

« DIS-MOI ORLEANS »  
**REPERAGE ET ANALYSE DE LA PERCEPTION D'UN LIEU DANS  
L'ORAL TRANSCRIT**

Iris ESHKOL-TARAVELLA

Laboratoire Ligérien de Linguistique (UMR7270), Université d'Orléans

Hélène FLAMEIN

Laboratoire Ligérien de Linguistique (UMR7270), Université d'Orléans

**Abstract (En):** The work is based on the methodologies of corpus linguistics and Natural Language Processing (NLP). The study concerns places, their perception and the expression of the latter by the different speakers in the ESLO2 corpus (Enquête sociolinguistique à Orléans – Sociolinguistic Investigation in Orléans). Place names are automatically detected thanks to a Named Entities Recognition tool. The manual correction that follows takes into consideration the variations in their denomination and the context of their use. The annotated places are studied quantitatively and qualitatively. The results enable us to make the first typology of the different processes that a speaker implements to express his/her perception of a place.

**Résumé (Fr) :** Le travail repose sur la méthodologie de la linguistique de corpus et du Traitement Automatique des Langues (TAL). L'étude porte sur les lieux, leur perception et l'expression de celles-ci par les différents locuteurs du corpus ESLO2 (Enquête Sociolinguistique à Orléans). Les lieux sont repérés automatiquement à l'aide d'un outil de reconnaissance d'entités nommées. La correction manuelle qui suit tient compte de la variation dans leur dénomination et du contexte de leur emploi. Les lieux annotés sont étudiés quantitativement et qualitativement. Les résultats permettent de faire la première typologie des différents processus que le locuteur met en œuvre pour exprimer sa perception du lieu.

**Keywords (En) :** place perception; place names; subjectivity; natural language processing; named entities; ESLO; oral corpus

**Mots-clés (Fr) :** Perception de lieux ; Désignation de lieux ; Subjectivité, lieux subjectifs ; Traitement Automatique du Langage ; Entités nommées ; ESLO ; Corpus oral

## **Introduction**

L'objectif de l'étude proposée est de présenter l'analyse de la perception des lieux par différents locuteurs dans le discours transcrit de l'oral en utilisant les techniques du Traitement Automatique du Langage (TAL). Ces techniques permettent de détecter automatiquement les lieux ainsi que les expressions qui les accompagnent dans les transcriptions. Le corpus traité est le corpus ESLO2 (Enquête SocioLinguistique à Orléans<sup>1</sup>). Il est composé d'une vingtaine de modules représentant chacun des situations de communication différentes, pour un total de 460 heures d'enregistrement.

Plusieurs étapes du travail seront développées. En premier lieu seront exposés les cadres théoriques dans lesquels s'inscrit cette recherche en rappelant les principes de la détection automatique des différentes désignations des lieux dans le discours. Dans un second temps, la méthodologie sera explicitée avec les différentes étapes de traitement avant de proposer une analyse des résultats obtenus.

---

<sup>1</sup> <http://eslo.huma-num.fr/>

## 1. Cadre théorique

Le travail décrit dans cet article entre dans le domaine du TAL – une discipline à la frontière de la linguistique et de l'informatique – et concerne la détection automatique de la perception des lieux dans le discours oral transcrit. La notion de perception est ici prise dans un sens large incluant tous les éléments subjectifs relatifs à un lieu.

Du point de vue du TAL, l'analyse des éléments subjectifs est souvent liée à la détection automatique des opinions et sentiments dans des corpus écrits et consiste la plupart du temps en l'attribution d'une polarité (positive, neutre, négative) aux énoncés analysés (MARCHAND, MESNARD, BESANCON et VINALT, 2014 ; PAK et PAROUBEK, 2010 ; VERNIER, 2011).

La subjectivité est étudiée dans de nombreux travaux en sciences du langage. Elle est notamment développée à travers la constitution de ressources lexicales composées d'adjectifs et de groupes adjectivaux (HATZIVASSILOGLOU et MCKEOWN, 1997), de verbes (VANDERVEKEN, 1988), d'adverbes évaluatifs (GUIMIER, 2002), de prédicats de sentiment (MATHIEU, 1999) et de mots d'affect (LE PESANT, 2011 ; BRADLEY et LANG, 1999). Du point de vue de l'analyse de discours, la subjectivité est abordée dans les travaux de KERBRAT-ORECCHIONI (1999) et de PLANTIN (2011).

Le travail présenté vise la détection automatique des « lieux subjectifs » introduits par DOMINGUÈS et ESHKOL (2013, 2015). Il s'agit des différents procédés lexicaux, syntaxiques et pragmatiques que le locuteur met en œuvre pour exprimer sa perception d'un lieu. Pour repérer les lieux, les méthodes symboliques utilisent des règles fondées sur le contexte de leur emploi dans le discours.

## 2. Présentation des données

L'étude est fondée sur le corpus ESLO (Enquête SocioLinguistique à Orléans<sup>2</sup>) (ESHKOL-TARAVELLA *et al.*, 2012), projet du Laboratoire Ligérien de Linguistique (UMR7270) de l'Université d'Orléans, qui met au cœur de son investigation les pratiques langagières dans la ville d'Orléans.

### 2.1. Historique du corpus

Au début des années 70, des universitaires britanniques ont initié le projet ESLO en enregistrant plusieurs centaines d'Orléanais dans leur vie de tous les jours. Cette première enquête constitue aujourd'hui le corpus ESLO1. Celui-ci compte environ 200 interviews agrémentées d'un ensemble de métadonnées sociolinguistiques relatives aux locuteurs enregistrés pour un total de plus de 300 heures d'enregistrement.

Quarante ans plus tard, une nouvelle enquête ESLO2 est lancée avec pour objectif de constituer un corpus comparable au premier et de cumuler près de 700 heures d'enregistrement. Le corpus comprend lui aussi plusieurs situations d'enregistrement : entretiens en face à face, interviews de personnalités,

---

<sup>2</sup> <http://eslo.huma-num.fr/>

enregistrements dans des cours de récréation, pendant des repas, etc. Cette seconde enquête constitue la base du travail présenté ici.

## 2.2. Données exploitées pour l'analyse

La transcription des enregistrements suit un protocole très précis et détaillé dans le *Guide du Transcripteur et du Relecteur des ESLOs*<sup>3</sup>. Chaque enregistrement est transcrit orthographiquement avec une distinction entre les tours de parole. La convention de transcription préconise de transcrire sans signes de ponctuation et sans majuscules au début des énoncés. Les exceptions en sont les points d'interrogation qui marquent les questions et les majuscules pour les noms propres.

Trois modules du corpus ESLO2 ont été sélectionnés : « Entretiens », « Entretiens jeunes » et « Itinéraires » (cf. Tableau 1). En considérant la nature des enregistrements et les trames qui ont servi à guider les entretiens, ces trois modules constituent le matériel le plus susceptible de contenir des expressions subjectives relatives aux lieux d'Orléans.

	<i>Entretiens</i>	<i>Entretiens Jeunes</i>	<i>Itinéraires</i>
<i>Description</i>	Discussion en face à face entre un chercheur et un locuteur témoin à partir d'une trame d'entretien	Discussion en face à face entre un jeune chercheur (moins de 25 ans) et un jeune locuteur témoin (18-25 ans) à partir d'une trame d'entretien	Enregistrements dans la rue de demandes d'itinéraire (Comment aller à la mairie d'Orléans ?)
<i>Protocole</i>	Prise de contact préalable : présentation du projet et prise de rendez-vous. Enregistrements réalisés chez les locuteurs témoins		Demandes d'itinéraire dans la rue. Le début est à micro discret ; la suite, à micro montré, donne lieu à une reformulation de l'itinéraire. La collecte a été effectuée dans divers endroits de la ville afin de couvrir des quartiers représentatifs de la diversité sociologique de la ville.
<i>Nombre de transcriptions</i>	84	9	91
<i>Nombre de mots total</i>	1 166 654	91 606	69 328

Tableau 1 : Description des modules sélectionnés dans ESLO2<sup>4</sup>

## 3. Annotation des lieux dans le corpus

Pour étudier la perception des lieux, nous utilisons la technique de l'annotation qui permet de mettre en évidence un lieu et son contexte d'utilisation.

L'annotation des lieux dans le corpus suit deux étapes : annotation automatique et correction manuelle (voir la Figure 1). Les fichiers de transcription sont d'abord

<sup>3</sup> <<http://eslo.huma-num.fr/index.php/pagemethodologie?id=71>>

<sup>4</sup> Les informations du tableau sont issues du site officiel du corpus : <<http://eslo.huma-num.fr/>>

soumis à l'outil qui repère automatiquement des lieux. Ils sont ensuite corrigés manuellement, ce qui permet de constituer un corpus de référence annoté en lieux.

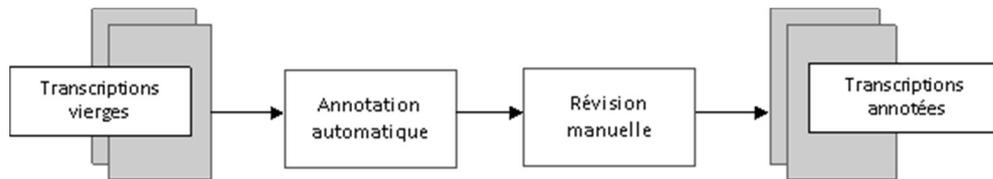


Figure 1 : Chaîne de traitement du corpus

### 3.1. Reconnaissance automatique des lieux

#### 3.1.1. Méthodologie

Pour repérer les lieux, nous utilisons les méthodes symboliques fondées sur des règles prenant en compte leur contexte d'emploi dans le discours. Ces méthodes ont montré leur efficacité dans la tâche de reconnaissance des entités nommées : lieux, personnes, organisations, etc. (EHRMANN, 2008 ; MAUREL *et al.*, 2011 ; NADEAU et SEKINE, 2009 ; NOUVEL *et al.*, 2015). Les entités nommées interviennent en TAL principalement dans une tâche d'extraction d'informations puisqu'elles sont considérées comme des objets textuels porteurs de sens permettant un accès rapide à son contenu.

La première étape du traitement est d'annoter les entités nommées et plus particulièrement celles catégorisées comme lieux par un module libre dédié à cette tâche : CasEN<sup>5</sup>. Ce module a été développé pour le logiciel Unitex<sup>6</sup> : une plateforme multilingue permettant de traiter des textes en langue naturelle au moyen de ressources linguistiques. Le module CasEN est donc spécialisé dans le repérage des entités nommées et répond à une approche symbolique. Afin de correspondre à la tâche traitée, l'outil est adapté pour n'annoter que les entités désignant les lieux dans les transcriptions sélectionnées.

#### 3.1.2. Évaluation

L'évaluation de l'annotation automatique vérifie à quel point l'annotation effectuée est proche de celle de la référence.

Le corpus de référence est un échantillon du corpus dans lequel toutes les mentions de lieux relatives à Orléans sont correctement identifiées et annotées par un humain. Cinq transcriptions ont été aléatoirement sélectionnées parmi les trois modules présentés précédemment (cf. 3.2) :

Transcriptions	Durée	Nombre de mots
ESLO ENT 1059	1:40:00	19 449
ESLO ENT 1002	1:37:00	14 791

<sup>5</sup> [http://tln.li.univ-tours.fr/Tln\\_CasEN.html](http://tln.li.univ-tours.fr/Tln_CasEN.html)

<sup>6</sup> <http://www-igm.univ-mlv.fr/~unitex/>

ESLO_ENT_1034	1:30:00	15 788
ESLO_iti_08_04	0:06:40	1001
ESLO_iti_02_09	0:04:00	299
<b>Totaux</b>	4:57:40	51 328

Tableau 2 : Volumes de données par transcriptions

Pour évaluer l'outil, on utilise les mesures de Précision (P), Rappel (R) et F-mesure (F). Le Rappel montre la part de détections pertinentes par rapport à l'ensemble des détections réalisées par le système. La Précision représente la part des détections pertinentes par rapport à la totalité des détections que le système est censé effectuer. Ces deux mesures prennent en compte le nombre d'éléments correctement (Vrais positifs – VP) et incorrectement (Faux positifs – FP) identifiés. La F-mesure combine le Rappel et la Précision pour informer sur les performances générales du système :

- $P = VP / (VP + FP)$
- $R = VP / (VP + FN)$
- $F = 2 \times (P \times R) / (R + P)$

	<b>Rappel</b>	<b>Précision</b>	<b>F-mesure</b>
ESLO ENT 1059	0,05	1,00	0,09
ESLO ENT 1002	0,15	0,97	0,27
ESLO ENT 1034	0,29	0,91	0,44
ESLO iti 08 04	0,35	0,90	0,90
ESLO iti 02 09	0,50	1,00	0,67

Tableau 3 : Mesures de Rappel, Précision et F-Mesure par transcription

Le Rappel indique le taux de non-détection par un outil d'un élément. Plus ce taux est proche de 0, moins les résultats sont satisfaisants. Suite aux résultats de l'évaluation (cf. Tableau 3), le système détecte peu de lieux présents dans le texte. La Précision illustre quant à elle la qualité des détections par le système. A l'inverse des résultats obtenus au niveau du Rappel, les résultats de la Précision, qui montrent le taux des lieux détectés correctement, sont très satisfaisants. On peut en déduire que le système reconnaît peu de lieux mais qu'il le fait de façon tout à fait acceptable. Les annotations ajoutées aux transcriptions sont donc de bonne qualité mais très insuffisantes en termes de quantité. Il est nécessaire de reprendre ces annotations pour corriger les éventuelles erreurs et surtout les compléter, comme dans l'exemple suivant :

(1) euh tu as la <loc.fac>cathédrale</loc> ou est-ce que se serait plus court ? ah ben si tu peux redescendre tu prends la tu prends la rue qui est là et tu vas tout au bout jusqu'à la rue de la Rép- tu vois où elle est ? la rue de la <org.adm>République</org>  
(ESLO2\_iti\_06\_11\_C)

où un nom de lieu (*cathédrale*) est correctement annoté alors que le lieu (*rue de la République*) est considéré comme une organisation administrative.. Le troisième

lieu (*rue de la Rép*) n'est pas pris en compte car c'est une version abrégée de *rue de la République*.

Ainsi, plusieurs erreurs ont été commises par le système. La première concerne les limites de la balise, la deuxième se rapporte à la typologie de l'entité et la troisième est une erreur de silence, où le lieu, présent sous forme abrégée, ce qui est un phénomène fréquent à l'oral, n'est pas reconnu par le système. La correction manuelle devient donc indispensable.

### 3.2. Correction et validation manuelle du corpus de référence

La troisième étape est la reprise manuelle des différentes transcriptions annotées automatiquement. L'exemple précédent corrigé a la forme suivante :

(1) *euh tu as la <loc.fac>cathédrale</loc> ou est-ce que se serait plus court ? ah ben si tu peux redescendre tu prends la tu prends la rue qui est là et tu vas tout au bout jusqu'à la <loc.oro>rue de la Rép-</loc> tu vois où elle est ? la <loc.oro>rue de la République</loc>*  
(ESLO2\_iti\_06\_11\_C)

Le lieu manquant est désormais annoté en tant qu'oronyme. Le lieu annoté incorrectement est rectifié et également annoté en tant qu'oronyme.

Le corpus de référence ainsi constitué est composé d'un ensemble de 5 transcriptions dans lesquelles 272 mentions de lieux sont correctement annotées. Ce corpus permet de faire une première typologie et une étude préliminaire de l'expression de la perception autour des lieux.

## 4. Typologie de la perception des lieux

Deux types de perception sont distingués : perception à travers la variation dans les dénominations de lieux par des locuteurs et perception manifestée dans le contexte d'emploi des lieux.

### 4.1. Variations des dénominations des lieux

Les entretiens enregistrés portent sur la ville d'Orléans, par conséquent les mentions de lieux y sont fréquentes. Les noms de lieux peuvent être désignés de manières très variées :

(2) *y a quoi par-là y a le cons- euh une partie du conservatoire y a oui si c'est Sainte-Croix quoi la cathédrale* (ESLO2\_iti\_10\_04\_C)

À l'oral, il est extrêmement rare que quelqu'un fasse allusion à la cathédrale en utilisant son nom complet. Son nom officiel est : *Cathédrale Sainte-Croix*. Dans l'exemple (2), le locuteur utilise les éléments du nom complet mais les donne dans le désordre. Il y fait allusion d'abord par le nom propre, et précise ensuite le type du monument. Dans l'exemple suivant :

(3) *donc c- je suppose que c'est la génération d'après et euh mais y avait quand même plein d'oies sur ce ce bout de d'île euh qui est euh pas très loin du pont euh euh du pont George Cinq oui c'est le pont Royal* (ESLO2\_ENT\_1034\_C)

le locuteur mentionne deux fois le même pont. Ce dernier a changé plusieurs fois de noms au cours de son histoire. *George Cinq* est le nom officiel actuel tandis que *Royal* était le nom qu'on lui avait attribué au moment de son inauguration en 1763. Malgré l'ancienneté de ce dernier nom, il est très régulier que les Orléanais se réfèrent à ce pont en utilisant son ancien nom. Enfin dans :

(4) *en gros euh sous les Arcades* (ESLO2\_ENTJEUN\_04\_C)

*les Arcades* correspondent à un surnom donné à la *rue Royale*, une rue centrale à Orléans. Celle-ci est bordée sur toute sa longueur par des galeries à arcades. Cette spécificité architecturale a conduit les Orléanais à se référer à cette rue en substituant à son nom officiel une appellation plus imagée. On observe ainsi une véritable réappropriation du nom de lieu. Faire allusion à une entité en utilisant un surnom est un cas de personnalisation, d'appropriation d'un lieu par un locuteur.

L'enjeu pour le TAL est de reconnaître un lieu mentionné dans un discours de différentes manières : par son nom officiel, son nom ancien ou son surnom. La difficulté principale dans le repérage automatique du lieu est de tenir compte de toutes les variations possibles dans sa dénomination et d'être capable de détecter la coréférence entre plusieurs mentions du même lieu, comme c'est le cas des exemples (3) et (4).

## 4.2. Analyse du contexte des mentions de lieux

La perception d'un lieu peut apparaître dans le contexte où ce lieu est employé.

### 4.2.1. Contexte proche

Le contexte proche est un contexte gauche ou droit de la mention du lieu.

Dans les exemples (5) et (6), le lieu identifié est la cathédrale et les deux locuteurs n'y font pas allusion de la même façon. Le premier s'y réfère de façon neutre en employant l'article défini *la*. Dans l'autre exemple, le locuteur emploie le déterminant possessif *ma*, indice de son appropriation du lieu.

(5) *et tu penses que ça peut enfin la cathédrale par exemple ça peut attirer du monde ? ou pas ?* (ESLO2\_ENTJEUN\_03\_C)

(6) *enfin je sais pas c'est mais à Orléans je le trouve moche le tram ils veulent le faire passer devant la cathédrale devant ma cathédrale* (ESLO2\_ENT\_1003\_C)

Dans l'exemple (7), le locuteur parle de la région et emploie l'adjectif *grande* qui est un indice sur la vision qu'il a du lieu auquel il fait référence.

(7) *b- c'est la grande région euh c'est la grande région euh Centre* (ESLO2\_ENT\_1034\_C)

L'emploi du lexique évaluatif autour du lieu est aussi observé dans les exemples suivants :

(8) *bon puis les bords de Loire sont magnifiques maintenant* (ESLO2\_ENT\_1070\_C)

(9) *Orléans c'est a- ça restera toujours la ville de mon cœur dans le sens où euh où euh où où où je recherche où Orléans c'est plus calme en fait ça me permet de me ressourcer euh de vraiment euh digérer Paris* (ESLO2\_ENTJEUN\_05\_C)

(10) *c'est très bruyant hein ce quartier-là* (ESLO2\_ENT\_1068\_C)

Les locuteurs s'expriment et donnent leur avis, leur sentiment sur les différents lieux d'Orléans.

#### 4.2.2. Contexte large

Le contexte large correspond à l'énoncé entier où est employé le lieu. Il ne se limite pas à un tour de parole comme dans les exemples précédents mais peut être continué sur plusieurs tours :

(11) *RL2: bah ce qu'il me plaît c'est que je trouve que la ville depuis quelques années a un autre aspect quand même elle est  
ch\_CD2: a bien changé oui  
RL2: beaucoup plus accueillante* (ESLO2\_ENT\_1002\_C)

L'avis du locuteur RL2 se trouve interrompu par le commentaire de son interlocuteur ch\_CD2.

Les méthodes symboliques peuvent actuellement tenir compte du contexte proche. Ce n'est pas le cas pour le contexte large.

### 5. Analyse quantitative

Au total, 272 mentions de lieux ont été identifiées dans le corpus de référence. Parmi ces mentions, 45% sont neutres et 55% se répartissent entre les différents niveaux de perception décrits précédemment.

Les cas de variation de noms de lieux sont les moins nombreux et ne représentent que 7% des annotations. Les éléments contextuels sont présents d'une manière significative, avec un équilibre entre les éléments des contextes proche et large.

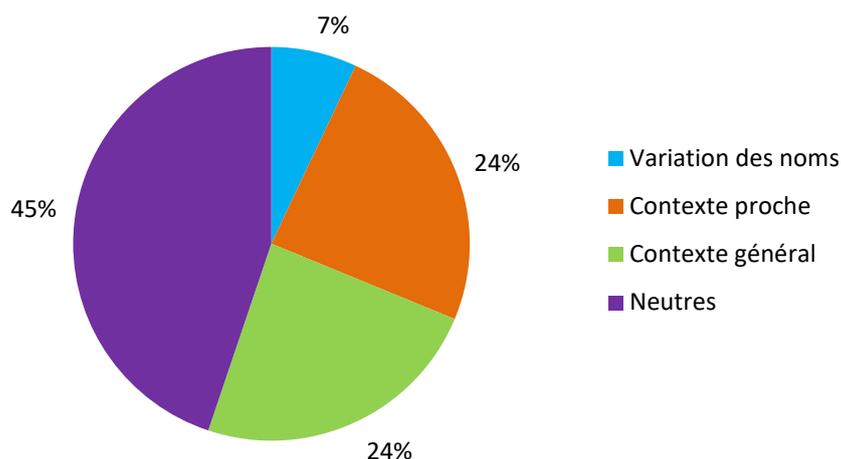


Figure 2 : Répartition des mentions de lieux en fonction du niveau de perception

La répartition des mentions de lieux peut s'expliquer en partie par le contexte d'enregistrement. Les locuteurs interviewés ne connaissent pas leurs interlocuteurs dans presque la totalité des cas. Ils préféreront, dans un souci de clarté, ne pas utiliser de mentions qu'ils pensent méconnues. C'est pourquoi on constate une fréquence non élevée de variations dans les dénominations de lieux. La perception apparaît plutôt dans le contexte d'emploi des lieux.

### Conclusions et perspectives

La dénomination d'un lieu est un processus social réapproprié subjectivement et est déterminée par la personnalité, l'histoire du locuteur. Un lieu peut être approprié, apprécié ou non par un locuteur. La perception d'un lieu peut aussi apparaître à travers l'avis, le jugement, le sentiment, la sensation, etc., exprimés par un locuteur et à travers sa dénomination. Repérer automatiquement un lieu et sa perception est une tâche difficile.

Les perspectives du travail présenté concernent le développement d'un module de reconnaissance automatique de lieux et de leur perception par des locuteurs dans le discours oral transcrit et d'un module de leur géoréférencement automatique. L'objectif est d'offrir une visualisation cartographique de la perception des Orléanais de leur lieu de vie. Les expressions identifiées dans les transcriptions analysées formeront la vision que des Orléanais ont de leur propre ville. Cette vision sera comparée avec celle proposée par les données issues des bases de données comme Linked Open Data ou Wikipédia. Le fait de prendre en compte les témoignages d'ESLO et d'y associer l'ensemble des informations neutres et objectives des bases de données donnera une dimension anthropologique et sociologique à la carte produite et permettra de constituer un véritable portrait de la ville d'Orléans.

## BIBLIOGRAPHIE

- BRADLEY M. Margaret ; LANG J. Peter (1999), *Affective norms for English words (ANEW): Instruction manual and affective ratings*, Technical Report C-1, The Center for Research in Psychophysiology, Florida, University of Florida, p. 1-45.
- DOMINGUÈS Catherine ; ESHKOL-TARAVELLA Iris (2015), Toponym recognition in custom-made map titles, *International Journal of Cartography* 1, Taylor & Francis.
- EHRMANN Michael (2008), Les entités nommées, de la linguistique au TAL : statut théorique et méthode de désambiguïsation, Thèse de doctorat, Paris, Université Paris 7.
- ESHKOL-TARAVELLA Iris ; BAUDE Olivier ; MAUREL Denis ; HRIBA Linda ; DUGUA Céline ; TELLIER Isabelle (2012), Un grand corpus oral « disponible » : le corpus d'Orléans 1968-2012, *Ressources linguistiques libres*, TAL 53, Orléans, Université d'Orléans, p. 17-46.
- GUIMIER Claude (2002), *Les adverbes du français*, Paris, Ophrys.
- KERBAT-ORECCHIONI Catherine (1999), *L'Énonciation. De la subjectivité dans le langage*, Paris, Armand Colin.
- HATZIVASSILOGLOU Vasileios ; MCKEOWN Kathleen (1997), Predicting the semantic orientation of adjectives, *Proceedings of The eighth conference on European chapter of the Association for Computational Linguistics*, Association for Computational Linguistics (ACL), Morristown, NJ, USA, p. 174-181.
- LE PESANT Denis (2011), Vers un thesaurus syntactico-sémantique des mots d'affect, *Cahiers de Lexicologie* 2(99), Centre National de la Recherche Scientifique, p. 117-132.
- MARCHAND Morgane ; BESANCON Romaric ; MESNARD Olivier ; VILNAT Anne (2014), Influence des marqueurs multi-polaires dépendant du domaine pour la fouille d'opinion au niveau du texte, *Actes de TALN 21*, Marseille, p. 1-12.
- MATHIEU Yvette Yannick (1999), Les prédicats de sentiment, *Langages* 33 (136), Paris, Larousse, p. 41-52.
- MAUREL Denis ; FRIBURGER Nathalie ; ANTOINE Jean-Yves ; ESHKOL-TARAVELLA Iris ; NOUVEL Damien (2011), Cascades de transducteurs autour de la reconnaissance des entités nommées, *TAL* 52(1), p. 69-96.
- NADEAU David ; SEKINE Satoshi (2009), a survey of named entity recognition and classification, in : SEKINE Satoshi, RANCHHOD Elisabete (éd.), *Named Entities : Recognition, Classification and Use*, Amsterdam, John Benjamins publishing company, p. 3-28.
- NOUVEL Damien ; EHRMANN Maud ; ROSSET Sophie (2015), *Les entités nommées pour le traitement automatique des langues*, London, ISTE.
- PAK Alexander ; PAROUBEK Patrick (2010), Construction d'un lexique affectif pour le français à partir de Twitter, *TALN 19-23 juillet 2010*, Montréal.
- PLANTIN Christian (2011), *Les bonnes raisons des émotions. Principes et méthode pour l'étude du discours émotionné*, Berne, Peter Lang.
- ROSSET Sophie ; GROUIN Cyril ; ZWEIGENBAUM Pierre (2011), Entités Nommées Structurées : guide d'annotation Quaero, *LIMSI 2011-04*, Orsay, p. 1-86.

VANDERKEN Daniel (1988), *Les actes de discours : essai de philosophie du langage et de l'esprit sur la signification des énonciations*, Bruxelles, Pierre Mardaga éditeur.

VERNIER Matthieu (2011), *Analyse à granularité fine de la subjectivité*, Thèse de doctorat, Nantes, Université de Nantes.