

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

ENJEUX LIÉS À LA DÉTECTION DE L'IRONIE

MÉMOIRE
PRÉSENTÉ
COMME EXIGENCE PARTIELLE
DE LA MAÎTRISE EN LINGUISTIQUE

PAR
SAMUEL LAPERLE

FÉVRIER 2024

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

La diffusion de ce mémoire se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.04-2020). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

REMERCIEMENTS

Je n'ose pas rédiger mes remerciements en majuscule, mais les minuscules semblent inadéquates pour exprimer ce qui m'habite en écrivant ces mots. Le contenu de ce mémoire est loin d'être aussi important que toutes les expériences et les relations que j'ai pu vivre autour et grâce à lui. J'ai beaucoup de reconnaissance à donner à beaucoup de gens et c'est difficile de garder le tout assez court.

J'aimerais, bien entendu, remercier mes parents qui n'ont jamais semblé douter de moi (même lorsqu'ils auraient dû).

Merci à mon directeur, Grégoire Winterstein, pour sa patience, ses conseils, ses enseignements, son argent, ses recommandations de livres et de films. Surtout, merci de m'avoir montré qu'il est tout à fait possible de rester soi-même et intègre tout en naviguant les pressions de performances intrinsèques au monde académique.

Merci à Michelle Stewart, Sklaerenn Le Gallo et Vicky Girard. Le projet de recherche Viral Populism est arrivé exactement au moment où je pensais que tout ce que je faisais ne servait à rien et j'ai adoré en apprendre plus sur le monde des communications, de la science politique et de la sociologie avec vous.

Merci aux collègues du SLIC et du département qui ont toujours été très gentils avec moi malgré le fait que je ressemblais de plus en plus à un des matelots fusionnés dans le mur du Hollandais Volant dans le film Pirates de Caraïbes 2.

Merci à ma conjointe, Roxanne, qui en plus d'être ma personne préférée avec qui écouter la télé, est une excellente contre partie pour mon absence chronique et probablement pathologique d'énergie.

Merci aux meilleurs colocs du monde, amis, collègues, guides, inspirations, bandits, Yoann et Audrée. J'aurais aimé habiter pour toujours. Vous êtes mes 2e et 3e partenaires préférés pour écouter la télévision.

Un dernier merci à mes ami·e·s terribles Cath, G. Stanley V., Gabrielle, Jacob, Marguerite, Marilyn, Luce, Lude, Solène, Théodène et Xav. Je suis toujours incroyablement surpris d'avoir scoré des bons chums comme vous autres. C'est à cause et malgré vous que je m'intéresse à la linguistique. Vous êtes les seules personnes qui me font rire comme une goule et la science a le droit de savoir les secrets derrière vos actes de sorcel-

lerie. Merci pour toute et je vous aime.

Finally, I dedicate this master thesis to Alanis Morissette. I hope this work brings you some answers.

TABLE DES MATIÈRES

TABLE DES FIGURES	vii
LISTE DES TABLEAUX	viii
RÉSUMÉ	ix
INTRODUCTION	1
CHAPITRE 1 L'IRONIE EN LINGUISTIQUE	5
1.1 Introduction	5
1.2 Historique du terme et premières théories linguistiques	5
1.3 L'ironie comme négation d'un sens littéral.....	6
1.4 Théorie de la pertinence et l'ironie.....	9
1.4.1 La théorie de l'écho.....	11
1.4.2 Critique de la théorie de l'écho	13
1.5 La théorie du faire semblant.....	15
1.6 L'ironie comme propos inapproprié mais pertinent	16
1.7 Définir l'ironie	18
1.7.1 Définition proposée	18
1.7.2 Typologie de l'ironie verbale	20
1.8 Autres caractéristiques importantes	27
1.8.1 Aspects évaluatifs de l'ironie.....	27
1.8.2 Le rôle du contexte dans la production et l'interprétation d'un énoncé ironique.....	30
1.8.3 Représenter le contexte	31
1.9 Conclusion	32
CHAPITRE 2 DÉTECTER AUTOMATIQUÉMENT L'IRONIE	34

2.1	Introduction	34
2.2	Modèles symboliques	34
2.2.1	Traits acoustiques	35
2.2.2	Traits lexicaux	36
2.2.3	Traits sémantiques	37
2.2.4	Traits pragmatiques	38
2.2.5	Traits cognitifs	39
2.2.6	Traits contextuels	40
2.2.7	Évaluation des traits	41
2.3	Apprentissage machine	42
2.4	Limitations	43
2.4.1	Problème de définition : ironie ou sarcasme ?	44
2.4.2	La polarité d'un énoncé ironique	46
2.4.3	Corpus.....	46
2.5	Qu'apprennent réellement les algorithmes d'apprentissage machine ?.....	48
2.6	Modèles pragmatiques computationnels	49
2.7	Conclusion	52
CHAPITRE 3 IRONISER AVEC BERT		54
3.1	Introduction	54
3.2	BERT.....	54
3.2.1	Fonctionnement	54
3.2.2	Performances	55
3.2.3	BERT et l'ironie	57

3.2.4	Artéfacts d'apprentissages	59
3.3	Méthodologie	61
3.3.1	Paramètres et ajustement de BERT	61
3.3.2	Corpus d'entraînement.....	61
3.3.3	Construction des items tests.....	62
3.3.4	Catégories d'énoncés ironiques.....	62
3.4	Résultats	66
3.5	Discussion	68
3.5.1	Patrons d'erreurs	68
3.5.2	Résumé	76
3.5.3	Construction du corpus SemEval 2018.....	76
3.6	Limitations et recherches futures	78
3.7	Conclusion	80
	CONCLUSION.....	81
	ANNEXE A CORPUS D'ÉVALUATION CONTENANT LES ÉNONCÉS IRONIQUES COLLECTÉS PAR BEALS (1995) ET LES RÉSULTATS DE BERT	84
	BIBLIOGRAPHIE	92

TABLE DES FIGURES

Figure 1.1	Types d'ironies échoïques et non-échoïques Seto [1998].....	15
Figure 1.2	Polarités et cibles possibles d'un énoncé ironique schématisée par Alba-Juez and Attardo [2014]	29
Figure 3.1	Pourcentage d'énoncés adéquatement classés comme étant ironiques par BERT	67

LISTE DES TABLEAUX

Table 1.1	Tableau 1 : Processus inférentiel selon Wilson and Sperber [2012].....	11
Table 3.1	Résumé des performances de BERT rapportées par Devlin et al. (2018)	56
Table 3.2	Performances de BERT sur les données de SemEval 2018	59
Table 3.3	Paramètres de BERT	61
Table 3.4	Résumé des catégories d'énoncés ironiques sélectionnées pour la tâche et ceux de Beals .	66
Table 3.5	Résultat de l'équipe THU_NGN à la tâche A de SemEval 2018	67

RÉSUMÉ

L'objectif de ce mémoire est d'évaluer la capacité d'un modèle de langue finement ajusté sur un corpus de données ironiques à généraliser ses apprentissages sur différents types d'ironie. Ce travail s'effectue en trois temps. Le premier chapitre traite des différentes approches linguistiques tentant de rendre compte de ce type de discours. Le deuxième chapitre aborde les différentes méthodes computationnelles ayant comme objectif de détecter automatiquement l'ironie et leurs limitations. Le troisième chapitre propose de discuter des performances de BERT finement ajusté sur les données de la tâche A de SemEval 2018 sur les données d'un corpus d'évaluation construit à partir d'énoncés ironiques provenant du travail de Beals [1995].

INTRODUCTION

Trouver une façon de représenter le langage de façon computationnelle impose son lot conséquent de défis et soulève des questions intéressantes autant en linguistique qu'en informatique. Parmi celles-ci, on retrouve tout ce qui touche à la modélisation et la détection d'usage figuratif du langage. Le présent mémoire se penche principalement sur les enjeux reliés à la détection automatique de l'ironie. Tenter d'arriver à une définition exhaustive et opérationnalisable de ce type de discours permet à la fois de mettre en évidence les limites des modèles computationnels, mais aussi, les différents débats théoriques sur ces questions présents en linguistique.

L'ironie pose une dichotomie intéressante entre ce que les locuteurs et les spécialistes comprennent du phénomène. Si ces derniers arrivent difficilement à proposer une définition exhaustive de ce type de discours, la population générale semble assez prompte à être en mesure d'identifier ce qui est ou n'est pas ironique. Souligner qu'il y a peu de consensus sur ces questions dans la littérature, ce n'est pas de dire qu'il n'y en a aucun. En effet, certaines caractéristiques sont bien documentées.

Par exemple, on retrouve parfois l'idée selon laquelle l'ironie est propre à certaines cultures et pas d'autres. Notamment, pendant un certain temps, on retrouvait sur le site de la BBC, un guide lié à la compréhension de l'ironie pour les nouveaux arrivants. L'utilisation de ce type de discours serait culturellement mise de l'avant au Royaume-Uni. En parallèle, on retrouve certains articles, toujours de la BBC, se questionnant sur la possibilité que les Américains ne puissent pas comprendre l'ironie¹. Lorsque l'acteur Tim Curry se fait questionner sur ce qui lui manque le plus du Royaume-Unis depuis qu'il habite aux États-Unis, il répond : "l'ironie"².

Évidemment, ce type de perception se trouve ailleurs. Par exemple, Muecke [2017] rapporte que, d'un côté, on retrouve des propos mentionnant que les Japonais ne peuvent pas comprendre l'ironie contrairement aux Chinois.³ De l'autre, il cite le journal du Goncourt (20 mars 1884) où on retrouve la citation : "les Japonais ont une aimable ironie, une ironie un peu à la française. (Muecke [2017], p.2)". Ce qui se dégage de ces mises

1. <http://news.bbc.co.uk/1/hi/magazine/3433375.stm>

2. Ibid.

3. "I had long been hearing, in the English colony at Tokyo, that no Japanese can understand irony (whereas the Chinese of course, use it all the time" (Muecke [2017], p.1 : New York Review of Books, 12 June 1975, p.37)

en opposition, c'est que certaines façons d'ironiser peuvent probablement dépendre de facteurs culturels. ⁴, mais que l'existence du phénomène en soi est universelle.

Néanmoins, il existe certaines conditions individuelles qui peuvent affecter la détection de ce type de discours. Parmi les différentes causes possibles liées à une compréhension plus difficile de l'ironie, on retrouve les lésions cérébrales traumatiques touchant principalement le lobe frontal chez les adultes (Martin and McDonald [2005]) comme chez les enfants (Dennis et al. [2013]). On retrouve des patrons similaires chez certains profils neuroatypiques comme chez les gens se situant sur le spectre de l'autisme. ⁵

Malgré le fait que tous ces patrons cognitifs soient extrêmement différents, il semblerait que la principale cause du traitement de ce type de discours soit la capacité de se représenter les états internes du locuteur par l'interlocuteur (a.k.a la théorie de l'esprit) (Martin and McDonald [2005]) ou dans "l'intégration de certains indices sociaux dans les interactions" (Deliens et al. [2018] : Chevallier et al. [2012]). De ce fait, parmi les autres facteurs individuels favorisant une interprétation d'un énoncé ironique dépend de la proximité du locuteur et de l'interlocuteur. Ainsi, pour Gibbs [2000], par exemple, on ironiserait plus avec des amis ou des proches. En ce sens, Pexman and Zvaigzne [2004] démontrent que des participants sont plus enclins à utiliser l'ironie dans des relations dites "solidaires" que l'inverse. Ils ajoutent aussi que :

irony also allows the speaker to highlight closeness to the addressee ; the speaker and addressee have mutual beliefs that make possible the second-order inferences involved in irony understanding. The solidary relationship is a cue to the addressee (or overhearer, which a is more accurate label for the participants in our study) that speaker and addressee mutually believe that each is likely to be able to infer the attitudes of the other. Our claim here is that the solidary relationship is a cue to irony and also a product of irony.

Comme mentionné plus haut, ce mémoire traite des difficultés liées à la détection automatique de l'ironie. Le choix de ce phénomène discursif n'est pas anodin. Il permet de mettre en évidence les liens qui tiennent les approches théoriques et expérimentales ensemble. En effet, le choix d'une définition théorique particulière pour expliquer et caractériser un objet précis affectera directement comment on tentera

4. Plus spécifiquement, comme il en sera question plus tard dans le mémoire, il est difficile d'ironiser sans partager un certain bagage d'informations partagées entre le locuteur et l'interlocuteur. Un contexte culturel peut favoriser ce terrain commun.

5. Ici, l'idée générale est de simplement montrer qu'il existe des variations individuelles dans le traitement de l'ironie. Dans tous ces cas, le traitement de ce type de discours est varié et la littérature sur ces questions est abondante. Cette très courte présentation ne prétend pas rendre justice à la complexité de ces questions.

d'évaluer empiriquement sa nature. De ce fait, les problèmes computationnels liés à la détection de l'ironie deviennent un terrain d'exploration intéressant pour évaluer les dynamiques entre ces relations. D'un côté, les questions relatives à l'élaboration d'une définition exhaustive et satisfaisante de l'ironie sont toujours en suspens. Il n'existe pas de consensus en linguistique permettant de bien cerner ce que c'est. De l'autre, les tentatives de réponses existantes affectent nécessairement comment on tente d'opérationnaliser ce type de discours pour façonner des algorithmes de détection automatique. Il n'existe pas d'algorithme étant en mesure de détecter systématiquement la présence d'ironie dans un énoncé. Ainsi, dans l'exposition et la mise en relation de ces défaites respectives, ce mémoire tente d'apporter une nouvelle perspective sur l'ironie, mais aussi, sur comment les approches interdisciplinaires sont nécessaires pour être en mesure de bien comprendre un phénomène précis. Les approches linguistiques et computationnelles ne sont pas en elles-mêmes suffisantes pour décrire et caractériser l'ironie, mais dans le choc des deux, on retrouvera nécessairement des éléments intéressants qui permettront de mieux comprendre ce phénomène ou, du moins, d'identifier certaines lacunes théoriques.

Pour bien refléter ces idées, les chapitres de ce mémoire s'organiseront d'une manière à exemplifier ce qui vient d'être décrit. Dans le premier chapitre, il sera question d'explicitier les caractérisations essentiellement linguistiques de ce type de discours. Plusieurs théories furent proposées pour tenter de définir l'ironie verbale. Fréquemment citées ensemble, on retrouve les perspectives de Grice [1975], Wilson and Sperber [1992] et Clark and Gerrig [1984] qui proposent respectivement que l'ironie soit une opposition entre le sens d'un énoncé et le sens voulant être transmis par le locuteur, qu'elle est une façon de faire écho à quelque chose de partagé entre deux entités ou qu'elle est une mise en scène où le locuteur joue un rôle. Ces trois positions mettent de l'avant des caractéristiques importantes de ce phénomène, mais en négligent d'autres. À partir de celles-ci, je présenterai d'autres définitions, celle de Beals [1995], et d'autres caractéristiques négligées par ces approches.

Dans le second chapitre, j'offrirai des exemples de différentes façons d'opérationnaliser certaines caractéristiques de l'ironie verbale pour élaborer un algorithme de détection automatique et les limitations conceptuelles qui en découlent. D'un côté, les approches à base de règles nécessitent une opérationnalisation figée de l'ironie qui ne peut détecter toutes les formes possibles de ce type de discours. De l'autre, les approches automatiques, elles, se basent sur une représentation non directement théorisée parfois basée sur des corpus d'entraînement ou sur l'élaboration de traits particuliers qui négligent certains aspects de ce type de discours.

Le troisième chapitre tentera de faire un pont entre les deux premiers en démontrant expérimentalement les conséquences directes du choix d'une définition particulière d'un phénomène sur les performances d'un algorithme. Concrètement, je testerai la capacité de BERT ajusté finement sur les données de SemEval 2018 à bien détecter l'ironie d'un sous-ensemble d'énoncés représentant des catégories de ce type de discours tel que proposé par Beals [1995].

CHAPITRE 1

L'IRONIE EN LINGUISTIQUE

1.1 Introduction

Un survol rapide de la littérature concernant l'ironie met en évidence les difficultés reliées à sa définition. Par exemple, les termes ironie et sarcasme sont parfois utilisés de façon interchangeable, ou encore, on s'explique mal les différences entre l'ironie verbale, exprimée de façon linguistique, et l'ironie situationnelle, la perception qu'un événement ou une situation est ironique. Parfois, on essaie de diminuer ce phénomène linguistique à l'une de ses caractéristiques pour être finalement confronté à des contre-exemples difficilement explicables. Évidemment, l'objectif de ce chapitre ne sera pas de finalement mettre fin à ces nombreux débats. Au contraire, le but de cette partie de mon mémoire est de proposer un état des lieux permettant de rendre compte des subtilités propres à la compréhension de l'ironie vue par les chercheurs du domaine de la linguistique. Typiquement, les questions relatives à l'ironie en sciences humaines concernent les processus cognitifs liés à son interprétation, les raisons sociales expliquant son utilisation et les caractéristiques linguistiques liées à sa production. L'idée est d'offrir une vue d'ensemble sur ces questions pour, au chapitre suivant, mettre en perspective les tentatives d'opérationnalisation de l'ironie pour la mise en place de modèle computationnel de détection.

Ce chapitre se divise en deux sections. La première mettra en perspective les grandes théories essayant de caractériser l'ironie soit la définition proposée par Grice [1975], la théorie de l'écho mise de l'avant par Wilson and Sperber [1992] et la théorie du faire-semblant par Clark and Gerrig [1984]. Il sera aussi question d'autres définitions comme celle de Beals [1995]. La deuxième partie soulignera d'autres caractéristiques importantes de l'ironie négligée par les conceptualisations précédentes.

1.2 Historique du terme et premières théories linguistiques

Pour Jankélévitch [1979], définir l'ironie apparaît être une tâche difficile en soi. De façon très pessimiste, il nous annonce que nous ne pouvons pas y arriver : les mots seraient insuffisants pour rendre compte de la réalité de ce concept. Ce serait comme essayer de décrire une douleur à quelqu'un d'autre. Évidemment, s'arrêter ici s'avérerait décevant.

Pour Muecke [1970], la difficulté relative à la définition de ce concept s'explique de par l'évolution de son usage à travers les âges. Évidemment, dans une perspective occidentale, il propose qu'on puisse dater les premiers usages du terme à Platon dans l'Apologie de Socrate pour qualifier les usages rhétoriques de ce dernier. En effet, que ce soit pour Muecke [1970] ou Jankélévitch [1979], une des caractéristiques centrales pouvant définir l'ironie c'est justement son utilisation dans un cadre argumentatif. On n'ironiserait que pour ridiculiser des points émis par son adversaire. Allant dans ce sens, Muecke [1970] nous rapporte que le mot grec employé, *Eironia*, se traduirait comme *a smooth low-down way of talking people down (p.15)*. Cet usage aurait ainsi un aspect foncièrement péjoratif. Muecke [1970] nous apprend aussi que ce serait au 18^e siècle que le terme ironie aurait été utilisé en anglais pour se rapprocher d'une définition plus classique soit exprimer le contraire de ce que quelqu'un pense.

1.3 L'ironie comme négation d'un sens littéral

Les premières définitions formelles de l'ironie traitent de ce phénomène comme étant l'expression inverse d'un énoncé produit. Un énoncé ironique serait une négation du sens littéral des mots employés par un locuteur. De façon similaire à Muecke [1970], Kerbrat-Orecchioni [1978], en se basant sur ce qu'écrivent le Petit Robert (1967) et Freud (1931), propose que l'ironie serait un acte illocutoire permettant de se moquer d'une cible. Comme l'antiphrase, elle transgresse une norme de sincérité. Kerbrat-Orecchioni [1978] propose que l'ironie serait une façon «d'exprimer le contraire de ce qu'on veut faire entendre». De Freud (1931), elle garde cet aspect paradoxal où, malgré tout, l'ironiste souhaite être compris par une audience.

D'autre part, elle souligne la distinction qui existe entre l'ironie situationnelle et l'ironie verbale. La première serait la mise en corrélation de deux événements externes au locuteur ayant un caractère ironique mis en évidence par ce dernier. Par exemple, l'ironie en (1) n'est pas exprimé verbalement par quelqu'un. Elle nous apparaît évidente dû à la relation entre les éléments produits. Plus précisément, selon Barbe [1993], les deux éléments mis ensemble sont ironiques par un rapport de coïncidence souligné par un agent externe à la situation.

1. Le patron de segway est mort en conduisant son segway en bas d'une falaise.

Au contraire, l'ironie verbale quant à elle serait la "mise en relation entre deux niveaux sémantiques [littéral et figuratif] attachés à une même séquence signifiante" [Kerbrat-Orecchioni, 1978]. Ce faisant, elle mettrait

en oeuvre un trio actanciel impliquant un locuteur émettant un propos ironique, un interlocuteur devant décoder ce propos et une cible. À noter qu'ici encore, on souligne l'aspect moqueur et dégradant de l'utilisation de ce type de langage figuratif. Toujours selon l'auteure, un énoncé d'ironie verbale s'exécute comme suit : un locuteur (L) produit un énoncé A voulant transmettre non-A. Par exemple, en 2, on pourrait imaginer un contexte où un locuteur souhaiterait transmettre l'opinion selon laquelle Alex est plutôt radin en affirmant ironiquement le contraire. Donc, contrairement à l'ironie situationnelle, on ne réagit pas sur un événement perçu comme étant ironique, mais on tente plutôt de transmettre une attitude qui se veut ainsi.

2. La plus grande qualité d'Alex, c'est sa générosité.

Cette définition de ce type d'énoncé figuratif se rapproche fortement de celle proposée par Grice [1975]. Ce dernier caractérise les rapports communicationnels par quatre maximes conversationnelles permettant de représenter le lien coopératif entre un locuteur et son interlocuteur. La maxime de quantité propose que le locuteur ne devrait exprimer ni moins ni plus que nécessaire. La maxime de manière souligne que nos propos ne devraient pas être ambigus ou obscurs, mais concis et précis. Dans le même sens, la maxime de pertinence ajoute, comme son nom l'indique, que le locuteur devrait émettre des énoncés jugés adéquats et convenables étant donné les buts admis et supposés de la conversation. Ainsi, selon Grice [1975], l'ironie se définit par une transgression de la dernière maxime ; celle de qualité. Selon celle-ci, un locuteur ne devrait pas dire ce qu'il croit être faux. Évidemment, un énoncé ironique étant généralement une dichotomie entre l'expression lexicale d'un énoncé et l'attitude souhaitant être transmise par le locuteur, il lui arrive d'y contrevenir directement. Concernant ce point, il écrit :

Irony is intimately connected with the expression of a feeling, attitude or evaluation. I cannot say something ironically unless what I say is intended to reflect a hostile or derogatory judgment or a feeling such as indignation or contempt. (Garmendia [2015] : Grice [1989])

Cette conception de ce phénomène langagier est enrichie par l'approche néo-gricéenne proposée par Dynel [2013]. La définition proposée par Grice [1975] explique difficilement différents types d'ironie. En effet, s'il est intuitivement raisonnable de qualifier cet usage comme étant une dichotomie entre le sens lexical des mots utilisés et le sens voulant être transmis par le locuteur, cette définition ne semble pas s'appliquer aux exemples suivants proposés par Dynel [2013] :

3. La personne était un peu saoule lorsqu'elle a décidé de vandaliser son propre appartement.

4. En entrant à la banque, un employé m'a gentiment fermé la porte au visage.
5. J'adore les enfants qui gardent leur chambre propre. [Exclamé dans la chambre non rangée d'un enfant.]
6. Peut-être que Superman viendra arranger nos problèmes.

L'ironie en (3) ne se manifeste pas par l'expression opposée entre deux sens, mais dans l'utilisation d'une litote. La personne en question était bel et bien saoule, mais on diminue verbalement son état d'intoxication de façon à créer un rapport d'incongruité entre ce dernier et son comportement irrationnel. De plus, selon Dynel [2013], bien qu'elle ne contrevienne pas directement à la maxime de qualité, elle n'exprime pas toute la vérité non plus. Comme pour l'utilisation d'hyperbole ironique comme en (7), sans être une négation complète du sens premier de l'énoncé, sa véritable interprétation implique un mouvement inverse vers des pôles évaluatifs. Dans ce cas ci, la peine de la personne mentionnée existe et est réelle, mais elle est exagérée.

7. Sa peine était infinie. [Pour décrire quelqu'un un peu déçu d'apprendre que le film qu'il attendait sort un peu plus tard que prévu.]

L'expression (4) contrevient à la maxime de qualité respectant la caractérisation proposée plus haut. Cette transgression prend naissance dans l'unique mot "gentiment". Le locuteur laisse clairement entendre le contraire. De cette forme de lexicalisation, Dynel [2013] propose qu'on puisse qualifier de la même façon les interjections ironiques (8) qui posent problème à cette approche. Selon le modèle proposé par Wilson and Sperber [1992] (voir les sections suivantes), les énoncés en (8) ne sont pas des propositions et, par le fait même, ne peuvent pas porter de valeur de vérité. Ainsi, ils ne peuvent contrevir à la maxime de qualité proposée par Grice [1975].¹ De plus, pour Dynel [2013], elles ne sont pas si problématiques que le laisse sous-entendre Wilson and Sperber [1992]. En effet, en se basant sur des éléments contextuels permettant d'enrichir le sens de ce type d'énoncé, l'interlocuteur devrait traduire ces énoncés en proposition plus complexe et comprendre l'implicature qui en découle. En (8), l'interlocuteur doit arriver à comprendre que ce que le locuteur tente de partager c'est sa déception causée par des décisions prises par le parti politique au pouvoir. Tout d'abord, il comprend l'énoncé en proposition du type "j'aime la CAQ". Puis, il infère l'inverse.

8. Ah, la CAQ!

L'exemple en (5) ne contrevient pas à la maxime de qualité. Techniquement, cet énoncé est véridique. L'iro-

1. Cette affirmation sur les valeurs de véracité portée par les interjections est remise en question par Ginzburg (2015).

nie prend forme dans la dichotomie entre les propos exclamés et le contexte. L'énoncé (6) quant à lui correspond au type d'ironie surréaliste. On ne peut extraire le sens voulant être transmis par le locuteur en n'inversant le sens lexical des mots employés ni en observant leur implicature. En effet, l'exemple (6) n'équivaut pas à l'énoncé (9).

9. Peut-être que Superman ne viendra pas arranger nos problèmes.

Ce type d'énoncé se construit généralement en utilisant un élément absurde compris dans les connaissances partagées issues "*d'informations encyclopédiques et de normes macrosociales*". Selon Dynel [2013], ce type d'ironie transgresse la maxime de qualité en proposant un sens littéral impossible ou dont les probabilités d'occurrences sont extrêmement faibles.

1.4 Théorie de la pertinence et l'ironie

Wilson and Sperber [1992] argumentent contre la perspective gricéenne sur l'ironie. Déjà, pour ces derniers, il est nécessaire de s'éloigner d'une modélisation de la communication par les maximes conversationnelles proposées par Grice [1975] pour se rapprocher d'une représentation plus proche de la cognition humaine. La théorie de la pertinence proposée par Sperber and Wilson [1995] a pour objectif de modéliser comment des locuteurs sont aptes à communiquer et déterminer le sens adéquat d'un énoncé parmi toutes les interprétations possibles. Par exemple, il est question de déterminer comment un locuteur peut arriver à comprendre qu'en (10) la géomorphologie du territoire de la Hollande n'est pas complètement plate, mais qu'elle l'est beaucoup plus que celles d'autres pays.

10. La Hollande est plate. C'est parfait pour y faire du vélo.

Comme sous-entendu, il est impossible de rendre compte parfaitement du monde et de ses subtilités en utilisant le langage. Il serait redondant de mentionner qu'évidemment la Hollande n'est pas complètement plate. Les locuteurs doivent donc nécessairement employer des mécanismes inférentiels pour réussir à déterminer le sens adéquat d'un énoncé. Pour les tenants de l'approche gricéenne, ce dernier peut être déterminé grâce aux maximes conversationnelles explicitées plus haut. Si une de ces dernières se trouve être transgressée, le sens transmis serait nécessairement proche de l'énoncé produit. Pour les tenants de la théorie de la pertinence, cette approche est trop rigide. Par exemple, pour Sperber et Wilson, la maxime

de qualité serait non-nécessaire pour traiter de la pragmatique du langage dû aux trop grands nombres d'exemples communicationnels où elle est transgressée. Par exemple, il y aurait, entre autres, le cas des blagues, de la fiction, des mensonges, des métaphores et de l'ironie. Pour les mensonges, le locuteur tente directement de transgresser de façon cachée la maxime de qualité où l'interlocuteur pense toujours qu'elle est respectée. Au contraire, pour la fiction et les blagues, elle serait ouvertement suspendue par les deux parties. Tandis que, pour l'ironie et les métaphores, elle serait transgressée ouvertement en laissant penser à l'interlocuteur que l'énoncé lui-même est faux, mais qu'il tente tout de même de partager une information véridique. Outre les usages moins littéraires du langage, la communication est parsemée d'énoncés qui ne respectent pas totalement la maxime de qualité comme en (10).

De plus, l'écart entre les propos produits et le sens voulant être transmis peut être considérable. Wilson and Sperber [2002] proposent que ces processus soient issus de mécanismes inférentiels guidés par la capacité intrinsèque du système cognitif de déterminer les éléments pertinents tels qu'explicités par les deux principes de base en (11) et en (12).

11. Principe cognitif de la pertinence Sperber and Wilson [1995] : la cognition humaine tend à être conçue pour trier de façon optimale les stimuli pertinents dans la scène cognitive.
12. Principe de la pertinence dans la communication Sperber and Wilson [1995] : chaque acte de communication présuppose sa propre pertinence optimale.

Ces chercheurs définissent la pertinence comme étant une caractéristique intrinsèque aux stimuli pouvant être perçus directement (représentations externes) ou indirectement (représentations internes) par un individu. Ces représentations sont considérées comme étant pertinentes s'il est possible de les connecter avec d'autres informations contextuelles. Wilson and Sperber [2002] précisent qu'on peut calculer la pertinence d'un stimulus : plus ce dernier crée un effet important sur la personne (ex. : un changement de croyance), plus sa pertinence est grande. Dans le même sens, plus un stimulus demande un effort pour être traité, moins il risque d'être perçu comme étant pertinent par le système cognitif. Ce rapport de saillance est évidemment variable et dépend du modèle interne issu de l'expérience de chaque individu dans un contexte donné. Appliqué à la communication, tel que mentionné par le principe en (12), on s'attend à ce qu'un énoncé produit soit implicitement pertinent et, se faisant, qu'il soit possible d'en tirer une interprétation adéquate. Concrètement, dans l'échange en (13), bien que B ne réponde pas directement à A, il lui est quand même possible d'inférer que, non B n'est probablement pas prêt pour l'examen.

13. A : Est-ce que tu es prêt pour l'examen d'aujourd'hui ?

B : Je suis hangover.

Le tableau 1, inspiré fortement de ceux proposé par Wilson and Sperber [2012], schématise comment se construit le sens d'un énoncé selon la théorie de la pertinence appliqué à l'exemple (13).

Table 1.1 Tableau 1 : Processus inférentiel selon Wilson and Sperber [2012]

(i) B dit à A : "Je suis hangover"	A doit décoder l'énoncé de B
(ii) L'énoncé de B est considéré optimalement pertinent pour A suivant le principe émis en (11)	L'attention de A est maintenu par la reconnaissance que l'énoncé de B est un acte de communication et qu'il implique nécessairement la présomption de pertinence
(iii) L'énoncé de B est pertinent car il répond indirectement à la question de A en laissant sous-entendre que B n'est pas dans le meilleur des états pour réussir un examen.	Sachant qu'en (ii) A croit que l'énoncé de B doit être pertinent, cette inférence est la plus pertinente
(iv) Passer un examen après avoir passer une nuit à boire n'est pas la façon la plus adéquate pour maximiser ses chances de réussites	Cette inférence répond aux attentes de A émises en (iii)

1.4.1 La théorie de l'écho

Cette prémisse de base permet à Wilson and Sperber [1992] de rendre compte des processus inférentiels pour décrire comment un interlocuteur peut interpréter adéquatement un énoncé ironique. Ils considèrent que l'approche gricéenne est inadéquate pour traiter de ce type de discours. Si, pour le cadre théorique gricéen, on peut caractériser les énoncés ironiques par la transgression de la maxime de qualité, Wilson and Sperber [1992] nous rapportent quatre exemples typiques où cette définition est non-suffisante. Parmi ceux-ci, certains de ces exemples font échos à ceux de Dynel [2013]. Outre les cas de litotes, d'hyperboles ou d'interjections ironiques (voir les exemples (3) à (8)), Wilson and Sperber [1992] rapportent le cas des citations ironiques qui contreviennent elles-aussi à la maxime de qualité. Typiquement, dans ce cas-ci ce n'est pas l'énoncé en lui-même qui porte l'ironie, mais c'est son rapport d'incongruence avec le contexte. Par exemple, quelqu'un qui chanterait la chanson en (14) en pleine canicule détournerait le sens premier de la chanson pour ironiser sur la température extérieur. Ce n'est pas l'énoncé lui-même qui est ironique, mais l'incongruence entre son sens premier et l'utilisation détournée par le contexte.

14. Baby it's cold outside

Outre les cas précédents exemplifiés par Dynel [2013] et Wilson and Sperber [1992], la définition d'énoncé ironique proposée par Grice [1975] surgénéralise sur d'autres types de discours. Par exemple, dans le cas des mensonges éhontés (*bald-face lies*), le locuteur est conscient que son interlocuteur sait qu'il ment, mais le fait tout de même pour des raisons qui lui sont propres. Dans un contexte judiciaire, un accusé pourrait continuer de plaider l'innocence malgré les preuves accablantes qui pèsent contre lui seulement par désespoir de cause. Ce type d'énoncé transgresse effectivement la maxime de qualité, mais n'est pas pour autant ironique. Stokke [2013] spécifie que cette distinction entre ces deux phénomènes s'effectue au niveau de l'espace de sens partagé (*common ground*) entre les deux locuteurs. Contrairement au menteur, un ironiste ne tente pas de changer cet espace d'informations, mais l'utilise seulement pour partager une certaine attitude.

De ces limitations propres à l'approche gricéenne, Wilson and Sperber [1992] proposent de caractériser l'ironie par la théorie de l'écho. Pour bien comprendre cette dernière, il est nécessaire de faire la distinction entre les concepts d'usage et de mention. Par exemple, en (15a), Mary ne répond pas à la question de Peter. Elle lui mentionne qu'elle ne peut pas lui parler. Se faisant elle fait usage des mots dans l'énoncé. Au contraire, en (15b), Mary mentionne les propos de Susan.

15. Peter : What did Susan say?

(a) Mary : I can't speak to you now.

(b) Mary : "I can't speak to you now".

Cette distinction entre mention et usage pourrait, selon Wilson and Sperber [1992], offrir une meilleure caractérisation de l'ironie que celle proposée par Grice [1975]. Cependant, pour ce type de discours, le terme de mention s'avère inadéquat pour bien en rendre compte. En effet, une mention est un rapport mot à mot d'un énoncé. Tandis que, l'ironie serait un type de citation indirecte transmettant l'attitude du locuteur concernant une cible. Ce type de discours ferait écho à une mention préalablement convenu entre le locuteur et l'interlocuteur. Par exemple, si on imagine l'énoncé (16) produit suite aux arrestations injustifiées des manifestants de Wet'suwet'en par un étudiant frustré par le travail des autorités, on comprend que Simon fait écho aux rôles des policiers tout en s'en dissociant.

16. Simon : Le travail des policiers est de protéger et de servir

Donc, en se basant sur le principe rapporté en (11), on peut construire l'interprétation d'un énoncé ironique qui suit le raisonnement proposé au travers de la théorie de la pertinence. L'interlocuteur de Simon présuppose que les propos émis par ce dernier en (16) sont optimalement pertinents. Son interlocuteur doit savoir que Simon ne possède pas une bonne opinion du corps policier. De plus, il doit savoir qu'effectivement, en théorie, le travail des policiers est de protéger et de servir. Ainsi, il fait écho aux rôles entendus de la police tout en transmettant une attitude, dans ce cas-ci, négative concernant le respect de ce rôle. De ce fait, son interlocuteur en vient à l'interprétation adéquate selon laquelle Simon ironise sur le travail de la police dans ce contexte précis.

1.4.2 Critique de la théorie de l'écho

Bien que la théorie de l'écho soit celle qui semble prendre le plus de place dans la littérature scientifique concernant l'ironie, il est néanmoins nécessaire de mentionner que pour certains néo-gricéens, il est possible de réadapter l'analyse des cas mentionnés plus haut par Sperber et Wilson pour qu'ils s'insèrent de façon adéquate dans la perspective des maximes conversationnelles de Grice [1975]. Par exemple, pour le cas des euphémismes ironiques, Dynel [2013] propose que s'ils ne transgressent pas directement la maxime de qualité, il serait plus adéquat de dire qu'ils transgressent celle de quantité. Cette dernière stipule que le locuteur ne devrait pas mentionner plus ou moins d'informations que nécessaire. En produisant un euphémisme ironique en (3), le locuteur transgresse cette maxime.

D'autre part, plusieurs auteurs ont manifesté des critiques envers la théorie de la pertinence. Si lors de sa conception Sperber et Wilson tentèrent de rediriger la place de la pragmatique dans les sciences cognitives, les variables proposées pour caractériser la pertinence sont difficilement opérationnalisables. Par exemple, Levinson [1989] critique la formule conceptuelle en (17) pour quantifier la pertinence d'un stimulus (R) en tenant compte du nombre d'effets contextuels (E) et de l'effort cognitif impliqué pour les obtenir (C).

$$17. R = EC$$

Il souligne l'impossibilité d'affecter des mesures objectives à ces variables. En effet, plusieurs éléments contextuels peuvent affecter (C) comme la fatigue de l'interlocuteur ou son état émotif. Plus spécifique-

ment, le point central de la critique apporté par Levinson va comme suit :

[The relevance theory] is more a polemical attempt to force a paradigm shift in favour of a particular approach to cognition than it is a detailed contribution to linguistic pragmatics. The approach requires the reader to buy a set of very specific, and to many minds implausible, assumptions : Fodorean modularity of a specific kind, with a restricted kind of deduction as the heart of central processing, a translation of all input perception into a simple canonical 'logical form', a governing cognitive principle that is not itself an object of processing, a subjective measure of effort being the main limit on processing, and so on. If you do not buy these assumptions, the theory will be untenable in its entirety. (Levinson, 1989, p.461)

D'autre part, il est difficile de bien comprendre comment la théorie de l'écho permet de rendre compte l'énoncé ironique non-échoïque. Par exemple, Seto [1998] souligne qu'il est difficile de savoir à quoi pourrait faire écho le locuteur de (18).

18. A : Jocelyn a emprunté ton char.

B : Ah, j'adore ça.

Plus précisément, Seto [1998] propose qu'on puisse classer les énoncés ironiques en deux groupes : les énoncés échoïques et les énoncés non-échoïques. Il précise qu'on peut caractériser cette classification en spécifiant que les énoncés ironiques échoïques se basent, évidemment, sur des citations directes ou indirectes. Tandis que, pour les énoncés ironiques non-échoïques, ils seraient, pour la plupart, des inversions sémantiques qui se rapprocheraient de la définition gricéenne rapportée plus haut. Seto [1998] ajoute que, dans les deux cas, les énoncés ironiques échoïques ou non-échoïques purs sont pratiquement inexistantes. Ils se basent généralement sur d'autres types de discours comme l'allusion ou la parodie pour le premier type et l'euphémisme ou l'oxymore pour le deuxième type. Il schématise cette conceptualisation dans la figure (2) :

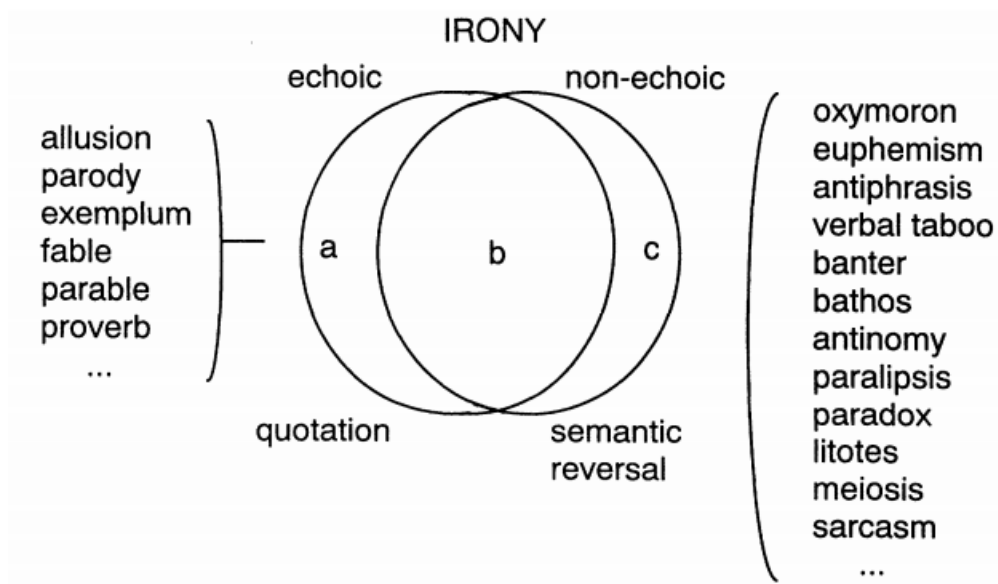


Figure 1.1 Types d'ironies échoïques et non-échoïques Seto [1998]

1.5 La théorie du faire semblant

Outre certaines des limitations de la théorie de l'écho mentionnées plus haut, Clark and Gerrig [1984] soulignent d'autres problèmes avec cette perspective. Pour eux, Sperber et Wilson auraient mal compris la définition proposée par Grice [1975]. Selon Clark and Gerrig [1984], l'essentiel de cette définition ne porterait pas sur le rapport contradictoire entre le sens d'un énoncé et l'attitude du locuteur, mais plutôt sur le rôle que celui-ci jouerait. Ils appuient cette idée avec une citation de Grice [1978] :

To be ironical is, among other things, to pretend (as the etymology suggests), and while one wants the pretense to be recognized as such, to announce it as a pretense would spoil the effect

Clark and Gerrig [1984] rationalisent la relation entre un locuteur ironique et son interlocuteur de la façon suivante :

- Un locuteur L parle à un interlocuteur I et un autre interlocuteur réel ou imaginaire I'
- En étant ironique, L prétend être quelqu'un d'autre soit L' qui communique avec ce I'.
- L ne croit pas ce qu'il dit en tant que L'
- I' ne perçoit pas le rôle joué par le locuteur et croit ce que L dit au travers de L'

- Au contraire, I doit être en mesure de comprendre la mascarade et, ainsi, comprendre l'attitude que tente de transmettre L sur L' en relation avec A'.

Cette schématisation peut sembler plus compliquée qu'elle ne l'est réellement. Concrètement, si on reprend un exemple précédent en (19), on peut en faire l'analyse suivante en se basant sur la théorie du faire semblant.

19. Simon : le travail des policiers est de protéger et de servir

En acceptant évidemment le fait que Simon est ironique, on comprendrait que Simon (L) joue le rôle de quelqu'un satisfait du travail des policiers qui croit littéralement la phrase produite en (19) (L'). Il énonce cette dernière à un interlocuteur avec qui il tente de partager une attitude précise (I) et un autre interlocuteur fictif ou réel (I'). L'interlocuteur (I) de Simon comprend qu'il joue un rôle. Il lui est alors possible de comprendre l'attitude réelle qu'il éprouve envers S' et I'.

Il peut s'avérer difficile de bien saisir quelles nuances la théorie du faire-semblant apporte à la théorie de l'écho. La perspective de Clark and Gerrig [1984] tente d'élargir les horizons de celle de Wilson and Sperber [1992] en proposant une vue d'ensemble plus globale du phénomène de l'ironie. De ce fait, les auteurs ne circonscrivent plus ce type de discours à la linguistique, mais élargissent leurs réflexions dans une perspective sociale et cognitive. L'objectif de Clark et Gerrig était de "rendre justice à l'ironiste" en soulignant qu'il fait plus que mentionner explicitement ou implicitement certains propos pour partager une attitude; il jouerait un rôle.

1.6 L'ironie comme propos inapproprié mais pertinent

Pour Attardo [2000], les modèles présentés jusqu'à maintenant sont insuffisants pour décrire adéquatement l'ironie. Néanmoins, ils nous permettent d'établir plusieurs points de départ qui semblent faire consensus parmi les différents travaux présentés. Par exemple, il est clair que ce soit pour Grice, Sperber, Wilson, Clark et Gerrig que l'ironie est un phénomène fondamentalement pragmatique basé sur un processus inférentiel. De plus, on s'entend sur le fait que ce processus dépend du principe de coopération. Ce dernier est transgressé volontairement par l'ironiste et devrait, selon Attardo [2000], être restauré suite à l'interprétation d'un énoncé ironique.

The reconstruction of the ironist's intended meaning is supposed to be based on a set of shared presuppositions : H knows that S cannot mean M, and S knows that H knows that, and therefore S can count on the fact that H will not stop at S's literal meaning M, but rather look for a more suitable meaning among the infinite set of other meanings which may have been implicated by S.

Comme mentionné plus haut, bien qu'Attardo ne soit pas entièrement d'accord avec la définition Grice [1975], sa proposition s'insère tout de même dans le cadre du principe de coopération établie par ce dernier. Ce principe qui serait inhérent à la communication joue un rôle primordial dans l'interprétation d'un énoncé ironique. En effet, selon Attardo [2000], l'ironie est un phénomène plus pragmatique que sémantique. Il souligne un point important soulevé par Schaffer [1982]; l'ironie n'est pas portée par les mots en eux-mêmes, mais elle prend naissance à partir des implicatures décodées par l'interlocuteur.

Ce rapport débute quand un interlocuteur s'aperçoit que les propos qu'on lui transmet transgressent des maximes conversationnelles. Sur ce fait, Attardo [2000] précise que la transgression du principe de coopération est temporaire. L'interlocuteur accepte inconsciemment que pour l'espace d'un énoncé, les maximes conversationnelles ne soient plus en vigueur et qu'elles le redeviennent par la suite. Attardo nomme cette caractéristique particulière comme étant le principe de *least disturbance*. Malgré la possibilité qu'un locuteur puisse exprimer un énoncé non-littéral, il est tout de même possible pour l'interlocuteur de l'interpréter adéquatement et de savoir jusqu'où s'étend cette transgression de maxime.

Donc, à partir de ces éléments précis, Attardo [2000] propose de définir un énoncé ironique (u) comme suit : (u) est contextuellement inapproprié et, à la fois, pertinent. En effet, le locuteur l'a produit intentionnellement en gardant en tête le fait que le contexte est inapproprié. Il souhaite que son interlocuteur puisse décoder son énoncé.

Le point central qui distingue la perspective d'Attardo c'est le fait de caractériser un énoncé comme étant approprié ou non. Cette caractéristique décrit la relation entre le contexte et l'énoncé. Un énoncé est considéré comme étant inapproprié lorsqu'il y a un écart entre le sens littéral de ce dernier et le contexte dans lequel il a été produit. Attardo [2000] propose la définition suivante pour rendre compte de ce concept :

20. an utterance u is contextually appropriate iff all presuppositions of u are identical to or compatible with all the presuppositions of the context C in which u is uttered (cf. the notion of 'common ground');

Clark [1996]), except for any feature explicitly thematized and denied in u.

Ce principe de convenance (*appropriateness* en anglais) se distingue du principe de pertinence établie par Sperber et Wilson en ce sens où le premier contrairement au deuxième est sensible aux valeurs de vérité des énoncés. Par exemple dans les énoncés (21) et (22), si (22) devient faux (21) devient inapproprié :

21. John devrait quitter la pièce

22. John est dans la pièce.

Ces deux caractéristiques ne sont pas mutuellement exclusives. Un énoncé peut être pertinent et approprié ou, pour un énoncé ironique, être pertinent, mais inapproprié. Ce principe de convenance fait écho à un aspect de la définition de l'ironie établie par Beals [1995] qui sera présentée dans la prochaine section.

1.7 Définir l'ironie

S'il y a bien une chose qui doit ressortir des analyses exposées jusqu'ici c'est la difficulté d'arriver à une définition de l'ironie verbale qui rend compte de façon exhaustive des ses manifestations. En effet, bien que tous les travaux présentés jusqu'à présent mettent en perspective différents aspects pertinents et caractéristiques de l'ironie verbale, différents aspects centraux de ce type de discours restent négligés.

Pour définir l'ironie verbale, Beals [1995] parcourt les différentes définitions proposées en soulignant les caractéristiques de l'ironie verbale qui sont décrites adéquatement et met en évidence celles qui sont omises. Cette section sera composée de trois parties. La première traitera des critiques des différents modèles soulignées par Beals [1995]. Le chemin pour arriver à une définition adéquate de l'ironie étant quelque peu tortueux, cette section permettra, en même temps, d'offrir une synthèse des points importants des autres cadres théoriques. La deuxième partie exposera la définition proposée par Beals [1995] et ses arguments principaux pour l'appuyer. Puis, la troisième section étalera les différents types d'ironie identifiés par Beals.

1.7.1 Définition proposée

La définition de Beals [1995] tente de pallier les failles théoriques des autres définitions exposées dans ce travail. En récapitulatif, comme démontré en 4.1 par Wilson and Sperber [1992], l'ironie n'est pas qu'une dichotomie entre le sens lexical et figuratif d'un énoncé tel que proposé par Muecke [1970], Kerbrat-Orecchioni

[1978] ou Grice [1975]. Néanmoins, ce type de discours ne fait pas non plus exclusivement échos à des informations partagés entre locuteurs comme le proposent Wilson and Sperber [1992]. Aussi, contrairement à la définition proposée par Clark and Gerrig [1984], l'ironiste ne joue pas nécessairement un rôle. Cette conceptualisation engloberait fautivement la parodie et la caricature.

Ainsi, Beals [1995] définit l'ironie verbale ainsi :

23. Ironie Verbale : l'utilisation d'une expression verbale pour faire semblant que quelque chose est vrai tout en soulignant quelque chose d'extrêmement faux.

Les termes *l'utilisation d'une expression verbale* se présentent comme étant une distinction par rapport à la définition proposée par Wilson and Sperber [1992]. En effet, contrairement à ce que proposent ces derniers, une expression ironique peut aussi être un usage. Sperber et Wilson présentent les notions d'usage et de mention comme étant des concepts mutuellement exclusifs ce qui n'est pas nécessairement le cas. Par exemple, comme Beals [1995] le démontre en reprenant l'exemple classique d'écho de Sperber and Wilson [1995] en (24), l'ironie ne prend pas racine que dans la mention de ce vers dans ce contexte particulier. Le locuteur fait aussi usage directement de ce propos. Beals [1995] ajoute que la proposition en (24) que l'Angleterre est un endroit où il fait bon être dans ce contexte « *are implied by the present speakers' citations, not by Browning and Boswell's original use of what is cited. Thus, both use and mention are involved : the verbal expression which each speaker uses in his ironic pretense is a citation.* »

24. [Uttered during a cold, wet, windy English Spring, a quote from Browning, "Home thoughts from abroad"] Oh to be in England, Now that April's here.

Concernant la partie de sa définition qui propose qu'un énoncé ironique serve à *faire semblant que quelque chose est vrai*, Beals [1995] répond aux critiques faites sur la théorie du faire-semblant de Clark and Gerrig [1984]. Elle rejette les propositions concernant le fait que le locuteur ironique jouerait un rôle et qu'il lui serait nécessaire d'avoir un public. Comme mentionné plus haut, cette affirmation est trop large pour décrire adéquatement ce type de discours. Elle maintient toutefois l'idée selon laquelle le locuteur ne croit pas les propos qu'il énonce et qu'il souhaite tout de même que son interlocuteur s'en aperçoive.

À propos de l'aspect de "*souligner quelque chose d'extrêmement faux*". Beals [1995] propose que cette par-

tie de sa définition prend pour acquis la fausseté de ce type de discours. Plus précisément, Beals propose de regrouper tous les énoncés qui sont inappropriés, non pertinents et, évidemment, non-véridique sous cette catégorie. De plus, elle défend l'utilisation du terme "*extrêmement*" en soulignant le fait que la plupart, si ce n'est pas tous les énoncés ironiques, peuvent sembler humoristiques. Ce serait un des aspects perlocutoires de ce type de discours. Un énoncé ironique comme (25) s'avère comique de par l'écart assez grand qu'il peut y avoir entre le terme génie et la perspective véritablement sous-entendue que Peter soit stupide. Donc, le terme *extrêmement* vient mettre en évidence cet écart exagéré entre ces deux sens.

25. Peter est un génie

Par le terme «*souligner*», elle indique que l'ironie permet de mettre en évidence une information que l'interlocuteur saurait à priori consciemment ou non. Par exemple, après avoir effectué une piètre performance au karaoké, un locuteur pourrait produire l'énoncé en (26). Selon cette partie de la définition de Beals [1995], le locuteur soulignerait ainsi une information jugée extrêmement fautive partagée avec son interlocuteur.

26. Je devrais faire les auditions en chant au conservatoire de musique.

1.7.2 Typologie de l'ironie verbale

En tentant de définir ce type de discours, Beals [1995] confronte systématiquement les théories qu'elle aborde dans son travail à différents types d'ironie. De ces exemples, elle en tire une typologie contenant 13 catégories.² Les sous-sections suivantes en présenteront une synthèse avec des exemples.

1.7.2.1 Opposition de sens

Dans cette catégorie, on retrouve les cas où le sens littéral et le sens véhiculé par l'énoncé sont en opposition. De cette catégorie découlant principalement de la définition de Grice, on peut retrouver des énoncés où le sens dépend d'une opposition polaire (27), de la négation de la proposition du sens littéral de l'énoncé (28) et où on retrouve plusieurs formes d'oppositions (29).

2. La majorité des exemples présentés dans ces sous-sections directement est tiré et traduit du travail de Beals. Lorsque nécessaire, le contexte est ajouté entre crochets.

27. [Concernant les ligues de sport pour enfant] La question demeure : insister pour qu'il soit échangé ? Ou le laisser aller dans les ligues mineures ? C'est la décision la plus importante que doit prendre une mère (Ryoko, 1968 : 21)
28. [Concernant une montre Mickey Mouse que le narrateur souhaite s'acheter pour lui-même] Le bracelet de la montre était son seul défaut. Il est large, rouge et en plastique. Tout ça est OK pour un enfant, mais un homme adulte ne devrait pas avoir ça sur sa montre Mickey Mouse (Ryoko, 1973 : 234)
29. Je suis toujours fasciné par le fait que certains ou certaines secrétaires puisse être assis à à peine trois pieds de son patron ou de sa patronne et ne pas savoir s'il est à son bureau ou pas. (Rooney 1981 : 142)

En 27, l'ironie provient de l'opposition entre le fait que choisir le futur sportif de son enfant n'est pas réellement la décision la plus difficile que peut prendre une mère. En 28, c'est la proposition mentionnée par le locuteur concernant le fait que c'est le bracelet rouge de la montre qui est enfantin et non Mickey Mouse. L'exemple en 29 contient plusieurs oppositions. La première concerne le fait que le locuteur n'est pas réellement fasciné par le fait qu'il présente et la seconde concerne le fait que les secrétaires ne savent pas si leur supérieur est présent ou non.

1.7.2.2 Expression de renforcement

Un autre type d'ironie implique l'utilisation d'expression de renforcement. L'ironie derrière ce type de phrases ne provient pas directement de ces expressions en soi. Elles ne font qu'augmenter l'effet de l'ironie déjà présente dans l'énoncé. Ce type d'expression prend des formes stéréotypiques comme celles en gras dans l'exemple suivant :

30. [Dans une lettre critiquant les politiques environnementales appliquées par la compagnie U.S. Steel]
Il est maintenant clair pour moi que mon raisonnement logique était trop simple et que j'ai été terriblement injuste...**Il est clair** qu'ils dépensent chaque sou qu'ils peuvent se permettre et qu'ils font de leur mieux.

Suivant le contexte, le locuteur ne pense pas vraiment que "son raisonnement logique" est "trop simple". Outre cette opposition de sens, les expressions "il est maintenant clair pour moi" ou "il est clair" viennent

"augmenter l'effet ironique" de l'énoncé (Beals [1995] : 14).

1.7.2.3 Énoncés ironiques avec des expressions irrealis

Selon Beals [1995] (p.199) certaines formes grammaticales comme les questions fermées ou les conditionnelles peuvent créer un environnement irrealis. Ce type de construction permet de décrire une situation qui ne va pas réellement se produire ou un fait qui n'existe pas (Bybee [1998]). Pour Beals, cette forme d'énoncé permet au locuteur de ne pas "asserter ni impliquer le contenu propositionnel d'un énoncé". Ainsi, l'ironie de ce type d'énoncé provient d'ailleurs.

31. Peut-être que ça pourrait être une bonne idée de réorienter professionnellement tous les anciens présidents en évêque méthodiste. (Mencken, 1960 :251)

Pour Beals [1995], l'espace irrealis en 31 découle de l'expression modale "peut-être". À cause de celle-ci, le locuteur n'asserte pas que l'idée qu'il propose en serait une bonne. Comme pour les expressions de renforcement, l'ironie de l'énoncé ne dépend pas de l'expression utilisée.³

1.7.2.4 Expressions idiomatiques

Ce type d'ironie se manifeste suite au rapport d'incongruité entre l'utilisation du sens convenu d'une expression idiomatique et le contexte dans lequel elle est utilisée.

32. [Suite à l'attribution des rôles à l'intérieur d'un orchestre] Grâce son oreille musicale, Alex s'est vu attribuer comme instrument de musique les cymbales.

L'ironie en 32 découle de l'opposition de sens émergeant de l'expression "avoir une bonne oreille musicale" et le fait de jouer de la cymbale dans un orchestre.

3. Il est très probable que l'utilisation du concept d'irrealis soit inadéquat. Dans ce cas-ci particulièrement, à première vue, l'utilisation d'expressions comme "peut-être" permet au locuteur de laisser sous-entendre plus clairement qu'il ironise à son interlocuteur en se désengageant de ce qu'il dit.

1.7.2.5 Expression utilisée pour des propositions inférables grâce à leurs contenus sémantiques

Ce type d'expressions contient une interprétation ironique grâce au contenu extrêmement faux présenté par ce que Beals qualifie comme étant la sémantique de la phrase.

33. Les armes à feu sont tellement un outil efficace pour combattre le crime que les États-Unis, avec le nombre le plus élevé de fusils par habitant, possèdent le plus haut taux de crimes au monde. (Ryoko, 1968 : 128)

L'ironie de cet exemple peut découler de deux niveaux d'interprétation. Le premier concerne une opposition entre le sens que le locuteur souhaite véhiculer et le sens littéral de l'énoncé. En effet, le locuteur partage l'idée selon laquelle les armes à feu sont inutiles pour "combattre le crime". Le second niveau découlerait à la proposition inférable de l'énoncé. Dans cet espace interprétatif, l'ironie prend forme si on comprend les propos du locuteur comme si ce dernier proposait qu'une mesure positive de l'efficacité des armes à feu en tant que dispositif de lutte contre le crime soit la quantité de criminalité dans les zones avec beaucoup d'armes à feu.

1.7.2.6 Expressions utilisées pour leurs présuppositions et leurs implicatures conventionnelles

Ce type d'énoncés ironiques se caractérisent par la présence d'éléments ajoutant une information contextuelle grâce à des formes de parenthèses⁴ (34), de clauses relatives, d'expressions référentielles ou d'implicatures conventionnelles.

34. [Se faisant observer par des clients d'un magasin Maxim à Paris] Ils pensent probablement que je suis en train de faire une sieste, ce qui serait la chose suave à faire.

Dans cet exemple, l'ironie prend forme dans la clause (parenthetical) "ce qui serait la chose suave à faire". L'énoncé est ironique parce qu'il ne serait pas suave de dormir devant un magasin.

4. Le travail Beals [1995] ne précise pas directement ce qu'elle entend lorsqu'elle utilise l'expression parenthétique. Si on se base sur les exemples qu'elle rapporte, ce sont des éléments venant préciser des éléments contextuels. Il n'y a aucune référence ou aucun lien avec les implicatures conventionnelles à la Potts [2002]

1.7.2.7 Expressions utilisées ironiquement pour leurs implicatures conversationnelles

Comme mentionné plus tôt dans ce travail, l'ironie peut prendre forme dans la transgression de toutes les maximes conversationnelles de Grice [1975]. Toutefois, Beals [1995] ne donne qu'en exemple des transgressions relatives à la maxime de qualité 35 et à la maxime de quantité 36. C'est au travers de cette dernière qu'on retrouve l'utilisation d'euphémisme ou d'hyperbole ironique.

35. Le remplacement de 5 membres du cabinet signifie un profond changement concernant l'attribution de places assises lors des diners officiels. (p.443 : Ryoko, 1982 : 135)
36. Les idées émises par les présidents américains au courant des 30 ou 40 années ont un tel caractère que la somme totale des connaissances humaines a été appréciablement enrichie par celles-ci. (p.444 : Mencken, 1960 : 48)

En 35 le locuteur ironise sur le fait qu'il pense qu'un tel changement au niveau des sièges apportera un réel changement politique. Dans l'exemple 36, l'interprétation ironique concerne l'enrichissement des connaissances humaines par les présidents américains.

1.7.2.8 Expressions utilisées pour généraliser des implicatures conversationnelles et conventionnelles

Pour Beals [1995] ce type d'expressions se caractérise par l'utilisation des adverbes comme "seulement" (only) et "même" (even) 37.

37. Pour ces vingt dernières années, U.S. Steel les profits ont seulement été de 4 700 000 000\$
38. Des tests ont démontré qu'un coup de poing avec assez de force pourrait, non seulement tuer Ali, mais peut-être même le rendre silencieux.

L'ironie véhiculée en 37 provient du fait que l'utilisation de seulement laisse sous-entendre que ce n'est pas beaucoup d'argent. Concernant l'exemple en 38, l'ironie découle du fait que le boxer Ali est reconnu pour son arrogance loquace et que la mort pourrait être la seule chose qui le rendrait silencieux.

1.7.2.9 Utilisation ironique de phrases non assertives

Dans certains cas, Beals utilise la nomenclature de Searle [1975] décrivant les actes illocutoires. Cette catégorie d'énoncé implique d'asserter une proposition. Au contraire, certains énoncés non assertifs comme les questions, les phrases expressives, les directives et les déclaratives peuvent véhiculer de l'ironie.

39. Encore une fois, j'en appelle à la NRA de nous joindre pour élargir les lois entourant les restrictions relatives aux armes à feu.

Dans cette phrase déclarative, le locuteur sait que la NRA ne participera jamais à la régulation des armes à feu. De ce fait, il ne les invite pas réellement à ce genre de tâche, mais souligne plutôt l'impossibilité de les voir s'impliquer dans cette direction.

1.7.2.10 Utilisation ironique d'acte de langage indirect

Toujours selon la nomenclature proposée par Searle, Beals [1995] pointe qu'il est possible d'ironiser en utilisant des actes de langage indirect. Selon Beals, *the irony in such acts is also dependent on implicatures from felicity conditions, but its source is not an implicature from a felicity condition of the speech act which the utterance is literally in the shape of*. Au contraire, c'est la transgression de la condition de félicité du sens littéral d'un énoncé qui permet l'interprétation indirecte du propos du locuteur. Les cas les plus fréquents de ce type d'ironie sont les questions rhétoriques. Dans l'énoncé 40, l'effet perlocutoire recherché par le locuteur n'est pas d'avoir une réponse à sa question. Au contraire, il utilise une interrogative de façon indirecte pour affirmer quelque chose.

40. Si la vérité ne peut pas être trouvée sur les étagères du *British Museum*, où, me demandais-je en prenant un calepin et un crayon, se trouve-t-elle.

En 40, le locuteur ne pose pas réellement une question nécessitant une réponse. Il affirme, en ironisant, que la vérité devrait nécessairement se trouver au *British Museum*.

1.7.2.11 Implicatures lousses

Beals [1995] souligne que certains énoncés ironiques ne découlent pas d'implicatures conventionnelles ou

conversationnelles ni des conditions de félicités, mais où le locuteur réussit tout de même à faire émerger la fausseté de son propos. Elle catégorise ce type d'ironie comme étant des implicatures "lousses". Ces dernières peuvent découler du contenu de l'énoncé et son interaction avec le contexte d'élocution. Par exemple, en 41, l'ironie provient du fait que le locuteur agit comme s'il pensait que son conseil en était un bon.

41. [Discussion avec Air Force concernant l'utilisation d'une nouvelle colle.] La seule chose que je peux faire pour aider Air Force, c'est leur donner mon meilleur conseil. Mon conseil c'est de trouver une autre solution.

1.7.2.12 Hyperbole ironique

Un douzième type d'ironie catégorisé par Beals concerne les hyperboles ironiques. Comme mentionné par Sperber et Wilson plus haut, ces dernières impliquent une transgression de la maxime de quantité et exagère le sens réel d'un énoncé.

42. Je crois qu'il faut porter les enjeux relatifs à la ponctuation dans la rue s'il le faut. (Pullum, 1991; Beals,463)

Ici, le locuteur exagère sa position concernant la ponctuation. Ainsi, bien que le locuteur présente une sensibilité concernant ces enjeux, il ne propose pas réellement de prendre la rue. D'un côté, il souligne le ridicule qui découlerait de ce genre de manifestation. De l'autre, il souligne tout de même son intérêt concernat les règles de ponctuation.

1.7.2.13 Suggestions ridicules

La dernière catégorie d'énoncés ironiques proposée par Beals concerne les suggestions ridicules qui, comme son nom l'indique, décrit les énoncés où l'ironie prend naissance dans le ridicule volontaire du sens de l'énoncé. Dans l'exemple en 43, ce que le locuteur propose est ridicule et il le sait.

43. Si je le pouvais, le code des bâtiments décrèterait qu'on devrait au moins retrouver trois poignées de porte sur chaque porte de garde-robe. Nous savons tous qu'une poignée de porte est le meilleur

endroit pour accrocher quelque chose.

Il est important de souligner que le travail de Beals [1995] n'était pas de créer une typologie exhaustive de l'ironie verbale. Si elle nous permet d'avoir un aperçu global des formes que peut prendre ce type de discours, certaines de ces catégories sont floues et d'autres empiètent les unes sur les autres. Dans le chapitre 3 de ce travail, j'aborderais plus spécifiquement ces critiques et comment je propose d'opérationnaliser les types d'ironie verbale proposés

1.8 Autres caractéristiques importantes

J'ai tenté dans cette première partie de chapitre de présenter un survol concis et clair de ces approches centrales lorsqu'il est question de ce type de discours. Bien qu'imparfaites, elles soulignent néanmoins plusieurs caractéristiques de l'ironie verbale offrant ainsi une perspective globale du phénomène. Néanmoins, certains aspects pouvant être particulièrement intéressants dans l'élaboration d'un système de détection sont négligés. Cette deuxième partie tentera d'en faire l'inventaire.

Tout d'abord, il sera question de présenter les aspects évaluatifs propres à l'ironie. S'il semble généralement convenu qu'un énoncé ironique peut permettre de transmettre une attitude négative sur une cible, il est tout aussi possible d'utiliser ce type de discours pour communiquer une attitude positive. D'autre part, je traiterai du rôle du contexte dans l'interprétation d'un énoncé ironique.

1.8.1 Aspects évaluatifs de l'ironie

Si le rapport d'opposition entre le sens lexical et sous-jacent à un énoncé ironique s'avère complexe, ces discussions mettent néanmoins en perspective le rapport évaluatif impliqué dans la production de ce type d'énoncé. Que ce soit pour Sperber and Wilson [1995] ou pour Grice [1975], il semble y avoir un consensus sur le fait qu'une production ironique doit s'effectuer pour apposer un jugement négatif ou critique sur une cible donnée. Comme présenté dans la première section de ce travail, Garmendia [2015] rapportait cette citation de Grice résumant bien cette caractéristique de l'ironie :

Irony is intimately connected with the expression of a feeling, attitude or evaluation. I cannot say something ironically unless what I say is intended to reflect a hostile or derogatory judgment or a feeling such as indignation or contempt.

En se basant sur cette intuition de Grice, Giora [1995] propose que l'ironie soit une forme de négation sans marqueur explicite. L'ironiste produirait généralement une phrase affirmative dans le but de transmettre un message différent de cette dernière. Ainsi, le sens de ce type d'énoncé n'annulerait pas son interprétation littérale, mais mettrait en perspective leurs différences. Giora [1995] propose l'exemple (44) où le locuteur tente de faire comprendre qu'il trouve cette célébration ordinaire.

44. Quelle fête extraordinaire !

Pour Giora [1995], la négation implicite permise par l'ironie lui offre la possibilité de ne pas être soumise aux mêmes limitations que la négation directe :

“Direct negation is further limited, where irony is not. Hedged statements, intensified statements, approximations (see also Horn, 1989), and (novel and rich) metaphors cannot be directly denied via negation “ (Giora, 1995)

Malgré tout, si l'aspect négatif de l'ironie semble être le plus communément accepté, il existe plusieurs instances où des locuteurs peuvent utiliser cette forme d'expression de façon positive. Par exemple, Dews and Winner [1995] étendent le champ d'action de l'ironie. Ils proposent que d'un côté on ait la possibilité de produire une critique ironique en déclarant un énoncé positif pour transmettre une attitude négative comme en (45). De l'autre, on pourrait transmettre un compliment ironique en produisant un énoncé négatif pour transmettre une attitude positive comme en (46).

45. Belle partie ! [Déclaré suite à une défaite]

46. Quelle horrible partie ! [Déclaré suite à une victoire]

Alba-Juez and Attardo [2014] poussent ce raisonnement plus loin. En effet, si on peut représenter le caractère évaluatif de l'ironie sur des extrêmes polaires, il est tout aussi possible que ce type d'énoncé soit neutre. Ils rapportent la citation du philosophe Blaise Pascal par en 47. Alba-Juez and Attardo [2014] argumentent que bien que la citation implique une forme d'évaluation en disant que la lettre est longue elle ne comporte que de l'amusement et non une évaluation négative ou positive.

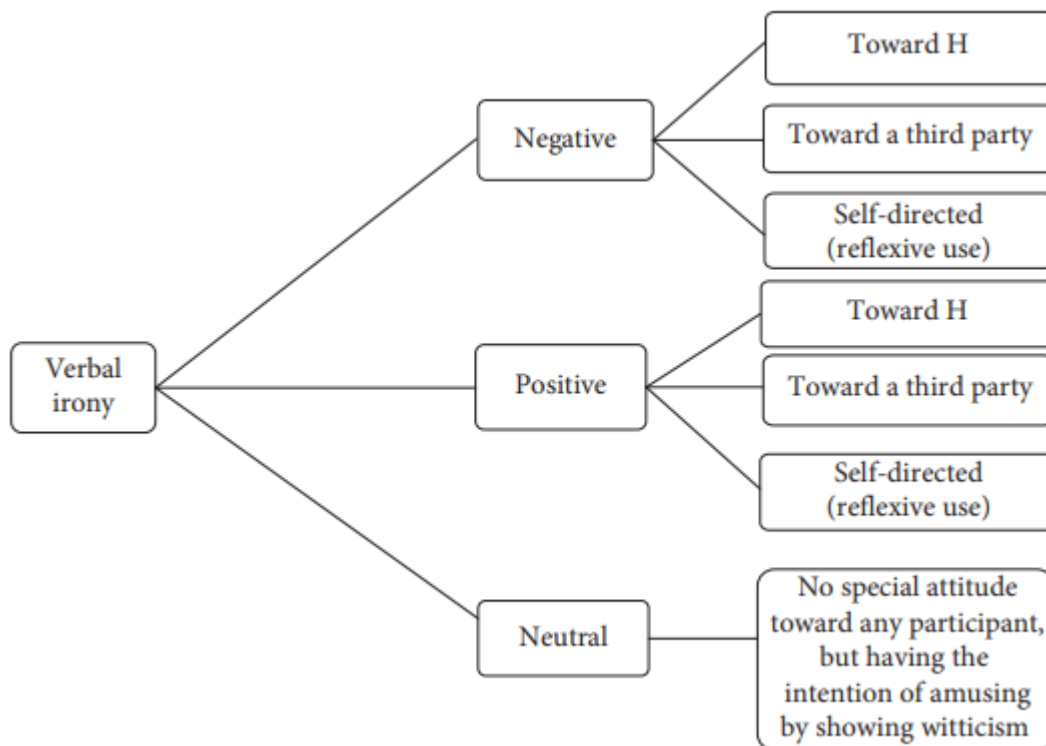


Figure 1.2 Polarités et cibles possibles d'un énoncé ironique schématisée par Alba-Juez and Attardo [2014]

47. The letter is longer than usual because I didn't have the time to make it shorter.

L'ensemble du spectre évaluatif proposé par Alba-Juez and Attardo [2014] est schématisé dans la figure 2.

Le caractère positif ou négatif d'un énoncé ironique se manifeste suivant les attitudes vécues et exprimées par un locuteur envers une cible. En effet, suivant cette perspective, il est possible de viser directement l'interlocuteur, une entité externe à la conversation ou soi-même. Au contraire, un énoncé ironique neutre ne vise rien en particulier; il ne servirait qu'à produire des jeux d'esprit amusants.

Bien que la figure 2 propose une représentation relativement simple du rapport interprétatif de l'ironie, les positions possibles que peut prendre ce type d'énoncé ne sont pas mutuellement exclusives. Alba-Juez and Attardo [2014] proposent l'exemple (48) où une actrice doutant de son talent dirait à son ami :

48. A : I'm a total disaster. I'm never going to make it in the theatre world. I'm a rather mediocre actress.

Pour qu'après la réception d'un prix soulignant son talent, cet ami lui réponde (49) :

49. F : Congratulations, dear friend ! You certainly ARE a mediocre actress. I don't know how they could give you this award !

Cet échange met en évidence la complexité sous-jacente à l'interprétation d'un énoncé ironique. D'un côté, on note que F transmet une évaluation positive du talent d'actrice de son ami tout en émettant une appréciation négative du jugement négatif qu'elle s'auto-imposait. Ces variations dépendent des cibles visées par le locuteur et du contexte de savoir partagé entre les individus. Ces exemples démontrent la versatilité que peut prendre un énoncé ironique en termes de communication d'attitude. Comme l'ont démontré Alba-Juez and Attardo [2014], on ne peut pas se contenter de définir l'ironie comme étant un type de discours qui nous permet de transmettre qu'une attitude négative.

1.8.2 Le rôle du contexte dans la production et l'interprétation d'un énoncé ironique

Il semblerait que nous soyons moins portés à utiliser l'ironie dans des contextes où nous connaissons peu nos interlocuteurs (Gibbs [2000]). En effet, de par sa nature, son utilisation peut nuire à nos objectifs communicationnels. Il est plus difficile de faire bien comprendre à son interlocuteur notre attitude véritable lorsqu'on ne possède pas un ensemble de connaissances partagées.

Comme Mateo [1995] le mentionne, même si certains facteurs métalinguistiques comme la prosodie peuvent nous renseigner sur la présence d'ironie dans un énoncé, sans le contexte il nous est impossible de bien comprendre ce que le locuteur essaie de transmettre. Toujours pour Mateo [1995], on pourrait résumer le contexte comme étant la relation entre un mot ou une expression et l'ensemble d'une situation ou d'un texte.

Cognitivement, ce qui nous permet de réussir à inférer adéquatement le sens d'un énoncé ironique, c'est notre aptitude à pouvoir nous mettre à la place de quelqu'un d'autre c'est-à-dire à pouvoir assumer son point de vue sur une certaine situation (Nilsen et al. [2011]). Dès l'âge de 6 ans, les enfants neurotypiques seraient aptes à déterminer si un locuteur croit les propos qu'il dit (Nilsen et al. [2011]) . Donc, contrairement, à ce que Mateo [1995] propose plus haut, le contexte ne se définit pas qu'en termes de relation textuelle avec une situation, il dépend aussi de la perception inférée par l'interlocuteur de la relation entre

l'expérience du locuteur et la situation qu'il évalue.

En d'autres mots, au-delà du contexte proposé par Mateo [1995], il y a l'ensemble de présuppositions qui s'effectuent entre le locuteur et son interlocuteur. Ce rapport est défini par Stalnaker [2002] comme étant un ensemble de connaissances partagées (*common ground*). En communiquant, de façon implicite, on sait certaines informations sur le locuteur et nous savons qu'il sait qu'on sait certaines informations qu'il sait. Ce rapport d'informations nous permet d'ajuster nos objectifs communicationnels et comment choisir les informations que nous souhaitons communiquer. Ces informations partagées sont plus précisément un ensemble de croyances dynamiques qui peuvent varier selon certaines circonstances. Il est possible pour l'ironiste de jouer sur ce rapport d'informations partagées. En se fiant sur la définition de Clark and Gerrig [1984], il est plus simple pour un locuteur de produire un énoncé ironique quand il sait que son interlocuteur sait au préalable sa véritable attitude sur un sujet discuté. Par exemple, si on change le contexte de (35) plus haut et que l'interlocuteur sait que le locuteur de cet énoncé est un fervent défenseur des autorités policières, il y a plus de chance d'interpréter adéquatement l'ironie derrière ce message en le produisant comme une citation ironique faisant écho à certains slogans anti-policiers.

1.8.3 Représenter le contexte

Allant dans le même sens que Stalnaker [2002], Cohn-Gordon and Bergen [2019] réhabilitent la perspective décrite par Clark and Gerrig [1984]. Un ironiste joue un rôle et il tente d'être adéquatement démasqué pour son interlocuteur. Pour ce faire, l'ironiste doit nécessairement avoir une bonne connaissance des informations partagée entre son interlocuteur et lui. Aussi, son interlocuteur doit avoir une bonne connaissance des croyances et des intentions de l'ironiste. Cohn-Gordon and Bergen [2019] proposent une caractérisation de l'ironie qui se veut être un compromis entre la perspective de Sperber et Wilson et celle de Clark et Gerrig.

Selon eux, il serait nécessaire de porter une attention particulière à la capacité de l'ironiste de prétendre un état du monde qui serait faux et à sa capacité de se mettre à la place d'autrui. En effet, comme dans la perspective de la théorie de l'écho, l'ironiste fait généralement référence (écho) à une information partagée entre lui et son ou ses interlocuteurs. Il joue aussi le rôle de quelqu'un qui ne partage pas réellement sa véritable opinion comme dans la perspective du faire semblant. Concrètement, Cohn-Gordon and Bergen [2019] proposent une représentation formelle de la relation entre un locuteur et son interlocuteur lors de la production d'un énoncé ironique en se basant sur le modèle des Rational Speech Acts (Frank and Goodman [2012]) qui sera développé plus en détail dans le chapitre 2. De façon globale, un locuteur serait

plus tenté de produire un énoncé qui ne ferait pas un portrait représentatif du monde réel s'il pense que son interlocuteur possède une connaissance du monde réel. Ainsi, cet interlocuteur imaginé par le locuteur serait favorablement porté à interpréter un énoncé proposant ou impliquant un état du monde s'avérant faux. Donc, si un locuteur possède une bonne connaissance des informations partagées et du monde réel, il peut utiliser l'ironie pour communiquer des informations qu'il croit déjà partagées.

De ce fait, Cohn-Gordon and Bergen [2019] proposent qu'au moins une des utilités de l'ironie soit justement de réaffirmer le contenu des informations partagées entre le locuteur et son interlocuteur. Par exemple, en (50), le locuteur met en évidence l'information partagée qu'il n'aime pas les pièces de théâtre expérimental.

50. Les pièces de théâtre expérimentales c'est ce que je préfère. [Dans un contexte où le locuteur n'aime pas les pièces de théâtres expérimentales]

Suivant ce que proposent Cohn-Gordon and Bergen [2019], ce type d'énoncé agit sur plusieurs fronts. Il sert en premier lieu à partager une attitude négative sur une cible précise, soit le théâtre expérimental, mais aussi, à recadrer l'espace d'informations partagé entre le locuteur et l'interlocuteur. En résumé, l'ironie dépend nécessairement du contexte pour être interprétée adéquatement au point où son absence trouble directement la détection de ce type de discours.

1.9 Conclusion

Comme mentionné au début de ce chapitre, l'idée n'était pas mettre fin aux débats entourant la conception d'une définition au concept d'ironie, mais pour paraphraser Muecke [1970], d'ajuster nos montres à la même heure. La plupart des éléments essentiels concernant l'ironie verbale ont été survolés. Si la majorité des travaux portant sur l'ironie mettent en perspective la définition de Grice [1975], celle de Wilson and Sperber [1992] et celle de Clark and Gerrig [1984], il est peu fréquent de tomber sur celle proposée par Beals [1995] en (26). L'ironie est un type de discours aux manifestations extrêmement larges et, comme le mentionne Jankélévitch [1979], circonscrire ce type de discours dans une définition simple et exhaustive peut s'avérer difficile. Néanmoins, la présentation de ces différentes théories nous permettra de bien comprendre sur quelles caractéristiques les méthodes de détections computationnelles peuvent s'appuyer.

Évidemment, la difficulté inhérente à la tâche de définir l'ironie se reflète nécessairement sur ces méca-

nismes de détection computationnels. Effectivement, les chercheurs qui s'y attaquent doivent réussir à joindre l'aspect théorique derrière ce type de discours et les technologies préalablement existantes pour tenter de surmonter ce défi. La section suivante proposera de faire un état des lieux sommaires sur ces différentes méthodes et leurs exécutions.

CHAPITRE 2

DÉTECTER AUTOMATIQUEMENT L'IRONIE

2.1 Introduction

Ce deuxième chapitre offre un survol des méthodes computationnelles de détection automatique de l'ironie verbale. Après avoir explicité les caractéristiques linguistiques de ce phénomène au travers du premier chapitre de ce travail, il sera question d'évaluer comment il est possible de concevoir des systèmes en mesure d'évaluer automatiquement la présence de ce type de discours dans un énoncé. Dans un premier temps, j'aborderai comment certaines facettes de l'ironie verbale ont été opérationnalisées dans des algorithmes à base de règles et d'apprentissage machine. Puis, les limitations respectives de ces deux approches seront elles aussi examinées. Finalement, j'offrirai un aperçu des propositions théoriques sur la modélisation de l'interprétation d'un énoncé ironique basée sur les statistiques bayésiennes et des Rational Speech Acts (RSA).

De ce fait, il est important de noter que l'objectif derrière cette section n'est pas de mettre directement en perspective les différents modèles existant en fonction de leurs résultats. Ce travail de synthèse a déjà été effectué par, entre autres, Van Hee [2017]. Par exemple, Van Hee synthétise bien les différents algorithmes marquants de ce domaine en présentant le corpus d'entraînement utilisé, l'approche choisie par les chercheurs, les traits visés par le système de détection et les résultats disponibles. Dans cette section, il sera principalement question des méthodes d'opérationnalisation des caractéristiques de l'ironie et de leurs limitations.

Il aussi est important de mentionner que certains travaux utilisent de façon interchangeable les expressions comme "sarcasme" et "ironie". Cet aspect particulier sera traité plus en profondeur dans la section 3.1.

2.2 Modèles symboliques

Dans un algorithme à base de règles, le traitement des données est codé par un être humain. On y expose explicitement et manuellement comment seront traitées les données d'entrée. Ce type d'approche se distingue des algorithmes d'apprentissage machine présentés dans la section 4.

Les règles caractéristiques de ces architectures se basent sur des traits préalablement encodés par les

concepteurs. La présence ou l'absence de ceux-ci peut permettre de discriminer un énoncé ironique d'un énoncé non-ironique. Les motivations derrière la conception ou l'intégration d'un trait en particulier se posent parfois sur des recherches préalables sur l'ironie verbale, mais pas exclusivement. Cette section présentera un portrait des différents traits utilisés par les chercheurs pour détecter l'ironie.

2.2.1 Traits acoustiques

Une des premières tentatives de détection automatique de l'ironie fut conçue par Tepperman et al. [2006]. Ces derniers proposent de détecter la présence de ce type de discours par l'utilisation de l'expression, "yeah right", en se basant sur des indices prosodiques et contextuels. Celle-ci se retrouverait fréquemment autant dans des contextes non-ironiques qu'ironiques.

Selon Tepperman et al. [2006], on pourrait classer ce type d'expression dans quatre catégories distinctes. D'un côté, il est possible de l'utiliser pour acquiescer comme en (51). De l'autre, on peut le retrouver dans un contexte ironique dans des rapports d'accord/désaccord. Par exemple, en 52, on peut imaginer un contexte où il est su que le locuteur B possède d'excellents résultats en syntaxe. Ce faisant, le "yeah right" devient nécessairement ironique. D'autre part, il est possible d'observer ce type de production dans l'énonciation d'une anecdote (53). Finalement, on peut entendre les mots "yeah" et "right" mis ensemble sans nécessairement en faire l'usage comme en (54).

51. A : Don't forget the american elections are coming in november

B : Yeah right.

52. A : You are going to fail your syntax class.

B : Yeah right.

53. He said to me "we have too many pets". I thought, "yeah right, tell me about it", you know ?.

54. A : Do you know where I can find the Park plaza suites ? B : Yeah, right across de street.

Tepperman et al. [2006] soulignent la présence d'indices contextuels permettant de déterminer si la production d'un "yeah right" à l'oral est ironique. Parmi ceux-ci, on aurait la présence de rire lors de l'énonciation. Ce type de locutions serait aussi plus souvent ironique lorsqu'elle est produite en de façon interrogative que lorsqu'elle est produite de façon affirmative. De plus, Tepperman et al. [2006] s'attendent à ce que ce

type de production soit énoncé plus rapidement qu'un énoncé littéral puisqu'on aurait tendance à observer de l'ironie dans des contextes de joute verbale.

Les traits phonétiques et contextuels proposés par Tepperman et al. [2006] mettent en lumière certaines caractéristiques intéressantes propre à la production d'un énoncé ironique. Néanmoins, comme ils le mentionnent, leur modèle est difficilement généralisable à des données écologiques. Bien entendu, l'ironie ne se circonscrit pas qu'à l'utilisation d'expressions figées comme celle étudiée dans leur travail. De plus, bien qu'il soit possible pour un locuteur d'utiliser diverses stratégies communicatives, comme faire varier sa prosodie pour faire comprendre à son interlocuteur qu'il est ironique, leur présence n'est pas essentielle à la production de ce type de discours. Notamment, on retrouve certains auteurs comme Bryant and Fox Tree [2005] qui affirment, qu'au contraire, il n'existerait pas de ton de voix particulier permettant de caractériser la présence d'ironie dans un énoncé. Selon ces derniers, les interlocuteurs se basent beaucoup plus sur des informations contextuelles pour interpréter adéquatement ce type de discours.

Finalement, avec l'expansion des réseaux sociaux majoritairement basée sur des productions écrites, la plupart des systèmes de détections automatiques de l'ironie qui ont été produits s'intéresse généralement à des données textuelles et non orales.

2.2.2 Traits lexicaux

Bien qu'il ait été précédemment établi que l'ironie ne prend pas son sens dans les éléments lexicaux d'un énoncé, Kreuz and Caucci [2007] proposent tout de même de recenser les expressions les plus fréquemment associées à ce type de discours. Pour ce faire, des participants devaient lire des énoncés préalablement récoltés par Kreuz and Caucci [2007] et évaluer la présence ou l'absence d'ironie dans ceux-ci. À partir des jugements des participants, Kreuz and Caucci [2007] notent que la présence d'interjections aurait été la variable permettant le mieux de prédire leur perception. D'autre part, Kreuz and Caucci [2007] rapportent qu'on retrouverait aussi dans des énoncés ironiques la présence de certaines formules convenues (e.g. : *merci beaucoup*, *beau travail*), des termes appartenant à d'autres langues, des énoncés rhétoriques ou la présence de répétitions.

Allant dans le même sens, Bouazizi and Ohtsuki [2015], quant à eux, proposent comme marqueurs lexicaux de l'ironie la présence de mots peu communs.

2.2.3 Traits sémantiques

Même s'il est difficile de représenter de façon computationnelle le sens complexe que peut prendre un énoncé, certains marqueurs sémantiques permettraient selon Reyes et al. [2012] d'identifier automatiquement la présence d'ironie. Dans leur algorithme, ils proposent d'évaluer la présence d'ambiguïté sémantique contenue dans la production d'un locuteur. Pour ce faire, ils utilisent un réseau lexical Wordnet leur permettant d'évaluer la dispersion sémantique du contenu d'une phrase. Ce type de base de données contient les liens d'hyperonymie, d'hyponymie, de méronymie, de synonymie, d'antonymie et d'holonymie entre les mots et permet d'évaluer la distance sémantique entre deux termes. Ainsi, Reyes et al. [2012] émettent l'hypothèse que l'ironie pourrait être caractérisée par la présence d'un plus grand écart sémantique entre les éléments d'un énoncé. Ils justifient ce choix en soulignant que plus cet écart est grand plus on peut s'attendre à une interprétation inattendue de ce dernier. Outre la présence de ces deux caractéristiques particulières, ils proposent de tenir compte des variations au niveau de la polarité des mots dans un énoncé pour évaluer la présence d'ironie verbale dans une locution.

Le rôle de ce rapport de polarité dans un énoncé pour détecter automatiquement l'ironie verbale a aussi été exploré par Joshi et al. [2015]. Pour arriver à une détection adéquate de ce type de discours, ces derniers proposent d'utiliser la présence de l'incongruité explicite comme en (55) ou l'incongruité implicite comme en (56) pour déterminer la présence d'ironie.

55. J'aime être malade.

56. J'aime tellement cette nourriture que je la donne à mon chien.

L'incongruité explicite est caractérisée pour Joshi et al. [2015] par la présence d'éléments lexicaux de différentes polarités. Cette dernière se base sur un lexique tiré du système Lingpipe SA Alias-i [2008]. Concrètement, on mesure cette variable en vérifiant le nombre de fois où un mot positif est suivi par un mot négatif (et vice versa), le nombre de mots positifs, le nombre de mots négatifs et la polarité globale du tweet. Par exemple, en (55), l'incongruité explicite prend forme dans les polarités contraires du mot "aime" qui est positif et du mot "malade" qui est négatif.

Pour l'incongruité implicite, les éléments lexicaux n'expriment pas directement la polarité sous-jacente à l'énoncé. Donc, pour évaluer cette dernière, Joshi et al. [2015] ont utilisé un algorithme préexistant permettant d'évaluer l'incongruité d'un énoncé en se fiant respectivement à la polarité d'un verbe et de son sujet. De ce fait, l'expression en (56) "donne à mon chien" serait taguée comme étant négative et il y aurait ainsi

une incongruité entre la polarité de cet élément et la polarité de l'expression "j'aime".

2.2.4 Traits pragmatiques

Comme le rapportait Attardo [2000], l'ironie est un type de discours essentiellement pragmatique. Ce dernier soulignait ainsi que l'ironie ne prend pas naissance directement dans les mots employés par un interlocuteur, mais dans les implicatures découlant d'un énoncé.

Pour rendre compte de cette caractéristique, Carvalho et al. [2009] proposent différents traits. Parmi ceux-ci, on retrouve la présence d'éléments visant essentiellement à souligner un rapport condescendant envers une cible. Par exemple, en portugais, il est possible d'utiliser des diminutifs des noms propres d'un individu pour exprimer de la tendresse et de l'intimité. Néanmoins, lorsque ce genre de production est utilisé pour désigner quelqu'un d'inconnu ou quelqu'un qui occupe une position d'autorité, le locuteur crée un faux rapport de familiarité pouvant être interprété comme du mépris ou une façon d'infantiliser sa cible. Donc, Carvalho et al. [2009] signale qu'on s'attendrait à ce qu'un tweet contenant l'expression "Socratezinho" pour nommer un des premiers ministres du Portugal, José Socrates, soit ironique. Allant dans le même sens, Carvalho et al. [2009] rapportent que la présence de déterminants démonstratifs devant des noms propres qui serait aussi caractéristiques de certains énoncés sarcastiques. Par exemple, en (57), le déterminant "este" dénote un rapport méprisant envers Socrates.

57. Este Socrates 'e muito amigo do Sr.Jack¹

Toujours dans cette optique d'opérationnaliser la condescendance d'un locuteur, Carvalho et al. [2009] soulignent qu'en portugais, l'utilisation de pronom désignant la deuxième personne du singulier varie selon certains contextes. Comme en français, le "tu" serait utilisé dans des rapports sociaux plus familiers. Au contraire, "você" serait utilisé comme marqueur de politesse dans des rapports formels. Ainsi, Carvalho et al. [2009] proposent d'utiliser la présence d'un "tu" dans un contexte où on retrouverait le nom de quelqu'un de connu pour dénoter la présence d'ironie dans un énoncé. En effet, suivant les conventions sociales, on s'attend à la production d'un discours formel lorsqu'on parle à une personnalité connue. De plus, leur système utilise comme trait la mise en guillemets d'adjectifs positifs et de nom propre. Ces traits s'appuient essentiellement sur la connotation négative de l'ironie qui peut être utilisée comme façon d'insulter une

1. Ce Socrates est très ami avec Mr.Jack.

cible.

D'autres parts, comme mentionné précédemment (Cohn-Gordon and Bergen [2019]), un locuteur aurait plus de chances de produire un énoncé ironique lorsqu'il connaît bien son interlocuteur et les informations qu'ils partagent. De ce fait, Bamman and Smith [2015] proposent d'utiliser des traits qui représentent l'auteur d'un tweet et l'audience qu'il posséderait. Les traits relatifs à l'auteur du tweet se manifestent au travers des termes les plus saillants dans son historique déterminé par la fréquence des termes utilisés par rapport à la fréquence inverse des documents présents (TF-IDF), les sujets les plus abordés par l'auteur, les informations personnelles de l'auteur (son genre, son nombre d'abonnés, l'âge du compte, le nombre moyen de publications par jour, le fuseau horaire et la présence du crochet bleu signifiant que le compte est vérifié par twitter) et l'historique de la valence des tweets de l'auteur². L'ensemble de ces traits sont mis en relation avec des traits relatifs à l'audience du locuteur. Ces derniers comprennent : la similarité des sujets qu'ils abordent et l'historique de communication entre ces deux entités.

Si les modèles présentés plus haut appliquent une opérationnalisation des traits pragmatiques qui leur sont propres, on note tout de même certaines généralités transcendant les différents algorithmes proposés. Par exemple, faisant écho à la perspective de Wilson and Sperber [1992], l'utilisation d'interjections dans un tweet est souvent adoptée comme un trait significatif pour la présence d'ironie dans un énoncé (Carvalho et al. [2009], Kreuz and Caucci [2007], Karoui et al. [2015], Gupta and Yang [2017], Bouazizi and Ohtsuki [2015], Farías et al. [2016]). On observe la même chose pour la présence de ponctuation exagérée comme plusieurs points d'exclamations, d'interrogations ou de suspensions (Carvalho et al. [2009], Kreuz and Caucci [2007], Karoui et al. [2015], Joshi et al. [2015], Gupta and Yang [2017]). Aussi, plusieurs systèmes de détection (Carvalho et al. [2009], Bamman and Smith [2015], Bouazizi and Ohtsuki [2015], Farías et al. [2016]) intègrent la présence de rire explicite (e.g. : *lol*, *haha*, etc.) ou d'émojis (e.g. : :p) pour discriminer la présence d'ironie dans un énoncé.

2.2.5 Traits cognitifs

Gupta and Yang [2017] proposent un système de détection automatique de l'ironie qui rendrait compte des aspects affectifs, cognitifs et sociolinguistiques de ce type de discours. Pour opérationnaliser l'affect, Gupta

2. Pour déterminer cette dernière caractéristique, la valence de l'ensemble du tweet est évalué en utilisant le Stanford Sentiment Analyzer (Socher et al. [2013]). Bamman and Smith [2015] évaluent aussi la valence au niveau des mots d'un énoncé en se basant sur le dictionnaire de Warriner et al. [2013] et de Thelwall et al. [2010]

and Yang [2017] se basent sur l'idée selon laquelle l'ironie tenterait parfois de partager une attitude négative au travers de termes positifs. Ainsi, ce trait comptabiliserait le nombre de mots positifs, le nombre de mots négatifs, leurs positions et l'ordre qu'ils occupent dans une phrase. Si cette approche semble similaire à celle de Joshi et al. [2015], en plus de la polarité des mots, ils proposent de vérifier la valence de ces derniers grâce à un lexique préalablement établi. Selon eux, la présence de mots extrêmement positifs pourrait être une des caractéristiques de l'ironie faisant ainsi écho aux types d'ironie prenant naissance dans des hyperboles. Pour représenter l'aspect cognitif, Gupta and Yang [2017] partagent l'idée qu'un énoncé ironique seraient nécessiteraient des fonctions cognitives plus complexes. De ce fait, cette complexité se refléterait dans le choix des mots d'un énoncé. Par exemple, des mots comme des conjonctions, des prépositions et des mots de plus de six lettres impliqueraient des processus cognitifs plus complexes selon Gupta and Yang [2017].

Se basant sur le principe mentionné plus haut que certaines conditions neurologiques peuvent nuire à une interprétation adéquate d'un énoncé ironique³, Gupta and Yang [2017] émettent l'hypothèse que la production d'un énoncé ironique serait plus complexe qu'un énoncé littéral. Cette complexité serait reflétée dans l'utilisation de mots liés à des processus (p.ex. : *because*), de conjonctions (p.ex. : *although*), de prépositions (p.ex. : *to*) et de mots ayant 6 lettres ou plus.

Farías et al. [2016] proposent aussi d'utiliser certains traits représentant l'affect d'un énoncé ironique. Pour déterminer l'émotion véhiculée par un tweet, ils se basent sur des bases de données préalablement construites. Pour avoir accès à d'autres informations affectives, ils utilisent le *Dimensional models of emotions* qui contient des éléments comme le niveau d'activation causé par des états affectifs, l'imagerie soit la capacité de se faire une image mentale d'un mot précis et le degré de plaisir causé par celui-ci.

2.2.6 Traits contextuels

Comme mentionné dans la section 8.2 du chapitre 1 de ce travail, le contexte joue un rôle primordial dans l'identification et l'interprétation d'un énoncé ironique. Néanmoins, cette caractéristique reste la plus dure à opérationnaliser en termes computationnels dans un algorithme de détection automatique de ce type de discours. En effet, pour faire écho à la notion d'informations partagées de Stalnaker [2002], il s'avère un peu improbable d'implémenter un ensemble dynamique pouvant représenter adéquatement ce type de données. Toutefois, certains chercheurs proposent des façons intéressantes d'affronter ce défi.

3. Le parallèle fait par Gupta and Yang [2017] ne rend pas justice à toute la littérature scientifique sur les processus psycholinguistiques associés à l'interprétation d'un énoncé ironique.

Par exemple, Karoui et al. [2015] proposent de capturer le contexte d'un tweet pour vérifier la présence d'ironie. En premier lieu, l'algorithme comporte un système de classification binaire qui tente de prédéterminer s'il est ironique ou non en se basant sur des traits de surface, des traits exprimant des sentiments, des traits vérifiant la présence de termes lexicaux exprimant des oppositions ou la présence de marqueurs d'opposition. Par la suite, si le tweet est considéré comme étant ironique, l'algorithme corrige ce résultat en dérivant le contexte externe en vérifiant si la présence de certains mots clés se retrouvent dans le titre d'articles de journaux. Par exemple, l'énoncé en 58 serait, dans un premier temps, segmenté en deux phrases. Ensuite, la négation et les mots fonctionnels sont retirés pour ne garder que "*Valls interior minister*". Ces mots sont utilisés pour faire une requête web où ils se retrouvent sur une page Wikipedia. Ainsi, l'algorithme catégorise cet énoncé comme étant ironique car, contrairement à ce qu'on retrouve dans l'énoncé original, Valls est le ministre de l'intérieur.

58. #Valls has learnt that Sarkozy was wiretapped in newspapers. Fortunaely he is not the interior minister.

2.2.7 Évaluation des traits

Les modèles présentés plus haut varient grandement de par leurs architectures. Il est difficile d'avoir un niveau de comparaison entièrement équivalent pour établir quels traits sont les plus ou les moins efficaces. Néanmoins, en gardant en tête que l'objectif de ce travail est de déterminer où se situent les faiblesses des systèmes de détection automatique de l'ironie, je crois qu'il est intéressant de faire l'inventaire de ces derniers pour vérifier lesquels offrent les meilleurs.

Pour Kreuz and Caucci [2007], c'est la présence d'interjections qui semblaient jouer un rôle déterminant pouvant permettre une identification adéquate de l'ironie dans un énoncé.

De leur côté, Carvalho et al. [2009] notent que les traits permettant de détecter ce type de discours seraient la présence de rire explicite et l'utilisation de guillemets. Néanmoins, cette caractéristique peut être problématique. Par exemple, on peut utiliser des guillemets pour nommer des entités physiques comme des villes sans être ironique. De façon similaire, les traits représentant les marqueurs stéréotypiques de la forme d'un énoncé sarcastique seraient les plus performants pour leur système de détection automatique.

Au delà de la forme, Reyes et al. [2012], quant à eux, ont observé que les traits liés au niveau du contexte émotionnel autour d'un énoncé faciliterait la détection automatique d'ironie. Plus particulièrement, ce serait des caractéristiques comme le niveau de plaisir (*pleasantness*), d'activation et de perplexité porté par les termes d'un énoncé. De plus, l'incongruité entre ces derniers et le contexte ferait aussi parti des traits utiles à la détection.

Des traits démographiques proposés par Bamman and Smith [2015], le fait d'être un homme aux États-Unis au compte twitter non-vérifié semblait être un indice de la présence d'ironie dans un énoncé.

L'ensemble de ces traits nous permettent d'avoir un portrait des trois composantes essentiels à la détection de l'ironie soit les marqueurs représentant la forme d'un énoncé, son contenu affectif et des caractéristiques représentant le locuteur qui produit l'énoncé.

2.3 Apprentissage machine

Les algorithmes d'apprentissage machine se distinguent des modèles présentés plus haut de par leur rapport avec les données d'entraînement. À partir de celles-ci, l'architecture de ce type de système est en mesure de faire émerger des structures lui permettant de réussir une tâche de classification. L'objectif derrière cette section est de présenter un survol du fonctionnement de ce type d'algorithme, de certains travaux tentant d'appliquer ce type de technologie à la détection de l'ironie et leurs résultats.

Typiquement, parmi toutes les architectures différentes d'apprentissage machine, on note, entre autres, les réseaux de neurones. De façon sommaire, comme le synthétisent Jurafsky and Martin [2021], ce type d'algorithme est caractérisé par différentes couches de traitement composées de plusieurs "neurones". Concrètement, les neurones sont des unités de calcul dont le résultat dépend d'une donnée en entrée multipliée par un vecteur représentant le poids. Les poids sont appris par une phase d'apprentissage sur les données. À ce produit, on ajoute la valeur d'un biais. À partir de données d'entraînement adéquates, on applique généralement des réseaux de neurones à des tâches de (re)production ou de détection.

Parmi ces types d'algorithme, on retrouve Poria et al. [2016] qui proposent un modèle basé sur un réseau neuronal convolutif (RNC) préentraîné pour extraire des traits concernant les sentiments, les émotions et la personnalité du locuteur. Concrètement, selon Poria et al. [2016], grâce à ce type de réseau, il est possible de former un vecteur englobant l'ensemble de traits locaux d'un énoncé lui permettant de créer une repré-

sentation adéquate du contexte lexical. En testant ce modèle sur différents jeux de données, ils arrivent à des scores-F1 supérieurs à 90% de détection adéquate de l'ironie verbale.

De leur côté, Ghosh and Veale [2016] utilisent un conjonction de différents types de réseaux neuronaux, soit un réseau de neurones composé d'un RNC, suivi d'un réseau de neurones récurrent (RNR). La première couche de cette architecture est celle des données langagières contenues dans un tweet qui est vectorisé. Ensuite, le résultat de ce traitement passe par une couche du RNC qui permettrait d'extraire des séquences de mots discriminatoires fournissant à la couche du LSTM les données adéquates. Ce dernier serait en mesure de créer une représentation sémantique. L'ensemble de ces couches arrivent à un score de précision de 91.9%, un score de rappel de 92.3% et un score-f de 92.1%. Malgré ces résultats, Ghosh and Veale [2016] notent tout de même que leur système de détection semble mal différencier des énoncés ironiques similaires comme "I just love Mondays !" et "Thank god it's Monday !".

Huang et al. [2017] quant à eux proposent une architecture composée de trois réseaux de neurones : un RNC, un RNR et un réseau neuronal récurrent attentif (RNRA). Ce dernier permet d'avoir une vue d'ensemble sur les données préalablement traitées. De ce fait, il est possible d'évaluer a posteriori quels mots ont reçu le plus d'attention de ce réseau et, donc, de déterminer les expressions lexicales les plus discriminantes dans l'interprétation d'un énoncé ironique. Dans le cadre de leur expérimentation, Huang et al. [2017] notent que leur RNRA semblait porter plus attention aux expressions positives, aux hyperboles et aux noms communs laissant sous-entendre que ce serait les caractéristiques de l'ironie verbale la plus commune pour le jeu de données tests qu'ils ont utilisé. Enfin, leur modèle arrive à un score de détection F1 de 89.8%.

2.4 Limitations

Les systèmes symboliques possèdent des limitations inhérentes à leur conception. Par exemple, le fait d'avoir à systématiquement écrire des règles découlant de cas de figure possibles nécessite une adaptation constante en cas d'exception. Cet ajout constant a un prix. On doit trouver un équilibre entre un nombre fini de ressources détenues par un ordinateur pour effectuer des calculs de façon optimale face à des règles pouvant être pratiquement infinies, mais pas opaques.

Les sous-sections suivantes tenteront de décrire plus précisément les lacunes des tentatives de modélisations précédemment exposées en les mettant en perspective avec les théories développées dans le chapitre 1 de ce travail. Se faisant, dans les sections suivantes, il sera question de caractéristiques importantes comme

la superposition des définitions concernant l'ironie et le sarcasme, l'opérationnalisation de la polarité d'un énoncé ironique, mais, aussi, de l'inconsistance entre les définitions de l'ironie verbale et la construction d'un corpus d'entraînement représentant ce type de phénomène.

2.4.1 Problème de définition : ironie ou sarcasme ?

Un des enjeux centraux relatifs à la détection automatique de l'ironie se situe au niveau de la définition même de ce concept. La majorité des travaux concernant ce sujet semble éviter d'aborder cette question. Paradoxalement, la façon dont on choisit d'appréhender cette question influence nécessairement les méthodes employées pour arriver à une détection adéquate de ce type de discours. Lorsqu'on en retrouve, elles sont essentiellement superficielles. Par exemple, Carvalho et al. [2009] caractérisent l'ironie verbale similaire comme :

[...] As the rhetorical process of intentionally using words or expressions for uttering a meaning different (usually the opposite) from the one they have when used literally.

Dans le même sens, Van Hee [2017] propose de décrire l'ironie comme :

an evaluative expression whose polarity (i.e. positive, negative) is inverted between the literal and the intended evaluation, resulting in an incongruence between the literal evaluation and its context.

Ces définitions ressemblent à celles proposées par Kerbrat-Orecchioni [1978] et Grice [1975]. Par conséquent, percevoir l'ironie verbale comme étant une négation du sens littéral ou comme un rapport de polarité inverse négligent un ensemble important d'énoncés ironiques s'exprimant autrement.

D'autre part, dans la littérature sur la détection automatique de l'ironie, on retrouve fréquemment une juxtaposition de ce terme avec celui désignant le sarcasme. On présente parfois ces deux concepts comme étant interchangeables. Par exemple, Davidov et al. [2010] écrivent :

Sarcasm (also known as verbal irony) is a sophisticated form of speech act in which the speakers convey their message in an implicit way.

D'autres fois, on présentera l'ironie comme étant une supracatégorie pouvant contenir des énoncés sarcastiques (sans expliquer comment) comme dans Farías et al. [2016] : "*irony is here considered an umbrella term that also covers sarcasm*".

Il est toutefois important de noter que, comme il a été démontré dans le chapitre 1 de ce travail, l'ironie verbale est extrêmement difficile à définir. De plus, les frontières entre le sarcasme et l'ironie sont difficiles à cerner. Dans le même sens, Van Hee [2017] rapportent :

[I]n fact, among others, Tsur et al. (2010) and Eisterhold et al. (2006) claim that there is no way of formally distinguishing between the terms, and many researchers consequently consider sarcasm and irony as strongly related (Hallmann et al., 2016). Another reason why researchers do not differentiate between irony and sarcasm is the observation of a shift in meaning between the two terms. Over time, the term 'sarcasm' seems to have gradually replaced what was previously designed by 'irony' (Nunberg2001). In their experimental study, Bryant and Fox Tree (2002) and (Gibbs 1986) both found evidence for this meaning shift, observing that student respondents understood the term 'sarcasm' better than 'verbal irony'.

Les points apportés par Van Hee [2017] sont valides. Toutefois, ils font fi de tout un pan de la littérature qualifiant le sarcasme qui possède une connotation plus négative que le terme ironie (Kreuz and Glucksberg [1989]). De plus, on n'évalue pas comment ces problèmes affectent indéniablement les processus de détection. Définir adéquatement le phénomène que l'on souhaite observer à des impacts qui influencent un ensemble de facteurs primordial. De ce fait, une attention particulière devrait être accordée à cette question.

Il est difficile de résumer mieux la situation que par les mots de Goddard [2018] :

So briefly, the terminological cycle goes something like this. (i) One starts with ordinary English words, poorly defined or undefined, then (ii) "technicalizes" them and extends their range, often making some formal adjustments along the way, e.g. using adjective + nominalization combinations or coining backformations such as a tease, or a bant. (iii) Subsequently there is uptake, i.e. different scholars begin to employ the terms, often using them in slightly different ways from the original authors.¹ (iv) Scholarly debate begins about what the new terms mean

or should mean, how the terms relate to one another, etc. It may be swimming against the tide to say so but in my view the confusion of terminology betokens a confusion of concepts.

2.4.2 La polarité d'un énoncé ironique

Comme mentionné dans le chapitre 1, la polarité d'un énoncé ironique n'est pas systématiquement négative. Alba-Juez and Attardo [2014] ont mis en avant les différents cas de figure où un énoncé ironique peut avoir une polarité positive, négative et neutre. De plus, parfois, un même énoncé peut être positif envers une cible tout en étant négatif envers une autre.

Ainsi, la plupart des algorithmes présentée dans la section 2.3 de ce chapitre négligeront nécessairement des énoncés ironiques comme celui-ci préalablement présentés dans le chapitre 1 :

59. Belle partie! [Déclaré suite à une défaite]

60. Quelle horrible partie! [Déclaré suite à une victoire]

En (59) et en (60), ce n'est pas la polarité de l'énoncé qui permet l'identification d'ironie verbale dans celui-ci, c'est la dichotomie qu'il manifeste relativement au contexte. Il est même possible de retrouver des énoncés ironiques qui sont neutres en apparence comme en 61. Ainsi, s'appuyer exclusivement sur cette première caractéristique en négligeant la seconde risque d'engendrer des erreurs de détection.

61. C'était une partie. [Déclaré à la suite d'une défaite extrêmement rapide.]

2.4.3 Corpus

Généralement, les corpus d'énoncés ironiques utilisés pour entraîner des algorithmes se construisent de deux façons (Farías et al. [2016]). Dans la première, on les retrouve directement et explicitement indiqués comme tels par les locuteurs les produisant. Concrètement, sur twitter, on aurait des tweets contenant des hashtags comme *#sarcasm*, *#sarcastic* ou *#nottrue*. Derrière l'élaboration de ce type de corpus, on prend pour acquis que le locuteur serait le mieux placé pour savoir si un énoncé qu'il produit est ironique ou non. De même, on assume que le locuteur le signale, et que les tweets ouvertement signalés comme étant ironiques auraient les mêmes caractéristiques que les tweets ironiques non ouvertement signalés comme

tels. Ce choix méthodologique amène certains problèmes. Effectivement, on sait peu de choses sur l'homogénéité des données. Concrètement, il est difficile de déterminer la fréquence d'énoncés similaires en termes de contenus ou de forme qui se retrouvent dans ce type de corpus. Ce manque de granularité entre les éléments collectés brouille nécessairement l'impact réel des traits choisis pour détecter la présence d'ironie contenue dans ces derniers. De plus, les hashtags mentionnés plus haut sont généralement utilisés dans des tweets où l'ironie n'est pas évidente à détecter pour les interlocuteurs. Ils servent de raccourcis permettant à ces derniers d'accéder à une interprétation adéquate de l'énoncé sans nécessairement avoir accès aux informations contextuelles requises. Par exemple, un tweet qui contiendrait la phrase "J'aime les chats" ne pourra pas être adéquatement traité comme étant ironique si les interlocuteurs n'ont pas accès à l'information préalable que l'utilisateur déteste cette espèce de félin. En ajoutant un #sarcasme, on évite la nécessité d'avoir recours au contexte, mais on se retrouve avec un énoncé qui possède peu (ou pas) de caractéristiques particulières à l'ironie. De même, il est possible de passer à côté d'énoncés ironiques qui ne nécessitent pas l'ajout de hashtag particulier tellement leur interprétation est évidente.

La deuxième façon de construire des corpus d'énoncés ironiques mentionnés par Farías et al. [2016] permet de résoudre en parti le problème mentionné plus haut. On demande tout simplement à des annotateurs de vérifier le contenu de corpus et de noter la présence d'ironie dans chacun des énoncés. Par exemple, dans l'algorithme de Bouazizi and Ohtsuki [2015], des participants devaient annoter le niveau de sarcasme sur une échelle à 6 options, d'extrêmement ironique à extrêmement non ironique, pour les 6000 tweets collectés automatiquement comportant le #sarcasm. Si cette méthode permet un certain contrôle concernant la qualité des tweets présents dans le corpus, on ne sait pas sur quels critères les juges se basent pour effectuer leurs jugements. À ce sujet, Farías et al. [2016] écrit : "*Mainly, the labeling process is carried out without any strict definition or guideline. Therefore, it represents a subjective task, where the agreement between annotators is often very low.* Ce rapport subjectif apparaît évident lorsqu'on regarde la définition de sarcasme utilisée par Bouazizi and Ohtsuki [2015] : "*[sarcasm is] a special form of irony by which the person conveys implicit information, usually the opposite of what is said, within the message he transmits.*"

Les problèmes méthodologiques sous-jacents à ces deux méthodes de collecte nous renvoient directement au coeur du problème évoqué par ce mémoire. L'ironie est un type de discours difficile à définir, mais il est nécessaire de s'attarder sur les caractéristiques qui lui sont propres pour être en mesure de développer un système de détection automatique qui fonctionne.

Cette relation entre la définition du concept étudié, l'élaboration d'un algorithme de détection automatique et le choix des éléments constituant les jeux de données permettant d'extraire des traits ou de tester leur justesse est particulière. Si l'élaboration de l'algorithme de détection de l'ironie dépend de la caractérisation de ce type de discours par les chercheurs, l'élaboration du corpus test doit nécessairement refléter cet aspect. Dans le cas contraire, on peut s'attendre à la présence de faux positifs ou de faux négatifs.

2.5 Qu'apprennent réellement les algorithmes d'apprentissage machine ?

Les approches se basant sur les techniques d'apprentissage machine nous offrent des résultats mitigés concernant la détection automatique de l'ironie. Comme pour les méthodes symboliques, elles réussissent généralement bien dans les conditions tests. Néanmoins, de par la nature de ces méthodes, il nous est impossible de savoir concrètement ce que ces algorithmes "apprennent". Aussi, on sait peu de choses sur la possibilité de généraliser ces apprentissages à un milieu plus écologique. De ce fait, le problème d'opacité des modèles fait surface. London [2019] souligne ce rapport paradoxal où l'architecture des modèles d'apprentissage machine sont bien compris, mais les processus permettant d'arriver à des résultats concluants sont plus difficiles d'accès.

Plus précisément, concernant la détection automatique de l'ironie, Wallace [2015] dira :

Current machine learning methods rely too heavily on shallow, unstructured, syntactic modeling of text to consistently discern ironic intent. Irony detection is an interesting machine learning problem because, in contrast to most text classification tasks, it requires a semantics that cannot be inferred directly from word counts over documents alone.

Même s'il est clair pour Wallace [2015] que l'objectif derrière ces modèles computationnels n'est pas de théoriser l'ironie verbale ou d'apporter une caractérisation complète de ce phénomène, il souligne tout de même plusieurs problèmes importants. Comme synthétisé par sa citation plus haut, la plupart des algorithmes d'apprentissages machines proposées jusqu'à maintenant semblent traiter les données langagières de façon superficielle. L'interprétation d'un énoncé ironique nécessite une extrapolation du sens qui dépasse la forme des éléments qui le constitue. Allant dans le même sens, Bender and Koller [2020] précisent que si le corpus d'entraînement d'un algorithme d'apprentissage machine ne se base que sur ce type de caractéristiques, il lui sera impossible d'inférer la relation entre la forme d'un énoncé et l'intention du locu-

teur. Concrètement, il est nécessaire d'utiliser ce type d'architecture conjointement à des traits considérés caractéristiques de l'ironie en se basant sur une définition claire et exhaustive du phénomène. Sinon, il nous sera impossible de savoir quels types ou quels formes d'ironies posent le plus de problème au niveau de la détection automatique.

2.6 Modèles pragmatiques computationnels

Les sections précédentes ont démontré les limites principales des systèmes de détection automatique de l'ironie. Ces dernières semblent intrinsèquement liées aux structures algorithmiques utilisées pour y arriver et aux corpus utilisés. En effet, d'un côté, les méthodes symboliques restent limitées quant aux traits utilisés. De l'autre, on ne sait pas réellement ce qu'apprennent les algorithmes d'apprentissage machine. À ce point-ci, il est tout à fait juste de se demander comment trouver une façon de bien rendre compte de ce type de discours tout essayant de trouver un équilibre entre les caractéristiques particulières à l'ironie et les limites d'informations que peut traiter une machine. Parmi les pistes de solutions intéressantes, il existe plusieurs travaux tentant de modéliser l'ironie en termes probabilistes.

Par exemple, Wallace [2015] propose une représentation formelle opérationnalisable computationnellement tenant compte de caractéristiques qu'il juge importantes pour l'élaboration d'un système de détection de l'ironie. Comme mentionné dans la section précédente, Wallace [2015] juge qu'une des grandes faiblesses de ce type d'architecture concerne l'absence de représentation concrète de la relation entre le locuteur et son environnement. Pour lui, les attentes de l'interlocuteur sont une des composantes essentielles permettant l'interprétation adéquate d'un énoncé ironique. Cet aspect, qu'il qualifie de "contexte pragmatique", contient l'ensemble des connaissances partagées par l'interlocuteur et le locuteur sur ce dernier. De même, le locuteur qui ironise doit posséder une bonne connaissance des éléments constitutifs de cet espace pour être en mesure d'émettre ce type de discours en sachant qu'il sera adéquatement interprété. Par exemple, si B entend de A qu'il est très satisfait des résultats d'une élection où le parti libéral aurait gagné, il lui sera possible d'interpréter l'ironie de cet énoncé seulement en possédant l'information que A déteste ce parti. De même, A aura plus de chance de produire cette phrase s'il sait que B connaît son allégeance politique. De ce fait, pour Wallace [2015], l'ironie prend ainsi forme dans le rapport d'incongruité entre ce contexte pragmatique et l'énoncé que le locuteur émet. Ce rapport est affecté par le niveau de confiance de l'interlocuteur dans sa représentation interne du locuteur : "*the more we know someone, the better able we are to infer when they are being ironic*" (p.10). Aussi, pour Wallace [2015], la présence

d'indices syntaxiques dans un énoncé ironique est non négligeable pour son interprétation adéquate. En effet, moins un interlocuteur a confiance en sa représentation interne d'un locuteur, plus il lui sera nécessaire d'utiliser des indices internes à l'énoncé pour interpréter l'ironie.

De ces éléments, Wallace [2015] propose une formalisation bayésienne. Concrètement, ce type d'approche propose de représenter le processus d'interprétation d'un énoncé en termes de probabilité. La règle de Bayes s'écrit comme suit :

$$P(H|E) = \frac{P(E|H) * P(H)}{P(E)} \quad (2.1)$$

La variable H rend compte d'une hypothèse et la variable E d'une observation. Cette règle permet de comparer l'hypothèse la plus probable après une observation donnée. Pour ce faire, on multiplie la vraisemblance de l'observation ($P(E|H)$) par la croyance à priori (H). Le tout est divisé par la probabilité de l'observation ($P(E)$). Cette dernière est constante.

Cette formule attribuée à Bayes nous permet de représenter les rapports inférentiels sous-jacents à l'interprétation sémantique ou pragmatique d'un énoncé. De ce fait, elle permet d'organiser les variables décrites par Wallace [2015] de la façon suivante :

$$P(I(u)) \propto \frac{P(I(u)|M_s(a))}{Var(M_s(a))} * P(I(u)|S(u)) \quad (2.2)$$

Pour Wallace [2015], cette modélisation permet de représenter la probabilité qu'un énoncé u soit interprété comme étant ironique ($P(I(u))$). Celle-ci varie de façon proportionnelle aux deux parties distinctes suivantes. D'un côté, on évalue la probabilité qu'un énoncé u soit ironique en fonction de notre connaissance du modèle interne du locuteur soit $M_s(a)$. La variable (a) représente les caractéristiques connues du locuteur. Par exemple, les positions politiques de ce dernier pourraient en faire partie. Le tout est divisé par la confiance inverse que l'interlocuteur a de ce modèle soit la variation possible de celui-ci $Var M_s(a)$. De l'autre, on tient compte de la probabilité qu'un énoncé u soit ironique en prenant en considération les indices syntaxiques soit $P(I(u)|S(u))$.

Wallace [2015] nous avertit que cette formalisation est incomplète. Elle sert avant tout de démontrer qu'il est possible de créer une représentation computationnelle du contexte partagé entre des locuteurs qui peut être opérationnalisés de façon optimale. Par exemple, il propose d'évaluer un scénario où on retrouverait une critique d'un restaurant BBQ comme en 62.

62. On sait tous à quel point j'aime la viande. Donc, on sait tous à quel point j'aime ce restaurant.

Dans cet énoncé, le contenu lexical en lui-même nous donne peu d'informations nous permettant d'évaluer avec certitude la présence d'ironie verbale. On note l'expression "on sait tous" qui fait référence à un ensemble d'informations partagées entre le locuteur et ses proches. Pour Wallace [2015], on pourrait opérationnaliser ce modèle interne du locuteur en prenant en compte le nombre d'informations disponibles concernant les opinions de ce dernier sur un sujet précis. Dans ce cas-ci, on pourrait vérifier les opinions exprimées du locuteur en lien avec la viande ou l'alimentation. Ainsi, si on apprend que le locuteur en 62 est végétarien, on peut donc inférer que l'énoncé est probablement ironique. En résumé, notre modèle interne du locuteur, $Ms(a)$, est représenté par les informations qui confirment nos hypothèses à propos de ce que pense réellement le locuteur et à l'inverse, les informations contradictoires viennent augmenter la valeur de notre confiance inverse en ce modèle. Dans l'éventualité où aucune information n'est disponible sur le locuteur, on doit se fier qu'aux informations lexicales et syntaxiques de l'énoncé.

Bien qu'intéressante, la composante contextuelle proposée par Wallace [2015] reste relativement floue. On y néglige, entre autres, l'aspect affectif sous-jacent à la production d'un énoncé ironique. Il existe d'autres représentations formelles similaires. Cependant, bien que ces dernières mettent en lumière des composantes intéressantes de l'ironie, elles n'ont pas comme objectif d'être implémentées directement dans des algorithmes.

Parmi ces dernières, on retrouve celle proposée par Kao and Goodman [2015]. Selon eux, pour modéliser l'ensemble d'informations partagées avec un locuteur, on doit nécessairement prendre en compte une caractéristique émotionnelle. En se basant sur le modèle des *Rational Speech Acts*, il leur est possible de conceptualiser une représentation plus fine des rapports inférentiels se produisant entre un locuteur et son interlocuteur. Ce type de formalisation tient compte du rapport récursif au niveau des informations partagées entre le locuteur et son interlocuteur. Kao and Goodman [2015] utilisent une extension de ce modèle, *question under discussion* (QUD), qui prend en compte la sujet de la conversation en cours. Ils proposent

d'utiliser l'incongruité entre l'état du monde réel et l'état affectif du locuteur pour évaluer la présence d'ironie dans un énoncé. Néanmoins, même si le travail de modélisation est extrêmement différent de celui issu du travail de Joshi et al. [2015], on arrive à des problèmes similaires où on se retrouve incapable de rendre compte adéquatement d'énoncés ironiques qui ne sont pas issus d'incongruité sémantique répétant ainsi les critiques mentionnées par Wilson and Sperber [1992] dans la section 4.1 du chapitre 1 de ce travail.

De ces limitations, Cohn-Gordon and Bergen [2019] s'inspirent des points apportés par les caractérisations de l'ironie par Wilson and Sperber [1992] et Clark and Gerrig [1984]. Comme rapporté à la section 8.3 du chapitre 1 de ce travail, pour Cohn-Gordon and Bergen [2019], le jeu du faire semblant permet à l'ironiste de mettre à jour ou de vérifier l'espace d'informations partagées entre lui et son interlocuteur. Concrètement, la modélisation proposée par Cohn-Gordon and Bergen [2019] s'inspire de celle de Kao and Goodman [2015]. Ils définissent un interlocuteur qui est en mesure de déceler que le locuteur joue un rôle. Ce dernier s'exprime sur le monde comme s'il n'en faisait pas partie. De ce fait, l'interlocuteur doit posséder une connaissance a priori des probabilités que le locuteur fasse semblant de croire à un état du monde différent du monde réel. Pour Cohn-Gordon and Bergen [2019], l'ironie verbale est une façon de contresigner (*countersignaling*) une information. Ce faisant, cette modélisation permet de rendre compte de plusieurs questions entourant l'ironie et ce qu'elle communique soit à la fois l'état du monde en utilisant du contresignement, mais aussi le contenu et l'état de l'espace d'informations partagées entre le locuteur et son interlocuteur.

2.7 Conclusion

Ce second chapitre offre un aperçu global des différentes approches computationnelles utilisées pour tenter d'identifier automatiquement la présence d'ironie verbale. Dans un premier temps, il fut question de cerner comment on pouvait opérationnaliser certaines caractéristiques de ce type de discours dans des algorithmes à base de règles. Pour ensuite, vérifier comment les nouvelles architectures à base d'apprentissage machines peuvent tenter de relever ce défi. Pour finalement, observer comment modéliser de façon formelle l'ironie verbale.

Les théories rapportées dans le premier chapitre permettent de mettre en perspective les méthodes computationnelles présentées dans cette deuxième partie de mon mémoire. Ce qui ressort principalement des limitations propres aux différentes approches, c'est un rapport inégal au niveau de l'attention portée vers l'architecture des systèmes plutôt qu'au phénomène étudié. Les modélisations pragmatiques rapportées

dans la section 5 de ce chapitre démontrent qu'il est tout à fait possible de concevoir des systèmes efficaces en se concentrant sur certaines caractéristiques de l'ironie. De ce fait, le prochain chapitre tentera d'évaluer quels types d'ironies est plus difficiles à détecter et pourquoi.

CHAPITRE 3

IRONISER AVEC BERT

3.1 Introduction

Dans le premier chapitre de ce travail, il a été question des différentes perspectives théoriques en linguistique tentant de rendre compte de la réalité complexe de l'ironie verbale. En effet, les différents points de vue rapportés démontrent qu'il est difficile de proposer une définition exhaustive et adéquate de ce phénomène. Dans le deuxième chapitre, un portrait général des différentes tentatives d'opérationnalisation computationnelles de l'identification automatique de ce type de discours a été exposé. Cette mise en perspective permet l'émergence de certains angles morts concernant des caractéristiques souvent négligées de l'ironie verbale. Concrètement, dans les algorithmes rapportés précédemment, plusieurs problèmes liés aux nuances entre l'ironie et le sarcasme, à la polarité que peut prendre un énoncé ironique et à l'élaboration d'un corpus d'entraînement ont été soulignés.

Ainsi, l'objectif de ce chapitre est de vérifier si ces constats théoriques peuvent se manifester dans certains patrons d'erreur d'un algorithme. Plus spécifiquement, ce chapitre a pour but de tester les performances de BERT, un algorithme d'apprentissage machine finement ajusté sur les données de SemEval 2018, sur la détection d'énoncés ironiques collectés et préalablement classés par Beals [1995]. Je cherche à évaluer sa capacité à généraliser ses apprentissages sur d'autres types d'ironie. Cette analyse qualitative permettra de déterminer quels types d'ironies sont plus difficiles à détecter. De ce fait, il sera possible d'évaluer si ces patrons d'erreurs sont causés par la nature même du corpus d'entraînement.

3.2 BERT

3.2.1 Fonctionnement

BERT, *Bidirectional Encoder Representations from Transformers*, est une architecture conçue par Devlin et al. [2018]. Contrairement à d'autres modèles mentionnés plus tôt, il ne se base pas sur des traits (feature-based) préencodés pour répondre à une tâche spécifique. BERT est pré-entraîné sur d'importants jeux de données comme le BooksCorpus (Zhu et al. [2015]) qui contient 800 millions de mots et sur l'entièreté des textes en anglais présents sur Wikipédia. Il est possible de réutiliser ses représentations préencodées à l'intérieur du modèle et de l'ajuster finement sur d'autres données pour exécuter une tâche précise. Si

BERT semble démontrer de bonnes capacités syntaxiques (Goldberg [2019]), la section suivante explicitera davantage ses capacités linguistiques en présentant ses résultats à des tests standardisés.

3.2.2 Performances

BERT s'est imposé comme modèle de langue de par son accessibilité et de par ses performances. Pour évaluer ces dernières, Devlin et al. [2018] l'ont soumis à une batterie de test standardisée comme le *General Language Understanding Evaluation* (GLUE) (Wang et al. [2018]), le *Stanford Questions Answering Dataset* (SQuAD) (Rajpurkar et al. [2016]) et le *Situation With Adversarial Generations* (SWAG) (Zellers et al. [2018]).

Le premier test, GLUE, contient un ensemble de tâches de compréhension du langage naturel (NLU). Parmi celles-ci, on retrouve trois types d'épreuves. Elles permettent d'évaluer si le modèle testé est en mesure de catégoriser adéquatement des phrases comme étant grammaticales ou non, s'il est en mesure de juger adéquatement de la similarité entre deux énoncés et s'il est capable de noter si une phrase en implique une autre. À ces tâches, BERT performe mieux que l'ensemble des modèles de langues existant avant lui en affichant un score de 82.1% pour BERTlarge.¹

Le second test, le SQuAD 1.1, est basé sur un jeu de données utilisé pour évaluer la compréhension de lecture. Construit à partir d'articles postés sur Wikipédia, chaque réponse provient d'un segment du texte d'où est tiré la question. Encore une fois, BERT performe efficacement manifestant un score de F1 0.932 pour la version BERTlarge.

Le troisième test, le SWAG, permet d'évaluer l'utilisation du sens commun (*common sense*) d'un individu ou d'un algorithme. Pour ce faire, les participants ou le modèle de langue sont confrontés à un début de phrase qui sert de contexte. Ils doivent choisir parmi quatre possibilités le reste de l'énoncé qui a le plus de sens. Par exemple, pour la phrase en 1, ce sera la réponse 1d qui sera la bonne.

1. Sur scène, une personne prend place devant un piano. Elle
 - (a) s'assoit sur un banc pendant que sa soeur joue avec une poupée.
 - (b) sourit à quelqu'un pendant que la musique joue.

(d) _____

1. une version de BERT qui contient 24 encodeurs et 16 têtes bidirectionnelles attentives. La version BERT_base contient 12 encodeurs et 12 têtes bidirectionnelles attentives.

(c) est dans la foule, en train de regarder les danseurs.

installe nerveusement ses doigts sur les touches.

À cette tâche, BERT performe mieux que les architectures précédemment testées. En effet, BERTbase possède un score F1 de 88.5%.

Ainsi, par ses résultats, BERT se positionnait à sa sortie comme l'état de l'art pour certaines tâches de traitement du langage automatisé. De plus, la possibilité de le régler finement en utilisant des corpus d'entraînement spécialement annotés permet de l'adapter facilement à différentes tâches. De ce fait, BERT semble mieux réussir que d'autres architectures sur des problèmes similaires. Par exemple, Emmerly et al. [2022] en entraînant ces trois architectures pour détecter la présence de harcèlement sur internet en utilisant des jeux de données spécifiquement conçus pour ce type de tâche soulignent les résultats supérieurs de BERT relativement à d'autres architectures, par exemple, basée sur des réseaux Long-Short term memory (LSTM).

Test	Modèle	Score F1
GLUE	BERT_base	76.6
	BERT_large	82.1
SQuAD 1.1	BERT_base	88.5
	BERT_large	90.9
SWAG	BERT_base	81.6
	BERT_large	86.6

Table 3.1 Résumé des performances de BERT rapportées par Devlin et al. (2018)

Au-delà de ses performances, il est possible d'utiliser une approche qualitative pour évaluer les forces et les faiblesses de BERT. Par exemple, Ettinger [2020] propose de mesurer les aptitudes de BERT en utilisant des tests psycholinguistiques standardisés qui se basent sur la complétion d'énoncés. Pour ce faire, il utilise le CPRAG-102 (*common sense and pragmatic inference*; Federmeier and Kutas [1999]) qui évalue les raisonnements nécessitant l'utilisation de bon sens (*common sense*) et les inférences pragmatiques, le ROLE-88 (*Event Knowledge and Semantic Role Sensitivity*; Chow et al. [2016]) qui évalue les connaissances événementielles et la sensibilité aux rôles sémantiques, le NEG-136 (*negation*; Fischler et al. [1983]) qui évalue la compréhension de la négation. BERT réussit bien dans les tâches où il est nécessaire de distinguer les éléments d'une même catégorie et où il doit de comprendre l'inversion des rôles dans une phrase. Il serait aussi

en mesure d'associer adéquatement des noms et leurs hyperonymes. Toutefois, il réussit moins bien dans des tâches nécessitant d'avoir accès au sens commun et de faire des inférences pragmatiques. Toutefois, comme le rapportent Kwon et al. [2019], BERT semble en mesure d'acquérir des nouvelles connaissances contextuelles issues des données utilisées pour l'ajuster finement. Ces capacités étant primordiales pour la perception adéquate de l'ironie présente dans un énoncé, elles risquent d'affecter directement les performances de BERT dans ce type de tâche.

3.2.3 BERT et l'ironie

Même si BERT réussit bien dans des tâches de compréhension langagière avec des jeux de données comme GLUE, SQuAD et SWAG, on sait peu de choses sur sa capacité à bien détecter des types de discours plus ambigus comme l'ironie.

On retrouve plusieurs travaux ayant ajusté finement BERT pour détecter automatiquement l'ironie. Par exemple, Baruah et al. [2020] ont présenté, dans le cadre du second workshop de l'ACL 2020 les résultats de leur modèle sur le traitement du langage figuratif dans la tâche relative à la détection du sarcasme. Au cours de cette tâche, les participants devaient produire un algorithme capable de classer un énoncé comme étant sarcastique ou non en utilisant le contexte des énoncés. Pour évaluer le rôle de ce dernier, l'entièreté de la conversation des utilisateurs contenue dans le jeu de données était disponible. Ceux-ci provenaient de Twitter et de Reddit. Baruah et al. [2020] comparent les résultats d'un Bidirectionnel Long Short-Term Memory (BiLSTM), de BERT et d'un classificateur Support Vector Machine (SVM). Pour évaluer l'influence du contexte conversationnel, ils ont présenté les phrases réponses contenant le sarcasme à ces différents modèles suivant 5 contextes différents : (1) en utilisant seulement la réponse sans le contexte (2) en utilisant la réponse avec la dernière phrase du dialogue la précédent (3) en utilisant la réponse avec les deux dernières phrases la précédent (4) en utilisant les trois dernières phrases la précédent (5) en utilisant toutes les phrases disponibles composant le contexte.

L'impact du contexte conversationnel dans ces tâches est mitigé. BERT a mieux performé que les autres modèles. Dans le cas des données Twitter, son meilleur score F1 de 0.743 s'est manifesté lorsqu'on retrouve la réponse et la dernière phrase le précédent. Dans le cas des données Reddit, les meilleures performances de BERT, F1 de 0.658, se présentent seulement lorsqu'on lui présente la réponse sans aucun contexte.

De leur côté, Adaikkan and Durairaj [2020] ont fait le même genre de travail que Baruah et al. [2020]. Ils

observent des résultats similaires. En effet, BERT performe mieux que les autres modèles. Il produit un score F1 de 0.722 pour le jeu de données de Twitter et de 0.679 pour le jeu de données issu de Reddit.

D'autres tâches ont créé des contextes pour tester les capacités de BERT à détecter l'ironie. Par exemple, lors de SemEval 2018, les participants devaient répondre à deux défis. Pour la tâche A, ils devaient déterminer si un tweet en anglais est ironique. Pour la tâche B, ils devaient identifier le type d'ironie de l'énoncé. Pour cette dernière tâche, les organisateurs de SemEval 2018 distinguent trois types d'ironie : l'ironie verbale causée par un contraste de polarité entre le sens souhaité et le sens littéral d'un énoncé, l'ironie situationnelle où une situation est ironique et d'autres types d'ironie verbale sans contraste de polarité. Par exemple, l'énoncé 1 est, selon Van Hee et al. [2018], un énoncé ironique par contraste de polarité. Ce dernier est causé par la dichotomie entre la polarité du mot "love" qui est positif et "migraine" qui est négatif. Pour exemplifier l'ironie situationnelle, Van Hee et al. [2018] propose l'exemple en 2, où l'ironie découle du rapport entre avoir un trouble du déficit de l'attention et ne pas être en mesure de se concentrer dans un cours portant sur ce sujet. Pour la catégorie contenant les autres types d'ironie, on retrouve des énoncés comme en 3 où l'ironie ne provient pas d'un contraste de polarité.

1. I love waking up with migraines #not
2. Most of us didn't focus in the #ADHD lecture. #irony
3. Human brains disappear every day. Some of them have never even appeared. <http://t.co/Fb0Aq5Frqs>
#brain #humanbrain #Sarcasm

Le plus haut score émergent de cette tâche en utilisant BERT provient de Zhang et al.. Ils arrivent à avoir un score F1 de 0.700 pour la tâche A. Néanmoins, les meilleurs résultats avec ce modèle qu'ils obtiennent pour la tâche B est un score F1 de 0.445. De façon intéressante, ils rapportent que leur version finement ajustée de BERT semble difficilement reconnaître l'ironie dite situationnelle (F1 de 0.382) et, mais surtout, qu'elle est incapable de reconnaître l'ironie classée dans la catégorie "autre".

De façon similaire, Agrawal et al. [2020] ont utilisé le modèle BERT-base finement ajusté sur les données de SemEval 2018. Pour la tâche A, ils arrivent à un score F1 de 0.70. Pour la tâche B, ils arrivent à un score F1 de 0.75.

D'autres versions tirées de BERT ont aussi été testées pour ce type de tâche. Par exemple, Potamias et al.

[2020] ont testé les capacités de RoBERTa Liu et al. [2019] qui est essentiellement une version de BERT préentraîné sur des mini-batches et des cycles d'entraînement plus larges². En effet, RoBERTa utilise 160 gigs de données d'entraînement v.s. 16g pour BERT, 500 000 epochs d'entraînement au lieu des 100 000 de BERT et 8 fois la largeur des batch de ce dernier. Ils ont aussi utilisé les données de SemEval 2018. Leur modèle leur donne un score F1 de 0.89.

Auteur·rices	Modèle	Score F1
Zhang et al. (2020)	BERT_large	Tâche A : 0.700
	BERT_large	Tâche B : 0.445
Agrawal et al. (2020)	BERT_base	Tâche A : 0.700
	BERT_base	Tâche B : 0.750
Potamias et al. (2020)	RoBERTa	Tâche A : 0.780
	n/a	n/a

Table 3.2 Performances de BERT sur les données de SemEval 2018

Les travaux présentés dans cette section suivent des tendances similaires. On y présente le modèle utilisé et les résultats qui en découlent. Les corpus d'entraînement sont sensiblement les mêmes. Bien que l'objectif principal de ces travaux soit d'évaluer les performances des modèles, on y voit peu de questionnement portant directement sur l'ironie. De ce fait, il devient difficile d'imaginer comment améliorer leur performance. De façon assez cohérente avec la littérature, Adaikkan and Durairaj [2020] diront qu'il serait pertinent d'avoir plus de données d'entraînement pour avoir de meilleurs résultats. Cependant, comme le mentionne Bender et al. [2021], l'augmentation de la taille des données d'entraînement implique nécessairement un coût social et environnemental et n'améliore pas systématiquement les performances des systèmes. De ce fait, il est primordial de s'interroger sur les données d'apprentissages et sur les apprentissages qui en résultent.

3.2.4 Artéfacts d'apprentissages

Certains patrons d'erreur particuliers ont été rapportés concernant les performances de BERT dans des tâches de *Natural Language Inference* (NLI). Ce genre de tâche demande à l'algorithme d'évaluer si une prémisse n'implique, ne contredit ou ne possède pas de lien sémantique avec une hypothèse. Les exemples

2. https://huggingface.co/docs/transformers/model_doc/roberta

rapportés en 4-7 (Gururangan et al. [2018], Bowman et al. [2015]) démontrent les manifestations possibles de ces trois cas.

4. **Prémisse** : Une femme vend des bâtons de bambous tout en parlant à deux hommes sur un quai d'embarquement
5. **Implication** : Il y a au moins trois personnes sur un quai d'embarquement
6. **Contradiction** : Une femme vend des bâtons de bambou pour faire vivre sa famille
7. **Neutre** : Une femme ne prend pas d'argent pour ses bâtons de bambou.

Gururangan et al. [2018] démontrent que dans ce type de tâche, il est possible pour des modèles comme BERT de se baser sur des traits superficiels pour classer des énoncés. Ce traitement inattendu des données découle directement de certains patrons qui émergent de l'élaboration des corpus. Gururangan et al. [2018] caractérisent ces éléments comme des artefacts d'annotation.

Concrètement, pour l'élaboration du corpus de NLI comme le *Stanford Natural Language Inference* (SNLI), des annotateurs devaient, suite à la présentation d'une description d'une photo servant de prémisse, écrire une phrase qui impliquait, qui contredisait ou qui restait neutre face au contenu de la description. Ainsi, Gururangan et al. [2018] rapportent que les annotateurs auraient entrepris certains raccourcis pour arriver à respecter chacune de ses conditions.

Notamment, dans le cas des implications, elles contiennent généralement des mots formant des supracatégories comme "animal, instrument, extérieur" permettant de généraliser sur le contenu de la description pouvant être plus restrictif. De même, les nombres présents dans la prémisse devenaient des quantificateurs généraux comme "some, at least, various, etc" dans la condition d'implication.

De façon similaire, Gururangan et al. [2018] ont retrouvé des mots négatifs comme "nobody, no, never, nothing, etc." plus fréquemment dans la condition de contradiction.

Pour la condition neutre, on retrouverait plus de modificateurs comme "tall, sad, popular, etc." et des superlatifs comme "first, favorite, most". Gururangan et al. [2018] émettent l'hypothèse que ce choix serait une stratégie permettant aux annotateurs d'ajouter de l'information sans que cela ne soit impliqué ni contredit par la prémisse.

Le travail de Gururangan et al. [2018] permet de souligner la possibilité de voir certaines tendances émerger au travers de la représentation des données d'un corpus et d'affecter directement le performance d'un algorithme. Ainsi, dans le cas qui nous intéresse, la tâche suivante permettra d'évaluer si le corpus d'entraînement choisi, celui de la tâche de SemEval 2018, aura un impact sur les types d'ironie que BERT détectera et si ce patron d'erreur provient du guide d'annotation utilisé lors de l'élaboration du corpus d'entraînement. De ce fait, contrairement à la limitation mentionnée par Adaikkan and Durairaj [2020] à la fin de la section précédente, l'ajout de plus de données d'entraînement n'est pas systématiquement suffisant pour augmenter les performances d'un modèle.

3.3 Méthodologie

3.3.1 Paramètres et ajustement de BERT

Pour ajuster finement BERT, j'utiliserai les paramètres ayant le mieux fonctionné pour l'évaluation du modèle de Zhang et al. dans la tâche A de SemEval 2018 soit en utilisant le modèle préentraîné de BERT large cased, un epoch de 4 et un batch size de 16 (voir 3.5).

Modèle	epoch	batch
bert-large-cased	4	16

Table 3.3 Paramètres de BERT

3.3.2 Corpus d'entraînement

Le corpus de SemEval 2018 pour la tâche A est constitué de tweets anglais annotés de façon binaire où 0 note l'absence d'ironie et 1 note sa présence. Il contient un total de 3 834 tweets. Les tweets ironiques le composant ont été collectés par une recherche twitter utilisant les hashtags #sarcasm, #irony & #not pour être annoté manuellement par la suite en suivant le guide de Van Hee et al. [2016]. Une discussion critique du corpus d'entraînement sera faite dans la section 3.5.

3.3.3 Construction des items tests

Dans cette tâche expérimentale, au lieu de me baser sur l'ensemble des caractéristiques d'énoncés ironiques proposées par Beals [1995], je propose d'en utiliser un sous-ensemble plus ciblé. Certaines catégories qu'elle propose se superposent parfois. D'autres catégories sont un peu floues ou nécessitent des éléments contextuels trop lourds pour être testés sur un modèle de langue comme BERT. Cette réorganisation me permettra de faire écho aux propositions théoriques sur la définition de l'ironie présentée dans le chapitre 1 de ce travail. Ainsi, je propose de synthétiser la typologie de Beals [1995] en 6 catégories.

Bien que l'inventaire de Beals n'a pas été fait ni conçu pour être utilisé en tant que corpus d'évaluation pour des modèles de langue ni pour rendre compte de façon exhaustive de toutes les catégories d'énoncés ironiques possibles, son ébauche me permet d'éviter d'avoir à construire et à évaluer moi-même un corpus d'énoncés en anglais.³

Le nombre d'éléments dans certaines de ces catégories est plus important que d'autres. De ce fait, j'ai sélectionné au hasard 10 énoncés dans les catégories de Beals suivante *cases where the literal and conveyed meanings are in semantic opposition* pour la catégorie d'opposition sémantique, *ironic hyperbole* pour la catégorie d'hyperbole ironique, *ludicrous suggestions* pour la catégorie suggestion ridicule, *ironic uses of unrealis expressions* pour la catégorie d'ironie unrealis. À cause de leur nombre plus restreint, j'ai pris l'entièreté des énoncés des deux catégories *rhetorical questions* pour la catégorie des questions rhétoriques et *ironic understatement* pour les euphémismes ironiques. Ces dernières contenaient respectivement 8 et 7 énoncés au total. On retrouve la totalité de ces énoncés et des résultats de BERT dans l'annexe de ce travail.

3.3.4 Catégories d'énoncés ironiques

3.3.4.1 Opposition sémantique

Parmi les différentes catégories utilisées pour ce corpus d'évaluation, on retrouve l'ironie par opposition sémantique qui reflète directement la définition classique proposée par Grice [1975]. Les énoncés de ce type transgressent la maxime conversationnelle de qualité et se manifestent par des énoncés généralement faux. Ils remplissent une des conditions de la définition de Beals [1995] qui dit que l'ironie implique de faire semblant que quelque chose est vrai comme en 8.

8. _____

3. De même, je crois qu'arriver à faire un inventaire complet des catégories permettant de rendre compte de toutes les formes possibles qui peut prendre l'ironie verbale serait un travail de maîtrise en soi.

This country has so many guns because guns are very useful.

3.3.4.2 Hyperboles ironiques

Les hyperboles ironiques englobent les énoncés ironiques qui, comme mentionnés dans la section 1.4.1 du chapitre 1 de ce travail, ne contreviennent pas directement à la maxime de qualité de Grice, mais où le locuteur ironise en exagérant ses propos. Par exemple, en 9, l'ironie du propos découle de l'exagération soulignée par le locuteur concernant la loquacité d'Alli. Ainsi, la phrase est ironique parce qu'il serait plus difficile de le faire taire que de le tuer.

9. Tests have shown that a punch with sufficient force would not only kill Alli, but might even silence him.

3.3.4.3 Euphémismes ironiques

Les euphémismes ironiques sont sensiblement similaires aux hyperboles ironiques. Toutefois, contrairement à ces dernières, le locuteur diminue la véritable portée d'une situation qu'il rapporte par ses propos. En 10, l'ironie provient de l'expression "a little intoxicated". En effet, pour le besoin de cet exemple, on peut supposer que Jude devrait être particulièrement intoxiqué pour agir de cette façon.

10. Jude was a little intoxicated when they decided to wreck their house.

3.3.4.4 Question rhétoriques

Les questions rhétoriques permettent, comme rapporté dans la section sur la typologie de Beals, d'utiliser une phrase non-assertive pour affirmer quelque chose. De ce fait, bien qu'elles soient techniquement des questions, elles n'appellent pas à une réponse, mais transmettent l'attitude du locuteur. Dans l'exemple en 11, le locuteur n'appelle pas réellement à une réponse. Au contraire, il souhaite plutôt que la population résidente aux États-Unis cesse de manifester leur mécontentement concernant la guerre du Vietnam.

11. Here this country has got its hands full trying to preserve freedom in Vietnam. How can we do that if our own people won't shut up and do what they're told?

Les questions rhétoriques ne sont pas nécessairement ironiques. En effet, il est tout à fait possible pour un locuteur de poser une question rhétorique sincère comme en 12.

12. Est-ce que tu vas finir par me payer ta partie du loyer ?

En 12, même si le locuteur n'attend pas sincèrement une réponse, il transgresse tout de même une condition de félicité. Selon Beals, bien que la présence de ce type de discours puisse dépendre de l'implicature causée par le bris d'une condition de félicité, elle n'en découle pas directement. Elle précise plutôt que la transgression d'une condition de félicité du sens littéral d'un énoncé permet à l'interlocuteur de comprendre que ce dernier l'utilise comme acte de langage indirect lui permettant de véhiculer ce sens « incroyablement faux » de l'ironie telle qu'elle la définit.

3.3.4.5 Suggestions ridicules

Les suggestions ridicules soulignent la composante de la définition de Beals [1995] qui dit qu'un énoncé ironique permet de souligner quelque chose d'extrêmement faux. C'est souvent au travers de cette caractéristique qu'un énoncé ironique peut être drôle. De plus, elle permet d'englober des énoncés ironiques surréalistes tels que présentés par Dynel [2013]. Dans ce cas-ci, l'ironie prend forme suite à l'absurdité assumée de la proposition du locuteur.

13. There probably ought to be a law passed saying you can't buy a new piece of clothing without throwing one out. Maybe the law could work like deposit bottles.

3.3.4.6 Ironie utilisant les expressions irrealis

Les expressions irrealis imposent un cadre intéressant qui se distingue de l'opposition sémantique, mais, aussi, des hyperboles et des euphémismes ironiques. En utilisant ce type de forme, le locuteur ironise en évitant de se commettre entièrement aux propos qu'il énonce.⁴

14. I actually heard the secretary [of State] say that America will help rebuild the economy of Iraq after this

4. On retrouve même des cas, où comme le démontre Martin [1998], l'utilisation de la particule "Paz" en Mocho, une langue maya, permettrait au locuteur de signaler qu'un énoncé est ironique en mentionnant qu'il n'a jamais eu lieu.

war. Mr Baker, **I just can't tell you** what enthusiasm your plan has inspired in Saudi Arabia. (O'Rourke 1992 : 194)

3.3.4.7 Les autres types d'ironie de Beals

Comme mentionné plus haut, certaines catégories proposées par Beals [1995] s'adaptent moins bien à ce type de tâche. Par exemple, on retrouve certains types qui, au final, peuvent être fusionnés à la catégorisation proposée plus haut comme les expressions de renforcement. Ces dernières s'ajoutent généralement à un énoncé déjà ironique et peuvent être autant des cas d'oppositions sémantiques, d'hyperboles ou d'euphémismes. On retrouve le même problème avec les ironies s'articulant autour des processus inférentiels comme les expressions utilisées pour des propositions inférables grâce à leur contenu sémantique, pour leurs présuppositions et leurs implicatures conventionnelles et conversationnelles et l'utilisation ironique de phrases non-assertives. Par ailleurs, l'ironie découlant des implicatures louses nécessite une trop grande connaissance d'éléments contextuels pour être vraisemblablement détectée par un système comme BERT.

Le tableau 3.4 synthétise les catégories de Beals [1995] et celles utilisées dans la tâche expérimentale.

Source	Catégorie
Beals (1995)	Opposition de sens
	Expression de renforcement
	Énoncé ironique avec des expressions irrealis
	Expression idiomatique
	Expression utilisée pour des propositions inférables grâce à leurs contenus sémantiques
	Expression utilisée pour leurs présuppositions et leurs implicatures conventionnelles
	Expression utilisée pour leurs implicatures conversationnelles
	Expression utilisée pour généraliser des implicatures conversationnelles et conventionnelles
	Utilisation ironique de phrases non assertives
	Utilisation ironique d'acte de langage indirect
	Implicature lousse
	Hyperbole ironique
	Suggestion ridicule
Corpus d'évaluation	Opposition sémantique
	Hyperbole ironique
	Euphémisme ironique
	Question rhétorique
	Ironie utilisant des expressions irrealis
	Suggestion ridicule

Table 3.4 Résumé des catégories d'énoncés ironiques sélectionnées pour la tâche et ceux de Beals

3.4 Résultats

À titre de comparatif et de mesures de base, les résultats de l'équipe ayant le mieux performé, THU_NGN (Wu et al. [2018]), à la tâche A de SemEval 2018 sont résumés dans le tableau suivant :

Précision (<i>accuracy</i>)	Précision (<i>precision</i>)	Rappel (<i>recall</i>)	F1
0.735	0.630	0.801	0.705

Table 3.5 Résultat de l'équipe THU_NGN à la tâche A de SemEval 2018

La mesure de précision (*accuracy*) sera le niveau de base pour comparer les résultats de BERT sur les données de Beals [1995]. Ces derniers sont résumés dans la figure 3.1.

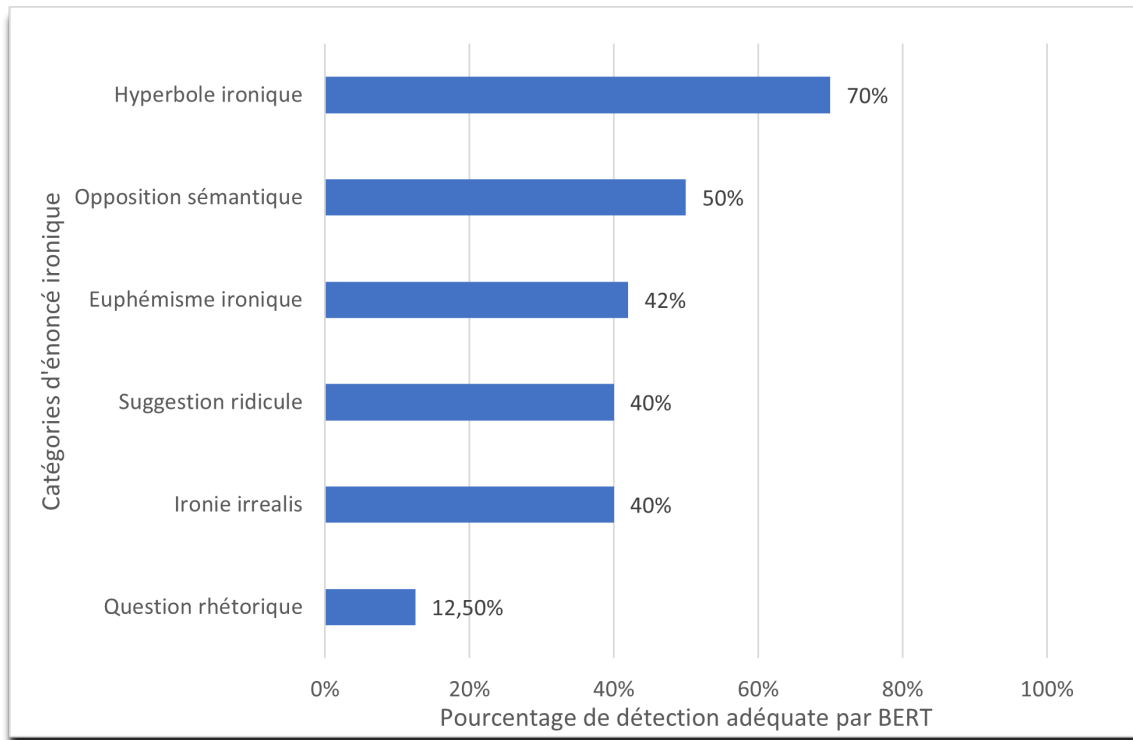


Figure 3.1 Pourcentage d'énoncés adéquatement classés comme étant ironiques par BERT

Les capacités de BERT ajusté sur les données de SemEval 2018 à généraliser sur les formes d'ironie provenant du travail de Beals [1995] semblent limitées. En effet, dans le graphique en 3.1, on ne retrouve qu'une catégorie d'énoncés ironiques où BERT a su adéquatement détecter la présence de ce type de discours à plus que 63% et plus soit les hyperboles ironiques, à 70%.

Parmi les catégories d'ironie ayant été perçues sous ce seuil par BERT, on retrouve l'ironie par opposition sémantique à 50%, les euphémismes ironiques à 42%, les suggestions ridicules à 40%, l'ironie irrealis à 40% et les questions rhétoriques à 12,50%.

3.5 Discussion

Il est important de souligner que les données d'entraînement du corpus de SemEval 2018 et les données du corpus d'évaluation du travail de Beals [1995] proviennent de sources bien différentes. Les premières sont essentiellement tirées d'annotation manuelle de tweets tandis que les seconds proviennent de différentes sources plus littéraires. De ce fait, certains référents qui s'y trouvent reflètent les différents contextes historiques desquels ces données proviennent. Par exemple, dans les données du corpus d'évaluation, on retrouve beaucoup de mentions de la guerre du Vietnam. Néanmoins, on s'attend tout de même à ce que BERT une fois finement ajusté sur un jeu de données adéquat soit en mesure de généraliser ces apprentissages sur de nouvelles données. Par exemple, dans le cas mentionné plus haut concernant le Vietnam, ce n'est pas l'identité du pays en question qui déclenche l'ironie verbale, c'est de proposer comme en 15, que de mettre le feu à un pays permettrait de le sauver.

Ainsi, la section suivante tentera de trouver des angles explicatifs pour les résultats collectés en présentant un exemple d'énoncé adéquatement détecté comme étant ironique et un exemple d'énoncé détecté comme étant non-ironique. Il sera d'abord question d'explicitier les caractéristiques des catégories d'énoncés ironiques pouvant rendre leur détection difficile. Ensuite, je soulignerai certains éléments liés à l'élaboration du corpus de SemEval 2018 ayant pu rendre la généralisation des apprentissages de BERT plus ardue.

3.5.1 Patrons d'erreurs

3.5.1.1 Question rhétorique

Certains types d'ironie sont particulièrement compliqués à définir et à détecter. Par exemple, comme mentionné plus haut, Beals [1995] décrit le rapport complexe entre l'ironie et les actes de langages indirects comme les questions rhétoriques. Malgré tout, le modèle a été en mesure de bien détecter l'ironie dans l'énoncé en 15 :

15. How can we make Vietnam free if we don't first burn it ?

Dans cette question rhétorique en 15, la seule que le modèle ait réussi à bien catégoriser, le locuteur ne souhaite pas réellement une réponse à sa question. Surtout, il ne cherche pas non plus à déclarer que

la seule façon de libérer le Vietnam c'est d'y mettre le feu. Au contraire, comme mentionné plus haut, il souligne plutôt l'absurdité de libérer un pays en l'incendiant. De même, cet énoncé ironise en utilisant une question rhétorique, mais il manifeste aussi un certain caractère hyperbolique.

Néanmoins, l'énoncé en 16 n'a pas été classé comme étant ironique. Plusieurs cas de figure sont possibles. Par exemple, il existe une lecture littérale de l'énoncé qui fonctionne où la position réelle du locuteur est opaque à un interlocuteur qui ne saurait pas qu'il est contre la guerre du Vietnam. Dans cette éventualité, il est nécessaire d'avoir accès au contexte et au contenu des informations partagées entre le locuteur et son interlocuteur pour bien interpréter l'ironie présente dans l'énoncé.

16. Here this country has got its hands full trying to preserve freedom in Vietnam. How can we do that if our own people won't shut up and do what they're told ?

Ce type d'ambiguïté nécessitant une bonne connaissance du contexte se retrouve dans les autres énoncés de cette catégorie considérés comme étant non-ironique par le modèle. Dans le cas en 17, l'interprétation ironique véhiculée par la question rhétorique à la fin de l'énoncé nécessite de comprendre que, généralement, personne n'est reconnaissant de recevoir des lettres qui nous incitent à donner de l'argent. En 18, Beals décrit l'ironie comme suit : *"the [speaker] pretends that the Protestants are behaving patriotically and that there is nothing more patriotic than fighting one's own countrymen to retain one's identity. Related to this verbal irony is the dramatic irony in the protestants' suggestion that they are behaving patriotically, and the event, irony of people fighting out of patriotism those whom they consider their own countrymen"*.

Dans ces deux cas, il est nécessaire d'avoir accès à des informations contextuelles, d'un côté, en lien avec les attentes partagées concernant les interactions sociales et, de l'autre, des rapports historiques entre deux communautés.

17. Some woman in New-Jersey - her name is Alice- save up all the mail she got last year asking for money and sent it to me. How can I ever thank you enough, Alice.
18. The Protestant "orange" or "loyalist" Irish have plenty of excuses themselves. They're patriots after all. "We'll fight anyone to stay British, even the British," they like to say. What could be more patriotic than that ?

3.5.1.2 Ironie irrealis

Les énoncés ironiques irrealis sont la deuxième catégorie ayant moins bien été détectée par BERT. Ces résultats sont cohérents sachant que ce type d'énoncés peut se construire sous la forme de questions rhétoriques. À noter que, selon les résultats de cette tâche, ce ne sont pas toutes les questions rhétoriques qui sont nécessairement catégorisées comme étant non-ironiques. À priori, ce serait seulement celles qui nécessitent une bonne connaissance contextuelle des informations partagées avec l'interlocuteur. De façon similaire à ces dernières, ce type d'énoncé permet au locuteur de ne pas asserter directement une proposition avec ses propos.

Comme le mentionne Beals : *"An irrealis environment is one whose propositional content is neither asserted nor implied by the speaker ; the sincere speaker, rather, leaves its truth in doubt or asserts that it is false"*.

Les énoncés en 19

20 ont été adéquatement détectés comme étant ironique par BERT, mais pas l'énoncé en 21.

19. I actually heard the secretary of State say that America will help rebuild the economy of Iraq after this war. Mr Baker, **I just can't tell you** what enthusiasm your plan has inspired in Saudi Arabia.

20. My proposals were simple : We must legalize the possession of sawed-off shotguns, portable and fixed machine guns, bazookas, flamethrowers, hand grenades, mortars and light artillery

21. Just trust me that I'm in one hundred percent agreement with you on these issues. **I can't imagine** how any right-thinking person could possibly believe other than we do.

La différence entre ces trois énoncés concerne la présence d'éléments contextuels permettant d'inférer que possiblement le locuteur ironise. Dans le premier énoncé (19), on retrouve la proposition rapportée par le locuteur que l'Amérique souhaiterait aider à reconstruire l'économie de l'Iraq après la guerre et que cette information risque d'inspirer de l'enthousiasme en Arabie Saoudite. Dans le second énoncé, le modèle est en mesure de comprendre l'aspect ridicule de la proposition mentionnée. Dans le troisième, (21), l'ironie provient du fait que, malgré ce qu'il avance, le locuteur peut imaginer comment n'importe quelle personne pourrait croire qu'ils peuvent.

Deux possibilités pourraient expliquer les différences d'annotation par BERT entre le premier énoncé et le second. Il est possible que BERT, grâce aux données sur lesquels il a été préentraîné, possède certaines représentations des liens sémantiques entre les États-Unis, l'Iraq et l'Arabie-Saoudite. De même, il est aussi probable que le manque d'informations contextuelles présentes dans l'énoncé lui-même en 21, rendent la perception de l'ironie impossible pour BERT.

Le même genre d'ambiguïté se retrouve en 22. Bien qu'on y retrouve une forme de suggestion ridicule, les éléments présents dans l'énoncé laissent place à une certaine ambiguïté pouvant empêcher le modèle de bien détecter l'ironie qui s'y trouve.

22. If a baby is chucked under the chin by everyone it meets, it could develop a callused jaw. So maybe chin-chucking is a little known abuse we ought to think about.

3.5.1.3 Suggestion ridicule

Concernant les suggestions ridicules, comme le mentionne Dynel [2013], ce type d'énoncé ne contrevient pas directement à la maxime de qualité et ne peut être pris en compte totalement par la perspective griécienne. Ainsi, comme pour plusieurs types d'ironie mentionnés jusqu'à présent, il est nécessaire d'avoir accès aux informations partagées entre les locuteurs pour savoir ce qui est absurde ou impossible. De ce fait, on peut expliquer la disparité entre la détection des énoncés provenant des exemples de Beals et ceux provenant d'autres sources.

Par exemple, dans l'énoncé 23 adéquatement catégorisé par BERT comme étant ironique, le locuteur propose d'ajouter 3 poignées sur chaque porte de garde-robe parce que ce serait plus facile d'y accrocher des choses. Dans l'énoncé suivant, en 24, aussi adéquatement détecté comme étant ironique, on compare l'ironie provient de la proposition faite par le locuteur d'appliquer le même système pour la récupération de bouteille que pour l'achat de vêtement. En 25, aussi bien détecté, l'énoncé est ironique parce qu'on y propose d'apprendre l'anglais grâce aux petites annonces dans le journal. Les trois énoncés suivants n'ont pas été adéquatement détectés comme étant ironiques par BERT. En 26 le locuteur propose de rendre les tomates et les melons illégaux pour certains mois de l'année. En 27, propose de faire des débats entre les candidats présidentiels et leurs conjointes. En 28, on y propose de changer la date de Noël.

Il est bien entendu difficile de savoir qu'est-ce que le modèle de langue perçoit et dans quelle mesure il

est apte à catégoriser le ridicule d'une proposition. Entre 23 et 26, il est plus facile d'imaginer un cas de figure où la première proposition est plus ridicule que la deuxième. En effet, bien que la proposition en 26 soit farfelue, il existe des cas où, pour des raisons de santé publique, certaines restrictions légales peuvent s'appliquer sur la distribution de nourriture (voir "la crise de la vache folle" causée par l'encéphalopathie spongiforme bovine). De ce fait, sans informations contextuelles, il est difficile de bien être en mesure de savoir si ce locuteur est ironique ou pas, contrairement en 13, où la proposition est originale et incongrue. Aussi, il est possible qu'une proposition soit ridicule, mais que le locuteur l'émette sans être ironique. Par exemple, en 28, on peut imaginer des contextes, comme un confinement suite à une pandémie, où proposer de déplacer Noël soit une réelle proposition non ironique. De ce fait, il manquerait un accès à d'autres informations contextuelles et au sens commun pour être en mesure de bien cerner la présence d'ironie dans l'énoncé.

23. If I had my way, the building code would decree that there be at least three door knobs on every closet door. We all know a door knob is the best and easiest place to hang something.
24. There probably ought to be a law passed saying you can't buy new piece of clothing without throwing one out. Maybe the law could work like deposit bottles
25. A course based on the classified advertising section of any newspaper would provide lots of examples of how we actually use the English language in practical situations. A study of the classifieds would also give students the same kind of mind-expanding exercise that translating French into English gives them now
26. If I'm ever elected to Congress, I think I'll try to get a law passed making tomatoes and melons illegal, except for certain months of the year.
27. I've got just one more idea for the last debate. We'll all now what the candidates think by them, and we are, after all, electing a First Lady, too. So the last debate will be different. For half an hour, each candidate will argue with his wife⁵
28. Christmas obviously should be moved. No one knows for sure which day Christ was born on. And he certainly wouldn't mind if we celebrated Christmas in February. . . . So here's what I'm suggesting. Move the whole month of January, including my birthday, so that it falls between August and September. Eliminate Labor Day altogether

5. Désolé pour les exemples misogynes

3.5.1.4 Euphémisme ironique

L'écart entre les résultats concernant les euphémismes et les hyperboles ironiques, respectivement 42% et 70% des énoncés bien détectés comme étant ironiques, soulève des questions intéressantes. À première vue, de par leur nature, on s'attendrait à un patron similaire entre les deux. Il est tout à fait possible que ces résultats soient causés par le corpus d'évaluation en lui-même. Par exemple, si ce mémoire s'intéresse à la granularité de l'ironie verbale, paradoxalement, il semblerait qu'il existe différents types d'euphémismes. Notamment, Farghal [1995] rapporte que, dans le cadre de son étude, les locuteurs de l'arabe peuvent s'appuyer sur quatre différentes façons pour produire ce type de discours soit en utilisant des expressions figuratives, des périphrases, en remodelant un énoncé ou en utilisant des antonymes. De plus, Farghal [1995] et Xin [2015] notent que les euphémismes ont la capacité de transgresser les quatre maximes conversationnelles de Grice. Il est alors possible que les énoncés utilisés dans le corpus d'évaluation se rapprochent des items ironiques par opposition sémantique. Par exemple, l'item issu du corpus d'évaluation en 29 transgresse la maxime de quantité parce que le locuteur diminue le niveau d'intoxication de Jude.

29. Jude was a little intoxicated when they decided to wreck their house.

BERT a su détecter l'ironie en 30 et 31, mais pas en 32 et 33. Pour Beals, l'énoncé en 30 est ironique parce qu'il porte une implicature de quantité qui dit que "[the] recent U.S. presidents may have contributed in less significant ways to the sum of human knowledge". De même, en 31, le modèle a perçu l'euphémisme ironique dans l'énoncé qui découlerait, selon Beals, du contexte lié au contexte historique des communautés allemandes et irlandaises à Chicago lors du 19e siècle. Il est très improbable que ce genre de représentation contextuelle soit accessible à BERT. Il est plus probable que la présence d'expression comme "*as everyone knows*" servent d'indice quant à la présence d'ironie dans un énoncé.

Quant à elle, l'interprétation de l'euphémisme ironique en 32 est plus complexe. D'un côté, elle porte une certaine forme proche des suggestions ridicules mentionnées plus haut. De l'autre, on peut en tirer une lecture littérale, car il est vrai que pendre les candidats perdants aux élections est illégal et (peut-être même) immoral. Dans ce cas-ci, la transgression de la maxime de quantité n'est peut-être pas assez forte pour qu'on puisse comprendre l'ironie véhiculée par le locuteur. De façon similaire, pour bien interpréter l'ironie en 33, il est nécessaire de comprendre que le British Museum n'est pas le seul endroit où on peut trouver des éléments qui seraient factuellement vrai. Toutefois, il est possible d'imaginer un locuteur littéral qui

pense réellement que le British Museum possède une certaine forme de vérité historique qui ne serait pas présente ailleurs.

30. The ideas emitted by Presidents of the United States during the past thirty or forty years have not been of such a character that the sum of the human knowledge has been appreciably enriched by them
31. And, as everyone knows, persons of German descent may now drink beer in Chicago and several Irishmen have been known to hold jobs on the police force.
32. Long ago I proposed that unsuccessful candidates for the Presidency be hanged, as a matter of public sanitation and decorum... On second thoughts, I withdraw my suggestion. It is probably illegal, and maybe even immoral.
33. If truth is not to be found on the shelves of the British Museum, where, I asked myself, picking up a notebook and a pencil, is truth ?

3.5.1.5 Opposition sémantique

Comme mentionné plusieurs fois dans ce travail, lorsqu'il est question d'ironie ou des algorithmes de détection automatique de ce type de discours, on en retrouve généralement une définition opérationnalisable en termes d'oppositions sémantiques. De ce fait, étant le type d'ironie le plus documenté, on s'attendrait à ce que ça soit la catégorie ayant le meilleur taux de réussite. Il est fort possible que ces résultats, 50% de détection adéquate, soient causés par les données du corpus d'évaluation créé à partir des énoncés de Beals. En effet, certains énoncés catégorisés comme étant non-ironiques par BERT semblent nécessiter une bonne connaissance des informations partagées entre locuteurs. Ces derniers ne peuvent pas être inférés directement des propos présentés à l'algorithme.

34. The men in these catalogs all look a lot like me, I think. They're handsome, young, rugged, about 150 pounds and never a hair out of place.
35. The strap was the only flaw. It was wide, red, plastic. That's OK for a kid, but a grown man shouldn't have that on his Mickey Mouse watch.
36. This country has so many guns because guns are very useful

37. Today's teenager has great ingenuity. He talks when he has nothing to say, rushes about when he has no destination and returns after having been nowhere. So it is perfectly natural that we should have a protest march when the marchers have no complaints

En 34, les éléments contextuels nous permettant de comprendre que les modèles décrits en question par le locuteur ne lui ressemblent pas sont absents rendant impossible la détection adéquate de l'ironie dans l'énoncé. Pour comprendre l'ironie en 35, ce que le modèle n'a pas été en mesure de faire, on doit comprendre que ce n'est pas la large ganse de plastique rouge qui serait inadéquate pour un adulte, mais l'image de Mickey Mouse. Pour comprendre celle en 36, ce que le modèle a réussi à faire, il faut comprendre que le locuteur ne croit pas que les armes à feu soit utiles. De même, en 37, on peut supposer que le contexte présent dans l'énoncé ait pu faciliter la détection de l'ironie par le modèle.

3.5.1.6 Hyperbole ironique

Le haut taux de réussite associé à la catégorie des hyperboles ironiques peut être causé par l'origine des données d'entraînement. Kreuz and Roberts [1995] soulignent le lien fort qui existe entre l'ironie et les hyperboles. Pour eux, ces deux types de discours partagent certains objectifs comme le fait d'être utilisés parfois pour être humoristique, pour mettre l'emphase sur quelque chose ou pour clarifier une position. Plus spécifiquement, ils rapportent que l'utilisation d'hyperbole permet, comme l'utilisation d'un certain ton de voix, de souligner de façon plus explicite qu'on ironise. Ce type de stratégie a pu être reflété par les données collecter sur Twitter où, généralement, les utilisateurs partagent moins d'informations communes et doivent, ainsi, être recourir à des stratégies conversationnelles plus transparentes pour communiquer l'ironie qu'ils souhaitent partager.

L'énoncé en 38, bien identifié par BERT comme étant ironique possède plusieurs éléments qui permettent de souligner l'exagération qu'il contient comme l'accumulation apportée par les mots "broken and disillusioned" et l'expression "curried tears". De même, 41, il n'est pas nécessaire de comprendre que le Boxer Mohamed Ali était reconnu pour parler beaucoup pour comprendre l'hyperbole ironique de l'énoncé.

Tandis que, en 40, bien que la maxime de quantité soit transgressée par l'expression "reciting latin noun paradigms", il existe tout de même une lecture littérale de l'énoncé. De façon similaire, 41, l'ironie, selon Beals, provient de fait que les banques ne sont pas si demandantes que le laisse supposer le locuteur de

l'énoncé. Dans ce cas-ci, ce type de jugement nécessite de bien savoir les informations partagées avec le locuteur pour interpréter l'énoncé.

38. Broken and disillusioned, we tottered into an Indian restaurant, where we pooled our curried tears. Between sobs, Mr. Steele remarked that we were part of America's national weakness- it's misplaced passion for immediacy.
39. Tests have shown that a punch with sufficient force would not only kill Ali, but might even silence him
40. On Realizing what I had done, I went into a paroxysm of agony and guilt that began a two-day episode of weeping, drinking, picking fights, smashing things, and reciting latin noun paradigms while banging my head against hard-surfaces.
41. Banks used to care about my business. They know me. I didn't have to bring my birth certificate, a copy of my listing in Who's Who, and four other pieces of positive identification to cash a check for twenty-five dollars.

3.5.2 Résumé

Donc, en résumé, dans les énoncés correctement identifiés comme étant ironiques, BERT semble en mesure de bien saisir des informations provenant du sens commun. Ces informations proviennent probablement du contenu tiré de ses données de préentraînement. De même, une certaine tendance se dégage des énoncés non adéquatement identifiés comme ironiques par BERT. En effet, parmi ceux-ci, il semble que les énoncés ironiques nécessitant une plus grande connaissance des informations partagées entre le locuteur et son interlocuteur soient plus difficiles à catégoriser adéquatement. Sans ces informations, des lectures littérales des énoncés sont généralement possibles sans que le contenu véhiculé soit complètement absurde comme en 26. Au-delà de cette caractéristique, certaines catégories d'énoncés ironiques semblent plus difficilement détectables que d'autres. Pour évaluer ces différences, la sous-section suivante tentera de souligner les caractéristiques du corpus d'entraînement pouvant avoir eu un impact sur les apprentissages de BERT.

3.5.3 Construction du corpus SemEval 2018

Comme mentionné plus haut, le corpus de SemEval 2018 (Van Hee et al. [2018]) a été construit à partir de recherches sur Twitter utilisant les termes #irony, #sarcasm, #not. L'ensemble des tweets récoltés a été

annoté par des évaluateurs suivant le guide rédigé par Van Hee et al. [2016]. Ce dernier leur indique la marche à suivre pour bien étiqueter les énoncés. Lorsqu'un tweet est considéré comme étant ironique, ils doivent indiquer si la présence d'un hashtag (irony, not, sarcasm) a été nécessaire pour son identification en plus du niveau d'impolitesse véhiculé par le message. Parmi les autres éléments que les annotateurs doivent recenser, on retrouve la polarité de l'attitude évaluative véhiculée par l'énoncé. Par exemple, en 42, le guide indique que les mots "hate", "no longer matters in life" et "#dislike" sont tous porteurs d'une polarité négative. Les annotateurs doivent aussi rapporter la présence de modificateur et l'élément modifié dans l'énoncé collecté. Concrètement, ce sont les éléments d'un énoncé qui augmente ou diminue l'intensité du sentiment véhiculé. En 43, l'adverbe "most" et la ponctuation "!!!" augmente l'intensité du mot "annoying". Finalement, toujours selon le guide d'annotation, il est nécessaire de rapporter la présence de cible précise d'un énoncé ironique et si la polarité visant cette dernière peut être inférée par le contexte, des connaissances générales ou le sens commun. Par exemple, en 44, la note "F-" est la cible de l'expression "did so well".

42. I hate it when my mind keeps drifting to someone who no longer matters in life. #irony #dislike

43. The most annoying kid lives next to my door!!!

44. I did so well on my history test that I got an F-!

Les tweets annotés peuvent être classés parmi les trois catégories mentionnées plus haut (1).

Dans le guide d'annotation, on retrouve plusieurs définitions de l'ironie verbale. La première se base sur celle de Grice [1975] : *Verbal irony is often defined as expressions that convey an opposite meaning* (p.1). Plus loin, le guide précise : *In accordance with this definition, we define irony as an evaluative expression whose polarity (i.e., positive, negative) is changed between the literal and the intended evaluation, resulting in an incongruence between the literal evaluation and its context.* Dans les deux cas, ces définitions rencontrent les mêmes problèmes que ceux soulevés au chapitre 1 à la section 1.4.1 et à la section 1.8.1 de ce mémoire. L'ironie peut se trouver dans des énoncés où leurs aspects évaluatifs varient selon la cible de ces derniers. De même, on ne peut résumer l'ironie verbale à de l'opposition de sens entre le sens littéral et celui véhiculé par le locuteur.

Paradoxalement, on retrouve plus loin une mention de ce type d'ironie impliquant les hyperboles et les euphémismes. Néanmoins, il n'est pas évident d'établir comment la définition proposée dans ce guide

concorde avec cette nuance sur cette caractéristique évaluative. Pour exemplifier ces types d'énoncés, le guide présente les phrases suivantes :

45. Oh no I love working in Baltimore #not
46. What a shock. Duck Johnson is hurt in an important game. #sarcasm #canes
47. So glad you'd rather read a book than acknowledge your own kid #not
48. Interesting visit with Terra Nova yesterday at Toneleigh, class tent.

De ce fait, les résultats concernant les hyperboles ironiques et l'ironie résultant d'une opposition sémantique ont plus de sens. En effet, ce sont des catégories explicitement visées par le guide d'annotation. Van Hee et al. [2016] souligne tout de même que : *examples of ironic hyperbole and ironic understatement are generally more difficult to perceive* (p.9).

En résumé, il semble y avoir quelques correspondances entre les types d'ironie non adéquatement identifiés par BERT et certain choix présenté dans le guide d'annotation. Cet écart nous ramène aux points abordés à la section 2.4 de ce mémoire où l'importance d'avoir une définition exhaustive d'un phénomène que l'on souhaite détecter est nécessaire.

3.6 Limitations et recherches futures

Les hypothèses concernant les causes probables des problèmes de détection soulevés dans la section précédente mettent en lumière un des problèmes centraux de la méthode expérimentale utilisée dans ce chapitre. Il nous est impossible de bien départager ce que le modèle a réellement appris et la responsabilité du corpus d'entraînement dans les erreurs de prédiction du modèle. De ce fait, plusieurs limitations liées aux résultats présentés plus haut nécessitent d'être exposés dans cette section.

La première limitation de ce projet concerne le corpus d'évaluation construit à partir des données du travail de Beals [1995]. Comme mentionné plus haut, ce dernier n'avait pas comme objectif d'être adapté pour ce type de tâche ni d'offrir une catégorisation exhaustive des types d'ironie possibles. De plus, l'écart entre les modalités des données d'entraînement et les données utilisées pour tester le modèle est non négligeable. D'un côté, les données du corpus de Van Hee et al. [2018] sont tirées de Twitter où on ne retrouve des utilisateurs qui s'adressent à tout le monde et à personne à la fois. De ce fait, la gestion des informations partagées entre le locuteur et son ou ses interlocuteurs suit une dynamique particulière différente d'autres modalités. De l'autre, le corpus de Beals [1995] provient de différentes sources plus littéraires. De ce fait,

le style d'expression est nécessairement différent dans les deux cas et les implications peuvent être nombreuses. On n'ironise pas de la même façon sur internet que dans un livre. De même, les cibles d'énoncés ironiques changent avec le temps. La nécessité d'avoir une bonne connaissance de l'espace d'informations partagées entre le locuteur et son interlocuteur peut circonscrire certains référents ou certaines cibles présents dans des énoncés ironiques à des époques précises. Par exemple, en humour, l'âge a un effet direct sur ce que l'on trouve drôle et sur ce qui nous fait rire (voir Schaiier and Cicirelli [1976], Stanley et al. [2014] ou Vitulli [2005]). Concrètement, il est tout à fait possible que du côté de Van Hee et al. [2018] peu ou aucun tweets ne concerne la guerre du Vietnam tandis que dans les données de Beals [1995], on retrouve certains énoncés qui y font référence. Malgré tout, on s'attend quand même à ce que BERT soit en mesure de généraliser sur des données d'apprentissages et qu'il puisse ainsi être en mesure d'exporter ses acquis sur de nouvelles données et, ce, malgré un certain fossé entre ces données et les données utilisées pour tester ses capacités.

À ce sujet, il existe de nouvelles façons de visualiser les apprentissages des modèles de traitement automatisé du langage. Différentes méthodes émergent présentement pour évaluer l'origine causale des abstractions faites par un réseau de neurones. Dans son travail Geiger et al. [2021] proposent une méthode efficace qui consiste à interchanger des neurones d'un modèle et de vérifier si cette opération affecte ses données de sortie. Cette procédure permet d'inférer adéquatement l'implication causale de certaines parties des couches neuronales d'un modèle d'apprentissage automatique et de pouvoir bien caractériser l'apport de chacune des couches dans ses prédictions.

Suivant ce type de procédure, de nouveaux outils comme Captum (Kokhlikyan et al. [2020]) permettent de visualiser et d'interpréter les apprentissages d'un modèle comme BERT. De ce fait, il est possible de voir les éléments qui ont permis au modèle de catégoriser adéquatement ou non un énoncé après avoir été finement ajusté sur un ensemble de données.

Ce faisant, pour le futur, il serait intéressant de construire un corpus d'évaluation avec une typologie des catégories d'énoncés ironiques avec des énoncés annotés et jugés par des locuteurs natifs de l'anglais. Par la suite, il serait pertinent d'appliquer un algorithme comme Captum pour bien comprendre ce que BERT a appris sur l'ironie et comment il applique ses acquis pour classer des énoncés.

3.7 Conclusion

Cette section a permis de tester et de caractériser les apprentissages du modèle BERT sur des énoncés ironiques préalablement caractérisés par Beals [1995] lorsque ce dernier est finement ajusté pour ce type de tâche. Bien que la méthode utilisée nous apporte des réponses quelque peu limitées sur l'origine des prédictions fautives commises par BERT, l'observation de ses patrons d'erreurs et l'étude du guide d'annotation permettant la confection du corpus d'entraînement de Van Hee et al. [2016] mettent l'emphase sur l'impact d'une définition exhaustive du phénomène que l'on souhaite étudier. Ce processus d'opérationnalisation impactera nécessairement les résultats et les limites du modèle conçu.

CONCLUSION

L'ironie est une forme de discours difficile à identifier, autant pour les humains que pour les modèles de langue. Néanmoins, pour paraphraser de façon plus nuancée le célèbre dicton "garbage in, garbage out", il est nécessaire de porter une attention particulière aux phénomènes que l'on souhaite étudier pour être en mesure de conceptualiser un corpus d'entraînement qui permettra au modèle de langue de bien pouvoir classer ce qui nous intéresse.

Dans ce travail, les difficultés théoriques entourant la conceptualisation d'une définition de l'ironie verbale ont été abordées. Par la suite, il a été démontré que ces dernières affectent autant l'élaboration d'algorithme se basant sur des modèles symboliques que la construction de corpus d'entraînement pour les modèles d'apprentissage machine. La dernière partie de ce mémoire tentait de vérifier si un modèle de langue, comme BERT lorsqu'il est ajusté finement sur un corpus d'énoncés ironiques, est en mesure de généraliser ses apprentissages sur différentes catégories d'ironie. Les résultats de cette expérimentation démontrent que certaines formes d'ironie, comme les questions rhétoriques, les ironies irrealis et les suggestions ridicules sont plus difficiles à détecter que d'autres. Ces erreurs proviennent à la fois de la conceptualisation du corpus d'entraînement et de l'accès impossible pour le modèle de langue à certaines informations contextuelles.

L'amélioration des algorithmes pour ce type de tâches est bénéfique pour plusieurs raisons. De prime abord, les recherches sur cette question nous permet d'en savoir plus sur l'ironie verbale et comment elle se manifeste chez des locuteurs. Dans le cadre de ce travail, la tentative de classification des différents types d'ironie soulève un point intéressant. Effectivement, un énoncé ironique peut s'interpréter de plusieurs façons. Par exemple, en 49, on retrouve un énoncé ironique qui est une question rhétorique, mais qui contient un certain aspect hyperbolique. Dans le cadre de ce mémoire, les catégories construites dans le chapitre 3 s'inspiraient fortement de celles de Beals pour des besoins d'opérationnalisation. Néanmoins, il serait nécessaire de se pencher plus en détail sur la possibilité de créer une façon adéquate de bien catégoriser les différents types d'ironie verbale.

49. How can we make Vietnam free if we don't first burn it

De même, bien que ces dernières n'aient pas été le point central de ce mémoire, les questions relatives

à l'élaboration d'une définition adéquate de l'ironie verbale et à comment les locuteurs arrivent à bien l'identifier lors de leurs conversations restent des problèmes linguistiques importants.

Au-delà de ces considérations fondamentales, on retrouve des applications concrètes qui sont importantes pour les questions entourant le traitement automatique du langage. Si, comme Gibbs [2000] le mentionne, l'ironie est fréquemment utilisée dans des conversations entre amis et qu'elle permet de réaffirmer des informations partagées entre des locuteurs (Cohn-Gordon and Bergen [2019]), elle s'observe aussi comme appareil rhétorique dans des sphères plus toxiques du web. Par exemple, DeCook [2020] rapporte que l'ironie, parfois qualifiée de *trolling*, est utilisée par certains pans de la droite extrême et radicale pour affirmer et partager indirectement des positions dangereuses sans pouvoir en être blâmés :

Hiding behind hoaxes, irony, edginess, and trolling, members of the alt-right and other extremist Internet cultures then engage in a kind of subversion that allows them to avoid taking any responsibility for real and violent attacks that occur as a result of their discourse. (DeCook [2020]).

Ce type de discours se retrouve sur plusieurs plateformes et nécessite un travail important de modération. Ce dernier peut être facilité par des outils automatiques permettant de détecter la toxicité. Néanmoins, comme le mentionne Risch and Krestel [2020], l'ironie, le sarcasme et les questions rhétoriques rendent la classification du contenu toxique et haineux difficile.

Bien que ces considérations pratiques soient nécessaires et importantes, l'étude des performances des modèles de langue ne doit évidemment pas se restreindre à des objectifs liés à l'amélioration de ces systèmes. Kodner et al. [2023] décrit adéquatement cette relation particulière où l'objectif des chercheurs devrait se distinguer de ceux des entreprises :

The LLMs of today are a corporate product, not a scientific one. Industry dominates the creation of LLMs due to the high financial and compute costs associated with their training (Ahmed et al., 2023), and the corporations releasing these LLMs are often cagey about the details of their implementation (Liesenfeld et al., 2023). [...] [T]he goals of the corporations creating the models (i.e., to increase profits) are not the same as the goals of the scientists trying to probe the models (i.e., to come to a scientific understanding of language).

Cette mise en perspective est importante. Au final, les performances des modèles de langue comme BERT ne servent pas à améliorer les performances de ce dernier dans une tâche comme la détection de l'ironie verbale. Au contraire, elles servent de levier pour souligner à quel point on sait peu de chose sur les capacités des individus à comprendre et à produire des énoncés ironiques dans des cadres conversationnels. Advenant un futur où des algorithmes seraient en mesure de performer parfaitement dans ce type de tâche, on ne comprendrait probablement toujours pas pourquoi deux individus en viendraient à utiliser ce type de discours pour communiquer et, plus important encore, le plaisir qui peut en découler.

ANNEXE A

CORPUS D'ÉVALUATION CONTENANT LES ÉNONCÉS IRONIQUES COLLECTÉS PAR BEALS (1995) ET LES
RÉSULTATS DE BERT

Catégorie	Énoncé ironique et leur contexte	Prédiction de BERT
Opposition sémantique	But the question is, which way to go : to insist that he be traded ? Or to let him go to the minor leagues? This is one of the most important decisions a mother has to make (Royko [1968])	Non-Ironique
	He didn't explain what the signifiacnce of people's shock was, but I gather it indicated most people are surprised when they see someone in a nun's habit smooching with a hippie at an airport : this is a profound discovery, indeed (Royko [1968])	Ironique
	Today's teenager has great ingenuity. He talks when he has nothing to say, rushes about when he has no destination and returns after having been nowhere. So it is perfectly natural that we should have a protest march when the marchers have no complaints (Royko [1968])	Ironique
	The gun is so effective as a crime-fighting device that the United States, with more privately owned guns than any other country, has the highest crime rate in the world (Royko [1968])	Ironique
	The men in these catalogs all look a lot like me, I think. They're handsome, young, rugged, about 150 pounds and never a hair out of place.(Rooney [1981])	Non-Ironique
	Personnally, I am in favor of the League- not that I am under any delusion about its intents and purposes, but precisely because I regard it as thumpingly dishonest ?Five years of it will see all the principal members engaged in trying to slaughter one another. In other words, it will make for wars- and I have acquired an evil taste in wars (Mencken [1960])	Non-Ironique
	This country has so many guns because are very useful (Royko [1968])	Ironique

Catégorie	Énoncé ironique et leur contexte	Prédiction de BERT
	U.S. Steel is pending a one third of one percent of total sales on air and water pollution. By the time they finished, there wasn't too much left over. (Royko [1968])	Non-Ironique
	The strap was the only flow. It was wide, red, plastic. That's OK for a kid, but a grown man shouldn't have that on his Mickey Mouse watch (Royko [1973])	Non-Ironique
	Royko tries to chat with a waiter at Maxim's in Paris about an ice cream shop run by Greeks at Pulsaki and Irving in Chicago. The waiter looked a little dazed. They're not used to somebody really worldly. Royko [1973]	Ironique
Hyperbole ironique	Broken and disillusioned, we tottered into an Indian restaurant, where we pooled our curried tears. Between sobs, Mr. Steele remarked that we were part of America's national weakness- it's misplaced passion for immediacy (Crisp [1984])	Ironique
	If they had spent twice as much, their profits would have been murdered (Royko [1968])	Ironique
	You really could be libeled by an example sentence. It could happen at any time. It might be happening right now in the pages of Natural Language and Linguistic Theory. Tremble, and read on. (Pullum [1991])	Ironique
	Tests have shown that a punch with sufficient force would not only kill Ali, but might even silence him. (Royko [1982])	Ironique
	I apologize to the tidy-minded for this navel-contemplative digression on the inner workings of the kinds of plagiarism that I employ when creative genius eludes me (Pullum [1991])	Ironique
	Even I, with my fingers on the very pulse of the linguistic community, could not have foreseen the extent of the upwelling of popular support for the campaign (Pullum [1991])	Ironique

Catégorie	Énoncé ironique et leur contexte	Prédiction de BERT
	And what a shock to read those abstracts! Where was the soup of whimsy, the twinkle in the commentorial eye, the stylish levity I imagined I had occasionally attained? The graceful, leaping gazelle of the prose I had recalled had been transformed into a squat, humorless, pygmy hippo (Pullum [1991])	Ironique
	On Realizing what I had done, I went into a paroxysm of agony and guilt that began a two-day episode of weeping, drinking, picking fights, smashing things, and reciting latin noun paradigms while banging my head against hard-surfaces (Pullum [1991])	Non-Ironique
	But the rest of England was straightforward missionary country. It was densely populated by aborigines who had never heard of homosexuality and who, when first they did, became frightened and angry. I went to work on them. (Crisp [1968])	Non-Ironique
	Banks used to care about my business. They know me. I didn't have to bring my birth certificate, a copy of my listing in Who's Who, and four other pieces of positive identification to cash a check for twenty-five dollars. (Rooney [1981])	Non-Ironique
Euphémismes ironiques	The ideas emitted by Presidents of the United States during the past thirty or forty years have not been of such a character that the sum of the human knowledge has been appreciably enriched by them (Mencken [1960])	Ironique
	Long ago I proposed that unsuccessful candidates for the Presidency be hanged, as a matter of public sanitation and decorum? On second thoughts, I withdraw my suggestion. It is probably illegal, and maybe even immoral (Mencken [1960])	Non-Ironique
	And, as everyone knows, persons of German descent may now drink beer in Chicago and several Irishmen have been known to hold jobs on the police force. (Royko [1967])	Ironique
	The war was on and I was almost the only roughly male person left with two arms and two legs (Crisp [1968])	Non-Ironique

Catégorie	Énoncé ironique et leur contexte	Prédiction de BERT
	The young man, who was twenty, indicated to Howe, who is thirty-five, that he was not not one of Howe's admirers. At Howe's age, people are usually set in their ideas about how the young should act, so Howe suggested that he shut up. (Royko [1967])	Non-Ironique
	His work had involved a sort of nation-level accountancy, developing useful statistical measures of the extent to which rich countries were getting richer and poor countries were getting poorer? It seems to me that such work hardly compares with the discovery of the background radiation from the birth of the universe that clinched it for the Big Bang Theory of cosmology (Pullum [1991])	Ironique
	Some have claimed to see hints of something autobiographical in it; after all, I have participated in collaborative work on grammatical theory with teams of various sizes, sometimes four. Pullum [1991])	Non-Ironique
Question rhétoriques	If truth is not to be found on the shelves of the British Museum, where, I asked myself, picking up a notebook and a pencil, is truth? (Wolf [1929])	Non-Ironique
	Here this country has got its hands full trying to preserve freedom in Vietnam. How can we do that if our own people won't shut up and do what they're told (Royko [1968])	Non-Ironique
	Napalm bombs on villages are one thing. How can we make Vietnam free if we don't first burn it? But infiltrating pool halls with social workers is an atrocity. (Royko [1973])	Ironique
	And if you could teach her to bark at burglars, what more could a man ask? (Royko [1982])	Non-Ironique
	Some woman in New Jersey- her name is Alice- saved up all the mail she got last year asking for money and sent it to me. How can I ever thank you enough, Alice (Rooney [1981])	Non-Ironique

Catégorie	Énoncé ironique et leur contexte	Prédiction de BERT
	There were other doctors present who didn't think alcohol was good for you and I personally tend to doubt it myself, but- what difference does that make? At least it's good news, and if it's good news, who cares if it's true or not- right America? (Rooney [1981])	Non-Ironique
	The New York Times, America's closest approach to a serious newspaper of record, has changed it's position on the snow-term count by over 50% within four years. And in the Science section. But hey : nine, forty-eight, a hundred, two hundred, who cares? (Pullum [1991])	Non-Ironique
	The protestant "orange" or "loyalist" irish have plenty of excuses themselves. They're patriots after all. "We'll fight anyone to say British, even the British," they like to say. What could be more patriotic than that? (O'Rourke [1992])	Non-Ironique
Suggestions ridicules	Their especial quality was Victorianism now put into a perverse form. I was only sorry that this was not enhanced by giving the painting such titles as "The Blight of the World" or "when did you last rape your father" (Crisp [1968])	Non-Ironique
	I've got just one more idea for the last debate. We'll all now what the candidates think by them, and we are, after all, electing a First Lady, too. So the last debate will be different. For half an hour, each candidate will argue with his wife (Rooney [1981])	Non-Ironique
	Christmas obviously should be moved. No one knows for sure which day Christ was born on. And he certainly wouldn't mind if we celebrated Christmas in February? So here's what I'm suggesting. Move the whole month of January, including my birthday, so that it falls between August and September. Eliminate Labor Day altogether (Rooney [1981])	Non-Ironique

Catégorie	Énoncé ironique et leur contexte	Prédiction de BERT
	If I had my way, the building code would decree that there be at least three door knobs on every closet door. We all know a door knob is the best and easiest place to hang something (Rooney [1981])	Ironique
	My advice is for them to consider some other way. Elastic bands, for example, are infinitely better than glue? String or rope is a possible alternative to glue (Rooney [1981])	Non-Ironique
	There probably ought to be a law passed saying you can't buy new piece of clothing without throwing one out. Maybe the law could work like deposit bottles (Rooney [1981])	Ironique
	A course based on the classified advertising section of any newspaper would provide lots of examples of how we actually use the English language in practical situations. A study of the classifieds would also give students the same kind of mind-expanding exercise that translating French into English gives them now (Rooney [1981])	Ironique
	If I'm ever elected to Congress, I think I'll try to get a law passed making tomatoes and melons illegal, except for certain months of the year (Rooney [1981])	Non-Ironique
	To be honest with you, I think everyone but me should be limited to driving fifty-five mph? If fifty-five is safe for a competent driver, we ought to have a lower limit for incompetent ones (Rooney [1981])	Non-Ironique
	The first one to torpedo the freighter and leave the scene without looking for survivors would be the winner (Rooney [1981])	Ironique
	Just trust me that I'm in one hundred percent agreement with you on these issues. I can't imagine how any right-thinking person could possibly believe other we do (Rooney [1981])	Non-Ironique

Ironie irrealis

Catégorie	Énoncé ironique et leur contexte	Prédiction de BERT
	I admitted that I sometimes make purchases at a Salvation Army resale shop. "Ah", the reporter said, "You do shop after all." Of course, I said, I have loved ones, and I have to get them something for Christmas and Birthdays. I'm no scrooge. (Royko [1989])	Non-Ironique
	I actually heard the secretary of State say that America will help rebuild the economy of Iraq after this war. Mr Baker, I just can't tell you what enthusiasm your plan has inspired in Saudi Arabia (O'Rourke [1992])	Ironique
	Maybe it would be a good idea to make every ex-President a Methodist bishop. (Mencken [1960])	Ironique
	There seems to be a need for another special unit to cope with Chicago's bombings. This new unit should be given the job of issuing fresh, new indignant statements about bombings and the crime syndicate (Royko [1967])	Ironique
	I may never eat again, now that I know how many people have had their hands in my mouth (Royko [1973])	Non-Ironique
	The doorman and the other employe were standing outside, craning their necks, and looking from side to side, as if trying to see where the action was. Or maybe they were looking for someone who had change for a quarter. (Royko [1973])	Non-Ironique
	My proposals were simple : We must legalize the possession of sawed-off shotguns, portable and fixed machine guns, bazookas, flamethrowers, hand grenades, mortars and light artillery (Royko [1982])	Ironique
	If a baby is chucked under the chin by everyone it meets, it could develop a callused jaw. So maybe chin-chucking is a little known abuse we ought to think about. (Royko [1982])	Non-Ironique

Catégorie	Énoncé ironique et leur contexte	Prédiction de BERT
	You can save a lot of energy by switching off lights and turning down the heat. Anyway here is what they say they will do to conserve energy. You might pick up some helpful tips. (Royko [1982])	Non-Ironique

BIBLIOGRAPHIE

- Kalaivani Adaikkan and Thenmozhi Durairaj. Sarcasm identification and detection in conversation context using BERT. In *Proceedings of the Second Workshop on Figurative Language Processing*, pages 72–76, Online, July 2020. Association for Computational Linguistics. doi : 10.18653/v1/2020.figlang-1.10. URL <https://aclanthology.org/2020.figlang-1.10>.
- Abhishek Agrawal, Abhishek Kumar Jha, Ashish Jaiswal, and Vinod Kumar. Irony detection using transformers. In *2020 International Conference on Computing and Data Science (CDS)*, pages 165–168. IEEE, 2020.
- Laura Alba-Juez and Salvatore Attardo. The evaluative palette of verbal irony. *Evaluation in context*, 242, 2014.
- Alias-i. Lingpipe, 2008. URL <http://alias-i.com/lingpipe>.
- Salvatore Attardo. Irony as relevant inappropriateness. *Journal of pragmatics*, 32(6) :793–826, 2000.
- David Bamman and Noah A Smith. Contextualized sarcasm detection on twitter. In *Ninth international AACL conference on web and social media*. Citeseer, 2015.
- Katharina Barbe. "isn't it ironic that..." : Explicit irony markers. *Journal of Pragmatics*, 20(6) :579–590, 1993.
- Arup Baruah, Kaushik Das, Ferdous Barbhuiya, and Kuntal Dey. Context-aware sarcasm detection using BERT. In *Proceedings of the Second Workshop on Figurative Language Processing*, pages 83–87, Online, July 2020. Association for Computational Linguistics. doi : 10.18653/v1/2020.figlang-1.12. URL <https://aclanthology.org/2020.figlang-1.12>.
- Katharine P Beals. *A linguistic analysis of verbal irony*. PhD thesis, University of Chicago, Department of Linguistics, 1995.
- Emily M Bender and Alexander Koller. Climbing towards nlu : On meaning, form, and understanding in the age of data. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5185–5198, 2020.
- Emily M Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. On the dangers of stochastic parrots : Can language models be too big?. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, pages 610–623, 2021.
- Mondher Bouazizi and Tomoaki Ohtsuki. Sarcasm detection in twitter : " all your products are incredibly amazing!!!"-are they really? In *2015 IEEE Global Communications Conference (GLOBECOM)*, pages 1–6. IEEE, 2015.
- Samuel R Bowman, Gabor Angeli, Christopher Potts, and Christopher D Manning. A large annotated corpus for learning natural language inference. *arXiv preprint arXiv :1508.05326*, 2015.
- Gregory A Bryant and Jean E Fox Tree. Is there an ironic tone of voice? *Language and speech*, 48(3) : 257–277, 2005.
- Joan L Bybee. " irrealis" as a grammatical category. *Anthropological linguistics*, pages 257–271, 1998.

- Paula Carvalho, Luís Sarmento, Mário J Silva, and Eugénio De Oliveira. Clues for detecting irony in user-generated contents : oh...!! it's" so easy";-. In *Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion*, pages 53–56, 2009.
- Coralie Chevallier, Gregor Kohls, Vanessa Troiani, Edward S Brodtkin, and Robert T Schultz. The social motivation theory of autism. *Trends in cognitive sciences*, 16(4) :231–239, 2012.
- Wing-Yee Chow, Cybelle Smith, Ellen Lau, and Colin Phillips. A “bag-of-arguments” mechanism for initial verb predictions. *Language, Cognition and Neuroscience*, 31(5) :577–596, 2016.
- Herbert H Clark. *Using language*. Cambridge university press, 1996.
- Herbert H Clark and Richard J Gerrig. On the pretense theory of irony. 1984.
- Reuben Cohn-Gordon and Leon Bergen. Verbal irony, pretense, and the common ground. 2019.
- Quentin Crisp. *The Naked Civil Servant*. Holt, Rinehart and Winston, 1968.
- Quentin Crisp. *How to go to the Movies*. St. Martin's Press, 1984.
- Dmitry Davidov, Oren Tsur, and Ari Rappoport. Semi-supervised recognition of sarcasm in twitter and amazon. In *Proceedings of the fourteenth conference on computational natural language learning*, pages 107–116, 2010.
- Julia Rose DeCook. Trust me, i'm trolling : Irony and the alt-right's political aesthetic. *M/c Journal*, 23(3), 2020.
- Gaétane Deliens, Fanny Papastamou, Nicolas Ruytenbeek, Philippine Geelhand, and Mikhail Kissine. Selective pragmatic impairment in autism spectrum disorder : Indirect requests versus irony. *Journal of autism and developmental disorders*, 48(9) :2938–2952, 2018.
- Maureen Dennis, Nevena Simic, Alba Agostino, H Gerry Taylor, Erin D Bigler, Kenneth Rubin, Kathryn Vannatta, Cynthia A Gerhardt, Terry Stancin, and Keith Owen Yeates. Irony and empathy in children with traumatic brain injury. *Journal of the International Neuropsychological Society*, 19 (3) :338–348, 2013.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert : Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv :1810.04805*, 2018.
- Shelly Dews and Ellen Winner. Muting the meaning a social function of irony. *Metaphor and Symbol*, 10 (1) :3–19, 1995.
- Marta Dynel. Irony from a neo-gricean perspective : On untruthfulness and evaluative implicature. *Intercultural Pragmatics*, 10(3) :403–431, 2013.
- Chris Emmery, Ákos Kádár, Grzegorz Chrupała, and Walter Daelemans. Cyberbullying classifiers are sensitive to model-agnostic perturbations. *arXiv preprint arXiv :2201.06384*, 2022.
- Allyson Ettinger. What bert is not : Lessons from a new suite of psycholinguistic diagnostics for language models. *Transactions of the Association for Computational Linguistics*, 8 :34–48, 2020.
- Mohammed Farghal. Euphemism in arabic : A gricean interpretation. *Anthropological Linguistics*, pages 366–378, 1995.

- Delia Irazú Hernández Farías, Viviana Patti, and Paolo Rosso. Irony detection in twitter : The role of affective content. *ACM Transactions on Internet Technology (TOIT)*, 16(3) :1-24, 2016.
- Kara D Federmeier and Marta Kutas. A rose by any other name : Long-term memory structure and sentence processing. *Journal of memory and Language*, 41(4) :469-495, 1999.
- Ira Fischler, Paul A Bloom, Donald G Childers, Salim E Roucos, and Nathan W Perry Jr. Brain potentials related to stages of sentence verification. *Psychophysiology*, 20(4) :400-409, 1983.
- Michael C Frank and Noah D Goodman. Predicting pragmatic reasoning in language games. *Science*, 336 (6084) :998-998, 2012.
- Joana Garmendia. A (neo) gricean account of irony : an answer to relevance theory. *International Review of Pragmatics*, 7(1) :40-79, 2015.
- Atticus Geiger, Hanson Lu, Thomas Icard, and Christopher Potts. Causal abstractions of neural networks. *Advances in Neural Information Processing Systems*, 34, 2021.
- Aniruddha Ghosh and Tony Veale. Fracking sarcasm using neural network. In *Proceedings of the 7th workshop on computational approaches to subjectivity, sentiment and social media analysis*, pages 161-169, 2016.
- Raymond W Gibbs. Irony in talk among friends. *Metaphor and symbol*, 15(1-2) :5-27, 2000.
- Rachel Giora. On irony and negation. *Discourse processes*, 19(2) :239-264, 1995.
- Cliff Goddard. "joking, kidding, teasing" : Slippery categories for cross-cultural comparison but key words for understanding anglo conversational humor. *Intercultural Pragmatics*, 15(4) :487-514, 2018.
- Yoav Goldberg. Assessing bert's syntactic abilities. *arXiv preprint arXiv :1901.05287*, 2019.
- H Paul Grice. Further notes on logic and conversation. In *Pragmatics*, pages 113-127. Brill, 1978.
- H Paul Grice. Logic and conversation. william james lectures, 1967. reprinted in : H. p. grice, ed. *Studies in the Way of Words*, pages 22-40, 1989.
- Herbert P Grice. Logic and conversation. In *Speech acts*, pages 41-58. Brill, 1975.
- Raj Kumar Gupta and Yinping Yang. Crystalnet at semeval-2017 task 4 : Using sarcasm detection for enhancing sentiment classification and quantification. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 626-633, 2017.
- Suchin Gururangan, Swabha Swayamdipta, Omer Levy, Roy Schwartz, Samuel R Bowman, and Noah A Smith. Annotation artifacts in natural language inference data. *arXiv preprint arXiv :1803.02324*, 2018.
- Yu-Hsiang Huang, Hen-Hsen Huang, and Hsin-Hsi Chen. Irony detection with attentive recurrent neural networks. In *European Conference on Information Retrieval*, pages 534-540. Springer, 2017.
- Vladimir Jankélévitch. L'ironie (1936 ?). *Paris, Flammarion*, pages 16-17, 1979.
- Aditya Joshi, Vinita Sharma, and Pushpak Bhattacharyya. Harnessing context incongruity for sarcasm detection. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2 : Short Papers)*, pages 757-762, 2015.

- Dan Jurafsky and James Martin. *Speech and language processing*. 2021.
- Justine T Kao and Noah D Goodman. Let's talk (ironically) about the weather : Modeling verbal irony. In *CogSci*, 2015.
- Jihen Karoui, Farah Benamara Zitoune, Véronique Moriceau, Nathalie Aussenac-Gilles, and Lamia Hadrich Belguith. Towards a contextual pragmatic model to detect irony in tweets. 2015.
- Catherine Kerbrat-Orecchioni. *L'ironie*. Presses universitaires de Lyon, 1978.
- Jordan Kodner, Sarah Payne, and Jeffrey Heinz. Why linguistics will thrive in the 21st century : A reply to piantadosi (2023). *arXiv preprint arXiv :2308.03228*, 2023.
- Narine Kokhlikyan, Vivek Miglani, Miguel Martin, Edward Wang, Bilal Alsallakh, Jonathan Reynolds, Alexander Melnikov, Natalia Kliushkina, Carlos Araya, Siqi Yan, et al. Captum : A unified and generic model interpretability library for pytorch. *arXiv preprint arXiv :2009.07896*, 2020.
- Roger Kreuz and Gina Caucci. Lexical influences on the perception of sarcasm. In *Proceedings of the Workshop on computational approaches to Figurative Language*, pages 1-4, 2007.
- Roger J Kreuz and Sam Glucksberg. How to be sarcastic : The echoic reminder theory of verbal irony. *Journal of experimental psychology : General*, 118(4) :374, 1989.
- Roger J Kreuz and Richard M Roberts. Two cues for verbal irony : Hyperbole and the ironic tone of voice. *Metaphor and symbol*, 10(1) :21-31, 1995.
- Sunjae Kwon, Cheongwoong Kang, Jiyeon Han, and Jaesik Choi. Why do masked neural language models still need common sense knowledge ? *arXiv preprint arXiv :1911.03024*, 2019.
- Stephen C Levinson. A review of relevance. 1989.
- Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta : A robustly optimized bert pretraining approach. *arXiv preprint arXiv :1907.11692*, 2019.
- Alex John London. Artificial intelligence and black-box medical decisions : accuracy versus explainability. *Hastings Center Report*, 49(1) :15-21, 2019.
- Ingerith Martin and Skye McDonald. Evaluating the causes of impaired irony comprehension following traumatic brain injury. *Aphasiology*, 19(8) :712-730, 2005.
- Laura Martin. Irrealis constructions in mocho (mayan). *Anthropological linguistics*, pages 198-213, 1998.
- Marta Mateo. The translation of irony. *Meta : journal des traducteurs/Meta : Translators' Journal*, 40(1) : 171-178, 1995.
- H.L. Mencken. *On Politics : A Carnival of Buncombe*. Malcolm Moos, 1960.
- Douglas Colin Muecke. *Irony*, volume 13. Methuen, 1970.
- Douglas Colin Muecke. *Irony and the Ironic*, volume 12. Taylor & Francis, 2017.
- Elizabeth S Nilsen, Melanie Glenwright, and Vanessa Huyder. Children and adults understand that verbal irony interpretation depends on listener knowledge. *Journal of Cognition and Development*, 12 (3) :374-409, 2011.

- P.J. O'Rourke. *Give War A Chance : Eyewitness Accounts of Mankind's Struggle against Tyranny, Injustice, and Alcohol-free Beer*. The Atlantic Monthly Press, 1992.
- Penny M Pexman and Meghan T Zvaigzne. Does irony go better with friends? *Metaphor and symbol*, 19 (2) :143–163, 2004.
- Soujanya Poria, Erik Cambria, Devamanyu Hazarika, and Prateek Vij. A deeper look into sarcastic tweets using deep convolutional neural networks. *arXiv preprint arXiv :1610.08815*, 2016.
- Rolandos Alexandros Potamias, Georgios Siolas, and Andreas-Georgios Stafylopatis. A transformer-based approach to irony and sarcasm detection. *Neural Computing and Applications*, 32(23) : 17309–17320, 2020.
- Christopher Potts. The syntax and semantics of as-parentheticals. *Natural Language & Linguistic Theory*, 20 :623–689, 2002.
- Geoffrey K. Pullum. *The Great Eskimo Vocabulary hoax and Other Irreverent Essays On the Study of Language*. University of Chicago Press, 1991.
- Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. Squad : 100,000+ questions for machine comprehension of text. *arXiv preprint arXiv :1606.05250*, 2016.
- Antonio Reyes, Paolo Rosso, and Davide Buscaldi. From humor recognition to irony detection : The figurative language of social media. *Data & Knowledge Engineering*, 74 :1–12, 2012.
- Julian Risch and Ralf Krestel. Toxic comment detection in online discussions. *Deep learning-based approaches for sentiment analysis*, pages 85–109, 2020.
- Andrew A. Rooney. *A few minutes with Andy Rooney*. Atheneum, 1981.
- Mike Royko. *Up Against it*. H. Regnery Co, 1967.
- Mike Royko. *I May Be Wrong But I Doubt It*. H. Regnery Co, 1968.
- Mike Royko. *Slats Grobnik and some Other Friends*. Dutton, 1973.
- Mike Royko. *Sez Who ? Sez Me*. Warner Books, 1982.
- Mike Royko. *Dr.Kookie, You're Right*. Dutton, 1989.
- Rachel Rhoda Schaffer. *Vocal cues for irony in English*. PhD thesis, The Ohio State University, 1982.
- Aron H Schaier and Victor G Cicirelli. Age differences in humor comprehension and appreciation in old age. *Journal of Gerontology*, 31(5) :577–582, 1976.
- John R Searle. *A taxonomy of illocutionary acts*. 1975.
- Ken-ichi Seto. On non-echoic irony. *PRAGMATICS AND BEYOND NEW SERIES*, pages 239–256, 1998.
- Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D Manning, Andrew Y Ng, and Christopher Potts. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 conference on empirical methods in natural language processing*, pages 1631–1642, 2013.

- Dan Sperber and Deirdre Wilson. *Relevance : Communication and cognition*. *Relevance : Communication and cognition*, 2, 1995.
- Robert Stalnaker. Common ground. *Linguistics and philosophy*, 25(5/6) :701–721, 2002.
- Jennifer Tehan Stanley, Monika Lohani, and Derek M Isaacowitz. Age-related differences in judgments of inappropriate behavior are related to humor style preferences. *Psychology and aging*, 29(3) :528, 2014.
- Andreas Stokke. Lying and asserting. *The Journal of philosophy*, 110(1) :33–60, 2013.
- Joseph Tepperman, David Traum, and Shrikanth Narayanan. " yeah right" : Sarcasm recognition for spoken dialogue systems. In *Ninth international conference on spoken language processing*, 2006.
- Mike Thelwall, Kevan Buckley, Georgios Paltoglou, Di Cai, and Arvid Kappas. Sentiment strength detection in short informal text. *Journal of the American society for information science and technology*, 61(12) :2544–2558, 2010.
- Cynthia Van Hee. *Can machines sense irony ? : exploring automatic irony detection on social media*. PhD thesis, Ghent University, 2017.
- Cynthia Van Hee, Els Lefever, and Véronique Hoste. Guidelines for annotating irony in social media text. Technical report, version 2.0. Technical Report 16-01, LT3, Language and Translation . . . , 2016.
- Cynthia Van Hee, Els Lefever, and Véronique Hoste. Semeval-2018 task 3 : Irony detection in english tweets. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 39–50, 2018.
- William F Vitulli. Humor and gender roles : does age make a difference ? *Psychological reports*, 97(1) : 167–168, 2005.
- Byron C Wallace. Computational irony : A survey and new perspectives. *Artificial intelligence review*, 43(4) :467–483, 2015.
- Alex Wang, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, and Samuel R Bowman. Glue : A multi-task benchmark and analysis platform for natural language understanding. *arXiv preprint arXiv :1804.07461*, 2018.
- Amy Beth Warriner, Victor Kuperman, and Marc Brysbaert. Norms of valence, arousal, and dominance for 13,915 english lemmas. *Behavior research methods*, 45(4) :1191–1207, 2013.
- Deirdre Wilson and Dan Sperber. On verbal irony. *Lingua*, 87(1) :53–76, 1992.
- Deirdre Wilson and Dan Sperber. *Relevance theory*. Blackwell, 2002.
- Deirdre Wilson and Dan Sperber. *Meaning and relevance*. Cambridge University Press, 2012.
- Virginia Wolf. *A Room of Ones Own*. Harcourt, Brace and World, 1929.
- Chuhan Wu, Fangzhao Wu, Sixing Wu, Junxin Liu, Zhigang Yuan, and Yongfeng Huang. Thu_ugn at semeval-2018 task 3 : Tweet irony detection with densely connected lstm and multi-task learning. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 51–56, 2018.
- WANG Xin. Cooperative principle in english euphemism. *English Language Teaching*, 3(1) :11–14, 2015.

Rowan Zellers, Yonatan Bisk, Roy Schwartz, and Yejin Choi. Swag : A large-scale adversarial dataset for grounded commonsense inference. *arXiv preprint arXiv :1808.05326*, 2018.

Cheng Zhang, Masashi Kudo, and Hayato Yamana. Evaluation of bert and xlnet models on irony detection in english tweets.

Yukun Zhu, Ryan Kiros, Rich Zemel, Ruslan Salakhutdinov, Raquel Urtasun, Antonio Torralba, and Sanja Fidler. Aligning books and movies : Towards story-like visual explanations by watching movies and reading books. In *Proceedings of the IEEE international conference on computer vision*, pages 19–27, 2015.