



**HAL**  
open science

# Ontology Adaptation for Opinion Mining in French Corpora

Valentina Dragos, Adrien Legros

► **To cite this version:**

Valentina Dragos, Adrien Legros. Ontology Adaptation for Opinion Mining in French Corpora. 26th International Conference on Knowledge Based and Intelligent information and Engineering Systems, Sep 2022, Verone, Italy. hal-03877255

**HAL Id: hal-03877255**

**<https://hal.science/hal-03877255v1>**

Submitted on 29 Nov 2022

**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 - NonCommercial - NoDerivatives 4.0 International License

26th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES 2022)

# Ontology Adaptation for Opinion Mining in French Corpora

Valentina Dragos<sup>a</sup>, Adrien Legros<sup>a</sup>

<sup>a</sup>ONERA - The French Aerospace Lab  
Chemin de la Hunière  
Palaiseau 91 123, France  
[valentina.dragos@onera.fr](mailto:valentina.dragos@onera.fr);

---

## Abstract

This paper presents the development of an ontology for opinion mining in French corpora. Since ontology construction from scratch is expensive and time consuming, the model is built by adapting an existing ontology modeling appraisal categories in English. The construction method consists of two main steps: first, the ontology of appraisals is translated in French by using a concept-to-concept approach; then different adaptation strategies are implemented to improve the result of this translation. Adaptation strategies are based on text mining, weakly supervised methods and the use of WordNet to refine the model. The goal of the adaptation phase is to cope with limitations of concept-to-concept translation, to integrate concepts and relations from external corpora and resources, and to build a model that captures various features of opinions. The ontology was designed and built to incorporate linguistic and extra linguistic information on the description of opinions in French. The model was validated by using both expert insights to validate the relevance of concepts and relations and formal criteria to describe the ontology qualities for practical use.

© 2022 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Peer-review under responsibility of the scientific committee of the KES International.

*Keywords:* Opinion mining; Social data analysis ; Appraisal theory

---

## 1. Introduction

Social data analysis has been actively studied in the past few years, and the Internet became an increasingly popular source of data, with mainly machine and deep learning models being developed for this task [22]. The studies keep the distinction between factual and subjective aspects and focus either on topics [30] and narrative detection [31] or subjectivity analysis [10], [24]. Both tasks are difficult because content and dynamics in social data varies, sometimes

---

\* Corresponding author. Tel.: +0-000-000-0000 ; fax: +0-000-000-0000.

*E-mail address:* [valentina.dragos@onera.fr](mailto:valentina.dragos@onera.fr)

in the same conversation. However, meaningful clues hidden in online data are often a combination of topics and subjective statements and their identification requires deep analysis of emotions towards specific targets [32].

More specifically, techniques developed for opinion analysis are based on co-occurrences of words or purely linguistic approaches and show variable accuracy levels when dealing with this online data. Those methods are far from being able to infer the subjective and cognitive information associated with social data [4]. For instance, *religious-belief* and *buy-Easter-chocolate* are two expressions with subjective connotations that come not from individual terms, but from putting those terms together. Knowledge models serving as description resources and emphasizing the concepts that are relevant to describe opinions are therefore needed to improve those methods. However, the performance of new for social data exploration depends heavily on the coverage of available resources, which poses a problem for languages other than English.

This paper presents the construction of an ontology for opinion detection in French corpora by adapting an existing ontology of appraisals. The appraisal theory is a cognitive framework explaining how emotions are extracted from continuous evaluations of situations, agents and events. For this work, the term ontology is used in the sense of Gruber [11] and refers to a formal representation of a shared conceptualization, including definitions of concepts, relations, and a set of constraints that ensure the logical consistency of the model. The ontology incorporates linguistic and extra linguistic information: concepts from the initial resource are specific to the appraisal theory but the model also incorporates terms associated to those concepts.

Building the ontology for opinion mining by adapting a model developed on top of the appraisal theory allows us to: 1) abstract over the different categories and systems introduced by the appraisal theory 2) facilitate reasoning between features of subjectivity and 3) enrich information extraction for social data analysis.

The remainder of the paper is organized as follows: section 2 discusses related approaches; section 3 presents the architecture and resources of the framework developed for ontology adaptation while section 4 describes the techniques used to support concept and relation extraction for ontology engineering. Ontology evaluation is discussed in section 5 and section 6 concludes the paper and sketches directions for future work.

## 2. Related work

The analysis of social data is an active research field developing approaches that can be classified into three main categories: lexicon-based methods [26], supervised machine learning and deep learning techniques [5].

Early approaches for social data analysis aimed at detecting polarity of reviews [25] [15], and the analysis was performed at sentence [7] or document level [29]. Classification algorithms were widely used for opinion detection [12], and those techniques rely upon feature engineering and manually defined rules and resources, such as dependency and causality relations. Those methods build numerical representations of textual data and convert the corpus into a term-document matrix, following several pre-processing techniques, such as normalization, stemming and part-of-speech tagging. Supervised approaches were also used to for opinion detection and Chen and colleagues used Hidden Markov models and conditional random fields to extract aspect and polarity from social data [6]. Supervised solutions achieve reasonable accuracy for sentiment analysis but also encounter difficulties in processing a mix of multi-domain data. Moreover, the robustness of those approaches is highly impacted by the use of manually tagged inputs [14].

Approaches using neural network techniques without feature engineering became popular for social data analysis [7] during the last years. Those solutions contain shallow semantic information and the algorithms rely upon embedded structures, such as low dimensional vectors of words.

To alleviate the need for large amounts of labelled data for training and annotation purposes, unsupervised methods based lexicons or ontologies [26] have been developed. Those resources have been designed to model concepts describing features of subjectivity, personal engagement, polarity and other attributes specific to opinion detection and sentiment analysis. Based on those resources, several techniques were developed which are able to detect instances of concepts within textual content and to emphasize on words that are relevant to aspects, emotions or opinions [2].

Although the research trend is dynamic, one of the main problem researchers face when dealing with languages aside from English is the lack of corpora, lexical resources and ontologies [27]. There are very few semantic resources for French in the more general area of social data analysis [1] and translating English resources does not offer good results, as demonstrated by Ghorbel and Jacot [9]. The authors translated English SentiWordNet entries into French

and showed that even if the translation of entries is correct, some parallel words do not always share the same polarities or meaning across both languages, simply due to differences in common usage of words.

This paper addresses the development of semantic resources to support the analysis of French contents and the main contribution of this work is the construction of an ontology for opinion detection in French corpora.

### 3. A framework for ontology adaptation

Two steps are carried out to build the ontology: first, the ontology of appraisals is translated in French by adopting a concept-to-concept translation. Then, the result of this phase is enriched by adding concepts and relations discovered by three components. This section describes the general architecture and processing methods, the ontology of appraisal categories and resources of the framework.

#### 3.1. Architecture and methods

A framework for ontology translation and adaptation is presented in fig. 1. In this framework three major components are defined, implementing methods based on text mining, weakly supervised techniques and lexical resources.

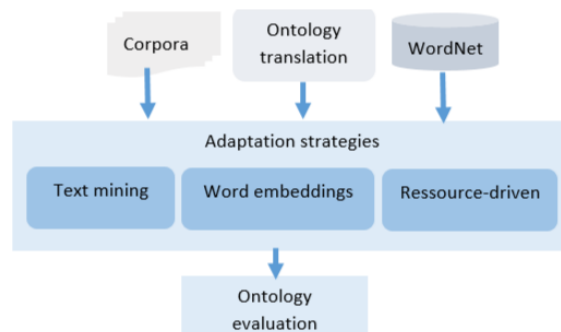


Fig. 1. Architecture for ontology adaptation

**The text mining component** uses various implementations of the Apriori algorithm [3] to extract association rules out of raw or part of speech annotated texts. Apriori is a well-known association rule learning algorithm designed for finding frequent items over transactional data sources. For this work the algorithm is used in order to discover associations of words that can further support relation acquisition.

**The weakly supervised component** uses word embeddings to infer the meaning of words by using the distribution of linguistic contexts they appear in. Word embeddings are distributed word representations mapping a given word to a vector in a pre defined N dimensional space. Word embeddings capture word relatedness for words appearing in similar contexts, encoding them such that they exist closer within the vector space. Words that have similar meaning also have similar vector representations, and for this work, this technique was used in order to support concept acquisition.

**The lexical component** accesses WordNet [21], a lexical data base for English, where each word is associated with one or more synsets. A synset consists of a identifier that represents a unique meaning. Words can have one or more synsets (*love* can be in reference to the verb or the noun and thus related to two different synsets) and one synset ca be related to one or several words (*nice* and *pretty* can be synonymous). The purpose of this component is to refine the ontology by adding new concepts or instances discovered by exploring the synsets.

The text mining and weakly supervised components require very little or even no background knowledge, but they may also be restricted to return only simple hints, like term associations and co-occurrences.

#### 3.2. The ontology of appraisal categories

The ontology of appraisal categories is the starting point of this work. The ontology is presented in [8], but we include in this section a brief description in order to facilitate the understanding of the following paragraphs. The

ontology was build on top of the appraisal theory, a cognitive framework claiming that people’s emotions are elicited by their personal and continuous interpretations, evaluations or appraisals of objects, events and situations. From a linguistic standpoint, the appraisal theory provides a way to express how humans interpret some particular event as positive or negative along with their position, support and engagement with respect to their own interpretation.

The appraisal theory structures appraisal expressions under three main basic systems describing attitude, engagement and graduation, as shown in fig. 2.

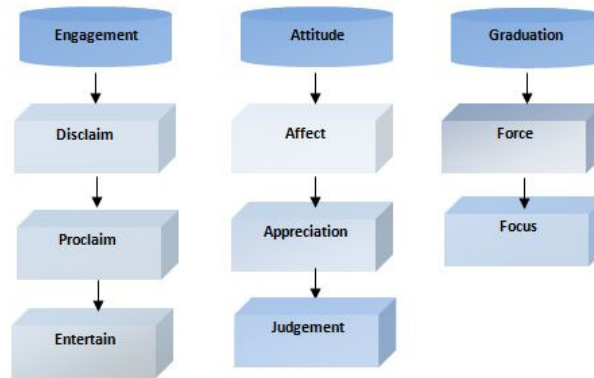


Fig. 2. Systems of the appraisal theory

**Attitude** is a system related to linguistic expressions conveying the current of authors at the time they write the text. The system covers three main subcategories: *affect*, *appreciation* and *judgement*.

*Affect* is related to linguistic expressions of author’s feelings such as happiness (e.g. Helping others makes me happy), joy, sadness, grief, etc.

*Judgment* highlights linguistic expressions conveying characterization of persons and behaviors by the author. Generally, it conveys opinions and personal tastes about objects, such as nice, ugly, beautiful, shy but also about interactions and behaviors in the social context: heroic (e.g. They are a *great* nation.), brave, open-minded, etc.

*Appreciation* captures the assessment and evaluations of entities, objects (The painting is *beautiful*), events and scenes.

**Engagement** system gathers linguistic expression specifying the author’s position with respect to his own statements. When reporting, writers often embed clues as to how strongly they support the content being conveyed and may indicate confidence, doubt, skepticism, conviction, etc., about the information reported. This system is closely related to features of trust, confidence, probability or possibility. Categories under this system encompass aspects related to denial (e.g. You *don’t* need to access the file.), concession, confirmation, endorsement (e.g. The reports clearly *show* he was involved in the accident), acknowledgement and distance (e.g. *Many* are claiming that he will not win).

**Graduation** is the last system of the appraisal theory and it is introduced in order to provide means to measure or at least estimate the orientation and various degrees of intensity associated to affect, appreciations, judgment and engagement. Main concepts are *Force* and *Focus*, and the graduation system also models linguistic modifiers such as intensifiers (very, enough, etc.) and downtoners (few, low, etc.).

### 3.3. Corpora and preprocessing

The Corpus of Parliamentary Debates [21] was used as the text input for two components : the text mining and the weakly supervised components. This corpus is made of transcripts of debates of the French parliament and captures a wide range of linguistic regularities, enabling the observation of variations in lexis, morphology, semantic and syntactic structures.

The corpus comprises 40 129 different documents, and contains around 6.5 million words and more than 100 000 different words. This corpus is a large and highly granular dataset of words occurring within different contexts, enabling the analysis of the evolution of word associations over a wide range of topics addressed by the French

parliament. The text was lowercased, special and numeric characters removed, and words tokenized. The texts have also been segmented into sentences and tagged for part-of-speech (POS) labels. The corpora of parliamentary debates is suitable for this task as the volume of data is large enough for approaches based on supervised learning and the content is of high quality, including argumentative debates and opinions.

#### 4. Ontology adaptation and engineering

The construction of the new ontology starts with the appraisal ontology, whose structure is shown in fig. 3.

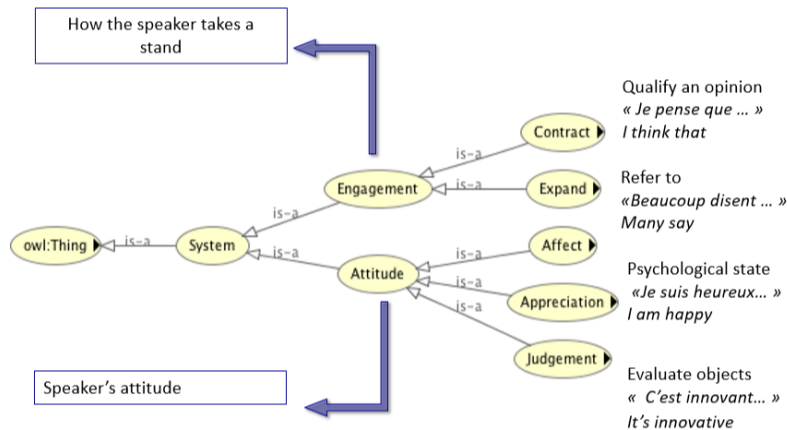


Fig. 3. Structure of the appraisal ontology

##### 4.1. Translation of the appraisal ontology

We translate this ontology and model two main concepts of the appraisal theory: *Attitude* and *Engagement*. The *Graduation* concept is not modeled.

The *Attitude* concept captures the state of mind of the author at the time of writing and has three subconcepts:

- *Affect* (*Affect*) corresponds to the deep, subjective feelings of the author ;
- *Jugement* (*Judgment*) is an assessment or characterization of individuals;
- *Appréciation* (*Appreciation*) is the appreciation of objects or entities.

The *Engagement* concept reflects the position of an author at the time of writing and it comprises two main subconcepts:

- *Référer* (*Refer*) concept is related to referring someone's opinions, directly or indirectly ;
- *Exprimer* (*Express*) concept describes the opinions of the author and it can be further divided into : *Rejeter* (*Reject*), when a statement is rejected or denied; *Accepter* (*Accept*) : when the statement is accepted or supported and *Envisager* (*Consider*), a concept that covers aspects of probability.

By creating the *Référer* concept, we wanted to be able to detect the fact that the author relies on statements that are not his own. This reference can be done directly, by citing another author, or indirectly, by taking some distance from the statements. The class *Direct* therefore captures references to another person while the class *Indirect* indicates the distance. According to the appraisal theory, verbs such as *penser* (*to believe*), "*dire* (*to say*) or *affirmer* (*to pretend*)" are instances of those concepts. We decided to model those verbs as individuals of the *Direct* concept and to add as

instances of the *Indirect* construction clearly indicating external references : *d'après* (according to), *selon* (according to), *en citant* (by citing).

During this phase we have identified several conflicts between the instances of the *Jugement* (*Judgement*) and *Appréciation* (*Appreciation*) concepts, as there is a set of words that are used in practice to describe both persons and objects. For instance, we can consider the adjectives *beau* (beautiful) and *fin* (thin) that can refer to persons as in the following examples *Un beau vieillard* (*A nice old man*) , *L'homme était grand et fin* (*The man was tall and thin*) or objects, as it is the case in the following sentence: *Un beau tableau* (*A beautiful painting*) and *Une surface fine* (*A thin surface* ). Consequently, one cannot make a clear distinction between the *Judgment* and *Appréciation* classes and these classes are not disjoint.

In order to cope with this ambiguity, we have decided to merge the *Jugement* (*Judgment*) and *Appréciation* (*Appreciation*) classes, under a new concept called *Evaluation* (*Assessment*) and this concept was further divided into 7 subconcepts inherited from the *Judgement* class : *Predictability*, *Emotivity*, *Importance*, *Authenticity*, *Complexity*, *Reliability*, *Causality* and also 3 concepts capturing the features of *Affect*: *Bonheur*, *Sécurité*, and *Satisfaction*. This representation is finer grained and thus those classes are disjoint, having distinct sets of instances.

This version of the ontology, although being in line with the appraisal theory, has a main weakness: a word can have synonyms, and the entire set of words covering a meaning is not attached to concepts. To overcome this limitation, this version of the ontology was enriched by using three approaches, as described hereafter.

#### 4.2. Text mining for relation extraction

The text mining component serves two main purposes: (1) detect sets of words that can be associated with concepts of the ontology and can be modeled and conceptual relations; (2) allows such information to be inferred by mining a general data set. This component implements the Apriori algorithm for association rules extraction. Apriori considers associations in the form of  $f : X \mapsto Y$  where  $X$  and  $Y$  are disjoint item sets, estimates  $\sigma(X)$  the number of transactions containing  $X$  and identifies association rules thanks to two criteria:

- Support of the rule:  $S(X \mapsto Y) = \frac{\sigma(X \cup Y)}{N}$ , where  $N$  is the number of transactions
- Confidence of the rule:  $C(X \mapsto Y) = \frac{\sigma(X \cup Y)}{\sigma(X)}$

Associations rules are associations having support and confidence above some empirical threshold values. Apriori was applied using the corpus of parliamentary debates but with different customized inputs, in an effort to detect various words associations stemming from the same input corpora. We implemented the Apriori algorithm by generating associations from row text corpus, POS annotated text and instances of lexical patterns. Association rules extraction is the appropriate choice for relation identification, because the goal of this component was to identify conceptual relations. The algorithm detects several words associated to concepts of the ontology by initially building associations of words and then repeatedly estimate their support and confidence while using threshold values for selection. An important advantage of using Apriori with a generic corpus is that associations are produced that progressively show how the concepts of the ontology are interconnected. From the resulting set of association rules, it is possible to identify a reasonable number of relations to be added to the ontology. This text mining procedure resulted in 72 association rule. Then we manually evaluated whether the word associations captured pertained to the ontology and added new relations to the model based on those associations. After manual validation, 48 relations were added to the ontology.

#### 4.3. Word embeddings for concept acquisition

The weakly supervised component utilizes word2vec [19] to detect similar words in order to support the acquisition of new concepts. The word2vec framework implements a shallow neural network to build word embeddings, The algorithm reduces the size of word embedding space and packs more information into each embedding, highlighting semantic and syntactic regularities within the content [20].

The word2vec framework can be used for two learning tasks, differing in their treatment of word-context relation. The first task is the continuous bag of words and the network tries to predict which word is most likely to occur given

its context. The second task is the skip-gram, when the network tries to predict the neighboring words (as context) which are most likely to occur, given the current word. Fig. 4 shows the architecture of the skip-gram model.

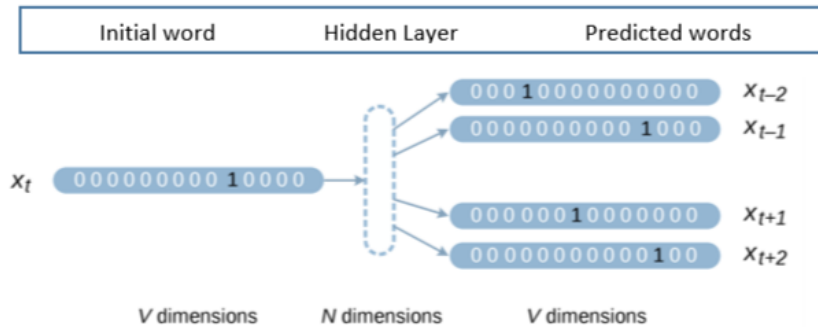


Fig. 4. Architecture of the skip-gram model [16]

The skip-gram architecture was used in order to extract words that are used in the same context as the word that are used to name concepts of the ontology. The algorithm requires the calibration of two parameters: the size of window and the learning rate. For this work, the window size was kept at 5 and the learning rate at 0.030. Fig. 5 shows the list of associations extracted for *horrible* (*horrible*) and *démontrer* (*to prove*), respectively, ranked by decreasing confidence.

<i>Horrible :</i>	<i>Démontrer :</i>
('sanguant', 0.8077459335327148)	('prouver', 0.8694282174110413)
('atroce', 0.8006289601325989)	('montrer', 0.8087137937545776)
('épouvantable', 0.7837381362915039)	('souligner', 0.7578847408294678)
('abominable', 0.7805442810058594)	('affirmer', 0.7564125657081604)
('meurtrier', 0.7672854661941528)	('rappeler', 0.7311943769454956)
('odieux', 0.7668964862823486)	('confirmer', 0.7178984880447388)
('jolie', 0.7533578276634216)	

Fig. 5. Results of the skip-gram model for *Horrible* and *Démontrer*

The results show that the skip-gram architecture of the word2vec model formed more distinctive word associations and extracted associations are accurate. At the end of this phase, 296 new words have been extracted thanks to the skip-gram model, out of which 77 are associated to concepts of the ontology. The ontology was manually updated by adding synonyms to 77 concepts and also by selecting 163 words to be integrated as new concepts.

#### 4.4. Ontology refinement with WordNet

This last refinement phase was implemented in order to improve the the robustness of the ontology. This phases uses WordNet [21], a lexical data base for English, where each word is associated with one or more synsets. A synset consists of a identifier that represents a unique meaning. A word can have one or more synsets (*love* can be in reference to the verb or the noun and thus related to two different synsets) and one synset ca be related to one or several words (*nice* and *pretty* can be synonymous). The refinement was carried out by following three steps:

1. Select the first WordNet entry for every concept of the ontology;



2. Select the most semantically related synsets of this entry. To estimate the similarity of synsets we used the Leacock Chodorov (LCH) similarity [13];
3. Select instances to be added to the ontology from the synsets previously retrieved.

Applying these steps yields a set of synsets through an extensive WordNet search using the concepts of the ontology as search terms. Synsets that deal simply with irrelevant meanings were filtered out, so only words that focused on opinion related aspects were retained. Fig. 6 shows the validation of synsets extracted for of *mauvais* (*bad*). For this example, one entire synset was rejected and from the remaining one, the instance *dangereux* (*dangerous*) was removed.

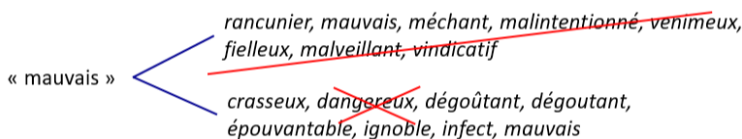


Fig. 6. Validation of synsets extracted for *mauvais* (*bad*)

This approach extracted 661 words associated to concepts of the ontology and several inconsistencies occurred as several words were associated to more than one concept. Tab. 1 shows examples of words associated to two concepts: *Emotivité* (*emotionality*) and *Bonheur* (*happiness*).

Table 1. Source of conflict: multiple word-to-concept associations.

Word	Concept1	Concept 2
assombri (darkened)	Emotivité (emotionality)	Bonheur (happiness)
funèbre (mournful)	Emotivité (emotionality)	Bonheur (happiness)
lamentable (lamentable )	Emotivité (emotionality)	Bonheur (happiness)
sinistre (grim)	Emotivité (emotionality)	Bonheur (happiness)

Those situations occur as the meaning of words in different contexts can be associated to different concepts. The manual validation allowed us to choose unique associations of words to concepts of the ontology. The validation highlighted 52 situations with multiple word-to-concept associations, with the following classes being affected by inconsistent associations: *SanctionSociale* (*Social Sanction*), *JugementPhysique* (*Physical Judgment*), *EstimeSociale* (*Social Esteem*), *Notoriété* (*Notoriety*), *Ressenti* (*Ressentiment*) and *Importance* (*Importance*). At the end of this step, 318 concepts were added to the ontology.

## 5. Ontology description and evaluation

The ontology was developed with Protégé (tool and desktop versions) [23] and at the end of the adaptation phase the ontology comprises 913 instances associated to 521 different concepts. The model also includes 76 relations. The ontology is defined in OWL [18].

The quality of the model was investigated by adopting a procedure for ontology verification and validation.

*Ontology verification* assesses the relevance of the real domain to the questions the model is expected to answer. This is a crucial aspect to be resolved during the conceptualization step. The quality of the ontology was ensured by carefully selecting the sources describing the doctrine and technical procedures and also by iteratively co-creating the initial ontology : a set of concepts was selected from the initial ontology of appraisals, their instances have been identified, hierarchical and functional relations have been modelled in a semi-supervised manner, based on manual interventions of experts.

Ontology verification impacted both the concepts and structure of the ontology. More specifically, automated reasoning was implemented to detect conflictual situations where words were assigned to more than one concept by different techniques (i.e. text mining and supervised approaches).

The analysis of conflicts identified several concepts having fuzzy boundaries. Conflicts were listed for further analysis and solved by interactions with experts. Fig. 7 illustrates the set of concepts being affected by inconsistent associations.

Concept	Number of conflicts
SanctionSociale (Social Sanction)	20
JugementPhysique (Physical Judgment)	17
EstimeSociale(Social Esteem)	14
Notoriété (Notoriety)	12
Ressenti (Feeling)	12

Fig. 7. Ontology verification: refining concepts

Ontology verification also refined the hierarchy of concepts, as illustrated in fig. 8.

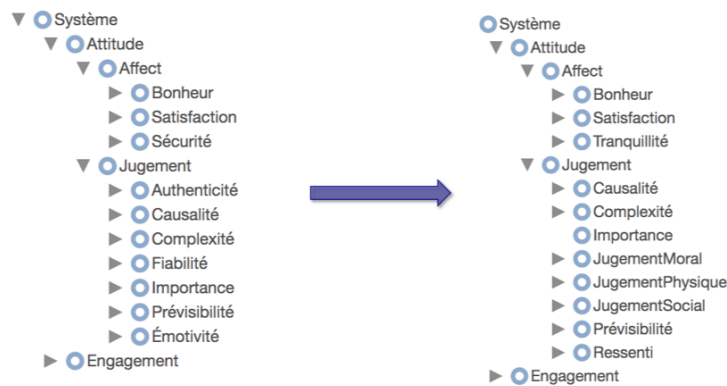


Fig. 8. Ontology verification: refining concepts

*Ontology validation* checks that the model covers relevant aspects of the real world on a credible manner. First, the ontology was validated by bringing on board domain experts able to check the model from a basic inspection perspective. This involvement allowed for a fresh set of eyes to consider the ontology and allowed to refine namely the functional relations and the axioms. Independently, three criteria were selected from [17] in order to analyze the quality of the model, in terms of completeness, consistency and redundancy, see tab. 2.

Table 2. Criteria to evaluate the ontology.

Criteria	Aspects verified	Conclusion
Inconsistency	Circularity errors	No cycles detected by the reasoner
Incompleteness	Incomplete definitions	No missing concepts according to experts
Redundancy	Semantic redundancy	Axioms indicate equivalent classes

Automated reasoning and expert analysis were used jointly to detect any logical inconsistencies in the ontology according to those criteria. When performing OWL reasoning for validation purposes, we used Hermit [28] reasoner to determine whether or not the ontology is consistent, and to identify subsumption relationships between concepts. The set of bi-directional relations was analyzed based on reasoning results and this allowed the modeling of inversed relations. During the validation step, any inconsistencies were traced back and rectified and this procedure concluded the check for logical consistency.

## 6. Conclusion

This paper describes the construction of an ontology for opinion mining in French corpora and contributes to the development of knowledge resources in French. As the volume of social data in French is increasing, this research direction deserves further development. The ontology was built by adapting an ontology of appraisals developed to explore English content. The adaptation followed three strategies, using respectively text mining, weakly supervised techniques and WordNet. Experiments show that the skip-gram model of word embeddings provided the most relevant results, although an important number of concepts and instances were also added to the ontology by querying WordNet. The ontology was validated by applying a proper qualification and validation schema and criteria.

Further research include several directions: first, improving the detection of relations, as the results of text mining techniques were rather disappointing and only a small number of associations were extracted from the corpora. Second, validation of the ontology from a practical perspective by designing an opinion detection task and analyzing the results of using the ontology in the light of gold standard inputs. Another perspective to consider is scaling up the accuracy of our model by using different lexical resources developed for French content analysis.

## Acknowledgements

This document has been produced in the context of the FLYER project funded by ASTRID, a research program supported by ANR (The French National Research Agency) and DGA (The French Defence Procurement Agency).

## Ontology availability statement

The ontology described in this paper is available on request from the authors.

## References

- [1] Abdaoui, A., Azé, J., Bringay, S., Poncelet, P., 2017. Feel: a french expanded emotion lexicon. *Language Resources and Evaluation* 51, 833–855.
- [2] Agarwal, A., Sharma, V., Sikka, G., Dhir, R., 2016. Opinion mining of news headlines using sentiwordnet, in: 2016 Symposium on Colossal Data Analysis and Networking (CDAN), IEEE. pp. 1–5.
- [3] Agarwal, R.C., Aggarwal, C.C., Prasad, V., 2001. A tree projection algorithm for generation of frequent item sets. *Journal of parallel and Distributed Computing* 61, 350–371.
- [4] Antonakaki, D., Fragopoulou, P., Ioannidis, S., 2021. A survey of twitter research: Data model, graph structure, sentiment analysis and attacks. *Expert Systems with Applications* 164, 114006.
- [5] Chakraborty, K., Bhattacharyya, S., Bag, R., Hassanien, A.E., 2018. Comparative sentiment analysis on a set of movie reviews using deep learning approach, in: International Conference on Advanced Machine Learning Technologies and Applications, Springer. pp. 311–318.
- [6] Chen, L., Martineau, J., Cheng, D., Sheth, A., 2016. Clustering for simultaneous extraction of aspects and features from reviews, in: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 789–799.
- [7] Deriu, J., Lucchi, A., De Luca, V., Severyn, A., Müller, S., Cieliebak, M., Hofmann, T., Jaggi, M., 2017. Leveraging large amounts of weakly supervised data for multi-language sentiment classification, in: Proceedings of the 26th international conference on world wide web, pp. 1045–1052.
- [8] Dragos, V., Battistelli, D., Kellodjoue, E., 2020. A formal representation of appraisal categories for social data analysis. *Procedia Computer Science* 176, 928–937.
- [9] Ghorbel, H., Jacot, D., 2011. Sentiment analysis of french movie reviews, in: Advances in Distributed Agent-Based Retrieval Tools. Springer, pp. 97–108.
- [10] Giachanou, A., Crestani, F., 2016. Like it or not: A survey of twitter sentiment analysis methods. *ACM Computing Surveys (CSUR)* 49, 28.
- [11] Gruber, T.R., 1993. A translation approach to portable ontology specifications. *Knowledge acquisition* 5, 199–220.
- [12] Hemmatian, F., Sohrabi, M.K., 2017. A survey on classification techniques for opinion mining and sentiment analysis. *Artificial Intelligence Review*, 1–51.
- [13] Leacock, C., Chodorow, M., 1998. Combining local context and wordnet similarity for word sense identification. *WordNet: An electronic lexical database* 49, 265–283.
- [14] Li, Z., Zhang, Y., Wei, Y., Wu, Y., Yang, Q., 2017. End-to-end adversarial memory network for cross-domain sentiment classification., in: IJCAI, pp. 2237–2243.
- [15] Liu, B., 2012. Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies* 5, 1–167.

- [16] McCormick, C., 2016. Word2vec tutorial-the skip-gram model. Apr-2016.[Online]. Available: <http://mccormickml.com/2016/04/19/word2vec-tutorial-the-skip-gram-model>.
- [17] McDaniel, M., Storey, V.C., 2019. Evaluating domain ontologies: clarification, classification, and challenges. *ACM Computing Surveys (CSUR)* 52, 1–44.
- [18] McGuinness, D.L., Van Harmelen, F., et al., 2004. Owl web ontology language overview. W3C recommendation 10, 2004.
- [19] Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J., 2013a. Distributed representations of words and phrases and their compositionality, in: *Advances in neural information processing systems*, pp. 3111–3119.
- [20] Mikolov, T., Yih, W.t., Zweig, G., 2013b. Linguistic regularities in continuous space word representations, in: *Proceedings of the 2013 conference of the north american chapter of the association for computational linguistics: Human language technologies*, pp. 746–751.
- [21] Miller, G.A., 1998. *WordNet: An electronic lexical database*. MIT press.
- [22] Mohammad, S., Bravo-Marquez, F., Salameh, M., Kiritchenko, S., 2018. Semeval-2018 task 1: Affect in tweets, in: *Proceedings of the 12th international workshop on semantic evaluation*, pp. 1–17.
- [23] Noy, N.F., Crubézy, M., Fergerson, R.W., Knublauch, H., Tu, S.W., Vendetti, J., Musen, M.A., 2003. Protégé-2000: an open-source ontology-development and knowledge-acquisition environment., in: *AMIA... Annual Symposium proceedings. AMIA Symposium*, pp. 953–953.
- [24] Ravi, K., Ravi, V., 2015. A survey on opinion mining and sentiment analysis: tasks, approaches and applications. *Knowledge-Based Systems* 89, 14–46.
- [25] Sahu, T.P., Ahuja, S., 2016. Sentiment analysis of movie reviews: A study on feature selection and classification algorithms, in: *2016 International Conference on Microelectronics, Computing and Communications (MicroCom)*, IEEE. pp. 1–6.
- [26] Salas-Zárate, M.d.P., Valencia-García, R., Ruiz-Martínez, A., Colomo-Palacios, R., 2017. Feature-based opinion mining in financial news: an ontology-driven approach. *Journal of Information Science* 43, 458–479.
- [27] Sanguinetti, M., Poletto, F., Bosco, C., Patti, V., Stranisci, M., 2018. An italian twitter corpus of hate speech against immigrants, in: *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.
- [28] Shearer, R., Motik, B., Horrocks, I., 2008. Hermit: A highly-efficient owl reasoner., in: *Owled*, p. 91.
- [29] Tang, D., Qin, B., Liu, T., 2015. Learning semantic representations of users and products for document level sentiment classification, in: *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pp. 1014–1023.
- [30] Wang, B., Liakata, M., Zubiaga, A., Procter, R., 2017. A hierarchical topic modelling approach for tweet clustering, in: *International Conference on Social Informatics*, Springer. pp. 378–390.
- [31] Zeng, J., Li, J., He, Y., Gao, C., Lyu, M.R., King, I., 2019. What you say and how you say it: Joint modeling of topics and discourse in microblog conversations. *Transactions of the Association for Computational Linguistics* 7, 267–281.
- [32] Zhou, Q., Zhang, C., 2017. Emotion evolutions of sub-topics about popular events on microblogs. *The Electronic Library* 35, 770–782.