



HAL
open science

TAL et analyse de l'activité en ergonomie : extraction d'informations spécialisées dans des transcriptions d'entretiens

Andréa Blivet

► To cite this version:

Andréa Blivet. TAL et analyse de l'activité en ergonomie : extraction d'informations spécialisées dans des transcriptions d'entretiens. 35èmes Journées d'Études sur la Parole (JEP 2024) 31ème Conférence sur le Traitement Automatique des Langues Naturelles (TALN 2024) 26ème Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RECITAL 2024), Jul 2024, Toulouse, France. pp.85-97. hal-04622989

HAL Id: hal-04622989

<https://inria.hal.science/hal-04622989v1>

Submitted on 26 Jun 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

TAL et analyse de l'activité en ergonomie : extraction d'informations spécialisées dans des transcriptions d'entretiens

Andréa Blivet¹

(1) SNCF DTIPG, 1-3 avenue François Mitterrand, 93210 Saint-Denis, France

andrea.blivet@gmail.com

RÉSUMÉ

L'ergonomie du travail est une discipline qui étudie les conditions de travail des individus. Son application se traduit, entre autres, par la réalisation d'entretiens d'analyse de l'activité qui ont pour objectif de faire émerger les impacts négatifs de la situation de travail sur la santé physique et morale des travailleurs. Notre étude consiste en la mise en place d'un système d'extraction automatique des informations relatives à ces impacts dans les transcriptions des entretiens réalisés. Ce système se fonde sur une approche hybride, associant ressources terminologiques et calcul de similarité contextuelle.

ABSTRACT

NLP and Activity Analysis in Ergonomics : Extraction of Specialized Information from Interview Transcriptions

The field of occupational ergonomics examines the working conditions of individuals, often through interviews analyzing their activities to identify negative impacts of the work environment on both physical and mental well-being. Our study focuses on implementing an automated system for extracting information pertaining to these impacts from interview transcriptions. The system adopts a hybrid approach, combining terminological resources with contextual similarity calculations.

MOTS-CLÉS : Extraction d'information, Marqueur, Ergonomie du travail, Corpus Spécialisé, Oral.

KEYWORDS: Information Extraction, Marker, Occupational Ergonomics, Specialized Corpus, Oral.

1 Extraction d'information en corpus spécialisé

En ergonomie du travail, les analyses de l'activité s'appuient en partie sur des interactions langagières, issues d'entretiens de terrain. Dans cette discipline, des approches linguistiques telles que l'analyse des pronoms personnels (Kebir *et al.*, 2021) ou encore l'analyse discursive des structures fonctionnelles des entretiens (Bonneau, 2008) ont déjà été utilisées pour proposer aux ergonomes de nouvelles manières de traiter les données d'entretien. Celles-ci nécessitent l'intervention de linguistes, et laissent aux ergonomes peu d'autonomie dans la réalisation de leurs analyses. Dans notre étude, nous proposons d'explorer les apports du TAL pour automatiser l'extraction des informations considérées comme pertinentes dans les données d'entretien tout en laissant la maîtrise de leurs analyses aux ergonomes. Les informations ciblées doivent être représentatives de la situation de travail de l'individu interrogé et s'inscrivent dans les thématiques suivantes (inspirées de la méthode ITMaMi¹) : l'individu,

1. https://bionet.scenari-community.org/Methodes_outils_PSE/co/ITMaMi.html

Les tâches qu'il effectue, le matériel qu'il utilise, l'organisation de son travail (soit l'organisation « humaine » et la planification des tâches), et son environnement physique. Définies de concert avec un ergonome expert, ces thématiques permettent à la fois de guider, structurer et filtrer l'extraction des informations qui y sont liées.

L'extraction d'information (EI) est une tâche classique du domaine du TAL, qui vise à dégager des segments textuels saillants pour un besoin donné en utilisant des méthodes de *text-mining* (règles lexico-syntaxiques, apprentissage automatique) ou des méthodes plus récentes en apprentissage profond (approches neuronales). Traditionnellement, les systèmes d'EI s'intéressent à la détection d'entités nommées (Nouvel *et al.*, 2015), à l'identification d'événements (Kodelja *et al.*, 2019), aux relations sémantiques entre les items (Wang, 2013). L'EI est utilisée dans divers domaines comme le domaine judiciaire pour l'analyse de dossiers d'enquêtes (Gianola, 2020) ou le domaine médical pour le traitement de compte-rendus médicaux (Lemaitre *et al.*, 2020), l'extraction de connaissances spécialisées (Trzmielewski & Gnoli, 2022) ou l'aide au diagnostic (Amato *et al.*, 2013).

Dans ces travaux, la délimitation des segments textuels considérés comme pertinents est définie en fonction du besoin informationnel. Les ressources (lexicales, ontologiques, etc.) et les outils (analyses syntaxiques, sémantiques, etc.) choisis pour leur extraction dépendent quant à eux de la nature des données dont sont extraits les segments. En effet, les données issues de l'oral sont particulièrement marquées par le caractère spontané de la situation d'énonciation, ne pouvant pas répondre à des normes aussi strictes qu'à l'écrit. L'oral spontané a ses propres spécificités syntaxiques, qui se traduisent par des structures moins rigides et moins explicitées que dans la forme écrite (Kurdi, 2003; Flamein, 2019). Le discours oral présente également des *extragrammaticalités*, comme des disfluences, des pauses ou encore des faux-départs, qui influent sur la prédictibilité du discours (Blanche-Benveniste *et al.*, 1990; Crystal, 2001; Kurdi, 2003) et en conséquence, sur le choix des outils et méthodes utilisés pour son analyse (Nasr & Béchet, 2009; Lacheret *et al.*, 2014). (Even, 2005) utilise par exemple des ontologies de termes spécifiques pour l'extraction d'information depuis des notes de l'oral quand (Tambellini, 2007) s'appuie sur des indices et des fenêtres contextuelles pour traiter des transcriptions d'entretiens d'embauche. Plus récemment, des approches neuronales ont été utilisées par (Parcollet *et al.*, 2019) pour l'identification de thèmes dans des conversations téléphoniques.

Dans notre étude, les informations ciblées par la tâche d'extraction sont dites « spécialisées » et se définissent comme des informations professionnelles (Chaudiron, 2000). Elles réfèrent à la fois au domaine ferroviaire et au domaine de l'ergonomie du travail, discipline guidant les analyses de l'activité. Les modèles d'extraction d'information classiques sur ce type de données, c'est-à-dire intégrant du vocabulaire technique ou des contraintes spécifiques à un domaine, présentent des lacunes avec une absence de données similaires à celles ciblées dans les données d'entraînement (Ramponi & Plank, 2020).

Sur la base des travaux mentionnés précédemment et des caractéristiques de nos données, nous proposons de décomposer la tâche d'extraction d'information en mettant au point un système capable de :

- **localiser** dans un texte les informations pertinentes à partir de ressources terminologiques,
- **délimiter** les informations afin de les rendre autonomes,
- **catégoriser** les informations selon les thématiques établies en amont,
- **suggérer** des informations qui pourraient être intéressantes pour l'utilisateur.

2 Corpus

2.1 Données

Les données que nous utilisons sont des entretiens d'analyse de l'activité réalisés en interne entre 2021 et 2023. Le corpus à disposition se compose de 52 entretiens oraux (soit 35h47 d'enregistrement) réunissant entre 2 et 5 locuteurs. Ces entretiens sont considérés comme relevant d'un genre et d'un domaine spécifiques pour plusieurs raisons. Tout d'abord, leur contexte de production correspond à des conversations dispensées à l'oral avec des intentions de communication spécifiques. Un locuteur (l'enquêteur) cherche à obtenir des productions langagières du type informationnel de la part d'un autre locuteur (l'agent) : les participants n'ont pas les mêmes intentions et l'échange peut être considéré comme asymétrique. Dans la typologie des échanges oraux proposée par (Walter, 1996), ces entretiens peuvent être considérés comme du « dialogue d'enquête ». L'autre spécificité concerne le contenu des entretiens : ces derniers contiennent des termes qui relèvent à la fois d'un domaine (le ferroviaire), et d'une discipline scientifique (l'ergonomie du travail). Les questions révèlent les enjeux de l'ergonomie : identifier des douleurs, des contraintes organisationnelles ou physiques, etc. Les réponses éclairent à la fois sur ces enjeux, et apportent également des éléments techniques sur le domaine ferroviaire, en décrivant des processus, ou des tâches par le biais d'outils utilisés, etc.

2.2 Pré-traitement des données

La première étape consiste en la transcription orthographique des enregistrements d'entretiens. Pour cela, les enregistrements sont d'abord segmentés en tours de parole à l'aide du système de reconnaissance de locuteur PyAnnote (Bredin *et al.*, 2020). Les tours de parole sont ensuite transcrits à l'aide du modèle Whisper² (Radford *et al.*, 2023), qui propose un premier niveau de normalisation en écartant de la transcription une partie des disfluences et en rétablissant certaines formes contractées (« aprem » devient « après-midi »). Enfin, un classifieur est entraîné pour reconnaître le rôle de chaque locuteur à savoir, enquêteur ou agent, à partir de descripteurs linguistiques (fréquence des pronoms, longueur des tours de parole, etc.). La Figure 1 représente l'ensemble des processus impliqués dans le pré-traitement et les données obtenues qui servent d'entrée au système d'extraction.

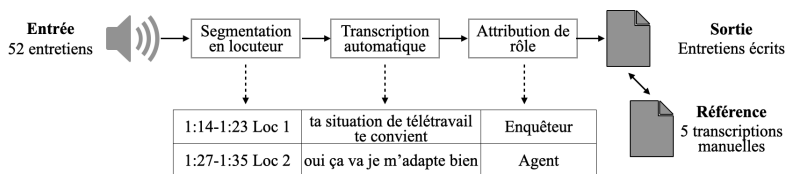


FIGURE 1 – Chaîne de traitement pour la transformation des données orales

En parallèle, 5 entretiens impliquant 2 locuteurs ont été transcrits manuellement pour servir de référence, représentant ainsi 41 minutes d'enregistrement et comportant 8280 tokens (dont 1166 différents) répartis en 418 tours de parole (soit 2% du corpus). Ainsi, la segmentation en locuteurs

2. Modèle Large-V2 <https://huggingface.co/openai/whisper-large-v2>

présente 85,41% d'exactitude, le *Word Error Rate* obtenu pour la transcription est de 11,3% et enfin, le classifieur pour l'attribution des rôles des locuteurs obtient un F-Score de 0,94.

3 Méthodologie : mise en œuvre d'une méthode hybride pour optimiser l'extraction d'information

La localisation, la délimitation et la catégorisation des informations sont des tâches réalisées à l'aide d'une approche supervisée, s'appuyant sur les connaissances transmises par les experts ergonomes internes. Ainsi, le système extrait dans un premier temps des segments textuels considérés *a priori* comme pertinents. Dans un second temps, un système de suggestions, s'appuyant sur une méthode non supervisée, a été mis en place, afin de pallier le manque de connaissances incluses dans les ressources externes.

3.1 Approche supervisée : utilisation de ressources lexicales

L'approche supervisée consiste à extraire les informations en s'appuyant sur les connaissances transmises par les experts, capitalisées dans des ressources externes, selon une méthode collaborative déjà réalisée dans des tâches d'extraction sur des données médicales (Lemaitre *et al.*, 2020). Les ressources utilisées ici sont des lexiques thématiques de marqueurs lexicaux présentés en Table 1. Ils se composent de termes lemmatisés et identifiés par des experts comme des pointeurs de la mention d'une thématique dans l'entretien. Par exemple, lorsque l'agent interrogé mentionne son épaule, le terme « épaule » est considéré comme marqueur d'une information portant sur la thématique de l'*Individu* et est ensuite répertorié.

Thématiques	Nombre de marqueurs	Exemples de marqueurs
<i>Individu</i>	42	pénibilité, nuque, position, posture
<i>Tâche</i>	30	méthode, manutentionner, cercler, procédure
<i>Matériel</i>	74	casque, chaise, outil, tournevis
<i>Organisation</i>	52	pause, mécanicien, aiguilleur, polyvalent
<i>Environnement</i>	37	sol, température, bruit, place
Total	235	

TABLE 1 – Répartition et composition des lexiques par thématique

La présence d'un marqueur permet de localiser une information dans le texte, leur localisation repose sur une recherche de correspondance sur une version tokenisée et lemmatisée³ de l'entretien. L'étape suivante consiste à délimiter le segment textuel qui correspond à l'information. D'après (Dupont *et al.*, 2002), une information est définie comme étant la combinaison d'entités ou d'événements avec leurs caractéristiques et relations. L'EI permet alors à formaliser les informations à partir d'analyses linguistiques comme l'analyse morphosyntaxique ou encore l'analyse sémantique compositionnelle. La stratégie que nous avons mise en place (cf.3 Partie A) repose sur cette hypothèse. Elle consiste dans un premier temps à récupérer les dépendants directs et indirects des marqueurs à partir d'une analyse

3. <https://spacy.io/models/fr> - modèle 'fr_core_news_lg'

syntaxique proposée par Stanza⁴ (Cao *et al.*, 2020). Concrètement, les marqueurs nominaux sont récupérés avec leurs dépendants directs et indirects (cf. Exemple a)), les adjectifs, quant à eux, sont mis en contexte en récupérant leur tête (cf. Exemple b)), enfin les arguments directs des marqueurs verbaux sont récupérés, soit les sujets et les compléments directs (cf. Exemple c)). Dans un second temps, les segments textuels extraits sont organisés selon les thématiques mentionnées en introduction. La thématique affectée à une information correspond à la thématique du marqueur ayant permis la localisation de cette même information.

- a) ta situation de télétravail te convient ?
- b) souvent on déplace des charges **lourdes**
- c) je ressentais des douleurs pendant la journée

Cette approche se nourrit uniquement des ressources constituées par les experts. Celles-ci ne peuvent cependant pas être exhaustives. Une approche complémentaire doit donc être intégrée.

3.2 Approche à base de modèles non-supervisés : proximité distributionnelle

Une approche non-supervisée, basée sur la proximité distributionnelle, a été utilisée pour compléter l'extraction et pallier aux limites de l'approche supervisée en proposant une alternative aux ressources lexicales.

La proximité distributionnelle est un principe qui considère que deux termes avec des contextes similaires peuvent être considérés comme sémantiquement proches, indépendamment de leur relation lexicale (Fabre, 2015; Firth, 1957). Pour accéder aux contextes des termes, nous utilisons des plongements lexicaux (pré-entraînés de SpaCy⁵) qui sont des représentations vectorielles des termes calculés, entre autres, à partir de leur environnement linguistique. La représentation par plongement permet de comparer la similarité des contextes entre deux termes pour définir leur niveau de proximité sémantique en calculant leur cosinus. Pour déterminer si deux termes sont considérés comme proches, il faut définir un seuil qui correspond au cosinus minimal attendu. Pour cela, un clustering hiérarchique a été effectué pour regrouper les marqueurs proches dans chaque lexique thématique. Les scores de silhouette des classes obtenus ont permis de confirmer ou non leur homogénéité. Ensuite, la moyenne du cosinus de similarité entre tous les marqueurs d'une même classe est considérée comme le seuil minimal de proximité nécessaire pour considérer un terme nouveau comme sémantiquement lié aux autres.

La Figure 2 illustre le processus de sélection, opérant sur l'ensemble des noms, verbes et adjectifs, à partir de l'exemple du terme *scoliose*, absent des ressources lexicales. Le cosinus de similarité est calculé entre ce terme et tous les autres marqueurs. Si le cosinus maximal est entre *scoliose* et *séquelle*, le terme est considéré comme candidat marqueur pour le système de suggestion d'informations puisque le cosinus est supérieur au seuil déterminé pour la thématique Individu dont est issu le terme *séquelle*. En revanche, si le cosinus maximal observé est entre *scoliose* et *motrice*, le terme *scoliose* sera exclu de la suggestion car le cosinus est inférieur au seuil pour la thématique à laquelle appartient le terme *motrice*.

4. <https://stanfordnlp.github.io/stanza/models.html> - modèle 'fr'

5. <https://spacy.io/models/fr> - modèle 'fr_core_news_lg'

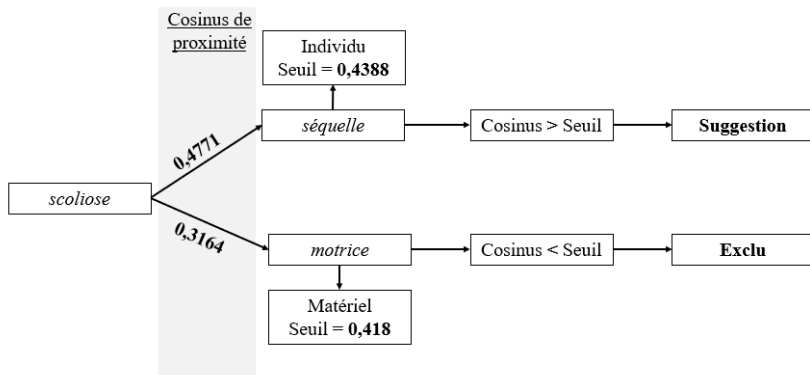


FIGURE 2 – Processus de sélection des marqueurs potentiels pour le système de suggestion

Les termes avec un score de proximité supérieur au seuil sont donc considérés comme des marqueurs potentiels. Le traitement appliqué est le même que pour les marqueurs avérés avec une délimitation des contextes à l'aide de l'analyse syntaxique (Figure 3 - Partie B). En revanche, contrairement aux informations obtenues à partir des ressources lexicales, un système de suggestion interactif est proposé à l'utilisateur pour qu'il puisse valider ou non la suggestion. Lui sont présentés, le marqueur, son contexte et la thématique associée (définie à partir de la thématique du marqueur avéré avec la proximité la plus élevée). L'utilisateur peut ensuite décider de valider l'information, de la rejeter ou bien d'attribuer une autre thématique. A la suite de cela, il est possible d'envisager un enrichissement de la liste des marqueurs en proposant un système d'amélioration continue, tel que présenté dans (Belkacem & Teissèdre, 2021). Ainsi, les marqueurs issus du système de suggestion qui auront été validés par les utilisateurs pourront être intégrés aux ressources.

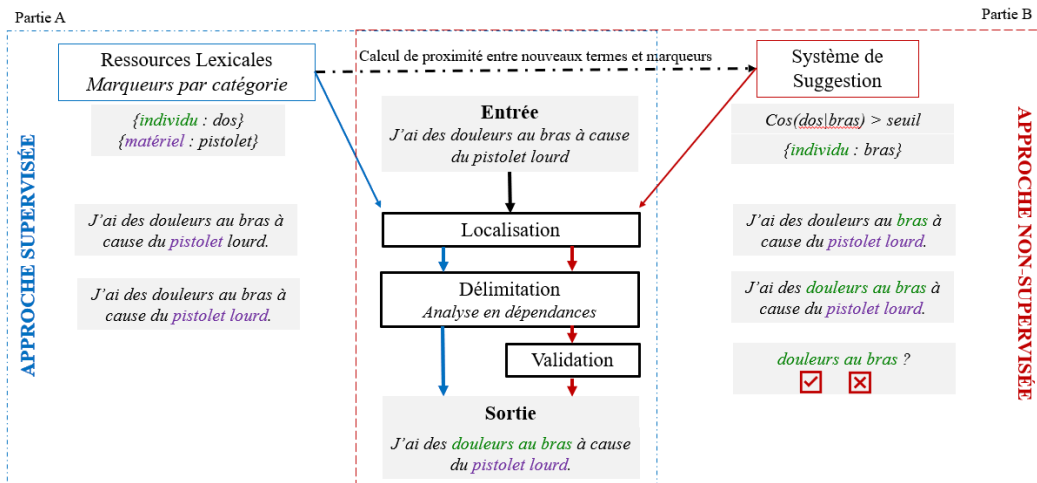


FIGURE 3 – Chaîne de traitement hybride pour l'extraction d'informations spécialisées - Partie A : Chaîne de traitement pour l'EI supervisée Partie B : Apport de l'approche non supervisée dans l'EI

4 Évaluation

L'objectif de l'évaluation est de déterminer les capacités du système à répondre aux attentes des utilisateurs et de mesurer l'exhaustivité de la double approche avec une confrontation des résultats de chacune.

Cinq experts SNCF ont été mobilisés pour participer à la phase d'évaluation. Les participants sont tous expérimentés et habitués des entretiens d'analyse de l'activité, ils ont également des connaissances approfondies sur l'organisation et les activités du groupe ferroviaire. L'évaluation est réalisée sur deux jeux de données. Le premier, illustré en Figure 4, se compose de cinq extraits issus de la transcription d'un même entretien. Ces cinq extraits ont ensuite été répartis entre les participants. Ils représentent au total 2901 tokens sur 124 tours de parole.

EN T'as dit que t'étais pas un grand sportif dans l'âme. Est-ce que tu aurais des restrictions d'aptitude ?
AG Non, pas de restriction.
EN Une limite de port de charge ?
AG Non, rien du tout.
EN T'as pas d'antécédent au niveau du dos ?
AG légère scoliose, voilà, pas une spécificité.

FIGURE 4 – Extrait de transcription à annoter

Le second jeu de données soumis aux évaluateurs se compose de 44 sorties du système de suggestions, correspondant à des informations potentielles détectées par le système, présentées avec leur contexte d'apparition et leur thématique assignée automatiquement (cf. Figure 5).

Suggestion	Contexte	Thématique
Douleur musculaire	je ressentais une petite douleur musculaire en bas du dos.	Individu

FIGURE 5 – Exemple de suggestion à évaluer

Dans un premier temps, les évaluateurs devaient relever dans un extrait de transcription d'entretien (premier jeu de données) les segments qui, selon eux, correspondent à des informations pertinentes pour représenter la situation de travail de l'individu interrogé. Ils devaient également assigner une ou plusieurs thématiques aux informations relevées. Dans un second temps, 20 suggestions issues de l'approche non-supervisée leur ont été présentées (second jeu de données). Pour chacune d'entre elles, les participants devaient : i) juger si elles renvoyaient à une information pertinente, le cas échéant ils devaient indiquer si la délimitation de cette dernière était correcte, en la corrigeant au besoin ; et ii) valider ou rectifier la thématique associée automatiquement à la suggestion.

L'évaluation prend ainsi en compte les sorties de l'approche supervisée et également l'ensemble de celles de l'approche non-supervisée, validant par défaut toutes les suggestions du système.

5 Résultats

Les résultats sont décomposés selon les tâches réalisées par le système, à savoir, localiser, délimiter, catégoriser et suggérer les informations, afin de les évaluer indépendamment. Cependant, la tâche de localisation a une influence sur la performance des autres tâches, puisque ces dernières sont réalisées à partir de la localisation. De plus, les approches mises en oeuvre ont un impact direct sur la localisation des informations. Ainsi, la Table 2 présente les résultats obtenus pour la localisation d'informations uniquement avec l'approche supervisée, puis combinée avec l'approche non-supervisée.

	Supervisée	Supervisée + non-supervisée
<i>Précision</i>	0,55	0,52
<i>Rappel</i>	0,32	0,67
<i>F-score</i>	0,40	0,58

TABLE 2 – Comparatif des résultats obtenus entre l'approche supervisée seule et combinée à l'approche non-supervisée

La différence notable dans ces résultats concerne le rappel, pour lequel l'alternative aux ressources lexicales de l'approche non-supervisée permet de le doubler, passant de **0,32** à **0,67**. A l'inverse, la précision est légèrement impactée par l'intégration de l'approche non-supervisée avec un score de **0,52** contre **0,55** avec uniquement l'approche supervisée. Dans la globalité, le *F-score* de **0,40**, obtenu uniquement avec l'approche supervisée, passe à **0,58** avec la combinaison des deux approches. Pour ce qui est de la délimitation, de la catégorisation et de la suggestion des informations, les résultats sont présentés en Table 3.

Tâches	Taux d'exactitude
Délimitation	5,82%
Catégorisation	55,28%
Suggestion	59,56%

TABLE 3 – Taux d'exactitude obtenus sur les différentes sous-tâches du système

Le taux d'exactitude pour la délimitation est de **5,82%**. Ce score très bas s'explique en partie par la méthode d'évaluation qui s'appuie sur une correspondance stricte, entre l'annotation experte et la sortie du système. Sur la totalité des erreurs de délimitation des informations, le placement de la frontière initiale de l'information est celle qui présente le plus de difficulté avec **56,47%** d'erreurs. A titre de comparaison, le placement de la frontière finale engendre une erreur dans **8,01%** des cas. Enfin, **20%** des erreurs de délimitation sont dues à un placement incorrect des deux frontières et prenant la forme d'informations enchâssées.

La catégorisation des informations, qui consiste à attribuer une thématique à une information retenue (indépendamment de la délimitation effectuée en amont), obtient un taux d'exactitude de **55,28%**. Le système assigne uniquement une seule thématique possible pour chaque information, tandis que les experts sont libres d'attribuer plusieurs thématiques. Les résultats ont permis de constater que seulement **3,01%** des désaccords s'expliquaient par cette différence de fonctionnement, c'est-à-dire les situations où l'expert attribue une thématique supplémentaire à celle sur laquelle il y a déjà un accord en l'expert et le système.

Enfin, la tâche de suggestion, uniquement réalisée par l'approche non-supervisée, présente un taux

d'exactitude de **59,56%**. Ce résultat signifie que presque 60% des suggestions proposées par le système sont considérées comme pertinentes par l'expert. Cette fonctionnalité participe donc largement à l'amélioration du rappel présenté plus haut.

Les annotations expertes permettent d'apporter des éléments de définition sur la notion d'information. En premier lieu, il a été observé que les informations se composent majoritairement de six tokens maximum. Seulement, **32%** des informations relevées comportent plus de six tokens. Ensuite, près de **90%** des informations expertes se terminent par un nom (*grosse semaine*), un verbe (*je me lève*) ou un adjectif (*charges lourdes*). En revanche, il est nécessaire de prendre en compte huit classes grammaticales pour représenter **90%** les frontières initiales d'informations (pronom, préposition, adverbe, déterminant, nom, verbe, conjonction et adjectif). Il apparaît donc que le début d'une information est soumise à une plus grande variabilité que sa fin. Enfin, les analyses montrent que quasiment l'intégralité des segments textuels considérés comme informations sont continus et que la plupart correspondent à des syntagmes nominaux ou a des triplets sujet-verbe-prédicat, comme « *une grande baie vitrée* » ou encore « *j'allonge ma pause repas* ». Cela confirme la stratégie de récupération des contextes basée sur l'analyse en dépendance.

6 Conclusion

Nous avons utilisé des méthodes d'extraction d'information supervisée (basée sur des ressources terminologiques expertes) et non-supervisée (avec de la similarité terminologique) dans l'objectif de proposer une structuration des informations abordées lors des entretiens d'analyse de l'activité. L'approche hybride que nous proposons permet d'obtenir des premiers résultats qui traduisent une certaine cohérence entre la méthode adoptée et les objectifs pré-définis. Cependant, ces résultats doivent être améliorés pour proposer par la suite un système utilisable et fiable.

6.1 Discussion

Les résultats de la délimitation des informations montrent que la méthode d'évaluation de cette tâche peut être repensée. En effet, une information est actuellement considérée comme validée par l'expert si la délimitation qu'il en fait est identique à celle du système au niveau des caractères. Cependant, cette méthode est trop contraignante et dévalue les capacités du système. En effet, l'information « *charges lourdes* » relevée par le système entraîne une erreur de délimitation quand l'expert a relevé « *des charges lourdes* ». Il apparaît pourtant dans cet exemple que l'intégration ou non du déterminant « *les* » influence peu, voire pas, la compréhension et l'interprétation sémantique de cette information. Pour améliorer la représentativité de l'évaluation sur le système, il est possible de proposer d'autres stratégies de comparaison entre deux segments, avec par exemple une prise en compte de la similarité à l'aide de coefficients tels que Dice (Cao *et al.*, 2020) ou Jaccard (Tapi Nzali, 2020), des métriques déjà utilisées en extraction d'information.

Toujours à propos de l'évaluation, certains extraits ont été annotés par différents experts. Il serait donc intéressant de prendre en compte le recouvrement de leurs annotations afin d'extraire un taux d'accord. En revanche, dans les cas de désaccords entre experts, il nous semble impossible d'être en mesure d'adjudiquer. Ainsi, l'ensemble des annotations seront tout de même considérées comme avérées, avec certaines qui pourront être qualifiées comme fiables lorsqu'elles font l'unanimité entre les experts.

Par ailleurs, l'approche non-supervisée permet d'améliorer les résultats et donc d'accroître les performances du système. De cette façon, le rappel est largement renforcé, passant de **0,32** à **0,67** sans pour autant impacter nettement la précision, qui perd uniquement 0,02 points. Ce rééquilibrage des scores est favorable aux attentes des utilisateurs qui ne souhaitent pas passer à côté des informations importantes (visant davantage le rappel), même si cela implique par la suite de trier certaines informations (impactées par la précision). Malgré la présentation des atouts de cette approche, il est important de prendre en compte certaines limites déjà identifiées. Dans un premier temps, l'approche non-supervisée doit être adaptée à la spécificité des données, en proposant, par exemple, un entraînement des plongements sur des données spécifiques, afin de prendre en compte la terminologie du domaine ferroviaire et de l'ergonomie, une étape mise de côté par le manque de données. Dans un second temps, les seuils établis pour l'identification des marqueurs potentiels doivent être retravaillés pour être plus discriminants et limiter le nombre de suggestions à présenter. Sur le même principe, les marqueurs issus des ressources lexicales peuvent être pondérés selon leur productivité, c'est-à-dire leur capacité à identifier une information réellement pertinente. En effet, certains marqueurs issus des ressources lexicales ne sont pas toujours intégrés dans les informations relevées par les experts. Ainsi, il semble que certains d'entre eux aient davantage un statut d'indice. Nous envisageons donc de distinguer les marqueurs (productivité forte, les marqueurs suffisent, seuls, à identifier une information) des indices (productivité moyenne, les indices doivent être associés à d'autres éléments lexico-syntaxiques) et espérons que cela puisse favoriser la précision du système.

D'autre part, la collaboration et les interactions avec les experts ont guidé et influencé positivement le développement du système et de la méthodologie. Les experts ont manifesté leur intérêt pour intégrer un tel système dans leur pratique professionnelle. L'explicabilité du système et le rôle de l'utilisateur en tant que juge final ont été des déterminants de la confiance accordée au système proposé. Le système ne se place pas en remplaçant de l'enquêteur mais propose une réelle collaboration pour mettre à profit les capacités d'interprétations para-verbales et les habilités à mener un entretien de l'enquêteur.

6.2 Limites et perspectives

Les ressources lexicales contiennent au total 235 marqueurs répartis en 5 thématiques. Enrichir ces ressources lexicales serait une première piste à explorer pour garantir la fiabilité du système et améliorer le calcul des seuils du système de suggestion. Cependant, faire appel aux experts pour obtenir de nouveaux marqueurs est un processus coûteux et compliqué à mettre en place. Ainsi, nous envisageons d'explorer les patrons syntaxiques autour des marqueurs identifiés pour faire émerger les structures lexico-syntaxiques propices à la présence de marqueurs.

Actuellement, nous nous sommes principalement concentrés sur le contenu textuel des données. Pour la suite, la spécificité concernant le contexte de production (l'oral) doit être envisagée. Par exemple, les structures grammaticales spécifiques de l'oral pourraient être prises en compte pour la délimitation des informations qui se base sur une analyse syntaxique des structures standards de l'écrit. Sur ce même principe, une étape de normalisation des formes contractées ou spécifiques de l'oral peut être intégrée au moment du pré-traitement des données orales afin d'accroître la possibilité de correspondance entre ces formes et leur représentation vectorielle dans les plongements lexicaux pré-entraînés. La modalité orale permet également de laisser la possibilité d'analyser les disfluences ou les hésitations langagières afin de proposer, dans le futur, des indicateurs sur le doute ou la confiance exprimés sur des informations énoncées.

Pour poursuivre ce premier travail, nous souhaitons proposer un système qui soit en mesure de s'adapter aux ressources lexicales et aux thématiques. Nous visons ainsi la mise en place d'un système qui soit utilisable dans d'autres cas d'usage ou d'autres disciplines, ayant recours à des entretiens semi-directifs.

Remerciements

Je tiens à remercier Hèlene FLAMEIN et Luce LEFEUVRE pour leur accompagnement et encadrement tout au long de ces travaux. Je remercie également Yonnel GIOVANELLI, expert ergonomiste à la SNCF à l'initiative de ce projet, ainsi que tous les experts en analyse de l'activité qui ont apporté leur contribution.

Références

AMATO F., LÓPEZ A., PEÑA-MÉNDEZ E. M., VAÑHARA P., HAMPL A. & HAVEL J. (2013). Artificial neural networks in medical diagnosis. *Journal of Applied Biomedicine*, **11**(2), 47–58. DOI : <https://doi.org/10.2478/v10136-012-0031-x>.

BELKACEM T. & TEISSÈDRE C. (2021). Outil interactif et évolutif pour l'extraction d'information dans des documents techniques. In *Actes de la 28e Conférence sur le Traitement Automatique des Langues Naturelles. Volume 3 : Démonstrations*, p. 12–14, Lille, France : Association pour le Traitement Automatique des Langues. HAL : [hal-03265912](https://hal.archives-ouvertes.fr/hal-03265912).

BLANCHE-BENVENISTE C., BILGER M., ROUGET C., VAN DEN EYNDE K., MERTENS P. & WILLEMS D. (1990). *Le français parlé (études grammaticales)*. Sciences du Langage.

BONNEAU J. (2008). Outils d'aide à l'exploitation d'entretiens semi-directifs : Etude de l'interaction entre intervieweur et interviewés sur un corpus ethnoécologique. In B. PINCEMIN & S. HEIDEN, Éd., *JADT 2008*, p. 219–232, ENS Lettres et sciences humaines, France : Presses Universitaires de Lyon. HAL : [hal-01362690](https://hal.archives-ouvertes.fr/hal-01362690).

BREDIN H., YIN R., CORIA J. M., GELLY G., KORSHUNOV P., LAVECHIN M., FUSTES D., TITEUX H., BOUAZIZ W. & GILL M.-P. (2020). Pyannote.audio : neural building blocks for speaker diarization. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Barcelona, Spain. HAL : [hal-02995345](https://hal.archives-ouvertes.fr/hal-02995345).

CAO D., BENAMAR A., BOUMGHAR M., BOTHUA M., OULD OUALI L. & SUIGNARD P. (2020). Participation d'EDF R&D à DEFT 2020. In R. CARDON, N. GRABAR, C. GROUIN & T. HAMON, Éd., *6e conférence conjointe Journées d'Études sur la Parole (JEP, 33e édition), Traitement Automatique des Langues Naturelles (TALN, 27e édition), Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition). Atelier DÉfi Fouille de Textes*, p. 26–35, Nancy, France : ATALA. HAL : [hal-02784739](https://hal.archives-ouvertes.fr/hal-02784739).

CHAUDIRON S. (2000). Nouveaux modes d'intermédiation de l'information spécialisée. In C. E. S. P. D. NITITENKO, Éd., *La publication en ligne*, p. 153–165. Hermès. HAL : [hal-02568788](https://hal.archives-ouvertes.fr/hal-02568788).

CRYSTAL D. (2001). *Language and the Internet*. Cambridge University Press. DOI : [10.1017/CBO9781139164771](https://doi.org/10.1017/CBO9781139164771).

DUPONT M., VUILLAUME J.-M., VICTORRI B., ENJALBERT P., MATHET Y. & MALANDAIN N. (2002). Nouvelles perspectives en extraction d'information. *Revue des Sciences et Technologies*

de l'Information - Série TSI : Technique et Science Informatiques, 1(21), 37–63. HAL : halshs-00009485.

EVEN F. (2005). *Extraction d'Information et modélisation de connaissances à partir de Notes de Communication Orale*. Theses, Université de Nantes. HAL : tel-00109400.

FABRE C. (2015). Sémantique distributionnelle automatique : la proximité distributionnelle comme mode d'accès au sens. *Études de linguistique appliquée : revue de didactologie des langues-cultures*, (4), 395–405. HAL : hal-02074874.

FIRTH J. R. (1957). Applications of general linguistics. *Transactions of the Philological Society*, 56(1), 1–14. DOI : <https://doi.org/10.1111/j.1467-968X.1957.tb00568.x>.

FLAMEIN H. (2019). *Etude de la perception d'une ville : Repérage automatique, analyse et visualisation*. Theses, Université d'Orléans. HAL : tel-04429419.

GIANOLA L. (2020). *Aspects textuels de la procédure judiciaire exploitée en analyse criminelle et perspectives pour son traitement automatique*. Theses, Université de Cergy-Pontoise. HAL : tel-02522680.

KEBIR Y., DELSART A., ARFAOUI S., AURIAC-SLUSARCZYK E. & SAINT-DIZIER DE ALMEIDA V. (2021). L'apport de la linguistique à l'ergonomie pour enrichir l'analyse de l'activité de consultation de suivi médicale. In *55ème Congrès de la SELF, L'activité et ses frontières. Penser et agir sur les transformations de nos sociétés*, p. 74–79, Paris, France. HAL : hal-02541962.

KODELJA D., BESANÇON R. & FERRET O. (2019). Modèles neuronaux pour l'extraction supervisée d'événements : état de l'art [neural models for supervised event extraction : state of the art]. *Traitement Automatique des Langues*, 60(1), 13–37.

KURDI M.-Z. (2003). *Contribution à l'analyse du langage oral spontané*. Theses, Université Joseph-Fourier - Grenoble I. HAL : tel-00005071.

LACHERET A., KAHANE S., BELIAO J., DISTER A., GERDES K., GOLDMAN J.-P., OBIN N., PIETRANDREA P. & TCHOBANOV A. (2014). Rhapsodie : un Treebank annoté pour l'étude de l'interface syntaxe-prosodie en français parlé. 8, 2675–2689. DOI : [10.1051/shsconf/20140801305](https://doi.org/10.1051/shsconf/20140801305), HAL : halshs-01061368.

LEMAITRE T., GOSSET C., LAFOURCADE M., PATEL N. & MAYORAL G. (2020). Deft 2020 - extraction d'information fine dans les données cliniques : terminologies spécialisées et graphes de connaissance. In *Actes de l'atelier Défi Fouille de Textes@JEP-TALN 2020 similarité sémantique et extraction d'information fine. Atelier DÉfi Fouille de Textes*, p. 55–65, Nancy, France : Association pour le Traitement Automatique des Langues. HAL : hal-02784742.

NASR A. & BÉCHET F. (2009). Analyse syntaxique en dépendances de l'oral spontané. In A. NAZARENKO & T. POIBEAU, Éd., *Actes de la 16ème conférence sur le Traitement Automatique des Langues Naturelles. Articles longs*, p. 21–30, Senlis, France : ATALA.

NOUVEL D., EHRMANN M. & ROSSET S. (2015). *Les entités nommées pour le traitement automatique des langues*. ISTE Group.

PARCOLLET T., MORCHID M. & LINARÈS G. (2019). Réseaux de neurones convolutifs de quaternions pour l'identification de thèmes de conversations téléphoniques. In *Conférence en Recherche d'Informations et Applications*.

RADFORD A., KIM J. W., XU T., BROCKMAN G., MCLEAVEY C. & SUTSKEVER I. (2023). Robust speech recognition via large-scale weak supervision. In *Proceedings of the 40th International Conference on Machine Learning, ICML'23*, p. 28492–28518, Honolulu, Hawaii, USA : JMLR.org.

RAMPONI A. & PLANK B. (2020). Neural unsupervised domain adaptation in NLP—A survey. In D. SCOTT, N. BEL & C. ZONG, Éd., *Proceedings of the 28th International Conference on*

Computational Linguistics, p. 6838–6855, Barcelona, Spain (Online) : International Committee on Computational Linguistics. DOI : [10.18653/v1/2020.coling-main.603](https://doi.org/10.18653/v1/2020.coling-main.603).

TAMBELLINI C. (2007). *Un système de recherche d'information adapté aux données incertaines : adaptation du modèle de langue*. Thèse de doctorat. Thèse de doctorat dirigée par Berrut, Catherine Informatique Grenoble 1 2007.

TAPI NZALI M. (2020). DEFT 2020 : détection de similarité entre phrases et extraction d'information. In R. CARDON, N. GRABAR, C. GROUIN & T. HAMON, Éds., *6e conférence conjointe Journées d'Études sur la Parole (JEP, 33e édition), Traitement Automatique des Langues Naturelles (TALN, 27e édition), Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition). Atelier DÉfi Fouille de Textes*, p. 91–96, Nancy, France : ATALA. HAL : [hal-02784745](https://hal.archives-ouvertes.fr/hal-02784745).

TRZMIELEWSKI M. & GNOLI C. (2022). Organisation des connaissances médicales : principes et recherche récentes. *Pratiques d'information et connaissances en santé*, p. 173–210.

WALTER H. (1996). Dialogue : unités lexicales et analyse du sens. *Onomázein : Revista de lingüística, filología y traducción de la Pontificia Universidad Católica de Chile*, (1), 167–179.

WANG W. (2013). *Extraction d'Information Non Supervisée à Partir de Textes – Extraction et Regroupement de Relations entre Entités*. Theses, Université Paris Sud - Paris XI. HAL : [tel-00998391](https://hal.archives-ouvertes.fr/tel-00998391).