour s'appliquer aux transcriptions automatiques bruitées. La figure 1 présente ces stratégies. Elles sont décrites brièvement dans les prochains paragraphes.

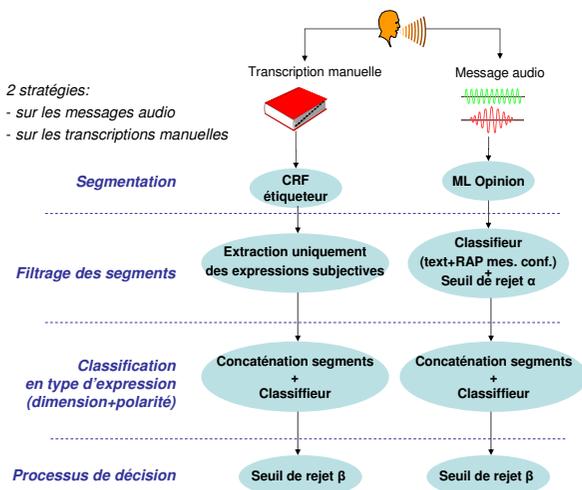


FIG. 1 – Stratégies de détection et de classification d'expressions subjectives pour les transcriptions manuelles et les messages audio

Modèles de Langage spécifiques à l’expression d’opinions. Nous avons montré dans (Camelin *et al.*, 2006) qu’un modèle de langage spécifique à la détection des opinions permettait d’obtenir de meilleures performances lors de la phase de classification qu’un modèle RAP standard de type bigramme. Ceci est dû principalement au protocole de collecte des messages (pas de contraintes d’élocution, encouragement à laisser des commentaires) qui implique une très grande dispersion dans les fréquences de distribution des mots du corpus collecté.

Ce modèle de langage spécifique permet ainsi d’obtenir le message en une suite d’hypothèses d’expressions subjectives séparées par un symbole représentant les segments considérés comme vides.

Un ensemble de mesures de confiance (acoustiques et linguistiques) est associé à chaque hypothèse. Sa probabilité d’être correcte est alors approximée sur le corpus d’apprentissage par régression logistique sur ses mesures de confiance. Comme présenté dans la figure 1 cette probabilité permet de filtrer les hypothèses peu fiables (seuil α de la figure).

Le principal avantage d’un tel modèle est de segmenter directement le flux audio en hypothèses d’expressions subjectives.

Segmentation des messages avec des Champs Conditionnels Aléatoires. Pour le traitement des transcriptions manuelles un segmenteur basé sur les Champs Conditionnels Aléatoires (ou *Conditional Random Fields* CRF) a été développé. Les CRF (Lafferty *et al.*, 2001) ont été utilisés avec succès dans de nombreuses tâches d’étiquetage telles que l’étiquetage morpho-syntaxique ou la détection d’entités nommées. L’avantage principal des CRF par rapport à des modèles génératifs tels que les Modèles de Markov Cachés est la possibilité d’utiliser l’ensemble des observations d’une séquence pour prédire une étiquette. Ce n’est donc pas le seul historique immédiat qui contraint l’attribution d’une étiquette à une observation mais potentiellement toutes les observations précédentes et suivantes.

Dans notre cas, le corpus d’apprentissage est formaté de manière à associer à chaque mot une étiquette indiquant s’il fait partie d’une expression subjective ou non. L’étiqueteur développé est basé sur l’outil *CRF++*¹ qui permet de représenter chaque mot selon différents niveaux. Ainsi un mot est représenté par : son lemme et une étiquette nommée *seed*. En effet, plusieurs études (Hatzivassiloglou & McKeown, 1997) utilisent un ensemble de mots (appelés *seeds*) qui expriment explicitement une opinion (*e.g. gentil, agréable, utile, efficace*).

Lors de l’analyse d’un nouveau message, les étiquettes posées par le CRF permettent d’extraire uniquement les hypothèses d’expressions subjectives.

Classification des opinions. Une fois la segmentation du message effectuée, les hypothèses d’expressions subjectives associées à un même label sont concaténées. Chaque segment ainsi obtenu est ensuite étiqueté par un classifieur basé sur l’algorithme *AdaBoost* (Schapire & Singer, 2000).

Deux modèles sont appris sur les expressions subjectives du corpus d’entraînement annotées manuellement. La transcription exacte est utilisée pour le modèle *ref* et la transcription automatique pour le modèle *asr*. Chaque expression subjective est représentée en entrée du classifieur par ses lemmes et seeds ainsi que le nombre de mots.

¹Toolkit CRF++ : <http://www.chasen.org/taku/software/CRF++/>

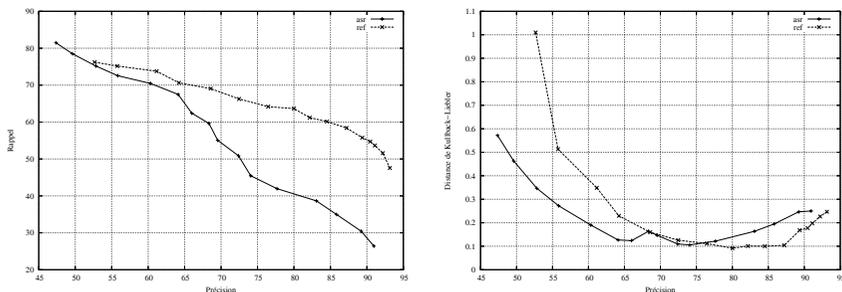


FIG. 2 – Courbes de précision/rappel sur les segments d'opinion (gauche) et de précision/divergence D_{KL} (droite) obtenues sur le corpus de test en faisant varier les seuils de rejet α et β pour les transcriptions manuelles (*ref*) et les sorties du module de RAP (*asr*)

Ainsi l'expression subjective (transcription manuelle) : *j'ai été très satisfaite euh de la communication avec euh l'interlocuteur que j'avais au téléphone donc euh j' attends des documents la personne était très gentille et très serviable et m' a bien dépannée* est représentée par : *il avoir être très satisfaisre de le communication avec le interlocuteur que il avoir au téléphone donc il attendre des document le personne être très gentil et très serviable et me avoir bien dépanné , satisfaisre communication téléphone attendre personne gentil serviable dépanné , 29.*

En sortie du classifieur, un score de confiance est attribué à chaque étiquette recherchée pour le segment considéré. Les étiquettes retenues sont celles dépassant le seuil β de la stratégie présentée dans la figure 1.

5 Expériences

Cette section permet de comparer les 2 stratégies présentées précédemment et ainsi d'évaluer l'influence des erreurs du module RAP. Une amélioration de la stratégie *asr* avec l'introduction de connaissances explicites est ensuite proposée. Enfin, les critères de choix d'une stratégie par rapport à une autre selon la problématique des sondages sont exposés.

Évaluation des stratégies. L'évaluation est faite sur le corpus de test par rapport à deux types de mesures : les mesures de précision/rappel sur la détection des expressions subjectives et la mesure de la distance de Kullback-Leibler (D_{KL}) entre la distribution de référence sur les opinions et celle estimée automatiquement. La figure 2 présente les courbes obtenues en faisant varier les seuils de rejet α et β .

Comme attendu la stratégie *ref* s'appliquant aux transcriptions manuelles donne de bien meilleurs résultats en terme de précision/rappel. Il est par contre particulièrement intéressant de constater que pour la mesure de la divergence les deux courbes atteignent les mêmes valeurs. Ce résultat valide notre approche consistant à sélectionner un sous-ensemble représentatif de messages pour lesquels les prises de décision sur l'attribution d'opinions sont fiables selon la stratégie implémentée.

Utilisation de connaissances explicites. Une étude manuelle des erreurs générées par la stratégie *asr* sur le corpus d'entraînement a permis de mettre en évidence que de nombreux messages rejetés contenaient des phrases idiomatiques selon une ou plusieurs des dimensions recherchées. Ces phrases ont été extraites du corpus d'entraînement puis manuellement généralisées sous formes d'expressions régulières. Ces expressions sont très générales, en petit nombre, non ambiguë, et sont relativement indépendante de l'application visée de service clientèle. La raison pour laquelle elles n'ont pas été capturées par le processus de classification automatique est due à la faible taille du corpus d'apprentissage. L'apport de connaissance explicite vise ainsi à pallier aux faiblesses des méthodes d'apprentissage automatique sur des données de taille réduite. A la suite de ce processus manuel huit expressions régulières ont été associées à la dimension *accueil*, deux pour *attente* et treize pour *efficacité*.

Pour pouvoir évaluer l'apport de l'utilisation de cette connaissance explicite à notre système, quatre stratégies sont proposées :

- La stratégie Ψ_1 est celle utilisée dans le système *asr*, sans l'apport de connaissances explicites.
- Pour la stratégie Ψ_2 , les expressions régulières ont été intégrées comme paramètre d'entrée de l'algorithme de classification *AdaBoost*.
- La stratégie Ψ_3 correspond à la fusion des hypothèses d'opinions obtenues par la stratégie Ψ_1 et celles obtenues en appliquant directement les expressions régulières sur les segments.
- Enfin la stratégie Ψ_4 correspond à une stratégie séquentielle : les expressions régulières ayant été apprises principalement sur l'ensemble des messages rejetés par la stratégie Ψ_1 , celles-ci sont appliquées uniquement sur l'ensemble des messages rejetés par Ψ_1 .

Toutes ces stratégies suivent la règle de rejet suivante : si aucun des segments du message n'a été associé à une étiquette d'opinion, le message est rejeté.

Pour chaque stratégie, la précision, le rappel et la F-mesure ont été calculés en faisant varier les seuils α et β . La figure 3, présentant la F-mesure en fonction de la précision, permet de mettre en évidence l'apport significatif de l'utilisation de connaissances explicites quelles que soit la stratégie choisie, et ce malgré le faible nombre d'expressions régulières rajoutées pour chaque dimension. La stratégie de fusion Ψ_3 est celle qui permet d'obtenir la plus forte valeur de F-mesure.

Choix de la stratégie et réglage du système. Pour les systèmes de détection d'entités, le choix de la meilleure stratégie ou le réglage de paramètres tels que les seuils de rejet est généralement fait sur des courbes de précision/rappel ou de F-mesure telles que la courbe 3. Dans cette étude, le choix de la stratégie à utiliser est fait selon la problématique des sondages d'opinions. En effet, il s'agit de trouver la stratégie qui conservera le mieux les distributions du corpus général. Pour cela la divergence de Kullback-Leibler entre les proportions réelles et celles estimées est calculée pour toutes les stratégies avec différentes valeurs pour les seuils α et β . Le point de fonctionnement du système est choisi comme celui qui minimise cette divergence. La figure 4 présente cette courbe pour les quatre stratégies développées.

Les stratégies Ψ_3 et Ψ_4 montrent une distance de Kullback-Leibler systématiquement plus faible que la stratégie Ψ_1 , la stratégie de fusion Ψ_3 apparaissant comme la plus performante.

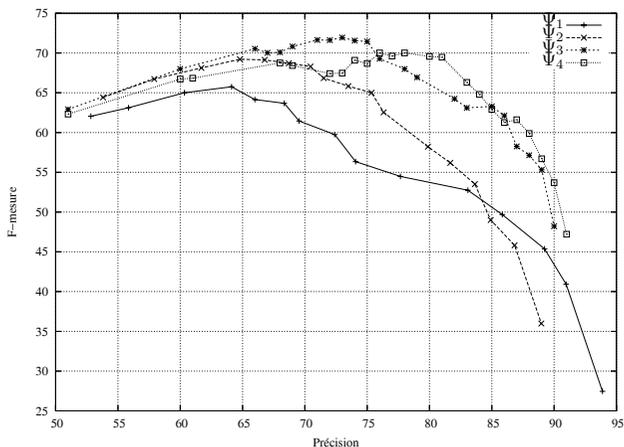


FIG. 3 – F-mesure obtenue par les 4 stratégies d'extraction d'opinions sur le corpus de test en faisant varier les seuils α and β .

6 Conclusion

Nous avons présenté dans cette étude la problématique de l'analyse automatique de sondages d'opinion à partir de messages oraux. Trois résultats originaux ont été obtenus :

1. Il est possible d'extraire de manière robuste de l'information à partir de transcriptions automatiques très bruitées (dues à l'extrême variabilité des corpus oraux collectés) si on accepte de filtrer et sélectionner les messages *fiabiles* selon un ensemble de mesures de confiance. Les résultats obtenus dans cette étude avec cette stratégie sont identiques à ceux obtenus sur des transcriptions exactes.
2. L'ajout de connaissances explicites peut améliorer un processus de classification automatique en permettant de généraliser certains phénomènes peu représentés dans le corpus d'apprentissage. Diverses stratégies sont proposées pour réaliser cet ajout, c'est la fusion des hypothèses qui s'est montrée la plus robuste dans notre étude.
3. Enfin le choix de la stratégie et son point de fonctionnement doivent être fait par rapport à la tâche visée. Dans le cadre de l'analyse de sondages, c'est la divergence entre la distribution de référence des opinions et celle estimée qui doit être minimisée, plutôt que la précision ou le rappel dans la détection des opinions.

Références

BÉCHET F., DAMNATI G., CAMELIN N. & DE MORI R. (2006). Spoken opinion extraction detecting variations in user satisfaction. In *IEEE/ACL Workshop on Spoken Language Technology*.

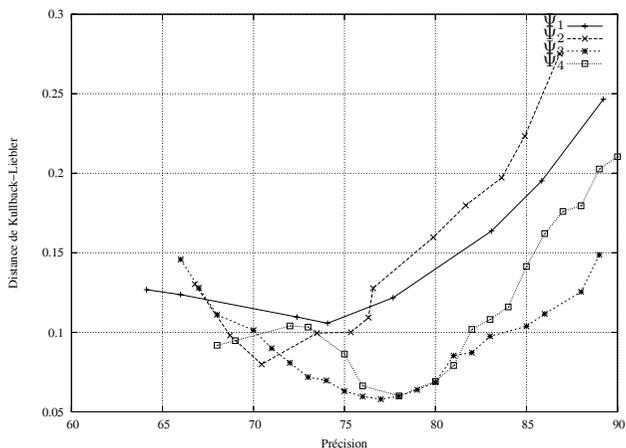


FIG. 4 – Distance de Kullback-Leibler obtenue par les différentes stratégies appliquées sur le corpus de test transcrit automatiquement. Les courbes sont obtenus pour différentes valeurs de α et β .

CAMELIN N., DAMNATI G., BÉCHET F. & DE MORI R. (2006). Détection automatique d'opinions dans des corpus de messages oraux. In *Journées d'Etude de la Parole*, France.

CHOI Y., CARDIE C., RILOFF E. & PATWARDHANN S. (2005). Identifying sources of opinions with conditional random fields and extraction patterns. In *HLT/EMNLP*, p. 355–362, Vancouver, Canada.

HATZIVASSILOGLOU V. & MCKEOWN K. R. (1997). Predicting the semantic orientation of adjectives. In *European chapter of ACL*, p. 174–181, Morristown, NJ, USA : Association for Computational Linguistics.

LAFFERTY J., MCCALLUM A. & PEREIRA F. (2001). Conditional random fields : Probabilistic models for segmenting and labeling sequence data. In *Proc. 18th International Conf. on Machine Learning*, p. 282–289 : Morgan Kaufmann, San Francisco, CA.

POPESCU A.-M. & ETZIONI O. (2005). Extracting product features and opinions from reviews. In *HLT/EMNLP*.

RILOFF E. & WIEBE J. (2003). Learning extraction patterns for subjective expressions. In *Proceedings of the 2003 conference on EMNLP*.

SCHAPIRE R. E. & SINGER Y. (2000). BoosTexter : A boosting-based system for text categorization. *Machine Learning*, **39**, 135–168.

WIEBE J., WILSON T. & CARDIE C. (2005). Annotationg expressions of opinions and emotions in language. In *Language Resources and Evaluation*.