

Prévention des risques liés à l'environnement de travail : constitution d'un corpus oral en vue de son traitement automatique

Sandra Cestic^{1,2} Iris Eshkol-Taravella¹

(1) LLL UMR 7270 CNRS, Université d'Orléans, 45100 Orléans, France

(2) ACATUS Informatique, 55 boulevard de Châteaudun, 45000, Orléans, France
sandra.cestic@acatus.fr, iris.eshkol@univ-orleans.fr

RESUME

Cette communication présente la constitution d'un corpus oral destiné à l'étude de l'expression verbale de la perception de facteurs physiques dans les environnements de travail. Cette étude a pour objectif d'apporter les connaissances nécessaires à la finalisation du développement d'une application informatique dédiée à la prévention des nuisances physiques au travail générées par le bruit et les ambiances thermiques. Nous abordons la méthodologie mise en œuvre pour collecter des données orales authentiques et constituer un corpus susceptible d'anticiper au mieux les procédures de traitement automatique pour l'extraction d'informations relatives aux risques. Le retour sur notre pratique embrasse toutes les étapes de la constitution de corpus pour répondre à l'objectif d'étude.

ABSTRACT

Prevention of risks related to the work environment: Constitution of an oral corpus toward natural language processing

This communication presents the constitution of an oral corpus for the study of the verbal expression of the perception of physical factors in the work environment. This study aims to provide the knowledge required to complete the development of a computer application dedicated to the prevention of physical harm at work generated by the noise and thermal environments. We discuss the methodology used for collecting authentic oral data and constitute a corpus likely to better anticipate the procedures of natural language processing for the extraction of information related to risk. Reflecting on our practice meets all steps of corpus constitution to reach the study objective.

MOTS-CLES : corpus oral, prévention des risques, environnement de travail, annotation de la perception, traitement automatique du langage naturel.

KEYWORDS: oral corpus, risk prevention, work environment, annotation of perception, natural language processing.

1 Introduction

Les risques professionnels sont une menace permanente pour les salariés. Leur santé peut s'altérer en raison de maladie professionnelle ou d'accidents du travail. Réduire ces risques afin d'assurer la sécurité des salariés et pour protéger leur santé physique et mentale est à la charge de l'employeur. Il doit prendre les mesures appropriées et les mettre en œuvre conformément aux principes généraux

de prévention énumérés par le Code du travail. En fonction de l'activité exercée, il doit évaluer les risques professionnels, consigner les résultats et mettre en œuvre des actions de prévention.

Cette communication présente une étude menée dans le cadre d'une thèse CIFRE associant le Laboratoire Ligérien de Linguistique et ACATUS Informatique et qui vise la prévention des nuisances physiques au travail générées par le bruit et les ambiances thermiques. L'application développée doit permettre l'évaluation de l'exposition des salariés aux différents facteurs considérés et celle de l'impact de ces facteurs physiques sur la santé des salariés. L'objectif de l'étude est de repérer et d'analyser la variation de l'expression de la perception du bruit et de la température dans le discours oral des salariés et comparer leur ajustement aux dispositions légales concernant l'environnement de travail. A partir des entretiens, on se propose d'approfondir les connaissances sur l'exploitation des ressources linguistiques actualisées dans l'expression de la perception auditive et de la perception thermique. En bref, il s'agit :

- d'apporter une expertise linguistique au partenaire industriel pour le développement de son application
- d'accéder à la perception des personnels sur leur environnement de travail
- d'extraire et d'analyser la perception des facteurs physiques dans les environnements de travail à partir du corpus constitué.

Le retour sur notre pratique embrasse toutes les étapes de la constitution de corpus pour répondre à l'objectif d'étude. Nous présenterons tout d'abord les cadres théoriques de notre recherche (section 2). Il s'agit de travaux appartenant aux différents champs de la linguistique (linguistique cognitive, sociolinguistique, linguistique de corpus, linguistique informatique) car notre recherche est avant tout pluridisciplinaire. Nous exposerons ensuite une méthode (section 3) que nous avons choisie pour constituer et traiter le corpus.

2 Cadres théoriques

2.1 Etudes sur la perception

Les études sur la perception de facteurs physiques comme le bruit et la température pour le domaine industriel emploient des questionnaires ou des entretiens pour collecter des données auprès des utilisateurs (références : Marry 2011, Guastavino 2003, Mzali 2002, Chevret 2013 et 2015, Moujalleb 2007, Mandrara 2011). Les données perceptives sont comparées aux mesures physiques effectuées en parallèle et confrontées aux normes imposées. Cependant l'expression de la perception n'est pas analysée sur le plan linguistique. C'est avec les travaux de (Delepaut 2007, Cance 2008) que la perception est abordée par le biais des sciences du langage.

(Delepaut 2007) analyse des discours de passagers sur le confort en train en mettant à profit la linguistique cognitive. Elle met en place une méthodologie pluridisciplinaire pour l'identification du concept de confort en train, alliant les plans cognitif (pour l'identification des représentations) et linguistique (pour le repérage des formes lexicales). Des questionnaires écrits ouverts et fermés sont utilisés. (Cance 2008) articule les connaissances et les méthodes propres aux champs de la linguistique et de la psychologie pour offrir des connaissances approfondies sur les ressources linguistiques et les procédés discursifs à l'œuvre dans l'expression du ressenti visuel.

A la différence de (Delepaut 2007), nous avons choisi l'entretien individuel à partir d'une trame de questions sans nous astreindre à la respecter à la lettre. L'entretien semi-directif paraît la forme d'entretien la plus adéquate pour recueillir ce type de données. Bien que les participants répondent assez spontanément et abordent librement les thèmes qui ne figurent pas explicitement dans la trame d'entretien, l'interaction « force les traits » et incite à une démarche réflexive sur la technique employée (Gadet 2003).

2.2 La linguistique et le TAL au service des risques et de la prévention

L'atelier "Linguistique et Traitement Automatique des Langues pour l'Aéronautique et l'Espace : Dimensions langagières du risque" du 2 juillet 2013 a permis de soulever les difficultés de traitement de la langue orale et écrite dans les domaines à risques humains et financiers, notamment parce que les communications en langue naturelle ne sont pas un moyen de transmission d'information parfait et du fait de la production importante de documents non structurés. Pour surmonter ces obstacles, linguistes, talistes et ingénieurs mobilisent des réponses en matière de gestion de la dimension langagière du risque : normes, guides de rédaction ou d'échanges conversationnels, outil d'aide à la rédaction ou à la vérification de données textuelles, outil d'extraction d'information, procédures de catégorisation des données pour normaliser et structurer l'information.

Nombreuses sont les applications de la linguistique et du TAL pour répondre à ces exigences de qualité d'information, d'efficacité et de rapidité de traitement et d'analyse. Dans (Tanguy et al., 2015), le TAL contribue à la classification automatique de rapports relatifs à la sûreté dans l'aviation. Quant aux travaux de (Neubig et al., 2011), les données utiles diffusées en ligne sur le réseau social Twitter lors du séisme survenu en 2011 au Japon sont exploitées via la création d'un système efficace pour la segmentation des mots, la reconnaissance des entités nommées et la classification des publications. Dans (Cohen et al., 2005) et (Dragos, Jaulent, 2009) le TAL participe à la mise en place de meilleures pratiques pour contribuer à l'exploitation de corpus du domaine biomédical (recherche d'information notamment). La thèse de (Tulechki, 2015) décrit les applications du TAL à la gestion des risques industriels. Il observe que, dans le domaine de l'aviation civile, le retour d'expérience (REX) génère de grandes quantités de données sous la forme de rapports d'accidents ou d'incidents. Le stockage et l'organisation sont mal adaptés à l'utilisation réelle de ces documents. Il démontre que le TAL peut répondre aux besoins d'experts en sécurité pour gérer la surabondance de données textuelles sur le retour d'expérience (REX). Les applications sur les données textuelles sont ainsi nombreuses, celles sur les données orales relatives à la prévention et aux risques sont plus rares. C'est à cet objectif que répond notre recherche qui vise le repérage et le traitement par les outils du TAL pour la prévention du risque à partir d'un corpus oral. Pour cela, le corpus de travail a été constitué. La méthodologie de sa constitution est décrite dans les sections qui suivent.

2.3 Constitution du corpus oral

Pour constituer le corpus, nous suivons les recommandations de Corpus oraux, guide des bonnes pratiques (Baude, 2006) et l'orientation préconisée par (Mondada, 2005). Partant des travaux de Sacks (ca 1970) qui ont codifié l'analyse conversationnelle, (Mondada, 1998) a problématisé la réflexivité nécessaire pour collecter des données sur le terrain. Notre enquête cherche à faciliter la parole du participant qui n'est d'ordinaire sollicité que brièvement et ponctuellement dans le cadre

professionnel pour évoquer son environnement de travail. La trame d'entretien est construite en fonction de ces principes, des contraintes du terrain et des objectifs de la collecte de données.

Suite à l'enquête de terrain effectuée auprès des salariés, les entretiens ont été transcrits. Nous utilisons les conventions de transcription d'ESLO (Baude, Dugua, 2011, Eshkol-Taravella et al, 2012) afin d'homogénéiser la présentation, assurer l'interopérabilité et faciliter le traitement automatique ultérieur tout en restant conforme à la nature orale des données (Blanche-Benveniste, Jeanjean, 1987).

2.4 Modélisation de la perception

Nous exploitons les méthodes de modélisation par jeu d'étiquettes pour appréhender la notion de perception (Cestic, Eshkol-Taravella, 2016), en nous inspirant pour partie des travaux en TAL portant sur l'analyse des opinions et sentiments (pour la polarité, voir les travaux de Pang et al., 2002).

Les premières observations effectuées sur notre corpus d'entretiens confirment la nécessité d'un processus d'identification de l'information en suivant les propositions avancées par (Boullier, Lohard, 2012) à propos de la détection de la tonalité dans les médias sociaux. Après avoir repéré l'expression du ressenti, nous identifions quel(s) sera(en)t le(s) facteur(s) à l'origine du ressenti exprimé. Nous évaluons ensuite le degré de polarité et d'intensité de cette expression. C'est ainsi que (Boullier, Lohard, 2012) abordent les principales classifications à opérer face à des enjeux de TAL. Le degré d'intensité est également évalué en regard du cadre réglementaire de la prévention des risques liés à l'exposition au bruit.

La question du ressenti a ceci de particulier qu'elle peut être corrélée à un facteur objectif, mais que son expression est subjective. De nombreux travaux en TAL mentionnés par (Boullier, Lohard, 2012) sont centrés sur la « cible », c'est-à-dire l'élément sur lequel porte l'opinion exprimée. Nous statuerons sur l'éventualité d'un jeu d'étiquettes spécifique pour l'expression du ressenti, avec pour alternative l'emploi d'étiquettes utilisées en détection d'opinion ou de sentiments qui pourraient être adaptées au ressenti.

Les cadres théoriques permettent de mettre en lumière comment l'étude présentée ici appelle le champ de la linguistique et du TAL pour traiter l'expression de la perception à des fins d'extraction d'information. Les études sur la perception nous montrent comment elles mettent en œuvre des moyens des sciences du langage pour le recueil de données langagières. Les domaines du risque et de la prévention font appel à l'ingénierie linguistique pour traiter une grande masse de données sensibles. La constitution d'un corpus oral s'avère nécessaire pour notre étude afin de pouvoir créer une modélisation de la perception fondée sur des données réelles. Enfin, l'extraction d'informations relatives aux risques appelle la modélisation de la perception au moyen des techniques utilisées en analyse d'opinion et de sentiments.

3 Méthodologie pour la constitution du corpus

Cette étude mobilise non seulement l'acquisition de données orales authentiques au cours d'entretiens enregistrés auprès de salariés et transcrits, mais aussi celle d'une méthodologie qui favorise le repérage, par un outil informatique, de la perception de facteurs physiques (bruit et

température) dans un environnement de travail à travers ces entretiens. Pour définir et mettre à l'épreuve cette méthodologie, un corpus oral est constitué. Il s'agit de provoquer l'expression des salariés à propos de la perception qu'ils ont du bruit et de la température sur leur lieu de travail. La spécificité du corpus tient à son oralité, avec toutes les particularités de ce mode d'expression, notamment la forme dialogique en sorte que la présence de l'enquêtrice doive être prise en compte dans l'analyse.

Une trame d'entretien est construite pour recueillir les réactions verbales des locuteurs. La technique d'entretien employée reste soucieuse et garante de la confidentialité et de l'anonymat des locuteurs. Un formulaire d'autorisation d'enregistrement est utilisé, ainsi que les principes d'anonymisation.

3.1 Sélection des témoins

L'application développée doit pouvoir cartographier un environnement de travail situé à l'intérieur d'un bâtiment et tenir compte à la fois des données physiques collectées au sein de ce bâtiment et des données langagières recueillies lors d'entretiens pour permettre leur traitement. L'objectif des entretiens étant de recueillir l'expression verbale des salariés concernant leur environnement de travail, nous avons choisi de solliciter des salariés d'entreprises dont les activités les obligent à être fréquemment présents à l'intérieur des locaux de l'entreprise. La restitution de l'expérience quotidienne de leur environnement de travail lors des entretiens sera pertinente en vue de l'étude et porteuse d'une réalité telle qu'elle est perçue par les salariés-témoins.

3.2 Technique d'entretien et création de la trame d'entretien

Pour mieux comprendre comment s'expriment les salariés à propos de leur environnement de travail, nous privilégions une technique qui recueille les réactions spontanées des témoins. L'entretien semi-directif est la technique sélectionnée. Une trame d'entretien est utilisée pour orienter l'échange autour de thèmes définis au préalable. Nous reproduisons ici quelques exemples de questions relatives au poste de travail et à l'environnement de travail posées lors des entretiens :

– *Poste de travail*

- A quel poste de travail êtes-vous installé ? Depuis combien de temps ?
- Qu'est-ce qui vous plaît ? Qu'est-ce qui vous déplaît ?
- Qu'est-ce qui est important pour que vous vous y sentiez à l'aise ?
- Est-ce qu'il y a quelque chose qui vous incommode/dérange lorsque vous êtes à votre poste de travail ?

– *Environnement de travail*

- Dans quel type d'espace travaillez-vous ?
- Quels sont les désagréments que vous percevez ?
- Y a-t-il des moments où c'est différent ?

L'ordre des questions et leur format sont respectés pour collecter un maximum d'informations.

3.3 Conduite de l'entretien

Pour des raisons pratiques, on utilise un bureau dans les locaux de l'entreprise accueillant l'enquête. Cela facilite le déplacement des salariés-témoins dans un lieu familier. Les caractéristiques de l'entretien sont les suivantes :

- L'entretien est individuel
- Réalisé dans un lieu calme
- Où l'on peut parler librement
- Où le témoin se sent à l'aise

Chaque entretien se déroule en suivant quatre étapes. La première étape vise la présentation de l'enquête avec les acteurs du projet, le besoin en informations pour l'amélioration des environnements de travail. Elle valorise le rôle du témoin et les bénéfices apportés à l'entreprise et aux salariés. La seconde étape aborde le protocole de l'entretien : les échanges entre l'enquêtrice et le témoin, l'autorisation d'enregistrement nécessaire, l'anonymat et la confidentialité. La troisième étape annonce les suites de l'entretien avec la communication aux témoins de l'enregistrement audio et la transcription de l'entretien. La dernière étape est déclenchée par la signature du formulaire d'autorisation d'enregistrement et s'ouvre sur l'entretien.

3.4 Transcription des entretiens et anonymisation

Suite à la passation des entretiens, ceux-ci sont transcrits à l'aide d'un outil de transcription manuelle Transcriber (Barras et al., 1998). Nous avons opté pour une transcription orthographique et sélective des parties de l'entretien où le locuteur mentionne l'objet d'étude ("Poste de travail", "Environnement de travail"). La segmentation des enregistrements dans Transcriber suit la trame de l'entretien. Les disfluences (« euh » d'hésitations, répétitions, autocorrections, reformulations, amorces, inachèvements...) sont conservées, mais elles peuvent être enlevées avec Distagger (Constant, Dister, 2010). La question de l'anonymisation des entretiens, qui se pose dans les corpus oraux, est en partie réalisée par l'indexation des entretiens au moyen d'un code (exemple : IND1004). Les faisceaux d'indices (Eshkol-Taravella et al., 2015) sont identifiés pour être manuellement traités. Nous montrerons comment nous les avons anonymisés.

3.5 Annotation manuelle

L'annotation des informations concernant l'expression de la perception permet leur extraction, leur analyse et leur comparaison avec les mesures objectives. Pour permettre l'extraction future des informations, nous procédons à un repérage manuel de l'expression de la perception des facteurs physiques. Grâce à ce repérage, nous déterminons quelques critères saillants de l'expression de la perception : cible de l'opinion (bruit ou température), sous-catégorie de la cible (bruit ambiant, bruit humain, bruit général, température chaude, température froide, température générale), polarité et intensité de l'expression. Pour chacun de ces critères, nous attribuons une étiquette. Le jeu d'étiquettes est construit en fonction des informations livrées par les témoins dans les entretiens et en regard de la législation en vigueur concernant la prévention des risques professionnels liés aux bruits et aux ambiances thermiques. Nous annotons l'expression du ressenti en identifiant :

- le facteur à l'origine du ressenti exprimé et le type de facteur considéré (l'attribut « target » prend pour valeur « noise » ou « temperature », puis dans le cas du bruit, l'attribut « kind » prend pour valeur « human », « ambient » ou « general », dans le cas de la température, l'attribut « kind » prend pour valeur « cold », « hot » ou « general »),
- la polarité employée par le locuteur (l'attribut « polarity » prend pour valeur « negative », « neutral » ou « positive »),
- l'intensité du ressenti exprimé (l'attribut « intensity » prend pour valeur « 0 » dans le cas d'une polarité neutre, puis « 1 », « 2 » ou « 3 » dans le cas d'une polarité négative ou positive).

Ces dernières valeurs sont attribuées en fonction, d'une part, de l'intensité de l'expression du ressenti et, d'autre part, de la réglementation française en vigueur concernant le bruit et la température au travail. Nous rappelons qu'il s'agit de permettre le repérage et l'extraction d'informations utiles à la prévention des risques dans les environnements de travail. Ainsi, à propos du bruit au travail, nous observons que la réglementation prévoit une valeur limite d'exposition (qui ne doit être dépassée en aucun cas), une valeur d'exposition supérieure déclenchant l'action (corrective) et une valeur d'exposition inférieure déclenchant l'action (préventive). La variation entre ces valeurs est concrètement observable grâce aux données physiques. Pour chacune de ces valeurs issues de la réglementation, nous mettons en vis-à-vis l'expression du ressenti du bruit et nous utilisons une valeur pour notre annotation :

- une expression négative faible à modérée de la perception du bruit obtiendra une valeur « 1 » : il s'agit d'expériences de perception du bruit ayant entraîné une faible gêne
- une expression négative forte de la perception du bruit obtiendra une valeur « 2 » : il s'agit d'expérience de perception de bruit ayant entraîné une forte gêne
- une expression négative extrême de la perception du bruit obtiendra une valeur « 3 » : il s'agit d'expérience de perception de bruit ayant entraîné un accident, une maladie
- une expression positive faible à modérée de la perception du bruit obtiendra une valeur « 1 » : il s'agit d'expériences de perception du bruit ayant entraîné un faible confort
- une expression positive forte de la perception du bruit obtiendra une valeur « 2 » : il s'agit d'expérience de perception de bruit ayant entraîné un confort élevé
- une expression positive extrême de la perception du bruit obtiendra une valeur « 3 » : il s'agit d'expérience de perception de bruit ayant entraîné confort extrême

Nous observons que, si la réglementation française concernant le bruit au travail est précise sur les seuils à respecter et nous permet donc de statuer sur l'annotation à exercer dans notre corpus, il en va autrement de la réglementation française concernant les températures de travail. En effet, le Code du travail ne donne pas d'indication de température minimale ou de température maximale. Nous devons donc, pour permettre la comparabilité des résultats, nous reposer seulement sur l'expression du ressenti de la température en tenant compte des mêmes critères linguistiques que ceux identifiés dans l'expression du ressenti du bruit. Pour autant, nous ne négligeons pas le fait que le Code du travail concernant les températures minimales et maximales ne nous permet pas de statuer fermement sur notre annotation, ni la nécessité d'observer attentivement l'expression de la perception du bruit et celle de la température pour mesurer combien il est possible d'utiliser les mêmes critères d'annotation. Nous émettons l'hypothèse qu'il peut exister des similitudes et des divergences entre l'expression de la perception du bruit et de la température qu'il faudra mettre au jour.

IND1004: <perception target="noise" kind="human" polarity="negative" intensity="1">les bru- les les les bruits humains c'est plus plus des bouts de conversation forcément on va on va avoir la concentration qui va plus vouloir qui va plus avoir tendance à aller à aller se greffer dessus

</perception><perception target="noise" kind="ambient" polarity="negative" intensity="1">après les bruits les bruits on va dire d'ambiance donc que ce soit ventilateur salle serveur travaux dans le dans le dans l'autre partie du bâtiment des trucs comme ça bon ça dépend l'intensité du bruit mais on peut on peut arriver </perception><perception target="noise" kind="human" polarity="negative" intensity="2">mais les les conversations ouais c'est souvent plus plus compliqué que le bruit ambiant à à supporter</perception>
[Cadre en informatique, homme, 24 ans – Durée de l'enregistrement : 0:26:42. Extrait : 0:13:36 > 0:14:04]

Dans cet extrait, le locuteur aborde la gêne provoquée par les bruits de conversation, puis par les bruits d'ambiance, puis revient aux bruits humains pour les comparer aux bruits d'ambiance. Si la polarité négative couvre l'ensemble de cet extrait, celle-ci est nuancée dans l'intensité. Le format d'annotation choisi est XML car c'est un format normalisé et interopérable. La DTD qui décrit la structure du corpus est créée.

4 Conclusion et perspectives

Nous avons présenté les cadres théoriques qui régissent ce travail ainsi que la construction d'un corpus oral dédié au recueil et à la recherche d'informations. Pour automatiser les futurs traitements, nous mobilisons et mettons à l'épreuve une méthodologie pour la constitution d'un corpus oral qui répond aux besoins de l'étude. Le retour sur notre pratique embrasse toutes les étapes de la constitution de corpus pour répondre à l'objectif d'étude : définition de la trame d'enquête, conduite des entretiens, convention de transcription, segmentation, anonymisation, métadonnées utilisées, annotation manuelle. L'annotation manuelle répond au besoin d'extraction d'informations relatives aux risques.

L'information annotée sera extraite, stockée et visualisée automatiquement pour pouvoir permettre l'analyse approfondie de l'expression de la perception du bruit et de la température dans le corpus constitué. Les résultats, qui aboutissent à une modélisation des données subjectives, seront rapportés aux données objectives obtenues par des mesures physiques. Nous souhaitons rendre possible la visualisation cartographique et graphique des données subjectives et des données objectives pour pouvoir les comparer et identifier les zones d'études qui nécessitent l'intervention. Ces zones d'études devront être des zones où les données physiques et/ou les données linguistiques rendent compte d'une situation à risques. Parmi les perspectives de ce travail, nous envisageons de tester l'annotation automatique en nous servant du corpus annoté manuellement comme d'un corpus d'apprentissage.

Remerciements

Ce travail est réalisé dans le cadre de la thèse CIFRE n° 2013/0209 ACATUS Informatique / Université d'Orléans.

Références

BARRAS C. ET AL. (1998) Transcriber, a free tool for segmenting, labeling and transcribing speech, Actes *Language Resources and Evaluation Conference, LREC'1998*, Grenade, Espagne, 1373-1376.

- BAUDE O., DUGUA C. (2011). (Re)faire le corpus d'Orléans quarante ans après : quoi de neuf, linguiste ? *Corpus 10, Varia*, pp. 99-118.
- BAUDE O., coord. (2006) *Corpus oraux, guide des bonnes pratiques*, Paris et Orléans, Editions du CNRS et Presses Universitaires d'Orléans.
- BOULLIER D., LOHARD A. (2012) *Chapitre 5. Détecter les tonalités : opinion mining et sentiment analysis* In : Opinion mining et Sentiment Analysis : Méthodes et outils [en ligne]. Marseille : OpenEdition Press, 2012 (généré le 30 octobre 2015). Disponible sur Internet : <http://books.openedition.org/oep/214>. ISBN : 978282182277. DOI : 10.4000/books.oep.198.
- BLANCHE-BENVENISTE C., JEANJEAN C. (1987). *Le français parlé. Transcription et édition*. Institut national de la Langue française. Didier Érudition, Paris.
- CESTIC S. (2015). *Quelques réflexions sur la constitution d'un corpus oral*. Communication présentée aux 1ères JOurnées jeunes chercheurs de la LInguistique de Corpus, 2-4 septembre 2015, Orléans.
- CESTIC S., ESHKOL-TARAVELLA I. (2016). *Perception de l'environnement de travail : analyse et modélisation de l'expression*. Communication présentée à la seconde édition du Colloque international de Sciences du Langage "La Perception en langue et en discours", 21-23 avril 2016, Université d'Opole.
- CHEVRET P. (2015). Utilisation d'un questionnaire pour l'évaluation du ressenti des salariés vis-à-vis de leur environnement sonore, *Congres des préventeurs ORANGE*, Paris, France.
- CANCE C. (2008). « Expériences de la couleur, ressources linguistiques et processus discursifs dans la construction d'un espace visuel : l'habitable automobile », Thèse de doctorat en sciences du langage, sous la direction de Danièle Dubois, Université de la Sorbonne nouvelle - Paris III, 494 p.
- COHEN B., FOX L., OGREN P., HUNTER L. (2005). Corpus Design for biomedical natural language processing. ACL-ISMB Workshop on Linking Biological Literature, Ontologies and Databases : Mining Biological Semantics, Detroit.
- CONSTANT M., DISTER A. (2010). Automatic detection of disfluencies in speech transcriptions. *Spoken Communication*. In Pettorino M., Giannini A., Chiari I., Dovetto F. M. (eds) (2010). Cambridge Scholars Publishing. pp. 259-272.
- DELEPAUT G. (2007). « Contribution de la linguistique cognitive à l'identification du confort : analyse des discours des passagers sur le confort en train, Thèse de doctorat en science du langage, sous la direction de Danièle Dubois, Université de la Sorbonne nouvelle - Paris III, 429 p.
- DRAGOS V., JAULENT M.-C. (2009). Construction d'un corpus biomédical pour la recherche d'informations. IN 8ème conférence internationale Terminologie et Intelligence Artificielle, Toulouse (France), 18-20 Novembre 2009
- DUBOIS D. (2009). *Le sentir et le dire, Concepts et méthodes en psychologie et linguistique cognitives*. Paris : L'Harmattan, 391 p.

- DURAND J., LAKS B., LYCHE C. (2009). Le projet PFC : une source de données primaires structurées. In J. Durand, B. Laks et C. Lyche (eds) (2009) *Phonologie, variation et accents du français*. Paris : Hermès. pp. 19-61.
- ESHKOL-TARAVELLA I., BAUDE O., MAUREL D., KANAAN-CAILLOL L. (2015). Recherche des indices permettant une identification : l'anonymisation des transcriptions du corpus ESLO. Actes de *TALN2015*, Caen, France.
- ESHKOL-TARAVELLA I., BAUDE O., MAUREL D., HRIBA L., DUGUA C., TELLIER I. (2012). Un grand corpus oral « disponible » : le corpus d'Orléans 1968-2012. *TAL : Ressources linguistiques libres*, vol. 52, n° 3, pp. 17-46.
- GADET F. (2003). Derrière les problèmes méthodologiques du recueil des données. *Texte !* juin-septembre 2003 [en ligne]. Disponible sur : <http://www.revue-texto.net/Inedits/Gadet_Principes.html>. (Consultée le 31 mars 2016).
- GUASTAVINO C. (2003). « Etude sémantique et acoustique de la perception des basses fréquences dans l'environnement sonore urbain », Thèse de doctorat en acoustique, sous la direction de Danièle Dubois, Université de Paris VI, 245 p.
- KERBRAT-ORECCHIONI C. (2000). « L'analyse des interactions verbales : la notion de 'négociation conversationnelle' : défense et illustration », *Lalies*, 20, pp. 63-141.
- QUIRK R., GREENBAUM S., LEECH G., SVARTVIK J. (1985). *A Comprehensive Grammar of the English Language*. New York: Longman.
- MANDRARA Z. (2011). « Impact thermique des revêtements en bois sur l'ambiance intérieure des bâtiments : application dans l'habitat traditionnel en France et à Madagascar », Thèse de doctorat en thermique, sous la direction de Gérard Guarracino et de Richard Cantin, Institut national des sciences appliquées de Lyon, 276 p.
- MARRY S. (2011). « L'espace public sonore ordinaire : les paramètres de la perception sonore dans les espaces publics : contribution à une connaissance de l'ambiance sonore », Thèse de doctorat en acoustique, Université de Grenoble.
- MONDADA L., LÜDI G. (éds). (1998). Linguistes en dialogue, *Acta Romanica Basiliensa (ARBA)*, 8.
- MONDADA L. (2005). Constitution de corpus de parole-en-interaction et respect de la vie privée des enquêtés : une démarche réflexive. Rapport sur le projet « Pour une archive des langues parlées en interaction. Statuts juridiques, formats et standards, représentativité » financé par le Programme Société de l'Information / Archivage et patrimoine documentaire, mars 2005, 43 p.
- MAHMOUDIAN M., MONDADA L. (éds) (1998). Actes du Colloque « Le travail du chercheur sur le terrain : Questionner les pratiques, les méthodes, les techniques de l'enquête », 13-14.12.1996, *Cahiers de l'ILSL*, 10, Université de Lausanne.
- MOUJALLED B. (2007). « Modélisation dynamique du confort thermique dans les bâtiments naturellement ventilés », Thèse de doctorat en thermique, sous la direction de Gérard Guarracino et de Richard Cantin, Institut national des sciences appliquées de Lyon, 330 p.

- MZALI M. (2002). « Perception de l'ambiance sonore et évaluation du confort acoustique dans les trains », Thèse en acoustique, sous la direction de Jean-Dominique Polack, Université Paris VI.
- NEUBIG G., MATSUBAYASHI Y., HAGIWARA M., MURAKAMI K. (2011). Safety Information Mining : What can NLP do in a disaster ? In: Proceedings of the 5th International Joint Conference on NLP (IJCNLP), volume 11, pages 965-973, Chiang Mai, Thailand.
- PANG B., LEE L., VAITHYANATHAN S. (2002). Thumbs up ? Sentiment Classification using Machine Learning Techniques. In : Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP).
- PIERRET, M., PARIZET, E., CHEVRET, P. (2013). Perception et évaluation des sources de bruit dans un bureau ouvert, *21st International Congress on Acoustics*, Montréal, Canada.
- TANGUY L., TULECHKI N., URIELI A., HERMANN E., RAYNAL C. (2016). *Natural language processing for aviation safety reports: from classification to interactive analysis*. Computers in Industry, 78, pp. 80-95.
- TULECHKI, N. (2015). « Natural language processing of incident and accident reports : application to risk management in civil aviation », Thèse de doctorat en linguistique, sous la direction de Ludovic Tanguy, Université Toulouse le Mirail - Toulouse II, 203 p.
- WIEBE J., WILSON T., CARDIE C. (2005). Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*, 39(2): p. 165-210.