

SENSE-LM : une synergie entre modèles de langage et représentations sensorimotrices pour la recherche de références olfactives et auditives dans des documents écrits

Cédric Boscher*, Christine Largeron**
Véronique Eglin*, Előd Egyed-Zsigmond*

*INSA Lyon, CNRS, Ecole Centrale de Lyon, Université Claude Bernard Lyon 1,
Université Lumière Lyon 2, LIRIS, UMR5205, 69621 Villeurbanne, France
prenom.nom@liris.cnrs.fr
<https://liris.cnrs.fr>

**Université Jean-Monnet Saint-Etienne, CNRS,
Laboratoire Hubert Curien, UMR5516, 42000 Saint-Etienne, France
prenom.nom@univ-st-etienne.fr
<https://laboratoirehubertcurien.univ-st-etienne.fr>

Résumé. Les cinq sens – la vue, le goût, l’odorat, l’ouïe et le toucher – influencent la perception humaine à travers diverses modalités. L’extraction de références sensorielles (expressions évoquant une expérience sensorielle) dans un corpus textuel est un sujet d’intérêt avec de nombreuses applications, notamment en neurosciences ou dans l’étude du patrimoine culturel. Cet article présente SENSE-LM, un système d’extraction d’informations sensorielles dans de grandes collections de documents textuels. En associant des modèles de langage et des représentations sensorimotrices, il automatise l’extraction d’informations sensorielles à gros grain (classification binaire de phrases) et à grain fin (extraction de termes sensoriels). L’évaluation de SENSE-LM sur deux sensorialités, l’odorat et l’ouïe, révèle des avancées significatives dans l’automatisation de ces tâches complexes.

1 Introduction

La sensorialité est un concept psychophysiologique permettant de modéliser la perception humaine du monde à travers les cinq fonctions sensorielles aristotéliennes : *la vue (VIS)*, *le goût (GUS)*, *l’odorat (OLF)*, *l’ouïe (AUD)* et *le toucher (HAP)*. Un sixième sens, l’interoception (INT), fait référence aux sensations intérieures au corps humain telles que la chaleur ou la faim. La linguistique sensorielle est l’étude de la relation entre le langage humain et les expériences sensorielles. Ce domaine de recherche a de nombreuses applications dans la vie réelle, comme les sciences cognitives, l’histoire culturelle ou encore l’urbanisme. Par exemple, Murphy (2019) a mis en évidence une corrélation entre le diagnostic de la maladie d’Alzheimer et la façon dont les expériences olfactives sont décrites par des patients. Pardoen (2019) se

SENSE-LM : Extraction de références olfactives et auditives dans des documents écrits.

concentre sur la découverte d'indices auditifs dans de grands corpus de documents d'archives pour concevoir une reconstruction réaliste de l'atmosphère sonore de la Ville de Paris au 19^{ème} siècle. Menini et al. (2022) s'intéresse au patrimoine sensoriel des odeurs entre le 16^{ème} et le 20^{ème} siècle, pour des applications concernant l'histoire de l'art ou la création d'expériences sensorielles immersives dans les musées.

Dans cet article, nous proposons *SENSE-LM*, un nouveau système qui combine les avantages de modèles contextuels tels que les modèles de langage (LM), des ressources linguistiques et des techniques de génération lexicale, afin de fournir une approche robuste pour détecter les informations sensorielles dans les grands corpus de texte, en identifiant les phrases et les mots ayant trait à la sensorialité. Contrairement aux travaux existants, notre système est compatible avec les cinq sens ; nous l'évaluons pour deux d'entre eux, l'ouïe et l'odorat. Pour compenser le manque de jeux de données de référence pour cette évaluation, nous avons également construit un jeu de données artificielles centré sur la sensorialité auditive, généré avec GPT-4 et annoté manuellement. Le code de notre système ainsi que les jeux de données utilisés sont mis à disposition de la communauté¹. Une description et une évaluation étendue du système sont proposées par Boscher et al. (2024).

2 Travaux Connexes

Approches Basées sur des Ressources Lexicales Les approches basées sur les ressources lexicales ont pour objectif de fournir une liste de termes ou une taxonomie liée à un domaine sensoriel spécifique, de manière pré-définie ou générée automatiquement à partir d'un petit échantillon de termes de référence. Lexifield (Mpouli et al., 2020), un système de construction automatique de lexiques par expansion sémantique, a été proposé et directement appliqué à la recherche de termes évoquant les fonctions sensorielles dans les œuvres littéraires. Cette solution produit des résultats probants, en enrichissant automatiquement un petit ensemble de mots graines de départ, à l'aide de techniques basées sur des plongements sémantiques et des ressources externes telles que des définitions de dictionnaires (Mpouli et al. (2019)). L'approche proposée s'est heurtée à des limites telles que la polysémie, mais a toutefois révélé de premiers résultats encourageants, bien que perfectibles, car elle exploite les plongements sémantiques sur la base d'hypothèses naïves.

Approches Basées sur des Modèles de Langage Menini et al. (2022) propose une première contribution portant sur l'extraction d'information sensorielle basée sur des modèles de langage, en s'appuyant sur MacBERTh (Manjavacas et Fonteyn, 2021), une variante de BERT pré-entraînée sur des textes historiques (1450–1950), pour une classification binaire de phrases sensorielles définie de la façon suivante : "*La phrase s contient-elle une référence à l'olfaction ?*". L'efficacité de ces solutions étant étroitement liée à la qualité d'annotation de la vérité terrain, leur bonne utilisation peut néanmoins requérir l'appui de spécialistes du domaine.

1. https://github.com/cfboscher/sense-lm_eacl2024

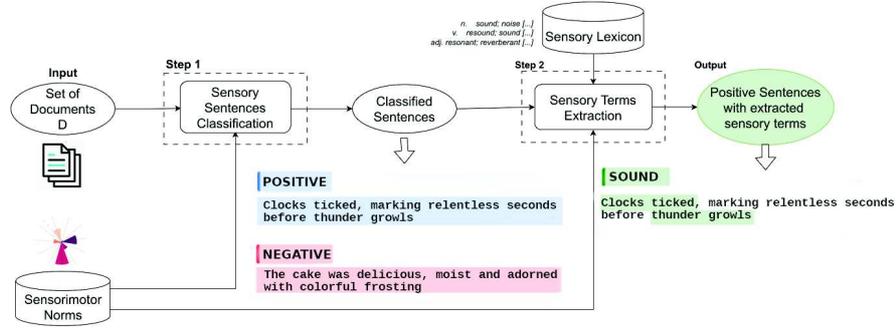


FIG. 1 – Flux de travail global de *SENSE-LM*, suivant l'exemple de la tâche de recherche d'information auditive.

3 Méthodologie

Le système que nous proposons, *SENSE-LM*, a pour objectif d'extraire les références sensorielles dans de grands corpus de documents texte, d'abord en classifiant les phrases, puis les mots qui font référence à une sensorialité donnée. La Figure 1 illustre le flux de travail global du système. L'étape 1 effectue une classification à gros grain pour identifier, dans un ensemble de documents D , l'ensemble des phrases évoquant une fonction sensorielle donnée. Ensuite, l'étape 2 réalise une classification à grain fin pour extraire les termes reflétant la sensorialité cible dans la phrase si elle est positive, *i.e.* si elle fait référence à la sensorialité donnée.

3.1 Étape 1 - Classification de Phrases Sensorielles

Définition du problème de classification de phrases

Nous considérons les fonctions sensorielles $\mathbb{M} = \{\text{OLF}, \text{GUS}, \text{AUD}, \text{VIS}, \text{HAP}, \text{INT}\}$. Nous définissons un corpus D composé de phrases, où chaque phrase s a une classe $C(s, m)$, positive (1) si elle fait référence à une fonction sensorielle m de \mathbb{M} , ou négative sinon (0). Par exemple, pour $m = \text{AUD}$, "Clocks ticked, marking relentless seconds before thunder growls" est positive, tandis que la phrase "The cake was delicious, moist, and adorned with colorful frosting" est négative. L'objectif de cette étape est d'apprendre une fonction ϵ qui associe chaque phrase s à une classe selon la fonction sensorielle m choisie : $\epsilon : D \rightarrow \{1, 0\}$ t.q. $\epsilon(s, m) = C(s, m), \forall s \in D$.

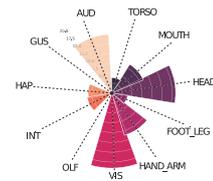


FIG. 2 – Représentation sensorimotrice de la phrase "Clocks ticked, marking relentless seconds before thunder growls".

Fonction de représentation sensorimotrice. Nous présentons ici le concept de représentation sensorimotrice, basé sur les normes sensorimotrices de Lancaster (Lynott et al., 2020). Cette ressource modélise 40 000 lemmes anglais, évalués par des annotateurs humains selon

SENSE-LM : Extraction de références olfactives et auditives dans des documents écrits.

leur appariement sémantique avec 6 fonctions sensorielles humaines (les cinq sens aristotéliens et l'interoception) et 5 fonctions motrices (utilisation des parties du corps). Chaque lemme est ainsi représenté par un vecteur à 11 dimensions avec des valeurs entre 0 et 5. Nous définissons le dictionnaire L_{SN} associant un ensemble de lemmes à leurs représentations sensorimotrices. Considérant une phrase s , nous recherchons, pour chaque mot w de s , la valeur associée à son lemme dans L_{SN} . Si le lemme de w n'est pas trouvé, nous recherchons la valeur associée à son plus proche *synset* *WordNet*, ou alors un vecteur nul à 11 dimensions. La représentation sensorimotrice globale s_{SN} de la phrase s est égale à la somme des représentations de l'ensemble des mots qui la composent. Nous préférons une somme à une moyenne, afin de refléter une quantité absolue d'information sensorielle, et ne pas pénaliser les phrases longues avec des mots non sensoriels.

Description du modèle de classification de phrases.

La première étape de *SENSE-LM* combine le plongement sémantique contextuel de la phrase produit par un modèle BERT, ainsi que la représentation sensorimotrice de cette même phrase décrite précédemment. Cette dernière représentation permet d'évaluer la similarité entre les termes à travers le prisme de la sensorialité, et apporte un niveau d'information complémentaire.

Comme le montre la Figure 3, *SENSE-LM* prend une phrase s en entrée. Sa première branche intègre les étapes successives de BERT : les couches Embedding, Transformers et Pooler, qui permettent d'extraire un plongement de s en 768 dimensions, noté s_B . La seconde branche du modèle transforme la phrase s en sa représentation sensorimotrice s_{SN} , produisant un vecteur de taille 11. Enfin, le modèle concatène s_B et s_{SN} en une représentation commune, fournie à une couche *Fully Connected* en sortie (dimension = 779), qui renvoie, soit la valeur 1, si la phrase s est considérée comme sensorielle par rapport à la fonction sensorielle m , soit 0 sinon. Toutes les couches de la composante BERT sont affinées sur un jeu d'entraînement.

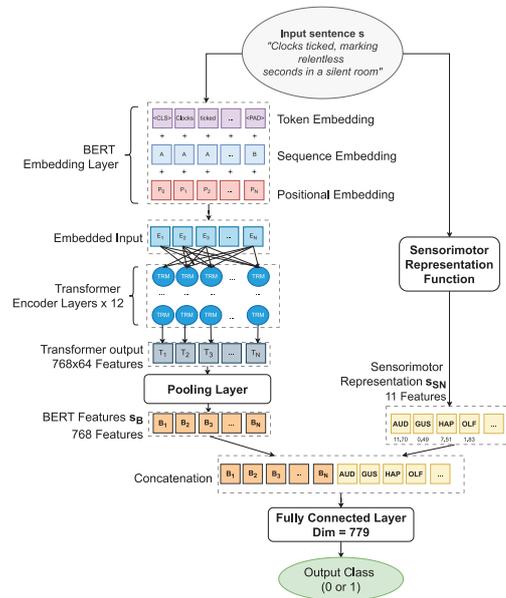


FIG. 3 – Architecture du modèle utilisé par *SENSE-LM* pour l'étape 1 (classification de phrases sensorielles)

3.2 Étape 2 - Extraction de Termes Sensoriels

Définition du problème d'extraction de termes sensoriels. L'objectif de la deuxième étape de *SENSE-LM* est d'extraire les tokens qui réfèrent à la sensorialité $m \in \mathbb{M}$ dans chaque phrase

s parmi celles classées positivement à l'étape 1. Pour chaque fonction sensorielle m et chaque phrase s de D_{pos} (phrases classées positivement), s est divisée en une séquence de tokens $t(s)$. On applique un padding avec une longueur l correspondant au nombre de tokens de la phrase positive la plus longue. Chaque token $i \in t(s)$, sauf les tokens de padding notés <PAD>, a une classe de vérité terrain $F(i, m)$: positive (1) si le token i fait référence à m dans s , et négative (0) sinon. L'objectif est d'apprendre une fonction γ qui prend $t(s)$ en entrée et renvoie un vecteur de prédictions, attribuant une classe (1 ou 0) à chaque token $i \in t(s)$, tel que :

$$\gamma : t(s) \rightarrow (1, 0, \forall i \in t(s)) \text{ avec } \gamma(t(s), m) = (F(i, m), \forall i \in t(s)), \forall s \in D_{pos}$$

Par exemple, pour $m = AUD$ et $s = \text{"Clocks ticked, marking relentless seconds before thunder growls."}$, nous obtenons $t(s) = (\mathbf{Clocks}, \mathbf{ticked}, \text{marking}, \text{relentless}, \text{seconds}, \text{before}, \mathbf{thunder}, \mathbf{growls}, \dots \text{<PAD>})$, où les tokens en gras sont positifs pour m . L'objectif est d'apprendre la fonction γ qui donne, dans ce cas : $\gamma(t(s), m) = (\mathbf{1}, \mathbf{1}, 0, 0, 0, 0, \mathbf{1}, \mathbf{1}, \dots \text{<PAD>})$.

Description du modèle d'extraction de termes. Pour réaliser cette tâche, nous proposons une approche combinatoire basée sur trois étapes successives :

Étape 2.1. Classification de termes avec RoBERTa. Nous proposons d'affiner un modèle de langage pour extraire des termes exprimant une sensorialité donnée, en utilisant une architecture BERT avec les paramètres pré-entraînés de RoBERTa, donnant les meilleures performances parmi plusieurs classifieurs évalués pour la classification des tokens sensoriels dans une phrase. Nous affinons toutes les couches du modèle. L'entrée est la phrase tokenisée $t(s)$, et la sortie est un vecteur $V(t(s), m)$, avec des 1 pour les termes positifs et des 0 pour les termes négatifs. $P_{pos}(s, m)$ désigne l'ensemble des mots positifs dans $t(s)$ et $P_{neg}(s, m)$ l'ensemble des mots négatifs.

Étape 2.2. Extension avec des ressources lexicales. Ensuite, nous nous appuyons sur une ressource lexicale, en l'occurrence Lexifield Mpouli et al. (2020), dans le but d'enrichir la liste des tokens sensoriels extraits au préalable à l'étape 2.1 et améliorer le rappel du système. Ce lexique, noté \mathbb{L}_m , contient un ensemble de mots appartenant au champ lexical de la fonction sensorielle cible m . Par exemple : $\mathbb{L}_{OLF} = \{\text{odeur (nom), sentir (verbe), ...}\}$. Pour chaque mot $w \in P_{neg}(s, m)$, on met la valeur correspondante dans $V(t(s), m)$ à 1 si $w \in \mathbb{L}_m$.

Étape 2.3. Heuristique basée sur le langage et le jugement humain. Afin de repêcher des faux négatifs, tout en évitant d'introduire des faux positifs, nous utilisons également une heuristique évaluant la distance entre les mots candidats et les mots positifs, respectivement basée sur les plongements sémantiques et les représentations sensorimotrices. Nous définissons \mathbb{E} comme un ensemble d'espaces de plongement sémantique et $\text{CosSim}_e(a, b)$ comme la similarité cosinus entre les mots a et b dans un espace $e \in \mathbb{E}$.

1. $w_{SN}(m) > T$, avec $T \in [0, 5]$,
2. $\exists e \in \mathbb{E}$, et $\exists x \in P_{pos}(s, m)$, t.q. $\text{CosSim}_e(w, x) > U$, avec $U \in [0, 1]$

SENSE-LM : Extraction de références olfactives et auditives dans des documents écrits.

| Méthode | Odeuropa Benchmark Dataset | | | Auditory Artificial Dataset | | |
|-----------------|----------------------------|---------------------|---------------------|-----------------------------|---------------------|---------------------|
| | Précision | Rappel | F1-Score | Precision | Recall | F1-Score |
| BERT | 91.51 ± 1.12 | 90.12 ± 0.61 | 90.80 ± 0.85 | 96.03 ± 0.31 | 96.14 ± 0.64 | 96.08 ± 0.45 |
| LR(s_{SN}) | 82.25 ± 1.51 | 72.33 ± 1.22 | 76.97 ± 1.36 | 87.64 ± 1.14 | 87.04 ± 1.32 | 87.23 ± 1.23 |
| GPT-4 | 91.59 ± 1.04 | 89.42 ± 2.21 | 90.4 ± 1.61 | N/A* | N/A* | N/A* |
| SENSE-LM | 94.09 ± 0.81 | 92.26 ± 0.72 | 93.16 ± 0.76 | 97.01 ± 0.15 | 97.22 ± 0.24 | 97.12 ± 0.19 |

TAB. 1 – Évaluation comparative de la tâche de classification binaire de phrases réalisée par SENSE-LM et par les approches de référence.

où $w_{SN}(m)$ est la dimension associée à m dans la représentation sensorimotrice de w . La condition 1 vérifie que le terme est cohérent avec la sensorialité m . La condition 2 assure que w est sémantiquement proche d’au moins un terme positif.

4 Expériences et Analyses

Nos expériences sont réalisées sur deux jeux de données :

Odeuropa : English Benchmark² (Menini et al., 2022). Ce jeu de données se focalise sur les expériences olfactives dans des textes historiques allant du 17^{ème} au 20^{ème} siècle. Il contient 2176 phrases, avec un ratio de phrases positives de 0,28 et 5530 occurrences de termes liés à l’odeur, répartis en 602 phrases.

Jeu de données artificiel auditif. En raison du manque de jeux de données disponibles pour d’autres fonctions sensorielles, nous avons créé un ensemble de données artificiel avec des phrases synthétiques générées par GPT-4, contenant des références à des sons. GPT-4 a été utilisé pour produire des phrases de structure et de longueur variable (400 phrases de 10 mots au maximum, 400 de 25-35 mots, et 200 de 35-50 mots), avec un ratio de 0,5 pour les phrases positives. Le protocole de génération est détaillé dans notre dépôt³.

Configuration expérimentale. Les jeux de données ont été séparés en ensembles d’entraînement et de test, avec un ratio de 0,2 pour le test. Nos modèles et ceux des approches de référence sont entraînés sur les mêmes données, avec une validation croisée à 10 plis pour 5 exécutions. Les plis de validation croisée utilisent une valeur de graine aléatoire fixée à 42. Nous utilisons l’optimiseur AdamW avec des hyperparamètres $lr = 2e^{-5}$ et $\epsilon = 1e^{-8}$, déterminés expérimentalement. Les modèles sont entraînés sur 30 époques.

Évaluation de l’Étape 1 — Classification de phrases sensorielles. Pour chaque jeu de données, nous utilisons des paramètres pré-entraînés optimisés pour la composante BERT de notre architecture : MacBERTh pour le jeu de données Odeuropa (textes historiques) et bert-base-uncased pour le jeu de données Auditory (textes contemporains).

Nous comparons SENSE-LM avec BERT, ainsi qu’une régression logistique entraînée sur une représentation sensorimotrice (11 caractéristiques) désignée par LR(s_{SN}), et GPT-4. Les protocoles détaillés de classification avec GPT-4 sont indiqués dans notre dépôt. Les résultats

2. https://github.com/Odeuropa/benchmarks_and_corpora.

3. https://github.com/cfboscher/sense-lm/tree/main/gpt4_prompts

| Méthode | Jeu de Données Odeuropa | | | Jeu de données Artificiel Auditif | | |
|---|-------------------------|---------------------|---------------------|-----------------------------------|---------------------|---------------------|
| | Précision | Rappel | F1-Score | Précision | Rappel | F1-Score |
| Lexifield (L_m) | 77.3 ± 1.33 | 43.53 ± 1.17 | 55.69 ± 1.25 | 43.25 ± 0.18 | 16.32 ± 0.27 | 23.69 |
| GPT-4 | 52.90 ± 2.11 | 70.99 ± 2.36 | 60.62 ± 2.24 | N/A* | N/A* | N/A* |
| <i>SENSE-LM</i> (Étape 2.1) | 80.01 ± 2.22 | 66.32 ± 1.13 | 72.52 ± 1.68 | 91.51 ± 2.84 | 89.25 ± 2.94 | 90.36 ± 2.89 |
| <i>SENSE-LM</i> (Étape 2.1 ∪ Étape 2.2) | 81.5 ± 2.11 | 72.7 ± 1.56 | 76.84 ± 1.74 | 91.75 ± 2.84 | 92.49 ± 2.75 | 92.11 ± 2.81 |
| <i>SENSE-LM</i> (Étape 2.1 ∪ Étape 2.3) | 80.48 ± 1.65 | 70.21 ± 1.87 | 74.99 ± 1.77 | 91.19 ± 2.76 | 92.32 ± 2.81 | 91.75 ± 2.79 |
| <i>SENSE-LM</i> (Toutes les étapes) | 82.01 ± 1.81 | 73.62 ± 1.56 | 77.58 ± 1.65 | 91.65 ± 2.72 | 93.01 ± 2.65 | 92.32 ± 2.70 |

TAB. 2 – Évaluation comparative de la tâche d’extraction des termes sensoriels par *SENSE-LM* par rapport aux approches de référence.

* Étant donné que les données auditives ont été générées par GPT-4 selon notre critère de classification, nous n’évaluons pas ce modèle pour éviter des résultats biaisés..

dans le tableau 1 montrent que *SENSE-LM* surpasse BERT, $LR(s_{SN})$, et GPT-4 en termes de précision, rappel et score F1 pour les deux ensembles de données. Cela peut être dû aux limites des modèles de langage comme BERT et GPT-4, qui ne bénéficient pas d’une représentation de concepts basée sur le jugement humain, contrairement à *SENSE-LM*.

Évaluation de l’Étape 2 - Extraction de termes sensoriels À l’Étape 2.1, nous configurons BERT avec les paramètres pré-entraînés de RoBERTa, et affinons le modèle sur notre jeu de données, dans le but de prédire une première série de mots pour chaque phrase candidate. À l’Étape 2.2, nous utilisons les lexiques de Lexifield (Mpouli et al., 2020) : 155 mots pour l’odorat et 551 pour l’ouïe. À l’Étape 2.3, notre heuristique considère 3 espaces sémantiques : ‘word2vec-google-news-300’, ‘glove-wiki-gigaword-300’ et la représentation sensorimotrice, avec des seuils $T = 3, 50$ et $U = 0, 65$ pour Odeuropa, et $T = 4, 50$ et $U = 0, 75$ pour le jeu de données Auditif.

Le Tableau 2 présente les résultats fournis par les approches de référence (en haut) et par *SENSE-LM*, avec une évaluation ablative de chaque composante (en bas). *SENSE-LM* affiche les meilleures performances globales. Notre raisonnement sur les limites de performance de GPT-4 pour l’Étape 1 demeure valable dans ce nouveau cas, et nos hypothèses sont même renforcées par le fait d’utiliser un échantillon encore plus petit, à savoir de 600 phrases. Dans un second temps, l’évaluation ablative de *SENSE-LM* met en évidence l’intérêt de combiner successivement ses 3 étapes, donnant les meilleures performances dans tous les cas.

5 Conclusion et travaux futurs

Dans cet article, nous avons présenté *SENSE-LM*, une méthode innovante d’extraction de références sensorielles, identifiant les phrases et mots sensoriels dans un corpus. *SENSE-LM* combine les représentations sensorimotrices avec les représentations contextuelles de BERT pour l’extraction d’informations sensorielles, applicable à tous les sens. Son évaluation sur deux jeux de données, caractérisant l’Odorat et l’Ouïe, montre des résultats encourageants par rapport aux solutions existantes. Notre approche pourrait être transférée à des tâches similaires impliquant le jugement humain, comme l’analyse des sentiments ou de la polarité politique, en adaptant la fonction de représentation sensorimotrice à d’autres domaines. Ce travail ouvre de nouvelles directions de recherche, incluant l’enrichissement des modèles de langage avec des modalités supplémentaires comme des images.

SENSE-LM : Extraction de références olfactives et auditives dans des documents écrits.

Remerciements

Nous remercions chaleureusement la Région Auvergne-Rhône-Alpes pour son soutien au projet *Symtesens* dans le cadre du programme *Pack Ambition Recherche 2020-2024*.

Références

- Boscher, C., C. Largeron, V. Eglin, et E. Egyed-Zsigmond (2024). SENSE-LM : A synergy between a language model and sensorimotor representations for auditory and olfactory information extraction. In Y. Graham et M. Purver (Eds.), *Findings of the Association for Computational Linguistics : EACL 2024*, St. Julian's, Malta, pp. 1695–1711. Association for Computational Linguistics.
- Lynott, D., L. Connell, M. Brysbaert, J. Brand, et J. Carney (2020). The Lancaster Sensorimotor Norms : multidimensional measures of perceptual and action strength for 40,000 English words. *Behavior Research Methods* 52(3), 1271–1291, doi: 10.3758/s13428-019-01316-z.
- Manjavacas, E. et L. Fonteyn (2021). Macberth : Development and evaluation of a historically pre-trained language model for english (1450-1950). In *Proceedings of the Workshop on Natural Language Processing for Digital Humanities*, pp. 23–36.
- Menini, S., T. Paccosi, S. S. Tekiroglu, et S. Tonelli (2022). Building a multilingual taxonomy of olfactory terms with timestamps. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pp. 4030–4039. European Language Resources Association.
- Mpouli, S., M. Beigbeder, et C. Largeron (2020). Lexifield : a system for the automatic building of lexicons by semantic expansion of short word lists. *Knowledge and Information Systems* 62(8), 3181–3201, doi: 10.1007/s10115-020-01451-6.
- Mpouli, S., C. Largeron, et M. Beigbeder (2019). Identifying sound descriptions in written documents. In *2019 13th International Conference on Research Challenges in Information Science (RCIS)*, pp. 01–06. IEEE.
- Murphy, C. (2019). Olfactory and other sensory impairments in alzheimer disease. *Nature Reviews Neurology* 15(1), 11–24.
- Pardoen, M. (2019). Projet Bretez : une pincée de son dans l'Histoire. *Digital Studies/Le champ numérique* 9(1), 11, doi: 10.16995/dscn.350.

Summary

The five human senses – vision, taste, smell, hearing, and touch – shape human perception through multiple modalities. Extracting references to sensory experiences in text is a complex task with broad applications. This paper introduces *SENSE-LM*, an information extraction system designed to extract sensory references in large text collections. By combining a language model, BERT, with linguistic resources like sensorimotor norms, *SENSE-LM* performs sensory extraction at both coarse-grained (sentence classification) and fine-grained (sensory term extraction) levels. Our evaluation on Olfaction and Audition centered texts shows *SENSE-LM* outperforms state-of-the-art methods in automating these tasks.