

# Table des matières

<b>Avant-propos</b> . . . . .	1
<b>Introduction</b> . . . . .	3
<b>Chapitre 1. De l'analyse d'opinion au traitement du langage figuratif</b> . . . . .	9
1.1. Introduction . . . . .	9
1.2. Définition de la notion d'opinion . . . . .	10
1.2.1. Multiples facettes de l'opinion . . . . .	10
1.2.2. L'opinion vue comme un modèle structuré . . . . .	11
1.2.3. Extraction d'opinions : principales approches . . . . .	12
1.3. Limites des systèmes d'analyse d'opinion . . . . .	16
1.3.1. Opérateurs d'opinion . . . . .	16
1.3.2. Dépendance au domaine . . . . .	17
1.3.3. Opinion implicite . . . . .	17
1.3.4. Opinion et contexte discursif au-delà de la phrase . . . . .	18
1.3.5. Présence d'expressions figuratives . . . . .	19
1.4. Définition du langage figuratif . . . . .	19
1.4.1. Ironie . . . . .	20
1.4.1.1. Ironie verbale . . . . .	21
1.4.1.2. Ironie de situation . . . . .	24
1.4.2. Sarcasme . . . . .	25
1.4.3. Satire . . . . .	26
1.4.4. Métaphore . . . . .	27
1.4.5. Humour . . . . .	28
1.5. Traitement automatique du langage figuratif : un défi pour le TAL . . . . .	28
1.6. Conclusion . . . . .	29

<b>Chapitre 2. Vers la détection automatique du langage figuratif . . .</b>	<b>31</b>
2.1. Introduction . . . . .	31
2.2. Principaux corpus existants pour le langage figuratif . . . . .	33
2.2.1. Corpus annotés en ironie/sarcasme . . . . .	34
2.2.1.1. Senti-TUT : corpus de tweets en italien . . . . .	34
2.2.1.2. Corpus de tweets en anglais et néerlandais . . . . .	36
2.2.2. Corpus annotés en métaphore . . . . .	38
2.2.2.1. Collecte du corpus . . . . .	38
2.2.2.2. Schéma d'annotation . . . . .	39
2.2.2.3. Résultats de la campagne d'annotation . . . . .	40
2.3. Détection automatique de l'ironie, du sarcasme et de la satire . . . . .	41
2.3.1. Approches surfaciques et sémantiques . . . . .	41
2.3.2. Approches pragmatiques . . . . .	43
2.3.2.1. Approches pragmatiques exploitant le contexte interne de l'énoncé . . . . .	43
2.3.2.2. Approches pragmatiques exploitant le contexte externe de l'énoncé . . . . .	53
2.4. Détection automatique de la métaphore . . . . .	55
2.4.1. Approches surfaciques et sémantiques . . . . .	55
2.4.2. Approches pragmatiques . . . . .	57
2.4.2.1. Approches pragmatiques exploitant le contexte interne de l'énoncé . . . . .	57
2.4.2.2. Approches pragmatiques exploitant le contexte externe de l'énoncé . . . . .	59
2.5. Détection automatique de la comparaison . . . . .	61
2.6. Détection automatique de l'humour . . . . .	62
2.7. Conclusion . . . . .	64
 <b>Chapitre 3. Un schéma multiniveau pour l'annotation de l'ironie dans les réseaux sociaux . . . . .</b>	 <b>67</b>
3.1. Introduction . . . . .	67
3.2. Corpus FrIC . . . . .	69
3.3. Schéma d'annotation multiniveau . . . . .	70
3.3.1. Méthodologie . . . . .	70
3.3.2. Schéma d'annotation . . . . .	73
3.3.2.1. Niveau 1 : classes du tweet . . . . .	73
3.3.2.2. Niveau 2 : types de l'ironie . . . . .	74
3.3.2.3. Niveau 3 : catégories de l'ironie . . . . .	76
3.3.2.4. Niveau 4 : marqueurs linguistiques de l'ironie . . . . .	80
3.4. Campagne d'annotation . . . . .	81
3.4.1. Présentation de l'outil Glozz . . . . .	81
3.4.2. Préparation des données . . . . .	83

3.4.3. Procédure d'annotation . . . . .	84
3.5. Résultats de la campagne d'annotation . . . . .	85
3.5.1. Résultats qualitatifs . . . . .	86
3.5.2. Résultats quantitatifs . . . . .	86
3.5.2.1. Fréquence des tweets selon les classes . . . . .	87
3.5.2.2. Fréquence des tweets selon le type de l'ironie . . . . .	87
3.5.2.3. Fréquence des tweets selon les catégories de l'ironie . . . . .	88
3.5.2.4. Fréquence des tweets selon les indices linguistiques . . . . .	89
3.5.2.5. Fréquence des relations dans les tweets ironiques avec contradiction explicite . . . . .	92
3.5.3. Corrélation entre les différents niveaux du schéma d'annotation . . . . .	92
3.5.3.1. Corrélation entre le type et les indices de l'ironie . . . . .	93
3.5.3.2. Corrélation entre l'activation et les indices de l'ironie . . . . .	93
3.5.3.3. Corrélation entre les différents indices de l'ironie . . . . .	94
3.5.3.4. Corrélation entre les catégories et les indices de l'ironie . . . . .	94
3.6. Conclusion . . . . .	94

## **Chapitre 4. Trois modèles pour la détection automatique de l'ironie . . . . .**

97

4.1. Introduction . . . . .	97
4.2. Corpus FrIC <sup>Auto</sup> . . . . .	98
4.3. Modèle SurfSystem : détection de l'ironie sur la base de traits surfaciques . . . . .	101
4.3.1. Traits utilisés . . . . .	101
4.3.2. Expériences et résultats . . . . .	102
4.4. Modèle PragSystem : détection de l'ironie sur la base de traits contextuels internes . . . . .	104
4.4.1. Traits utilisés . . . . .	105
4.4.1.1. Traits de surface . . . . .	105
4.4.1.2. Traits de sentiment . . . . .	105
4.4.1.3. Traits pour les modifieurs . . . . .	106
4.4.1.4. Traits pour les modifieurs de sentiment . . . . .	107
4.4.1.5. Traits de contexte . . . . .	107
4.4.1.6. Traits d'opposition . . . . .	107
4.4.2. Expériences et résultats . . . . .	108
4.4.2.1. Étude de la pertinence de traits au processus d'apprentissage . . . . .	109
4.4.2.2. Apport de chaque groupe de traits au processus d'apprentissage . . . . .	109
4.4.2.3. Résultats pour les meilleures combinaisons de traits . . . . .	109
4.4.3. Discussion . . . . .	115
4.5. Modèle QuerySystem : vers un modèle pragmatique contextuel pour la détection automatique de l'ironie . . . . .	117

4.5.1. Approche proposée . . . . .	117
4.5.2. Expériences et résultats . . . . .	120
4.5.3. Évaluation de la méthode à base de requêtes . . . . .	121
4.6. Conclusion . . . . .	122

## **Chapitre 5. Vers un système multilingue pour la détection automatique de l'ironie . . . . . 125**

5.1. Introduction . . . . .	125
5.2. Ironie dans les langues indo-européennes . . . . .	126
5.2.1. Corpus . . . . .	126
5.2.1.1. Collecte du corpus anglais . . . . .	126
5.2.1.2. Collecte du corpus italien . . . . .	127
5.2.2. Résultats de la procédure d'annotation . . . . .	128
5.2.2.1. Résultats quantitatifs de la procédure d'annotation en ironique/non ironique . . . . .	130
5.2.2.2. Résultats quantitatifs de la procédure d'annotation sur les types de l'ironie . . . . .	130
5.2.2.3. Résultats quantitatifs de la procédure d'annotation sur les catégories de l'ironie . . . . .	131
5.2.2.4. Résultats quantitatifs de la procédure d'annotation pour les indices de l'ironie . . . . .	132
5.2.2.5. Résultats quantitatifs de la procédure d'annotation des relations de contradiction . . . . .	135
5.2.2.6. Corrélation entre les différents niveaux du schéma d'annotation . . . . .	135
5.2.3. Synthèse . . . . .	137
5.3. Ironie dans les langues sémitiques . . . . .	137
5.3.1. Spécificités de la langue arabe . . . . .	139
5.3.2. Corpus et ressources . . . . .	140
5.3.2.1. Collecte du premier corpus arabe pour l'ironie . . . . .	140
5.3.2.2. Ressources linguistiques . . . . .	142
5.3.3. Détection automatique de l'ironie dans les tweets arabes . . . . .	143
5.3.3.1. Traits utilisés pour la détection de l'ironie . . . . .	143
5.3.3.2. Expériences et résultats . . . . .	143
5.3.3.3. Discussion . . . . .	144
5.4. Conclusion . . . . .	146

## **Conclusion . . . . . 147**

## **Annexe. Catégories de l'ironie étudiées dans la littérature linguistique . . . . . 151**

## **Bibliographie . . . . . 163**

## **Index . . . . . 179**