

Prévention des risques liés à l'environnement de travail : constitution d'un corpus oral à des fins de traitement automatique du langage naturel

Sandra Cestic

LLL UMR 7270 CNRS, Université d'Orléans, 45100 Orléans, France
ACATUS Informatique, 55 boulevard de Châteaudun, 45000, Orléans, France
sandra.cestic@acatus.fr

Cette communication présente une étude menée dans le cadre d'une thèse CIFRE 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 du travail est de repérer et d'étudier la variation de l'expression de la perception du bruit et de la température dans le discours oral des salariés. 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. Nous présenterons les méthodes que nous avons choisies pour constituer le corpus. Nous exposerons l'élaboration de la trame d'entretien, la procédure de sélection des témoins et la passation de l'entretien, l'enregistrement et les obligations légales relatives aux enregistrements, la convention de transcription etc.

Le projet 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 favorisera 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 suivant les recommandations de *Corpus oraux, guide des bonnes pratiques* (BAUDE, 2006) et l'orientation préconisée par (MONDADA, 2005). 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 pris 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. 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. L'ordre des questions et leur format sont respectés pour collecter un maximum d'informations. 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, F. 2003).

Suite à l'enquête de terrain effectuée auprès des salariés, les entretiens ont été transcrits. Nous utilisons les conventions de transcription d'ESLO (BAUDE, O. 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 C., JEANJEAN C. 1987).

Les premières observations effectuées sur notre corpus confirment la nécessité d'un processus d'identification de l'information. Après avoir repéré manuellement l'expression du ressenti, pour permettre l'extraction future des informations, nous procédons à une annotation manuelle grâce à un jeu d'étiquettes dont nous expliciterons la définition et l'utilisation. L'annotation des informations concernant l'expression de la perception permet leur extraction et leur analyse. 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 expliciterons l'intérêt de la création de sous-catégories de facteur et la prise en compte de la réglementation en vigueur à propos de la prévention des risques professionnels.

Pour automatiser les futurs traitements, nous mobilisons des méthodes et des outils actuels. La méthodologie pour parvenir à notre objectif sera explicitée. 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.

[Ce travail est réalisé dans le cadre d'une thèse CIFRE n° 2013/0209 ACATUS / Université d'Orléans]