Avant-propos

L'analyse des sentiments est un domaine de recherche extrêmement actif en traitement automatique des langues (TAL). En effet, ces dernières années ont vu se multiplier les sources de données textuelles porteuses d'opinion disponibles sur le web : avis d'internautes, forums, réseaux sociaux, enquêtes consommateurs, etc. Devant cette abondance de données, l'automatisation de la synthèse des multiples avis devient cruciale pour obtenir efficacement une vue d'ensemble des opinions sur un sujet donné. L'intérêt de ces données est considérable pour les sociétés qui souhaitent obtenir un retour client sur leurs produits comme pour les personnes souhaitant se renseigner pour un achat ou un voyage.

Depuis les années 2000, un grand nombre de travaux ont été publiés sur le sujet, faisant de l'extraction d'opinion un domaine très actif dans la recherche en TAL. Globalement, les systèmes actuels ont obtenu de bons résultats sur la classification automatique du caractère subjectif ou objectif d'un document. En revanche, ceux obtenus sur la tâche d'analyse de polarité (qui consiste à classer le document sur une échelle de subjectivité allant du plus positif au plus négatif) restent encore peu concluants. La raison principale de cet échec est l'incapacité des algorithmes actuels à comprendre toutes les subtilités du langage humain, telles que l'usage du langage figuratif. Contrairement au langage littéral, le langage figuratif exploite des dispositifs linguistiques tels que l'ironie, l'humour, le sarcasme, la métaphore et l'analogie qui entraînent une difficulté au niveau de la représentation linguistique ainsi qu'au niveau du traitement automatique du langage figuratif. Dans le cadre de cet ouvrage, nous nous focalisons sur l'ironie et le sarcasme dans un type particulier de données, à savoir, les tweets.

Dans ce cadre, nous proposons une approche par apprentissage supervisé afin de prédire si un tweet est ironique ou pas. Pour ce faire, nous avons suivi une démarche en trois étapes. Dans un premier temps, nous nous sommes intéressés à l'analyse des phénomènes pragmatiques utilisés pour exprimer l'ironie en nous inspirant des travaux

en linguistique afin de définir un schéma d'annotation multiniveau pour l'ironie. Ce schéma d'annotation a été exploité dans le cadre d'une campagne d'annotation d'un corpus formé de 2 000 tweets français. Dans une deuxième étape, en exploitant l'ensemble des observations faites sur le corpus annoté, nous avons développé un modèle de détection automatique de l'ironie pour les tweets en français qui exploite, à la fois le contexte interne du tweet à travers des traits lexicaux et sémantiques, et le contexte externe en recherchant des informations disponibles sur le web. Enfin, dans la troisième étape, nous avons étudié la portabilité du modèle pour la détection de l'ironie dans un cadre multilingue (italien, anglais et arabe). Nous avons ainsi testé la performance du schéma d'annotation proposé sur l'italien et l'anglais et nous avons testé la performance du modèle de détection automatique à base de traits sur la langue arabe. Les résultats obtenus pour cette tâche extrêmement complexe sont très encourageants et sont une piste à explorer pour l'amélioration de la détection de polarité lors de l'analyse de sentiments.

Introduction

I.1. Contexte et motivations

De nos jours, le web est devenu une source d'information incontournable grâce à la quantité et à la diversité des contenus textuels porteurs d'opinions exprimées par les internautes. Ces contenus sont multiples : blogs, commentaires, forums, réseaux sociaux, réactions ou avis, de plus en plus centralisés par les moteurs de recherche. Devant cette abondance de données et de sources, le développement d'outils pour extraire, synthétiser et comparer les opinions exprimées sur un sujet donné devient crucial. L'intérêt de ce type d'outils est considérable, pour les entreprises qui souhaitent obtenir un retour client sur leurs produits ou leur image de marque comme pour les particuliers souhaitant se renseigner pour un achat, une sortie, ou un voyage. Actuellement, les instituts de sondage s'intéressent à ces outils également pour l'évaluation d'un produit sur le marché ou pour prévoir les résultats lors des élections présidentielles, par exemple.

C'est dans ce contexte que l'analyse d'opinions (communément appelée *sentiment analysis* ou *opinion mining* en anglais) a vu le jour. Les premiers travaux de recherche en extraction automatique d'opinion remontent à la fin des années 1990 avec en particulier les travaux de Hatzivassiloglou et McKeown (1997) traitant de la détermination de la polarité d'adjectifs, et ceux de Pang *et al*. (2002), Littman et Turney (2002) sur la classification de documents suivant leur polarité positive ou négative. Depuis les années 2000, un grand nombre de travaux a été publié sur le sujet, faisant de l'extraction d'opinion un domaine très actif dans la recherche en traitement automatique des langues (TAL) (Liu 2015, Benamara *et al*. 2017). De nombreuses campagnes d'évaluation sont également consacrées à ce sujet, telles que la campagne TREC (*Text Retrieval Conference*) (Ounis *et al*. 2008), la campagne DEFT (Défi fouille de textes)

pour le français avec une première édition en 2005 (Azé et Roche 2005) et la campagne SemEval (*Semantic Evaluation*) avec une première édition en 1998[1].

Globalement, les systèmes actuels ont obtenu de bons résultats sur la tâche d'analyse de subjectivité qui consiste à déterminer si une portion de texte véhicule une opinion (c'est-à-dire qu'elle est subjective) ou ne fait que présenter des faits (c'est-à-dire qu'elle est objective) (Turney 2002). Par exemple, l'utilisation de lexiques de subjectivité couplés éventuellement à des techniques de classification permet de détecter le fait que l'auteur exprime une opinion positive envers le Premier ministre dans la phrase (I.1) (*via* l'utilisation de l'adjectif *excellent* de polarité positive) :

(I.1) Le Premier ministre a fait un excellent discours.

En revanche, les résultats des systèmes d'analyse d'opinions sur la tâche d'analyse de polarité, qui consiste à déterminer la polarité globale et/ou le score de l'opinion effectivement véhiculée par une portion de texte que l'on sait subjective, restent encore peu concluants. Les trois exemples ci-après, extrait de Benamara (2017), illustrent parfaitement la difficulté de la tâche :

(I.2) [J'ai acheté un iPhone 5s d'occas il y a trois mois.]$_{P1}$ [La qualité d'image est exceptionnelle.]$_{P2}$ [En revanche, la protection en verre trempée n'est pas de bonne qualité]$_{P3}$ [et la batterie m'a lâchée au bout de 15 jours !!]$_{P4}$

L'exemple (I.2) contient quatre propositions, délimitées par des crochets. Seules les trois dernières sont porteuses d'opinions (en bleu). Parmi ces opinions, les deux premières sont explicites, c'est-à-dire repérables par des mots, symboles ou expressions subjectives du langage, comme l'adjectif *exceptionnelle*. La dernière est cependant implicite car elle repose sur des mots ou groupes de mots qui décrivent une situation (fait ou état) jugée désirable ou indésirable sur la base de connaissances culturelles et/ou pragmatiques communes à l'émetteur et aux lecteurs.

Les exemples (I.3) et (I.4) ci-après, où l'auteur utilise du langage figuratif pour exprimer son opinion, illustrent aussi la difficulté de la tâche d'analyse de polarité. En effet, ces derniers expriment des opinions négatives bien que les auteurs utilisent des mots d'opinion positifs (adorer, merci, magnifique) :

(I.3) J'adore la façon dont votre produit tombe en panne dès que j'en ai besoin.

1. www.senseval.org/.

(I.4) Merci une fois de plus la SNCF. Ça annonce une magnifique journée ça encore.

Parfois, les opinions implicites peuvent s'exprimer ironiquement, ce qui complique davantage l'analyse de polarité. Dans le tweet (I.5), extrait du corpus FrIC (Karoui 2016), l'utilisateur emploie une fausse assertion (texte souligné) qui rend de ce fait le message très négatif envers Valls. On remarquera ici le recours au hashtag #ironie qui permet d'aider le lecteur à comprendre que le message est ironique :

(I.5) #Valls a appris la mise sur écoute de #Sarkozy en lisant le journal. Heureusement qu'il n'est pas ministre de l'Intérieur #ironie

Il est important de noter que bien que l'extraction des opinions dans ces exemples soit d'une simplicité presque enfantine pour un humain, son extraction automatique est extrêmement complexe pour un programme informatique. En effet, au-delà de la détermination d'expressions subjectives du langage, le problème de la distinction entre opinions explicites/implicites ou encore l'identification de l'usage du langage figuratif est encore non résolu du fait de l'incapacité des systèmes actuels à appréhender le contexte dans lequel les opinions sont émises.

Dans cet ouvrage, nous nous proposons de travailler sur la détection automatique du langage figuratif, un phénomène linguistique extrêmement présent dans les messages postés sur les réseaux sociaux. Depuis quelques années, la détection de ce phénomène est devenu un sujet de recherche extrêmement actif en TAL, principalement en raison de son importance pour améliorer les performances des systèmes d'analyse d'opinions, (Maynard et Greenwood 2014, Ghosh *et al*. 2015).

I.2. Vers la détection du langage figuratif

Contrairement au langage littéral, le langage figuratif détourne le sens propre pour lui conférer un sens dit figuré ou imagé, comme la métaphore, l'ironie, le sarcasme, la satire et l'humour. L'ironie est un phénomène complexe largement étudié en philosophie et en linguistique (Grice *et al*. 1975, Sperber et Wilson 1981, Utsumi 1996). Globalement, l'ironie est définie comme une figure de rhétorique par laquelle on dit le contraire de ce que l'on veut faire comprendre (voir exemples (I.3) et (I.4)). En linguistique computationnelle, l'ironie est un terme générique employé pour désigner un ensemble de phénomènes figuratifs incluant le sarcasme, même si ce dernier s'exprime avec plus d'aigreur et d'agressivité (Clift 1999).

Chaque type de langage figuratif a ses propres mécanismes linguistiques qui permettent de comprendre le sens figuré. L'inversion de la réalité/vérité pour exprimer

l'ironie (Grice *et al*. 1975), la présence des effets amusants pour exprimer l'humour (Van de Gejuchte 1993, Nadaud et Zagaroli 2008), etc. Dans la plupart des cas, l'ensemble des phénomènes figuratifs nécessite le recours au contexte de l'énonciation afin que le lecteur ou l'interlocuteur réussisse à interpréter le sens figuré d'un énoncé donné. Par conséquent, il est important de pouvoir inférer des informations au-delà des aspects lexicaux, syntaxiques voire même, sémantiques d'un texte. Ces inférences peuvent varier selon le profil du locuteur (comme le genre) ou encore son contexte culturel.

La majorité des travaux en détection de l'ironie en TAL concerne des corpus de tweets car les auteurs peuvent explicitement indiquer le caractère ironique de leurs messages en employant des hashtags spécifiques, comme #sarcasme, #ironie, #humour. Ces hashtags sont alors utilisés pour collecter un corpus annoté manuellement, ressource indispensable pour la classification supervisée de tweets comme ironiques ou non ironiques. Les travaux de l'état de l'art concernent majoritairement des tweets en anglais, mais des travaux existent également pour la détection de l'ironie et/ou du sarcasme pour l'italien, le chinois ou encore le néerlandais (Farias *et al*. 2015, Jie Tang et Chen 2014, Liebrecht *et al*. 2013).

Globalement, les approches qui ont été proposées reposent presque exclusivement sur l'exploitation du contenu linguistique du tweet. Deux principales familles d'indices ont été utilisées :

– indices lexicaux (n-grammes, nombre de mots, présence de mots d'opinion ou d'expressions d'émotions) et/ou stylistiques (présence d'émoticônes, d'interjections, de citations, usage de l'argot, répétition de mots) (Kreuz et Caucci 2007, Burfoot et Baldwin 2009, Tsur *et al*. 2010, Gonzalez-Ibanez *et al*. 2011, Gianti *et al*. 2012, Liebrecht *et al*. 2013, Reyes *et al*. 2013, Barbieri et Saggion 2014b) ;

– indices pragmatiques afin de capturer le contexte nécessaire pour inférer l'ironie. Ces indices sont cependant extraits du contenu linguistique du message, comme le changement brusque dans les temps des verbes, l'usage de mots sémantiquement éloignés, ou encore l'utilisation de mots fréquents *versus* mots rares (Burfoot et Baldwin 2009, Reyes *et al*. 2013, Barbieri et Saggion 2014b).

Ces approches ont obtenu des résultats encourageants[2]. Nous pensons cependant que ce type d'approche, bien qu'indispensable, n'est qu'une première étape et qu'il est primordial d'aller plus loin en proposant des approches plus pragmatiques, qui permettent d'inférer le contexte extra-linguistique nécessaire à la compréhension de ce phénomène complexe.

2. Par exemple, Reyes *et al*. (2013) ont une précision de 79 % pour des tweets anglais. Voir chapitre 2 pour un état de l'art détaillé et résultats des approches existantes.

I.3. Contributions

Dans ce cadre, nous nous focalisons pour la première fois sur des tweets en français et proposons une approche par apprentissage supervisé afin de prédire si un tweet est ironique ou pas. Nos contributions peuvent être résumées en trois principaux points :

1) un modèle conceptuel permettant d'appréhender les phénomènes pragmatiques mis en œuvre pour exprimer l'ironie dans les messages postés sur Twitter. En nous inspirant des travaux en linguistique sur l'ironie, nous proposons le premier schéma d'annotation multiniveau pour l'ironie. Ce schéma, publié dans l'atelier ColTal@TALN2016, a été exploité dans le cadre d'une campagne d'annotation d'un corpus formé de 2 000 tweets français (Karoui 2016). Une version étendue de ce corpus a été utilisée comme données d'entraînement dans le cadre de la première campagne d'évaluation sur l'analyse d'opinion et le langage figuratif DEFT@TALN 2017[3] (Benamara *et al*. 2017). Le schéma d'annotation ainsi que les résultats quantitatifs et qualitatifs de la campagne d'annotation sont décrits dans le chapitre 3 ;

2) un modèle computationnel permettant d'inférer le contexte pragmatique nécessaire à la détection de l'ironie. En exploitant l'ensemble des observations faites sur le corpus annoté, nous avons développé un modèle de détection automatique de l'ironie dans les tweets en français qui exploite à la fois le contexte interne du tweet à travers des traits lexicaux et sémantiques et le contexte externe, en recherchant des informations disponibles dans des ressources externes fiables. Notre modèle permet, en particulier, de détecter l'ironie qui se manifeste par des fausses assertions (voir exemple (I.5)). Ce modèle, qui a été publié à TALN 2015 (Karoui *et al*. 2015) et ACL 2015 est présenté dans le chapitre 4 ;

3) étude de la portabilité à la fois du modèle conceptuel et computationnel pour la détection de l'ironie dans un cadre multilingue. Nous avons d'abord testé la portabilité de notre schéma d'annotation sur des tweets en italien et en anglais, deux langues indo-européennes culturellement proches du français. Nos résultats, publiés à EACL 2017, montrent que notre schéma s'applique parfaitement sur ces langues (Karoui *et al*. 2017). Nous avons ensuite testé la portabilité de notre modèle computationnel pour la langue arabe où les tweets sont à la fois écrits en arabe standard et en arabe dialectal. Nos résultats montrent que notre modèle, là encore, se comporte bien face à une famille de langue différente. La portabilité de nos modèles est discutée au chapitre 5.

Avant de détailler nos contributions, nous commençons par présenter dans les deux premiers chapitres de cet ouvrage un état de l'art complet sur les approches linguistiques et computationnelles de détection de l'ironie. En fin d'ouvrage, une conclusion synthétise les résultats obtenus et ouvre des pistes de recherches futures.

3. https//deft.limsi.fr/2017/.