

Opinion Mining in an Informative Corpus: Building Lexicons

Patrice Enjalbert, Lei Zhang, and Stéphane Ferrari

GREYC lab - CNRS UMR 6072

University of Caen Basse-Normandie

FirstName.LastName@unicaen.fr

Abstract

This paper presents first steps of an ongoing work aiming at the constitution of lexicons for opinion mining. Our work is corpus-oriented, the corpus being of informative nature (related to avionic manufacturers) rather than opinion-oriented (as in current works dealing with social networks). We especially investigate the question of interrelation between factual information and evaluative stance. Another aspect concerns the intensity of expressed opinions. Lexicons for adjectives and adverbs have been built, based on the given corpus, and we present the principles and method used for their construction.

1 Introduction

The present work takes place inside the interdisciplinary project *Ontopitex*¹, involving computer scientists and linguists from three different laboratories as well as industrial partners. The application task is provided by the company *TecKnow-Matrix* and related to competitive intelligence.

The corpus concerns avionic technologies and more precisely Boeing and Airbus companies. It consists in 377 journalistic texts from economic and technical press in French language, representing approximately 340 000 words. As expected, opinion expression is not the main characteristic of such texts. However, even “objective” facts are commonly accompanied with some evaluative stance, either by connotation (“efficient”, “noisy”, “active”...) or denotation (“attractive”, “welcome”, “good”...). Inside this context the

¹Supported by the Agence Nationale de la Recherche.

present paper concerns the constitution of a lexicon of adjectives and adverbs that take part in evaluative acts, either because they support by themselves an evaluation (Section 2) or because they contribute to its intensity (Section 3).

While most of current studies relative to lexicon generation for opinion mining focus on automated methods (Hatzivassiloglou and McKeown, 1997; Turney, 2002), our lexicon was built “manually”, by corpus observation. Several reasons motivate such an enterprise: 1) Participate in the conceptual and linguistic study of the phenomenon of evaluation; 2) Take into account corpus specificities (for example, relative to ambiguity); 3) Provide a reference in order to evaluate automated procedures; 4) Provide a bootstrap for automated analyses, with special interest on negation and intensity. Despite interesting efforts (Vernier et al., 2009)² such resources are especially missing in French.

2 Evaluative lexicon

A firm opposition is often drawn between so-called “objective” and “subjective” sentences or terms (Wiebe and Mihalcea, 2006). However we can easily observe that both are often combined: a property (or fact) is presented for its own but in a manner which inherently associates a positive or negative evaluation. This is especially true in our journalistic corpus, We say that such assertions provide an *axiologic* evaluation of the denoted fact or property: the evaluation comes from intrinsic (“objective”) properties of its target together with a specific domain-oriented axiology.

²See <http://www.lina.univ-nantes.fr/?Ressources-disponibles-sous.html>

Apart from these, we distinguish opinion judgement that are closely related with the speaker's relation with the target, hence expressing directly his/her "subjectivity": we call them *estimations*. These two categories might be associated with the traditional connotation/denotation paradigm, but with a different theoretical background. In the following we will show how these ideas may be applied to help the constitution of lexicons. For sake of brevity we will focus on adjectives but the method applies to adverbs as well, with minor adaptations. We will put stress on linguistic tests that allow to detect such terms inside texts. Note that the proposed method was in fact first elaborated in order to detect all kinds of evaluative segments in texts, for example in order to establish reference annotations.

2.1 Axiologic (intrinsic) evaluation

We call *axiologic-evaluative* an adjective that fulfills the two conditions:

1. It implies a necessary or contingent property;
2. This property can be considered as desirable or on contrary regrettable. This qualification is relative to an axiology (or norm), reflecting some goals in a given situation.

A *linguistic test* readily comes from this definition. In order to know if an occurrence of an adjective is axiologic-evaluative, insert the proposition *comme il est souhaitable/regrettable* "as it is desirable/regrettable" in the text. One of the alternatives should provide a clear contradiction, and the other a strengthening of the evaluative force of the assertion. Besides deciding of the evaluative qualification, the polarity is inferred in the obvious way. Otherwise, the insertion is irrelevant and the term occurrence non evaluative. Examples³:

(1) *It [Boeing 787] is particularly innovative* [as it is desirable (OK) / regrettable (Contradiction)].

(2) *But the production of this new plane is so complex* [as it is desirable (Contradiction) / regrettable (OK)] *that the 787 adds problems to problems.*

(3) *The ins and outs of this contract are complex and varied* [as it is desirable (Irrelevant) / regrettable (Irrelevant)]. *They are not confined to . . .*

Note that the formulation of the test may have to be slightly accommodated to the embedding

³All examples come from the French corpus. For sake of brevity, we only mention the English translation.

sentence, for example in terms of tense and verbal mode. Also observe that these definitions primarily apply to occurrences. An application of the test *in abstracto*, without context, is always possible: *X is Adj [as it is desirable/regrettable]* but with attention to the possible ambiguities. For example, *complexe* "complex", that our test reveals as evaluative in (2) but not in (3).

A deeper analysis can show various *axiologic dimensions*. After corpus examination, we propose a set of dimensions closely related with the discourse domain (which is mainly economic/commercial and technical) such as: activity-reactivity (of a person or institution), quality (of a product or device), commercial performance, etc. Such corpus-oriented evaluative dimensions appear more relevant, and easier to assign to each lexical item, than generalist and rather psychologically oriented ones as in (Bednarek, 2009) or (Martin and White, 2005).

2.2 Estimations (extrinsic evaluation)

An *estimation* is a statement expressing a positive/negative appreciation of an entity, but which does not imply any intrinsic property of the qualified entity. Two (major) classes of adjectives seem to fall in that category:

- qualifications that implicitly or explicitly imply a relation between the speaker and the entity. Here we find common terms such as: *acceptable* "acceptable", *bienvenu* "welcome", *décevant* "disappointing" . . .
- qualifications expressing an overall judgement of the entity, so-to-speak summing-up a bunch of facts or properties evaluated according to definite axiologies⁴: *bon/mauvais* "good/bad", *beau* "nice, beautiful", *brillant* "brilliant", *célèbre* "famous" . . . Hence again, the speaker is present as the one who has made the integration of different features⁵.

The linguistic test has to stage that enunciative characteristic (presence of the speaker). We pro-

⁴Indeed, this kind of judgement is often accompanied in the corpus by informations that provide its motivation.

⁵This class of terms is especially subject to ambiguities and context sensitivity. For example *good* may also be an intensifier or catch a specific aspect of the qualified entity as in *a good dish*. To be more precise if some specific property emerges from the sentence, this results from the combination with the qualified name, not from the adjective alone.

pose to insert the following comment: "To say so reveals a good/bad appreciation of the speaker but does not mention any specific property of it". Examples (we leave the reader fill the sentence with the comment):

(4) *Le titre continue de nous paraître très attractif* "the share still seems very attractive"

(5) *Les résultats ne sont guère plus brillants pour British Airways* "The perspectives are no more brilliant for British Airways"

2.3 Experiments

Lexicon building The construction of the lexicons was performed according to the following steps, same for evaluative terms and intensity modifiers.

Step 1. Collect the terms tagged as adjective or adverb by our p.o.s. tagger⁶ (1657 and 494 respectively) and clean this list according to tagging errors, removing terms that are certainly non evaluative in any context (ethnonyms, logical connectives ...). 625 adjectives and 140 adverbs remain as "possibly evaluative or intensifier".

Step 2. A concordancer is constituted with these terms and a labeling is performed according to the above mentioned categories, using the linguistic tests.

Step 3. Four lexicons are established in XML format (adjectives or adverbs, evaluative or modifier). Due to ambiguities a same item may appear in different lexicons. An item is stored as soon as it possesses one evaluative occurrence.

On the overall we retained 415 adjectives (283 "axiologic", 92 "estimations" and 40 modifiers) and 86 adverbs (36 evaluative and 50 modifiers). An evaluative entry contains the following informations: lemma, p.o.s., subclass (axiologic/estimation), polarity, intensity. The question of "intensity" is addressed in Section 3.2.

Evaluation Two tests were performed as a first attempt to evaluate the reliability of our lexicons. A first one consisted in a projection on a similar corpus. Place lacks for details, but the overall result is a good preservation; notably, 90% of entries presents in both corpora have the same labeling.

The second test took advantage of an experiment concerning negation (cf. Section 3). It relies on a corpus of articles from the French journal *Le Monde* concerning similar topics as in our main

⁶Due to the second industrial partner *Noopsis*.

one (about 3.9 M words). 25 thousand occurrences of our adjectives in a negative context were obtained by syntactic patterns, such as: "n'est pas ADJ" (*is not ADJ*). From this result, we selected a sample of 125 sentences (no more than 2 for each lemma). Then two annotators had to decide for each occurrence: (a) if it is an evaluative statement and if so (b) if it is and axiologic one (in the above sense) or an estimation and (c) the polarity (including the effect of negation).

The agreement was about 93% for test (a), 82% on test (b) and 96% on test (c) (respectively raised to 94, 86, and unchanged after coordination). The disagreement over the polarity is mainly due to some extreme adjectives in negative context such as "n'est pas catastrophique" *is not catastrophic* (further studied in Section 3.2). These results (cautiously) advocate for reliability of the lexicon, including polarity; concerning the axiologic/estimation distinction, the notion appears as quite relevant and mostly consensual, with a fuzzy zone as expected.

3 Intensity

Semantic orientation, i.e. polarity and intensity, of evaluative segments is a key issue in opinion mining (Turney, 2002). It is determined by informations in the evaluative lexicon combined with negation and intensity modifiers (sometimes together called *valence shifters* (Zaenen and Polanyi, 2004)). We present here the part of our work devoted to intensity: first the construction of a lexicon of modifiers, and second a procedure to assign an intensity to evaluative terms.

3.1 Intensity modifiers: lexicon building

Information on intensity is notably supported by both adjectives and adverbs. Concerning adverbs, we distinguish a closed list of grammatical items (*très* "very", *peu* "a little"...) and true lexical items, which can be seen as a subclass of manner adverbs, and concentrate on the latter.

Extracting such a lexicon from the corpus is rather easy. Adjectives are applied to nouns denoting a gradable entity, or more generally that possess some gradable feature: in the present context it will be a graded evaluative value, as in *une victoire complète* "a complete victory", *une belle réussite* "a nice success" ... Hence, we can apply the following simple test: "replace the Adj by *important/petit* "important/small" (or

close variants”). In fact *important/small* appears as a weak synonymous in the sense of implication: *Adj N* implies *important/small N*. The situation is similar for adverbs. They can be replaced with little meaning loss by *très (adj) / beaucoup (verb phrase)* “very/much” or close terms⁷.

Following the same procedure as in Section 2.3 we get 40 “intensity” entries for adjectives and 50 for adverbs (out of 415 and 86 respectively). The XML format codes for the following features: lemma, p.o.s., and two slots to describe the role as intensity operators: a *direction* - “ascending” for intensifiers and “descending” for moderators - and *force* - “standard” or “extreme”. Force was first determined by a test of compatibility similar as the one described now for evaluative adjectives.

3.2 Intensity of evaluative adjectives

The question here is to assign intensity to evaluative lexical terms. A first decision to be taken is the number of such values to be considered. In a first step, we fixed this number to 2, “medium” and “extreme”, for lexical items, leading to 5 values for evaluative statements by combination with negation or modifiers (Zhang and Ferrari, 2012). In our opinion, the concept of “extreme intensity” is quite relevant and useful. First we may observe that in the experiment presented Section 2.3 disagreement over polarity is due to some adjectives with “extreme force”. But the main argument is that it provides a firm, linguistically motivated, ground for assignment of intensity values to lexical items, as described now.

Our test is based on the hypothesis that extreme qualities cannot vary in intensity; in other words, *extreme* adjectives (or adverbs) are *non-gradable*. First we build a lexicon of grammatical intensifiers (proposed by Charaudeau (1992) and Noailly (1999)) which consists in 5 classes: **low** – *un peu* “a little”, **moderate** – *moyennement* “fairly”, **high** – *très* “very”, **extreme** – *extrêmement* “extremely” and **relative** – *trop* “too much”. Then, we count the frequency of co-occurrences respecting the pattern “intensifier + evaluative adjective”⁸ in a big corpus of articles from the french journal *Le Monde* (20 years). For each intensifier class *int* we gather the set \emptyset_{int} of adject-

⁷The intensity value can be combined with an evaluative one, as in *un beau contrat* “a nice contract”.

⁸60 intensifiers and 283 axiological adjectives.

ives of frequency 0 or 1, i.e. adjectives that can hardly or not at all be varied by these intensifiers. We observe the following properties:

- $|\emptyset_{low}| = 37$, $|\emptyset_{moderate}| = 78$, $|\emptyset_{high}| = 22$, $|\emptyset_{extreme}| = 78$, $|\emptyset_{relative}| = 42$.
- $\emptyset_{high} \subset \emptyset_{extreme}$, $\emptyset_{relative} \subset \emptyset_{extreme}$, $\emptyset_{low} \subset (\emptyset_{extreme} \cup \emptyset_{moderate})$.
- $nonGradable = \bigcap_{int} \emptyset_{int}$, $|nonGradable| = 7$

From these observations, we conclude first that the “non-gradable” criterion is too strict, since it retains a set of only 7 items, clearly too small. We observed then that the words contained in $\emptyset_{moderate}$ and not in $\emptyset_{extreme}$ cannot be qualified as “extreme”. Moreover we found adjectives that cannot be varied by extreme intensifiers but do so by all others, such as *terrible* “terrible”, *extraordinaire* “extraordinary”, *catastrophique* “catastrophic”, etc. It appears then that $\emptyset_{extreme}$ gathers all plausible candidates and we decided to check carefully these 78 words. For semantic-morphologic reasons the subset of words derivated from verbs or nouns are not extreme⁹, and in addition, we found a few moderate words and technical words. Finally we retained about half of this set (34) as genuine extreme adjectives.

We conclude that our method allows to perform an efficient filtering for human validation: in our case 78 candidates were automatically selected out of 283 adjectives.

4 Conclusion

In this paper we proposed a method for the creation of corpus-dependant, manually curated, lexicons for opinion adjectives and adverbs in French. We especially investigated situations where factual information and evaluation are interrelated, and the question of evaluative intensity. The method heavily relies on several linguistic tests. First experiments are encouraging regarding the quality of the created resources, presently in use in the *Ontopitex* project (and freely available on demand from the authors). Future work notably includes an extension of the method to deal with nouns and verbal expressions, and experimentations on new corpora.

⁹suffixes: *-ible, -able, -ant, -é, -aire, -teur*

References

- M. Bednarek. 2009. Dimensions of evaluation: cognitive and linguistic perspectives. *Pragmatics and Cognition*, 17(1):181–193.
- P. Charaudeau. 1992. *Grammaire du sens et de l'expression*. Hachette Paris.
- V. Hatzivassiloglou and K.R. McKeown. 1997. Predicting the semantic orientation of adjectives. In *Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics*, pages 174–181, Morristown, NJ, USA. Association for Computational Linguistics.
- J.R. Martin and P.R.R. White. 2005. *The Language of Evaluation: Appraisal in English*. Palgrave.
- M. Noailly. 1999. *L'adjectif en français*. Editions Ophrys.
- P. Turney. 2002. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. In *Proceedings of the 40th Annual Meeting of the ACL (ACL'02)*, pages 417–424. Philadelphia, Pennsylvania, USA, ACL.
- M. Vernier, L. Monceaux, B. Daille, and E. Dubreil. 2009. Catégorisation des évaluations dans un corpus de blogs multi-domaine. *Revue des Nouvelles Technologies de l'Information*, pages 45–69, nov.
- J. Wiebe and R. Mihalcea. 2006. Word sense and subjectivity. In *ACL-44: Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, pages 1065–1072.
- A. Zaenen and L. Polanyi. 2004. Contextual valence shifters. In *Proc. AAAI Spring Symposium on Exploring Attitude and Affect in Text*, pages 106–111.
- L. Zhang and S. Ferrari. 2012. Opinion analysis: the effect of negation on polarity and intensity. In *PATHOS 2012*.