

Description de quelques procédés linguistiques de l'ironie, par le biais des tweets sur les transports en commun en français et en polonais

Linguistic description of irony in French and Polish corpora
of tweets about public transport

Aude Grezka

CNRS / Laboratoire LIPN-Université Paris 13, France

grezka@lipn.univ-paris13.fr

Małgorzata Niziołek

Université Pédagogique de Cracovie, Pologne

mniziolek1@gmail.com

Davide Buscaldi

CNRS / Laboratoire LIPN-Université Paris 13, France

buscaldi@lipn.univ-paris13.fr

Abstract

Irony is a complex phenomenon. Even though the numerous theories of irony differ from one another, they all seem to agree on the fact that irony implies certain contradiction / discrepancy / inconsistency between what is said and the reality. The purpose of this paper is to describe the language criteria of irony. The studies are contrastive character (French and Polish corpora of tweets about public transport). In the first phase of the study based on contrastive linguistic analysis, we will answer the following questions: What are the language criteria of irony? Are they identical for both languages? The research will help to expand the knowledge of automatic processing of natural language and, more precisely, in the analysis of text data and searching in texts.

Keywords: irony, French and Polish contrastive studies, tweets, automatic detection, corpus

1. INTRODUCTION

Ces dernières années, l'extraction d'opinion dans les textes s'est beaucoup développée et surtout depuis l'expansion du web social (Facebook, Twitter, etc.) qui permet aux internautes de manifester leurs opinions. L'ironie en est reine. Kerbrat-Orecchioni souligne que « l'ironie est de tous les tropes celui qui nage le plus volontiers dans les eaux troubles de l'ambiguïté. [...] Le sens dérivé dans l'ironie ôte toute pertinence au sens littéral : le principal intérêt de ce trope réside dans le brouillage sémantique et l'incertitude interprétative qu'il institue » (1986, p. 105). L'ironie joue donc sur l'implicite : le destinataire doit savoir lire entre les lignes pour comprendre que le locuteur pense le contraire de ce qu'il dit. La complexité de ce trope constitue un vrai défi pour la détection automatique. L'extraction automatique de l'ironie demande de délimiter et de définir linguistiquement ce trope.

En effet, certaines propriétés des langues naturelles constituent des obstacles au traitement automatique (inférence, figement, polysémie, etc.) (Grezka, 2006, 2009 ; Gross, 2015). Ces obstacles montrent que le langage n'est pas réductible à une machine et qu'une machine n'est pas encore susceptible de régler complètement les problèmes de modélisation du comportement humain, notamment du comportement langagier. Traiter du langage chez l'homme est un vaste projet se trouvant à l'intersection de plusieurs disciplines, sciences sociales, neurosciences, anthropologie, communication, etc. La complexité des langues naturelles, associée à la complexité du langage humain amène inévitablement à une difficulté de modélisation. Le langage figuratif pose ainsi de nombreux problèmes. En effet, ce type de langage attribue au sens propre des énoncés des sens supplémentaires. Le lecteur ou l'interlocuteur doit donc découvrir voire déchiffrer la véritable signification. Le langage figuratif est un déplacement du sens : il superpose au sens propre des énoncés une ou plusieurs couches de sens supplémentaires, phénomène qu'on peut illustrer entre autres par l'ironie, le sarcasme, l'humour, la métaphore ou encore les jeux de mots.

L'objectif de cet article est de décrire quelques procédés linguistiques français / polonais de l'ironie et d'exposer les différents problèmes liés à son extraction. Pour cela, nous nous appuyerons sur les tweets français et polonais qui ont pour sujet les transports en commun. Tout ce travail s'inscrit dans le cadre d'un vaste projet qui a débuté en 2018¹.

2. PROJET

Tout développement scientifique est intégralement lié au développement de moyens de communication, donc par ailleurs au développement de technologies

¹ Projet ayant obtenu le soutien du MEAE et du MESRI (projet financé par Campus France, Pologne, 2018-2019).

linguistiques. Ce projet cherche à avancer dans le domaine du traitement automatique des langues naturelles comme le français et le polonais.

Avec les réseaux sociaux qui permettent aux internautes d'émettre aisément des opinions, l'extraction d'opinion dans les textes s'est multipliée durant ces dernières années. Il existe plusieurs approches pour l'extraction allant de représentations par sac de mots à des modèles plus complexes qui traitent de phénomènes dépendant du contexte ou du niveau discursif. Les systèmes actuels ont des résultats relativement bons sur des tâches de classification du type objective / subjective, mais l'analyse de polarité (positif ou négatif) doit encore être affinée pour pouvoir prendre en compte, notamment, des formes figuratives comme l'ironie.

Sur les réseaux sociaux comme Twitter, la rhétorique est omniprésente. La palme revient à celui qui aura su sortir le trait d'esprit le plus abouti. Bien souvent, l'humour passe par l'ironie ou le sarcasme, cette façon de « railler avec méchanceté » (ironie mordante). L'analyse de tweets, notamment pour des tâches de classification, est un domaine en forte expansion que ce soit en linguistique, en Analyse de Données Textuelles, en Fouille de Textes ou en Traitement Automatique des Langues. Dans le cadre de ce projet franco-polonais, nous avons trois objectifs :

- Délimiter et définir linguistiquement les critères de l'ironie en français et en polonais : en proposant des ressources multilingues (dictionnaires d'intensité, d'opposition, etc.).
- Etiqueter des corpus français et polonais de tweets (comme ironique / non ironique ; positif, négatif ou neutre) (i) pour constituer une base d'apprentissage bilingue afin de pouvoir classifier automatiquement de nouveaux tweets ; (ii) pour proposer un corpus de tweets étiquetés à la communauté linguistique et informatique.
- Monter un algorithme le plus précis possible pour détecter automatiquement l'ironie.

3. ETAT DE L'ART SUR L'IRONIE

Comme le signale Bres (2011, p. 695) : « L'ironie fait partie, à l'instar de la métaphore, de ces plus vieux objets linguistiques du monde qui stimulent la réflexion sans jamais l'épuiser : depuis Platon, Aristote, Quintilien, l'ironie est un objet de recherche qui traverse les âges... sans prendre une ride ».

Si nous nous basons sur la classification de Schoentjes (2001), l'ironie peut se diviser en quatre axes : ironie socratique, ironie de situation, ironie verbale et ironie romantique. Les travaux sur la seule ironie verbale sont nombreux. Un parcours de la littérature permet de distinguer trois grandes approches, selon lesquelles l'ironie apparaît comme un phénomène : (i) rhétorique (Grice, 1979 ; Kerbrat-Orecchioni, 1978, 1980, 1986) ; (ii) argumentatif (Amossy, 2003 ; Eggs, 2009) ; (iii) énonciatif

(Sperber et Wilson, 1978, 1998 ; Berrendonner, 1981, 2002 ; Ducrot, 1984 ; Clark et Gerrig, 1984 ; Perrin, 1996 ; Recanati, 2000, 2004 ; Wilson, 2006). L'ironie verbale, qui nous intéresse ici, exprime une contradiction entre la pensée du locuteur et son expression : c'est-à-dire un énoncé dans lequel ce qui est dit diffère de ce qui est signifié. L'ironie peut être produite de différentes manières, dont certaines correspondent à des figures de style classiques (antiphrase, hyperbole, litote, etc.).

L'ironie est souvent amalgamée au sarcasme. Le sarcasme est en fait une forme d'ironie verbale, mais qui est intentionnellement insultant. À l'inverse de l'ironie, le sarcasme est plus difficile à déceler. En effet, être sarcastique c'est dire le contraire de ce que l'on pense, sans montrer qu'on pense le contraire de ce que l'on dit. Bien que le sarcasme soit considéré comme une « ironie mordante », il comporte plusieurs prédicats de définition qui ne sont pas présents dans l'ironie, tels que l'insulte, la méchanceté, la cruauté ou l'agression. Nous gardons simplement l'idée que l'ironie est une intention et que le sarcasme se veut l'une de ses verbalisations possibles (Loevenbruck et al., 2013 ; Attardo, 2000 ; Jorgensen, 1996).

Enfin, l'humour constitue avec d'autres pratiques linguistiques l'une des formes les plus complexes que peuvent prendre les interactions verbales. Il est donc source de nombreuses difficultés théoriques. Une simple lecture des dictionnaires et de leurs définitions montre qu'il est difficile de nous en remettre à leurs dénominations : amusant, comique, drôle, ridicule ; dérision, fantaisie, ironie, moquerie, plaisanterie, raillerie, etc.

Dans le cadre de ce projet, nous prenons actuellement l'ironie au sens large, pour couvrir les différentes formes du langage figuratif². Nous nous limitons à l'étude de l'ironie verbale en ne faisant aucune distinction entre ironie, sarcasme et humour étant donné que la frontière linguistique entre ces trois notions est souvent très floue, aussi bien dans la littérature, pourtant très abondante, que dans les dictionnaires (difficulté de classement, définitions approximatives, prolifération des dénominations, renvois synonymiques en boucle, etc.)³. Comme l'explique Charau-deau (2006, p. 20-21) :

On peut se moquer et tourner en ridicule par ironie, dérision, loufoquerie, etc. ; on peut ironiser par dérision, faire de la dérision de façon ironique, railler avec ironie, à moins que ce ne soit ironiser en raillant. Si, en plus, on combine ces termes avec des qualificatifs du genre mordant, ravageur, caustique, cinglant, acerbé, âpre-badin, anodin, léger, bénin, ou si l'on rajoute d'autres dénominations du genre boutade, vacherie, bouffonnerie, etc., on n'est guère éclairé.

² Peut-être serait-il judicieux à terme d'employer le mot « humour » pour désigner une notion générique qui ensuite peut faire l'objet de diverses catégorisations (ironie, sarcasme, etc.) ? Les questions restent ouvertes en l'état d'avancement du projet.

³ Les propositions de Kerbrat-Orrechioni (1978) s'attardent peu sur la question du sarcasme.

[...] En fait, c'est surtout l'ironie qui, dans la tradition rhétorique, a fait l'objet d'une catégorisation, à partir de la définition (minimale) qu'en propose Aristote, lequel s'en tient à la décrire comme une antiphrase qui consiste à dire le contraire de ce que l'on pense. Dans cette filiation, on voit se confronter divers points de vue qui ne clarifient guère les choses. Dans le *Dictionnaire de poétique et de rhétorique* d'Henri Morier (1981), ironie et humour sont présentés comme des catégories distinctes. La première s'opposerait à la seconde en ce qu'elle joue plus particulièrement sur l'antiphrase, alors que l'humour jouerait sur des oppositions qui ne seraient pas antiphrastiques ; de plus, l'ironie enclencherait le rire, alors que l'humour n'enclencherait que le sourire. Pour d'autres, au contraire, humour et ironie sont confondus ou du moins enchâssés l'un dans l'autre. C'est le cas de Robert Escarpit (1987, p. 115) qui met le paradoxe ironique au cœur même de tout processus humoristique « par la mise en contact soudaine du monde quotidien avec un monde délibérément réduit à l'absurde ». Ainsi l'ironie serait-elle destructrice, alors que l'humour serait ce qui la « corrige [...] par un clin d'œil complice ». [...] Enfin, souvent, ironie et raillerie sont mis dans le même panier, à commencer par César Chesneau Dumarsais et Pierre Fontanier (1967) qui écrivent que « l'ironie consiste à dire par manière de raillerie, tout le contraire de ce qu'on pense ou de ce que l'on veut faire penser aux autres ».

4. LE CORPUS

Nous avons construit notre corpus à partir des *posts* ou *tweets* déposés sur le réseau Twitter (www.twitter.fr). Twitter est un réseau informationnel, à mi-chemin entre le réseau social et le réseau professionnel, qui met des utilisateurs en relation grâce aux informations qu'ils postent eux-mêmes, les « tweets »⁴. Chaque compte Twitter permet de poster, de suivre des informations ainsi que d'en relayer. Ces informations peuvent être aussi bien des articles liés à l'actualité, aux centres d'intérêt, que des billets d'humeur. Twitter fonctionne donc comme un condensateur ou un accélérateur d'informations puisqu'il permet, très rapidement, de relayer une information et de la suivre. Ces posts sont devenus un outil de communication incontournable et redoutable, ...voire un défouloir.

Les domaines de discussion sont variés. Dans la première phase de notre travail, qui est la constitution d'un corpus d'analyse, nous avons choisi de restreindre notre corpus en ne prenant que les tweets ayant pour thème Poutine (l'objectif étant de trouver une thématique productive dans les deux langues). Mais les résultats n'ont pas été concluants. L'ironie y était pratiquement absente. Les internautes qui commentent sont soit favorables aux actions de Poutine, soit expriment leur mécontentement. Il est rare (en tout cas dans notre corpus) d'y trouver de la moquerie. Les échanges entre internautes sont plutôt sérieux, aussi bien en France qu'en Pologne.

⁴ Un tweet doit comporter au maximum 140 caractères, lien Internet compris.

Comme le corpus recueilli ne nous a pas fourni assez d'exemples pour analyser l'usage de l'ironie dans les deux langues, nous avons donc décidé de travailler sur les tweets en rapport avec les transports en commun.

Pour le corpus français, nous avons utilisé celui du DEFT 2018 (<https://deft.limsi.fr/2018/>) sur la recherche d'information et l'analyse de sentiments dans les tweets sur les transports en Île-de-France⁵. Ce corpus est « contient 76.732 tweets sélectionnés parmi 80.000 tweets annotés manuellement. Les messages sont issus d'une sélection à base de mots-clés et de mesure d'entropie pour filtrer les doublons et les messages dépourvus de texte intelligible » (Paroubek et al., 2018, p. 220).

Pour la construction du corpus de tweets liés aux transports en Pologne, nous avons interrogé le moteur de recherche de Twitter avec les requêtes suivantes (nous avons mis leur interprétation en français) :

- Polskie Koleje Państwowe (PKP – chemins de fer polonais),
- Polskie Linie Kolejowe (société responsable de la maintenance des voies ferrées),
- PKP Intercity (service intercités de PKP),
- PKP Przewozy Regionalne (réseau de transport régional),
- TLK (un autre gestionnaire de transports sur voies ferrées),
- ZIKiT Kraków (transports publics de Cracovie),
- opóźnienie (retard).

Nous avons choisi de retenir les résultats en polonais sur la période comprise entre le 1^{er} septembre 2017 et le 30 juin 2018 car nous avons considéré la période hivernale comme celle la plus critique pour le système des transports publics polonais (plus d'utilisateurs, plus de risques météorologiques, etc.). Les résultats ont été concluants puisque nous avons obtenu 9.465 tweets sur la période prise en compte. Pour chaque requête, nous avons obtenu le nombre de tweets suivants :

Requête	Nombre de Tweets
Polskie Koleje Państwowe	104
Polskie Linie Kolejowe	250
PKP Intercity	133
PKP Przewozy Regionalne	77
TLK	1745
ZIKiT Kraków	2048
opóźnienie	5359

⁵ Créé en 2005, le Défi Fouille de Textes (DEFT) est une campagne d'évaluation francophone qui propose annuellement de confronter les méthodes de plusieurs équipes de recherche sur une thématique régulièrement renouvelée dans le domaine de la fouille de textes. La treizième édition DEFT2018 a porté sur la recherche d'information et l'analyse de sentiments dans des tweets rédigés en français, sur la thématique : les transports en Île-de-France. La campagne a utilisé un corpus annoté manuellement produit dans le cadre du projet REQUEST (Programme d'Investissement d'Avenir, appel Cloud computing & Big Data, convention 018062-25005) en collaboration avec ELDA (<http://www.elda.org>).

Dont 250 tweets qui sont apparus dans les résultats d'au moins deux recherches différentes.

Que ce soit pour le polonais ou le français, la thématique « transport en commun » est prolifique. Tantôt lieu de défouloir, tantôt espace d'entraide, Twitter est surtout l'occasion pour les usagers de s'exprimer. Les voyageurs n'hésitent pas en effet à indiquer en temps réel les retards. Nombre d'entre eux en profitent également pour demander directement leurs renseignements aux comptes officiels (SNCF, RATP, PKP, TLK, etc.). Mais les usagers ne sont pas les seuls à s'exprimer. Chef de bord et autres personnels y prennent part également. L'occasion pour eux d'expliquer, parfois, les raisons de tel ou tel retard, mais aussi et surtout de raconter leur quotidien.

5. CRITÈRES LINGUISTIQUES

Comme nous venons de le voir, l'ironie est un phénomène linguistique complexe largement étudié. Même si les théories diffèrent sur la définition, elles s'accordent sur le fait que l'ironie implique une incohérence entre ce qui est dit et la réalité. Ainsi, dans des tweets ironiques, nous retrouvons souvent une opposition d'au moins deux propositions, que nous nommerons P1 et P2. Elles peuvent être dans le même énoncé (opposition explicite), ou bien l'une est présente et l'autre est implicite (opposition implicite) (Karouai et al., 2015).

L'opposition *explicite* peut impliquer une contradiction entre des mots de P1 et des mots de P2 qui ont des polarités opposées :

- (1) J'aime quand **il n'y a plus de train** et que tu dois aller au boulot. **Merci** la SNCF !
- (2) J'aime quand les écrans de la ratp qui indiquent le temps **ne fonctionne pas**.
- (3) Pozdrowiałam dzisiaj @ZIKiTKrakow? A **gratulowałam** wczorajszego **paraliżu** miasta? pic.twitter.com/YyjdgkEik8.
[Est-ce que j'ai déjà salué aujourd'hui ZIKIT ? Est-ce que je vous ai **félicité** pour la **paralysie** de la ville ?]
- (4) Powrót z Warszawy **nowoczesnym** pociągiem. Najpierw **opóźnienie** przez wykolejenie pod Oleśnicą, a teraz **awaria** składu.
[Le retour à Varsovie dans un train **moderne**. D'abord le **retard** à cause d'un déraillement et maintenant la **panne** du train.]
- (5) Mój pociąg do Poznania jest **opóźniony** 20 minut i jego opóźnienie może się zwiększyć. **Dobrze się zaczyna dzień**.
[Mon train est **retard** de 20 min et son retard peut augmenter. **La journée commence / s'annonce bien**.]

L'opposition *implicite* quant à elle, se produit quand il y a une opposition entre une proposition P1 décrivant un événement ou un état et un contexte pragmatique externe à l'énoncé qui souvent nie P1 ou son existence. Cette opposition est clairement illustrée par ce billet humoristique posté par une internaute :

Pourquoi j'aime le bus, le RER, le métro... et l'ironie...

7 raisons véridiques.

- 1) Ils sont toujours à l'heure. C'est un pur bonheur. D'ailleurs, il suffit d'attendre à peine une ou deux minutes pour en voir un arriver. Avec ce type de transport, c'est sûr, on est toujours à l'heure.
- 2) Les lieux sont très propres, particulièrement les couloirs du métro et du rer. Plus encore, ces transports grouillent d'originalité et de couleurs : du rose, du bleu, du vert. Il n'y a jamais de gris.
- 3) Ils sentent bon. On se croit à la campagne, ou bien à la plage. Pas d'odeur de sueur, de saleté ou de pisse, non, plutôt des parfums envoûtants, différents à chaque coin de tunnel. Un vrai bonheur, entre la parfumerie et la boulangerie.
- 4) Le métro aime le RER et le RER apprécie le bus : la preuve, les changements se font toujours sans jamais rater le suivant. Ils sont très proches et conservent une logique à toute épreuve.
- 5) Les gens sont heureux dans le métro, le bus et le rer. Ils sourient, ils rient, ils discutent entre eux et n'ont pas peur de se regarder. Quelle ambiance !
- 6) On est libre dans le métro. Pas de publicité. Un vrai sas de liberté.
- 7) Pas de raison de se sentir oppressé : on n'est jamais serré⁶.

Certains tweets que nous avons classés comme ironiques ne présentent pas de signes explicites ou repérables. L'ironie naît du contraste qui existe entre ce qui est dit ou écrit et le contexte, la situation ou encore ce que l'on sait ou ce qu'on imagine. Comme l'ironie joue sur l'implicite, le destinataire doit savoir lire entre les lignes pour comprendre que le locuteur pense le contraire de ce qu'il dit. En général, une critique sarcastique est jugée moins négative qu'une critique littérale. Inversement, encenser ironiquement sera jugé moins positif qu'un compliment direct.

L'analyse de notre corpus a permis d'extraire un certain nombre de critères linguistiques, stylistiques, syntactico-sémantique, etc. pour détecter l'ironie⁷. Ce qui est intéressant, c'est que nous retrouvons les mêmes marqueurs dans les tweets français et polonais ; la prosodie semble identique.

5.1. Les hashtags #ironie, #sarcasme, #humour, #sarkazm, #ironia, #humor, #zart

Nés sur Twitter, les hashtags « # » sont devenus un des éléments célèbres de la culture technologique d'aujourd'hui et dans le monde entier. Le terme « hashtag »

⁶ <http://unesemaineunchapitre.com/index.php?post/2008/06/12/Pourquoi-jaime-le-bus-le-rer-le-metro-et-lironie>.

⁷ Un premier travail avait déjà été réalisé dans le cadre d'un défi de fouille de texte (DEFT, 2017 ; TALN / RECITAL, 2017) sur un corpus imposé de tweets (Karoui et al., 2017 ; Buscaldi et al., 2017).

a d'ailleurs été ajouté en 2015 au dictionnaire Le Petit Robert⁸. Ce symbole, suivi d'un mot ou d'une expression, signifie qu'un sujet est fortement commenté par de nombreux utilisateurs. L'objectif de son utilisation est diverse : faciliter la lecture, faire de la veille, lancer une opération de communication, etc. Enfin, situé en fin de phrase, il permet le plus souvent de donner une connotation différente à la phrase : faire de l'humour, de la dérision, en un nombre de caractères qui reste limité. Dans le cadre de l'ironie, les utilisateurs ont tendance à utiliser des hashtags spécifiques (#ironie, #ironique, #sarcasme, #humour, etc. ; #sarkazm, #ironia, #żart, etc.) pour aider les lecteurs à comprendre que leur message est ironique⁹ :

- (6) Le TGV français est le train qui arrive le plus vite en retard. #SNCF #**Ironie**
- (7) Super mon train a été supprimé. C'est trop cool ça. #**Ironique** #Sncf
pic.twitter.com/FpHv1SIQm5
- (8) SNCF je t'aime ! #**sarcasme**
- (9) Une belle journée en perspective. #**ironie** #**sarcasme** #sncf #joiedevivre #periodedessai #mercilasncf
- (10) To na pewno nie ma nic wspólnego ze stanem infrastruktury, który, jak wiadomo, jest wzorowy. #**sarkazm**
[Ça n'a sûrement rien à voir avec l'état de l'infrastructure qui est exemplaire. #**sarcasme**]

D'autres types de hashtags ont également été relevés dans notre corpus : « #JDCJDR » pour *Je dis ça, je dis rien* (hashtag ironique pour faire une remarque à quelqu'un) ; « #Lesgens », hashtag ironique également, qui vise à critiquer un groupe de personne :

- (11) Les gens qui racontent leur vie sur Twitter. #**Lesgens**
- (12) #**lesgens** qui préfèrent faire 3 changements et se rallonger de 2h30 juste pour pas changer à Paris... #SNCF
- (13) Quand ce n'est pas les grèves, ce sont les avaries qui bloquent le trafic de la SNCM¹⁰. On devrait envisager une fusion avec la SNCF. #**jdcjdr**

⁸ Hashtag : mot-clé précédé du signe dièse permettant de retrouver tous les messages d'un microblog qui le contiennent.

⁹ Dans certains tweets de notre corpus, le hashtag peut ne pas apparaître, mais ce cas n'est pas le plus courant :

A z okazji piąteczku – opóźnienie Kolei Śląskich, mróz i ja w sukience. Dzień dobry, **Ironio**.
[A l'occasion du vendredi – le retard de Kolei Śląskich, du gel et moi en robe. Bonjour, **Ironie**.]

IRONIE ON : Y'a pas à dire. Les (moutons de) Français formatés par les médias et les journalistes connaissent mieux les sujets afférents aux #cheminots et à la #SNCF que nous même. **IRONIE OFF**.

¹⁰ La Société nationale Corse-Méditerranée (SNCM) est une compagnie française de navigation maritime.

Dans le cadre de notre travail, ces hashtags jouent le rôle d'étiquette de référence pour la détection de l'ironie. Mais comme le signale Karoui (2016), un grand nombre de travaux exploitent ces hashtags sans nécessairement recourir à d'autres critères. Pourtant l'utilisateur, dans la plupart des cas, ne fournit pas ces hashtags aux lecteurs. Les procédés linguistiques de l'ironie sont beaucoup plus subtiles.

5.2. Les émoticônes

Les tweets ont un défaut majeur puisqu'en l'absence d'intonation et d'indices corporels, ils sont un faible vecteur d'émotions. Il peut parfois être difficile d'interpréter les réelles intentions d'un message. L'usage d'une émoticône¹¹ tend à modifier la perception de l'ironie par le récepteur. Le fait de rajouter à son tweet un « clin d'œil » permet de repérer correctement l'ironie mais également d'amplifier la portée ironique du message, qu'il soit positif ou négatif (Filik et al., 2016). Si l'émoticône accompagne un énoncé dont le sens exprime le contraire de la signification attribuée à l'émoticône, il y a de fortes chances d'y trouver de l'ironie. Les émoticônes deviennent le nouveau langage de l'émotion. Dans les exemples suivants, les émoticônes visent la proposition de l'énoncé qui précède pour en modifier l'interprétation littérale :

- (14) On **aime** les trains que la RATP met sur les rames et au bout de 4 stations fonctionne plus. 😞😞
- (15) Jestem na dworcu w Pasewalk, wschodnie Niemcy. Jeżdżę tą trasą dość często i za każdym razem jest przynajmniej 10-minutowe opóźnienie. Widać, to jakaś **zaraza**, która przedostała się za granicę. ;) [Je suis à la gare de Pasewalk, Allemagne de l'Est. Je fais ce trajet assez souvent et à chaque fois il y a au moins 10 minutes de retard. Ça doit être un fléau qui s'est propagé à l'étranger. ;)]
- (16) Wah **j'adore** la sncf. 🙄

Deux caractéristiques principales des émoticônes peuvent poser problème au repérage automatique : (i) les émoticônes ont de nombreuses formes, même si la banque emoji tend à être employée de plus en plus et à standardiser les usages ; (ii) la question de leur sens se pose. Une émoticône de sourire n'est pas nécessairement l'indice de la joie du locuteur : elle peut être ironique. Pour parvenir à des

¹¹ Une émoticône est représentée par une suite de caractères typographiques bien précise. A l'origine, il s'agit d'un visage humain souriant :-) ou triste :- (qu'on voit en penchant la tête de 90° vers la gauche (les deux points figurent les yeux, le tiret représente un nez et la parenthèse fermante une bouche souriante). Avec le temps, les émoticônes se sont multipliées. L'appellation d'origine vient du mot anglais « smiley ».

résultats significatifs, il faut modéliser correctement le fonctionnement des émoticônes et le sens qu'elles ont selon les situations discursives. Les émoticônes sont donc de nouveaux modalisateurs comme le sont les interjections ou les adverbes d'énonciation, c'est-à-dire qu'ils reflètent la position de l'énonciateur par rapport à ce qu'il dit.

5.3. La ponctuation

La ponctuation est un indice dans le repérage de tweets ironiques. Parmi les signes les plus usités, on retrouve les guillemets. Son principal usage est de mettre en relief une expression, un terme ou une citation. Comme on peut le voir à travers les tweets, les guillemets sont parfois utilisés pour indiquer que le terme ou l'expression mis en exergue n'a pas sa signification littérale ou habituelle. On parlera alors de guillemets d'ironie. Autrement dit, la mise entre guillemets peut représenter, selon le cas, un type d'ironie. Maingueneau précise d'ailleurs que « L'énonciateur qui use de guillemets, consciemment ou non, doit se construire une certaine représentation de ses lecteurs pour anticiper leurs capacités de déchiffrement : il placera des guillemets là où il présume qu'on en attend de lui (ou qu'on n'en attend pas s'il veut créer un choc, surprendre). Réciproquement, le lecteur doit construire une certaine représentation de l'univers idéologique de l'énonciateur pour réussir le déchiffrement » (2012, p. 140) :

- (17) @MPK_Krakow, @ZIKiTKrakow tradycyjnie zorganizowały „dzień pieszego pasażera” – Lipska, brak zasilania w sieci.
[MPK a organisé traditionnellement « **la journée du passager à pied** » – Lipska, pas d'alimentation électrique sur le réseau.]
- (18) C'est la « **bonne** » réponse et pourtant c'est vraiment inouï (ironie inside) une multiplicité de tarifs, des cartes qui ne fonctionnent pas. Et on veut favoriser le déplacement en transports en commun ?
- (19) Le départ « **imminent** » a été annoncé ya 6 min mais ça correspond à quoi en langage sncf ? 10 min.

5.4. L'intensité

Le locuteur qui recourt à l'intensité cherche avant tout à augmenter l'efficacité de son message : il s'agit pour lui soit de le renforcer, soit de l'atténuer. Notre objectif dans cet article n'est pas de décrire tous les moyens qui permettent d'augmenter l'expressivité d'un message mais de décrire quelques procédés, les plus rencontrés, présents dans notre corpus¹².

¹² Pour plus d'informations sur l'intensité voir Romero (2017), Anscombe et Tamba (2013), Kleiber et Schnedecker (2007), Romero et Niziołek (2016), Wróblewska-Pawlak et Kieliszczyk (2015).

Pour faire de l'ironie, les adverbes d'intensité comme *très*, *trop* ou *plus* sont très productifs en français (un peu moins pour le polonais). Ils sont souvent employés pour produire de nouveaux énoncés figuratifs. En ce sens, on comprend que l'ironie puisse être classée comme figure de mots car ce sont les mots employés pour leurs antonymes, qui produiront l'ironie :

- (20) **Trop bien** le renouvellement navigo imagine R de la @SNCF ! Je saisis toutes mes infos, je met mon certificat de scolarité, tout se déroule impec jusqu'au paiement. Impossible d'accéder à la signature électronique / paiement en ligne. On me demande d'appeler l'agence...
- (21) To i tak jest nic – wczoraj TLK #Ogiński jechał jeszcze **bardziej** punktualnie. pic.twitter.com/gdXbTv00RF
[Ce n'est encore rien – hier TLK était **encore plus** à l'heure.]

Cependant, pour que le figuratif soit perçu à ce niveau, il faut que le tweet ne corresponde pas à l'observation du réel, c'est-à-dire révèle une contradiction par rapport au réel. Car, il va sans dire, que les adjectifs mélioratifs peuvent être utilisés en leurs sens propres, référentiels et dénotatifs. Dans l'exemple ci-dessous, l'étudiant est-il content de rater son examen ? :

- (22) Jv être en rtard pr mon exam cause dla sncf ahahah **trop bien**.

Pour savoir s'ils sont employés dans leurs sens réels ou pas, c'est la dimension pragmatique qui va jouer un rôle important (voir la partie 6.4).

Autre forme intéressante que nous retrouvons régulièrement dans notre corpus français pour marquer l'intensité, l'adjectif / adverbe « grave ». L'adjectif *grave* prend parfois des tournures d'adverbe. Il est sémantiquement proche d'*extrêmement* ou d'*énormément*, il dit l'intensité. L'adjectif *grave* peut intensifier un prédicat nominal, adjectif ou substantif :

- (23) Whouaa les bus de la sncf il y'a **grave** de la place pour les jambes. #ironie

Il est à noter qu'un groupe important de marqueurs linguistiques intensifs de l'ironie appartient à la catégorie des exclamatives (conjonctions ou locutions exclamatives). Ce sont des constructions bien spécifiques au français que l'on retrouve habituellement à l'oral : *que*, *comme qu'est-ce que*, *comment*, *combien* ; etc. Ils vont apparaître en tête de phrase et porter sur le verbe :

- (24) Ah la la... et dire **que** je suis heureux quand mon train arrive « seulement » avec 10 min de retard, sans excuses du conducteur, évidemment... 😂😂 #SNCF #Privatisation

On relève également dans notre corpus une construction propre à la langue parlée, dans laquelle à la place de *que* ou de *comme*, on utilise fréquemment la construc-

tion *qu'est-ce que* ou sa variante *ce que*. De même, à la place de *que de*, on emploie couramment *qu'est-ce que... comme* ou *ce que... comme*¹³ :

- (25) Grève des RER et TER à Noël et au Nouvel An. **Qu'est-ce qu'on** aime la SNCF et la RATP.
- (26) Cet hiver, on aura pas vu la neige... Par contre, **qu'est-ce qu'on** aura vu **comme** grèves ! Normal ?

Le marqueur *combien* s'utilise, quant à lui, de la même manière que *que*, mais il insiste davantage sur la quantité concrète¹⁴. Alors que *que* exclamatif est proche de *très* ou *beaucoup*, *combien* est plus proche de *à quel point*, *en grande quantité* :

- (27) Pas de prise en 2nd, elles ne fonctionnent pas en 1e. Passer son trajet ds des chiottes puant la pisse pour te dire **combien** je t'❤️ @SNCF 2/2.

La locution *mais c'est que* en tête de phrase peut également avoir une valeur exclamative. Cette tournure exprime généralement l'admiration / le dégoût ou la surprise, et est aussi assez souvent employée de façon ironique :

- (28) **Mais c'est que** vous avez de l'humour ! (Ironie). Les policierEs. Les agentEs ratp et **snf**. Pauvre homme obliger à respecter autrui... Vraiment c'est pô juste la vie.
- (29) **Mais c'est que** la @SNCF serait (presque) à l'heure ! 🙌😭

Enfin, parmi les moyens d'intensité, nous pouvons classer les interjections. Cette catégorie rassemble des formes très hétéroclites, allant des onomatopées aux substantifs. Dans notre corpus, les interjections repérées apportent de l'intensité à l'énoncé¹⁵. Au plan discursif, l'interjection est fréquemment rattachée à un prédicat qu'elle vient intensifier :

- (30) TLK dali i nawet znalazłam miejsce, **szok** !
[Et j'ai même trouvé une place, tiens ! (grande stupéfaction !)]
- (31) @snf on bloque le rer b pour un telephone vole ? **Bravo !!!**

Une des interjections les plus rencontrées dans nos tweets est le substantif *merci*, suivi ou non d'un point d'exclamation. Cette formule de politesse utilisée habituellement pour témoigner sa gratitude ou pour signifier un refus, poliment est le plus souvent utilisée ironiquement par les internautes :

¹³ À la forme négative, la construction est souvent disloquée et donc plus difficile à repérer.

¹⁴ Comme *que*, le mot *combien* peut se combiner à *de* pour former un déterminant complexe. Dans ce cas, contrairement à *que de*, qui exprime soit l'intensité soit la quantité, *combien de* exprime unique-ment la quantité, et on préfère généralement utiliser *combien de* pour déterminer un nom pluriel.

¹⁵ Romero souligne que l'intensité ne concerne qu'une partie des interjections et seulement dans certains emplois (2017, p. 99).

- (32) **Merci** la @SNCF et @ouisncf ! 🙄 Je prends rarement le train, mais à chaque fois il y a un problème ! Pas de train pour Troyes avant 17h00 minimum... **MERCI** !
On vous aime, ne changez pas.
- (33) **Dziękuję** @pkp_intercity po raz kolejny za długie opóźnienie.
[De nouveau, Merci @pkp_intercity pour ce long retard.]

5.5. Les mots en lettres capitales

Les lettres capitales¹⁶ sont une autre marque de l'ironie et de l'humour. Elles permettent de mettre un mot en vedette. Elles servent l'expressivité de l'énonciateur et traduisent son intention stylistique, qui est ou non immédiatement perceptible par le lecteur. L'emploi des capitales ajoute à l'écrit une information paralinguistique et revêt une fonction expressive intense (peut-être plus intense que dans un contexte écrit traditionnel). Les capitales peuvent marquer également la prononciation alphabétique. Les mots en lettres capitales peuvent être interprétés comme éléments exprimés en criant, et elles refléteraient alors « l'exacerbation d'une émotion – négative comme un sentiment de colère, ou positive comme la joie » (Tatossian cit. in : Cougnon, 2015, p. 56) :

- (34) Payer pour arriver tout le temps en retard, ON **ADORE**.
- (35) Opóźnienie 10 minut, **SUPER**.
[Le retard de 10 minutes, **SUPER**.]

5.6. Les expressions comparatives

Dans le corpus, nous avons repéré quelques exemples avec les conjonctions de coordination *comme* et *jak*. Mais par rapport aux autres marques linguistiques de l'ironie, la comparaison avec *comme* / *jak* n'est pas la plus représentée. La construction qui est à la base de cette comparaison est *V comme SN*. Dans les phrases repérées, l'effet ironique naît de l'incompatibilité des éléments de la comparaison : du comparant et du comparé. Ces comparaisons véhiculent également l'intensité :

- (36) Wymyśliłem nową obelgę: jesteś solidny **jak** TLK Kociewie, ewentualnie porzekało czekać na coś **jak** pasażer na TLK Kociewie.
[J'ai inventé une nouvelle insulte : t'es fiable **comme** TLK Kociewie, éventuellement un dicton attendre qch **comme** un passager TLK Kociewie.]

¹⁶ Germoni précise la différence entre les lettres majuscules et les capitales : « quoiqu'elles aient la plupart du temps la même graphie, la majuscule et la capitale ne doivent pas être confondues. Si l'entièreté d'un mot peut être capitalisée, seule sa première lettre peut revêtir une majuscule. En outre, tandis que l'usage de la majuscule est presque toujours contraint, celui des capitales relève d'un choix de composition typographique qui traduit bien souvent le désir d'expressivité du scripteur » (2013, p. 35).

- (37) Ceux qui sont bloqués **comme dans *Un jour sans fin***¹⁷ à cause de la SNCF ou de Air France... ;)
- (38) @RERC_SNCF encore la cata ce matin. Vity de 7:30 a Gravigny Balizy supprimé. Celui de 7:46 retardé... a 8:05 toujours rien ? ! C tous les jours... pourquoi ? Heureusement une voisine m'a déposé a Savigny pour que je puisse être transporter **comme du bétail à Paris !!** 😡😞

Dans l'exemple ci-dessous, on relève une séquence figée :

- (39) Bravo @snCF plus de places assises, tous debout serrés **comme des sardines**.

5.7. Les mots / expressions : d'opposition, d'opinion positive (mélioratif, appréciatif, laudatif) et d'opinion négative (dépréciatif, péjoratif)

On constate par rapport aux tweets du corpus que le figuratif est plutôt produit par les adjectifs et les adverbes (ou locutions adverbiales) que par les substantifs et les verbes. Ces derniers sont a priori moins employés de façon ironique. Ainsi, dans des tweets figuratifs, les termes mélioratifs ne sont jamais employés dans leurs sens usuels, mais simplement pour signifier leurs contraires : par exemple, *beau*, *bel*, *belle*, *joli(e)* et *beauté*, pour *laid(e)*, *moche*, *mauvais(e)* et *laideur* ; *bon(ne)* et *bonté* pour *méchant(e)* ou *vilain(e)* et *méchanceté* ; *chanceux(euse)* et *chance* pour *malchanceux(euse)* et *malchance* ; *charmant(e)*, *agréable*, *adorable* pour *désagréable* :

- (40) @MPK_Krakow @ZIKiTKrakow brawo! **Świetna** organizacja kom autobusowej pod cmentarzami. Krótkie autob. Ludzie sobie na głowach stoją.
[@MPK_Krakow @ZIKiTKrakow bravo! Une très **bonne** organisation de la communication devant les cimetières. Des bus courts. Les gens se marchent dessus.]

Pour faire de l'ironie, il est possible d'employer tous les adjectifs mélioratifs – *admirable* / *cudowny*, *agréable* / *przyjemny*, *amusant* / *zabawny*, *beau* / *ładny*, *bon* / *dobry*, *capable* / *zdolny*, *captivant*, *chanceux*, *compétent* / *kompetentny*, *fort* / *silny*, *généreux* / *hojny*, *habile* / *zwinny*, *heureux* / *szczęśliwy*, *intelligent* / *inteligentny*, *qualifié* / *wykwalfikowany*, etc. – dans leurs sens contraires, pour désigner des adjectifs péjoratifs. Les discours ironiques, pour disqualifier quelqu'un ou quelque chose, exploitent massivement les termes mélioratifs. Il faut souligner que l'inversion sémantique (de sens) qui caractérise la plupart des énoncés ironiques ne se produit que dans une seule direction : le positif a une valeur de dépréciatif mais pas le contraire, en d'autres termes, on critique en faisant semblant de louer, mais on ne loue pas en critiquant, en faisant recours aux termes valorisants (Kerbrat, 1978).

¹⁷ *Un jour sans fin* est un film américain.

Ainsi, les qualificatifs *beau, fort, grand*, avec leurs variables en genre et en nombre, sont très productifs, car ils participent de beaucoup d'expressions ironiques et d'antiphrases :

- (41) Bydłowóz zwany TLK z Poznania do Warszawy symuluje kolej w Indiach – **piękna** sprawa @PKPIntercityPDP!
[Une bétailière appelée TLK de Poznań à Varsovie simule le chemin de fer en Inde – une **belle** affaire.]

En ce qui concerne les verbes, on relève à de nombreuses reprises l'utilisation de verbes marquant l'affection, la tendresse, l'amitié ou bien la passion comme *aimer, adorer*, etc... mais employés ironiquement :

- (42) **J'aime** les trains supprimés dès le matin en début de semaine ça fait du bien on est dehors c'est bien surtout qu'à la base l'incident devait s'arrêter à 7h bref merci pour ces bonnes infos et les bonnes galères.
- (43) Donc rien que là j'ai 1h15 à attendre pour avoir mon train qui est en retard tandis que je verrais ma correspondance s'envoler au loin, **j'aime** la SNCF, vraiment. **J'adore** attendre, c'est ma passion première.
- (44) **Kocham** podróżować TLK.
[**J'aime** voyager avec TLK.]

6. DIFFICULTÉS POUR LA DÉTECTION AUTOMATIQUE DE L'IRONIE

Les difficultés d'analyse viennent du support même que sont les tweets puisque les codes d'écriture sont bien différents de ceux que l'on peut trouver dans la littérature traditionnelle (poésie, roman, etc.). L'ironie est linguistiquement différentes dans les tweets et dans les textes littéraires. L'exemple ci-dessous, qui décrit l'état pitoyable du train, illustre parfaitement ce décalage :

- (45) Stukot źle przykręconej deski wrywa mnie z letargu. Pociąg TLK, długi i zimny jak islandzkie noce, rozcina skrywający Polskę śniegu całun. Pocieram zdrętwiałe uda dłońmi, zgrabiały mi jak konary zmarzniętej wierzby. Czy konduktor zapuka? Papieru ostał się jeno listek. #ogorkizm

L'ironie résulte d'une stylisation du message en texte poétique / littéraire. Le registre utilisé, les choix stylistiques mettent en valeur l'absurdité de la description ci-dessous. Le style n'est pas approprié au support. Le décalage produit provoque le sourire du lecteur.

L'analyse automatique du langage figuratif comme l'ironie dans les tweets est donc compliquée. Nous sommes confrontés à de nombreux problèmes d'ordre

sémantique, syntaxique et morphologique. Toutes les subtilités du langage ne peuvent être reconstituées sous forme d'algorithmes pour être reconnues par un système informatique.

6.1. La sémantique

La première difficulté relative à la sémantique est la polysémie des mots, qui peut rendre ambiguë toute analyse du sens et créer des incompréhensions. La sémantique d'un mot et donc sa tonalité est amenée à varier lorsqu'elle est utilisée dans un contexte différent. Ainsi, si certains mots expriment à eux seuls un sentiment positif (*super, bien, sympa, etc.*) ou négatif (*inacceptable, énervé, honteux, etc.*), identifier leur présence dans un tweet ne suffit pas à donner la tonalité de cet énoncé. Prenons l'exemple de l'adjectif *positif*. Cet adjectif qui semble avoir une polarité positive, mélioratif (*cette femme est toujours positive, c'est agréable*)¹⁸ peut dans un tout autre contexte avoir une polarité contraire (*vos résultats d'analyse sont positifs, je suis désolée*)¹⁹. Si n'importe quel locuteur français peut faire la différence entre le sens de ces deux emplois, comment une machine peut-elle prendre mesure de la nuance entre les deux ?

6.2. L'orthographe

De même, les données textuelles sont soumises à des formes orthographiques particulières. Les fautes d'orthographe, fréquentes dans les réseaux sociaux, ne font que compliquer l'analyse automatique d'un texte :

(46) Je fais trop la meuf « han gneugneu la sncf i font des retard » alors que jsuis h24 en retars faut que jarrete la mauvaise foi.

La faute orthographique « retars » au lieu de « retard » revient très régulièrement dans notre corpus. On retrouve le même problème avec les abréviations (*jsuis, jre-garde, jsp, ptdr, psq, etc.*), puisque sur Twitter il faut respecter la limite des 140 caractères :

(47) **Jsuis grv** en retars **psq** on a vu le busil venai pas on a fait les go on prend le train et la il est passer se **fdp**.

(48) Le départ « imminent » a été annoncé **ya** 6 min mais ça correspond à quoi en langage sncf ? 10 min.

¹⁸ Définition : qui repose sur quelque chose, d'assuré, dont la réalité ne peut être mise en doute, par opposition à négatif (Larousse).

¹⁹ Définition : qui montre la présence de l'élément ou de l'effet recherché (Larousse).

- (49) **Ptdr** la snf pardon pour le « leger retard » de 50 minute ahaha j'adore un plaisir.
- (50) Cholera spieszylam się bardzo a **pociung** ma opóźnienie ponad 10 min **thx** a lot koleje śląskie.
[Merde, je me suis dépêchée et le train a un retard de 10 min **thx** a lot koleje śląskie.]

De par cette multiplication des formes orthographiques, la reconnaissance des unités lexicales pour l'analyse de l'ironie n'en est que plus difficile.

6.3. La syntaxe

Au niveau syntaxique, on relève de nombreuses formes syntaxiques hétérogènes, ne répondant pas le plus souvent aux normes grammaticales habituelles. Le langage utilisé par certains twittos est spontané et peut parfois être désordonné. Les mots ne sont pas toujours employés dans leur forme originale. Les internautes n'hésitent pas à modifier la structure des phrases (absence de verbes, phrases incomplètes, etc.) et reproduisent à l'écrit certaines caractéristiques liées à l'oral :

- (51) Je suis dans train qui est senser partire a 17h55 il 18h30 **ya aucun probleme !!!**

L'utilisation de formes négatives, habituellement reconnaissables par l'utilisation conjointe de l'adverbe *ne* et d'une particule *pas*, *aucun*, *plus*... n'est plus forcément un critère. Dans certains emplois, la forme négative peut être réduite à l'utilisation d'une seule particule.

Cette simplification d'emploi par les twittos rend l'analyse d'autant plus difficile puisque les « phrases » ne sont pas toutes construites de la même manière et ne répondent pas toutes aux mêmes règles. Pour analyser n'importe quelle structure de phrase, il faudrait alors prévoir la reconnaissance d'une multitude de formes syntaxiques, ce qui serait trop complexe sachant que les usages de la langue évoluent régulièrement.

6.4. La pragmatique

Ce niveau linguistique implique une connaissance générale du contexte de la situation, et pas seulement du contexte induit par l'énoncé lui-même. Cela englobe souvent des éléments extérieurs au langage, à savoir différentes informations sur les locuteurs, des repères spatio-temporels, etc.

L'interprétation du ton d'un message est intuitif pour un lecteur humain. Sans un contexte explicite ou une compréhension de la culture, le mot *super*, par exemple, sera potentiellement catégorisé comme un sentiment positif par une machine, ce qui serait correct dans un autre contexte mais pas dans notre exemple à propos du retard

- (52) Super la ponctualité des trains !

Dans l'exemple ci-dessous on retrouve deux éléments dont l'identification dans le texte par la machine, devrait nous mener vers l'ironie : *opóźnienie* / *retard* (négatif) et *uwielbiam* / *j'adore* (positif). Mais dans ce contexte, il faut interpréter le retard en terme positif (grâce au retard du train, quelqu'un a pu le prendre). Il n'y a pas d'ironie parce qu'il n'y a pas d'opposition :

- (53) Pociąg ma **opóźnienie** 10-minutowe, przez co zdążyłam na niego, **uwielbiam** takie momenty w moim życiu. XD
 [Le train a un retard de 10 minutes, c'est pourquoi j'ai réussi à le rattraper, j'adore les moments comme ça dans ma vie. XD]

L'autre difficulté est liée cette fois-ci au contexte textuel. En effet, les commentaires sont souvent accompagnés d'une photographie, d'un dessin humoristique, etc. Pour détecter l'ironie, il faudrait que la machine puisse décrypter le contenu de l'image... chose qui n'est pour le moment pas possible de faire. Sans l'image, ce tweet est difficilement compréhensible, il manque une partie de l'information :

- (54) 12 lat temu ktoś wpadł na pomysł wysłania na wycieczkę pociągiem uczestniczek konkursu Miss Świata odbywającego się w Polsce. Najlepsze dostępne wagony zapewniły PKP Przewozy Regionalne. Tak to wyglądało. :) @PKP_SA @POLREGIO_PL #kolej #PKP pic.twitter.com/OmfAGAC7MC
 [Il y a 12 ans quelqu'un a eu l'idée d'envoyer en train un groupe de participantes au concours de Miss Monde qui se déroulait en Pologne. Les meilleurs wagons ont été fournis par PKP Przewozy Regionalne. C'était comme ça. :)]

Dans l'exemple ci-dessous, sans le hashtag *#ironie*, il est impossible pour la machine de détecter l'ironie. En dehors de l'image qui accompagne le tweet (copie d'écran d'un sms reçu sur le téléphone de l'utilisateur), le hashtag est le seul indice linguistique qui permette d'identifier l'ironie :

- (55) Bon week end ! #ironie #sncf #intercite #normandie

aujourd'hui 14:36

Alerte SNCF :
 mouvement social.
 Votre train 3130 du
 14/12 ne circulera pas.
 RDV en point de vente
 pour échanger ou
 annuler votre billet.
 Merci de votre
 compréhension.

7. CONCLUSION

Notre analyse franco-polonaise nous a permis de faire émerger un motif quant à l'association d'un type ironique avec certaines thématiques. Dans les exemples analysés, le motif ironique est particulièrement mobilisé pour aborder la thématique des retards et de l'état des transports. Le recours à l'ironie contribue, entre autres, à dénoncer l'état souvent pitoyable des services des transports. Le double sens ou le sens connoté d'une expression s'avèrent être des stratégies rhétoriques pour économiser des caractères lors de la rédaction d'un message. Des oppositions peuvent être explicites ou implicites. Dans le cas de l'opposition explicite, c'est le contraste entre le positif et le négatif. Dans le cas de l'ironie implicite, c'est le contraste entre ce qui est dit et le contexte pragmatique. Ce n'est pas le sens de tel ou tel mot qu'il faut inverser en cherchant son contraire, mais la signification globale de l'énoncé. Il ne s'agit pas de transformer une phrase en son contraire, mais un acte de parole en un acte de parole opposé ou contradictoire.

Ce travail nous a également conduits à découvrir qu'en polonais il y avait très peu d'outils d'analyse automatique de la langue. L'étape suivante de nos recherches portera sur la construction d'un dictionnaire de polarité négative / positive qui permettra ainsi d'analyser les corpus polonais. Un de nos objectifs est de créer des outils pour le traitement automatique du polonais.

Enfin, il faut souligner que le recours au texte, au contexte, reste en dernière instance nécessaire pour décider si un énoncé est ironique ou non. Certains traits mis en valeur dans notre étude peuvent servir à extraire un premier ensemble de données mais le recours aux textes montre que ces indices peuvent tromper, l'énoncé n'étant pas ironique mais authentiquement laudatif, par exemple. Inversement, certains énoncés ironiques peuvent ne pas offrir de signaux ou de séquences de repérage probable. L'automatisation, ou la semi-automatisation, du repérage de l'ironie demande encore beaucoup de travail.

BIBLIOGRAPHIE

- Amossy, R. (2003). *Les ironies balzacziennes*, chapitre Les fonctions argumentatives de l'ironie balzacziennne. Pirot : Saint-Cyr sur Loire, 143–154.
- Anscombre, J.-C., Tamba, I. (ed.) (2013). « L'intensification ». In : *Langue française*, 177.
- Attardo, S. (2000). « Irony as relevant inappropriateness ». In : *Journal of Pragmatics*, n° 32, 793–826.
- Benamara, F., Grouin, C., Karoui, J., Moriceau, V., Robba, I. (2017). « Analyse d'opinion et langage figuratif dans des tweets : présentation et résultats du défi fouille de textes 2017 ». In : *Actes de l'atelier DEFT de la conférence TALN 2017, 26 juin 2017, Orléans*, 1–11. Caen, France : Association pour le Traitement Automatique des Langues.
- Berrendonner, A. (1981). « De l'ironie ». In : *Éléments de pragmatique linguistique*, 173–239. Paris : Editions de Minuit.
- Bres, J. (2011). « L'ironie, un cocktail dialogique ? ». In : *VNU Journal of Science : Foreign Studies*, 27 (3).
- Buscaldi, D., Grezka, A., Lejeune, G. (2017). « Tweetaneuse : Fouille de motifs en caractères et plongement lexical à l'assaut du DEFT 2017 ». In : *13e édition du Défi Fouille de Texte, Actes du colloque TALN 2017, 26-30 juin 2017, Orléans, France*, 65–76.
- Charaudeau, P. (2006). « Des Catégories pour l'Humour ? ». *Questions de communication*, n° 10. Nancy : Presses Universitaires de Nancy.
- Cougnon, L.-A. (2015). *Langage et sms. Une étude internationale des pratiques actuelles*. Louvain : Presses Universitaires de Louvain.
- Ducrot, O. (1984). « Esquisse d'une théorie polyphonique de l'énonciation. Le dire et le dit ». In : *Rhétorique et argumentation : de l'ironie*. Paris : Editions de Minuit, 171–233.
- Eggs, E. (2009). « Rhétorique et argumentation : de l'ironie ». *Argumentation et Analyse du Discours*, 2 (3).
- Filik, R., Ūrcan, A., Thompson, D., Harvey, N., Davies, H., Turner, A. (2016). « Sarcasm and emoticons : Comprehension and emotional impact ». In : *The Quarterly Journal of Experimental Psychology*, 69:11, 2130–2146. DOI : 10.1080/17470218.2015.1106566.
- Germoni, K. (2013). *Majuscules, abréviations et symbols*. Grevisse coll. Langue française.
- Grezka, A. (2006). *Les prédicats de perception. Traitement de la polysémie (Les sens des sens)*. Thèse de doctorat en Sciences du Langage, Université Paris XIII, Villetaneuse.
- Grezka, A. (2009). *La polysémie des verbes de perception visuelle*. Collection Sémantiques. Paris : L'Harmattan.
- Gross, G. (2015). « Traitement automatique de la polysémie ». In : *Studia Romanica Posnaniensia*, Vol. 42, Issue 1, 15-33.
- Jorgensen, J. (1996). « The functions of sarcastic irony in speech ». In : *Journal of Pragmatics*, 26 (5), 613-634.
- Karoui, J. (2016). « FrIC : Un corpus et un schéma d'annotation multi-niveaux pour l'ironie dans les tweets ». In : *Atelier Communautés en ligne : outils et applications en TAL (COLTAL 2016) dans le cadre de la conférence JEP-TALN-RECITAL 2016, 4 juillet 2016 – 8 juillet 2016, Paris, France*.
- Kerbrat-Orecchioni, C. (1978). *L'ironie*. Travaux du centre de recherches linguistiques et sémiologiques de Lyon, 2. Lyon : Presses Universitaires de Lyon.
- Kerbrat-Orecchioni, C. (1986). *L'implicite*. Paris : Armand Colin.
- Kleiber, G., Schnedecker, C. (ed.) (2007). *L'intensité*. Travaux de linguistique 55.
- Lievenbruck, H., Ameer Ben Jannet, M., D'imperio, M., Spini, M., Champagne-Lavau, M. (2013). « Prosodic Cues of Sarcastic speech in French : Slower, higher, wider ». In : *Inter-speech*, 3537–3541.

- Maingueneau, D. (2012 rééd.). *Analyser les textes de communication*. Paris : Colin, coll. I. Com.
- Muryn, T., Niziołek, M., Prażuch, W., Hajok, A. (2015). « Ce café me réveille ou l'intensité inférée ». In : K. Wróblewska-Pawlak, A. Kieliszczyk (ed.), *L'intensification et ses différents aspects*. Varsovie : Université de Varsovie, 85–100.
- Niziołek, M., Romero, C. (ed.) (2016). *L'intensité, entre langue et discours*. Synergies Pologne, n° 13.
- Paroubek, P., Grouin, C., Bellot, P., Claveau, V., Eshkol-Taravella, I., Fraisse, A., Jackiewicz, A., Karoui, J., Monceaux, L., Torres-Moreno, J.-M. (2018). « DEFT2018 : recherche d'information et analyse de sentiments dans des tweets concernant les transports en Île de France ». In : *Actes de la conférence Traitement Automatique des Langues, TALN 2018, 1–11, Rennes, France* <hal-01839407> .
- Romero, C. (2017). *L'intensité et son expression en français*. Ophrys.
- Sperber, D., Wilson, D. (1998). « Relevance Theory : Applications and Implications ». In : *Irony and Relevance : a reply to Seto, Hemmamoto and Yamanashi*, 283–293. Carston R., Uchida S. Amsterdam : John Benjamins.
- Wilson, D. (2006). « The pragmatics of verbal irony : Echo or pretence ? ». *Lingua*, 116, 1722-1743.
- Wróblewska-Pawlak, K., Kieliszczyk, A. (ed.) (2015). *L'intensification et ses différents aspects*. Varsovie : Université de Varsovie.