

Modélisation des terminologies sensorielles du discours expert et non expert pour la plateforme de recommandation hybride SmartAd

Cyril Nguyen Van^{1,2}, Christophe Cruz², Laurent Gautier^{1,3}

MSH Dijon¹, USR CNRS - uB 3516, Dijon

Le2i², CNRS FRE 2005, Arts et Métiers, Université Bourgogne Franche-Comté, Dijon

Centre Interlangues Texte, Image, Langage³ (EA 4182), Université Bourgogne Franche-Comté, Dijon

I. Introduction

La lisibilité et la compréhension du discours expert par un consommateur non-expert sont sources de difficultés impactant l'acte d'achat [1]. Cette réalité est particulièrement vraie dans le domaine du vin où les propriétés organoleptiques et sensorielles des vins décrits par les institutions et experts portent une charge sémantique complexe difficilement appréhendable par les amateurs. D'après [2], au moins trois facteurs sont responsables de cette tension :

- 1) la densité terminologique du discours et l'importance de celle-ci pour son degré de cohérence ;
- 2) le potentiel sémantique du terme, tant du point de vue de sa mise en discours que de son interprétation lors de la lecture ;
- 3) et dans une moindre mesure, la morphologie du terme.

Ainsi, la communication proposée visera à présenter des travaux interdisciplinaires associant linguistique de corpus, sémantique formelle et informatique et ambitionnant le développement d'un système générique de recommandation hybride d'items [3, 4, 5] présentant une dimension sensorielle : des vins ou des mets (goût), des fragrances ou des parfums (odorat), de la musique (ouïe), des matières en aménagement intérieur (toucher) ou encore des perceptions multimodales complexes (ambiances urbaines, traces touristiques).

La plateforme SmartAd, et le projet sous-jacent AdWine spécifiquement centré sur le vin, s'inscrivent ainsi dans le contexte stratégique de la transition numérique de l'économie et sont orientés vers le conseil et l'aide aux consommateurs. Ils permettent le partage pédagogique et intuitif des expériences sensibles, à la croisée entre terminologies des experts et professionnels pour décrire les items et terminologies adoptées par les consommateurs incluant, par-delà la dénomination technique, des dimensions évaluative, expressive et émotionnelle [6, 7, 8, 9].

A partir de ces travaux en cours, la communication présente une solution méthodologique et technique pour combler le gap sémantique entre la connaissance descriptive du discours expert des œnologues/sommeliers/prescripteurs incarnés par les guides et revues de référence, celle du discours institutionnel des interprofessions, et celle du discours de l'amateur néophyte. En l'occurrence, soit le non-expert reprend les discours experts sans nécessairement en avoir la maîtrise, engendrant ainsi une confusion sémantique, soit il crée son propre discours incluant une terminologie propre construite à partir d'expériences passées mais débouchant sur une couverture sémantique réduite à dimension technique minimale (i.e.

minéralité [10][11]), ou une spécialisation sémantique réduite à dimension évaluative maximale (i.e. *buvable/buvabilité*).

L'objectif de la solution discutée ici est donc de produire des services de recommandation d'items, et d'exploiter ces connaissances dynamiques dans un processus d'extraction d'information dans la mesure où des formes d'expression du web 2.0 (forums, blogs) déplacent l'étalon définitoire des discours prescriptifs.

II. La plateforme SmartAd

1. Les systèmes de recommandation

Les systèmes de recommandation aident les utilisateurs à prendre une décision. Sur la base des recommandations d'autres personnes ou autorités, ou du contenu descriptif de ces dernières, les utilisateurs peuvent choisir sans même une connaissance adéquate des solutions alternatives [12]. Une vue d'ensemble et une taxonomie des différents types de systèmes peuvent être trouvées dans [13], les deux approches principales étant le filtrage basé sur le contenu et le filtrage collaboratif. Toutefois, des techniques sémantiques ont été développées et introduites dans ces systèmes de recommandation. Elles sont généralement basées sur des ontologies [14], en particulier des ontologies de domaine spécifique pour créer une structure et contrôler la représentation.

Le système se propose donc d'exploiter les vocabulaires contrôlés des experts et des non-experts avec un enrichissement sémantique grâce aux liens entre ces vocabulaires contrôlés. Ainsi, les termes experts liés aux termes non experts sont exploités dans un système de recommandation hybride intégrant une ontologie basée sur le contenu et sur la base des vecteurs de mots-clés.

2. Les données

Les données se présentent sous deux formats : les vocabulaires contrôlés décrivant les items, puis les items en eux-mêmes définis en utilisant ces vocabulaires contrôlés.

Le contexte opérationnel est constitué d'un ensemble de langages documentaires et de vocabulaires contrôlés modélisant les différents niveaux d'expertise des utilisateurs dans le système de recommandation (expert, initié, novice). Ces vocabulaires ([15]) constituent le noyau dur terminologique constitutif des discours experts/institutionnels visés qui présentent la particularité, dans le domaine sensoriel, d'être rétif à une saisie sémantique stricte en conditions nécessaires et suffisantes pour mieux tolérer une saisie en termes de prototype.

Ils sont intégrés dans le système de recommandation par une approche basée sur le contenu. On peut ainsi profiler tout type d'items de tout domaine nécessitant de la recommandation sur des objets sensoriels dont la dimension prioritairement subjective pour le non-expert pose jusqu'à maintenant problème pour la modélisation sémantique. Les items sont décrits à l'aide de facettes de description et chacune représente un vocabulaire contrôlé. Ceci a pour avantage de permettre la gestion de domaines complexes tout en conservant des descriptions aisément accessibles pour un humain. Chaque facette est composée d'une ressource terminologique de type liste, taxonomie ou thésaurus. Ces ressources ont en premier lieu été pensées pour une

utilisation humaine, et sont donc facilement accessibles pour des humains. Le modèle permet la réutilisation de ressources terminologiques existantes.

3. Méthodologie

Le premier enjeu méthodologique est de nature sémantique. Ces discours concernent les propriétés organoleptiques, visuelles, tactiles et olfactives de divers produits appréhendables par les sens. Cela se matérialise par une description lexicale dans les deux univers de discours (experts/institutions vs consommateurs) et par le calcul de la distance sémantique entre les univers des deux terminologies sous-jacentes – sans le rapprochement desquels la communication et la recommandation, ne peuvent se faire. L’aspect novateur du projet réside dans la saisie non seulement du sens terminologique expert pour la bonne catégorisation de l’item, mais aussi dans la modélisation des trois couches subjectives propres aux utilisations consommateurs : évaluation, expressivité, émotion.

C’est pourquoi le système vise à proposer des préconisations basées sur l’expertise métier et les pratiques de consommation afin d’obtenir des recommandations pertinentes quel que soit le niveau d’expertise de l’utilisateur/consommateur final. Il s’agit de procéder à une extraction, puis à une modélisation du vocabulaire utilisé par les consommateurs non professionnels, par exemple dans les avis/pages forums, permettant d’une part de le relier au vocabulaire « officiel »/institutionnel/expert et d’autre part de déboucher sur la définition automatique des valences associées (positive, négative ou neutre) en tant que révélateurs de la recommandation ou non. C’est là que se fait l’articulation avec les travaux sur les valences et les tonalités jusque-là peu mises en œuvre dans le domaine.

III. Discussion

Compte tenu du fait que la dimension sensorielle n’est que très peu traitée par la communauté scientifique dans son approche terminologique, nous souhaitons montrer les multiples apports qu’offre une approche linguistique associée à une classification ontologique informatisée de descripteurs sensoriels. Une ouverture concrète est la recommandation en ligne qui offre des possibilités de personnalisation pour la vente au détail plus influente que les sources de recommandation plus traditionnelles telles que «experts humains» et «autres consommateurs». Le type de produit en vente a une influence significative sur la propension à suivre les recommandations de produits. Ainsi, le gap entre discours expert et non-expert impacte les ventes et l’alignement des vocabulaires dans un contexte de système de recommandation, permettant ainsi de guider la consommation en fonction de ses expériences passées et de ses goûts.

Dans des travaux futurs est envisagée l’exploitation de textes de forums pour l’enrichissement de la description des vins par l’utilisation d’un algorithme de plongements de mots (word embeddings) pour résoudre les tâches de traitement du langage naturel et obtenir des représentations distribuées de haute qualité des mots pour un large corpus textuel.

Références

- [1] Senecal, S., & Nantel, J. (2004). The influence of online product recommendations on consumers' online choices. *Journal of retailing*, 80(2), 159-169.
- [2] Collet, T. (2014). Obstacles lexico-sémantiques à la lecture réussie d'un texte de spécialité, *Revue TTR*, 27(1), 123–148.
- [3] Thomas, H., Werner, D., Bertaux, A., & Cruz, C. (2014). Profile Refinement in Ontology-Based Recommender Systems for Economical e-News. *The 14th International Conference on Informatics in Economy*, Bucarest.
- [4] Werner, D., & Cruz, C. (2013). A Method to Manage the Precision Difference between Items and Profiles: In a Context of Content-Based Recommender System and Vector Space Model. *International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, 337-344.
- [5] Werner, D., Cruz, C., & Bertaux, A. (2014). Evaluation de la pertinence dans un système de recommandation sémantique de nouvelles économiques. *EGC*, 549-552.
- [6] Leroyer, P., & Gautier, L. (2015). Le corpus oral au cœur de la lexicographie professionnelle du vin. *Estudios de Lexicografía*, 4, 99-110.
- [7] Gautier, L., & Hohota, V. (2014). Construire et exploiter un corpus oral de situations de dégustation : l'exemple d'Oenolex Bourgogne. *Studia Universitatis Babes Bolyai - Studia Philologia*, LIX(4), 157-173.
- [8] Leroyer, P., & Gautier, L. (2015). Extracting (good) discourse examples from an oral specialized corpus of wine tasting interactions. *ENeL-working papers*, Hersmontceux 2015.
- [9] Temmerman, R., & Dubois, D. (Eds) (2017). *Food and terminology. Expressing sensory experience in several languages*, Special issue of *Terminology*, 23(1).
- [10] Gautier, L., Le Fur, Y., & Robillard, B. (2015). La « minéralité » du vin : mots d'experts et de consommateurs. In : Gautier, L., & Lavric, E. (Eds). *Unité et diversité dans le discours sur le vin en Europe*, Frankfurt/Main, Peter Lang, 149-168.
- [11] Deneulin, P., Le Bras, G., Le Fur, Y. & Gautier, L. (2014). Minéralité du vin : représentations mentales de consommateurs suisses et français. *Revue suisse Viticulture Arboriculture Horticulture*, 46(3), 174-180.
- [12] Resnick, P., & Varian, H. R. (1997). Recommender systems. *Communications of the ACM*, 40(3), 56-58.
- [13] Burke, R. (2007). Hybrid web recommender systems, *The AdaptiveWeb*, 377-408.
- [14] Guarino, N., Oberle, D., & Staab, S. (2009). What is an Ontology? *Handbook on ontologies*. Berlin, Springer, 1-17.
- [15] Bach, M (2017). *Prototypicité discursive dans les discours de vente du vin. Etude contrastive français-allemand en Bourgogne et Rhénanie-Palatinat* [dir. Pr Gautier], Dijon, Université de Bourgogne [Mémoire de Master 2].