

Encoding Adjective Scales for Fine-grained Resources

Cédric Lopez*, Frédérique Segond*, Christiane Fellbaum**

* Viseo Technologies, 4 avenue Doyen Louis Weil, Grenoble (France)

** Princeton University, Computer Science Department, 35 Olden Street, Princeton (USA)

E-mail: cedric.lopez@viseo.com, frederique.segond@viseo.com, fellbaum@princeton.edu

Abstract

We propose an automatic approach towards determining the relative location of adjectives on a common scale based on their strength. We focus on adjectives expressing different degrees of goodness occurring in French product (perfumes) reviews. Using morphosyntactic patterns, we extract from the reviews short phrases consisting of a noun that encodes a particular aspect of the perfume and an adjective modifying that noun. We then associate each such n-gram with the corresponding product aspect and its related star rating. Next, based on the star scores, we generate adjective scales reflecting the relative strength of specific adjectives associated with a shared attribute of the product. An automatic ordering of the adjectives “correct” (*correct*), “sympa” (*nice*), “bon” (*good*) and “excellent” (*excellent*) according to their score in our resource is consistent with an intuitive scale based on human judgments. Our long-term objective is to generate different adjective scales in an empirical manner, which could allow the enrichment of lexical resources.

Keywords: adjectives scales, resource, opinion mining

1. Introduction

Adjectives modify entities (expressed by nouns or noun phrases) by providing additional information about the entities’ properties. Some properties (or attributes) are gradable. For example, degrees of “coldness” can be expressed by modifying the adjective “cold” with an adverb, as in “rather cold”, “very cold”, *etc.* But languages usually also have adjectives that encode such degrees; English distinguishes, “arctic,” “icy” and “chilly”. Typically, such adjectives fall on two scales, each centered about a polar adjective (e.g., “hot and “cold”); scales with polar adjectives can also be united into a single scale. We can reasonably assume that speakers know where on the scale of “coldness” and “hotness” the relevant English adjectives fall; at least, they are likely to agree on the relative position of two adjectives. But can one empirically determine this?

Our work focuses on a subset of gradable adjectives, evaluative adjectives such as “good,” “great,” “terrible” and “awful.” Our long-term objective is to generate different adjective scales in an empirical manner, which could allow the enrichment of lexical resources. Such a resource could be valuable for opinion mining, especially for the prediction of rating scores of reviews (Ifrim & Weikum, 2010) (Liu & Seneff, 2009).

2. Related Work

Evaluative adjectives express a “sentiment” of the speaker/writer with respect to the entity to which the adjective is applied. Thus, “a great car” expresses a positive sentiment, while “a lousy car” expresses a negative sentiment. Sentiment Analysis is often performed with the help of a lexical resource where entries are annotated with sentiment values. For example, SentiWordNet (Baccianella *et al.*, 2010) is the result of

automatically annotating all WordNet synsets (Miller, 1995) (Fellbaum, 1998); Hu and Liu (2004) is a list of 6,800 words annotated for sentiment. A simple model may determine the sentiment of a document based on the number of “positive” and “negative” words.

(Sheinman & Tokunaga, 2009) propose a method for automatically extracting partial orderings of adjectives, using lexical semantic patterns such as “X, even Y” and “If not Y, then at least X,” where Y is a more intense adjective than its scalemate. (Sheinman *et al.*, 2013) propose a model for integrating scalar adjectives into WordNet and representing their relative ordering to one another on a scale of intensity.

3. Data and Method

Our data (in French) come from the domain of cosmetics. We focus on adjectives associated with product and brand names reviews, building on previous work aimed at identifying product names, brand names, and related entities in the cosmetic domain, where we took a symbolic approach (Lopez *et al.*, 2014).

The proposed resource of scalar adjectives is developed on the basis of data that was provided by users of cosmetic products and that has been manually annotated. Our approach focuses on linguistic patterns identified in the product reviews and their associated rating scores.

Briefly, our method involves four steps:

- 1- Acquiring reviews and associated “star” ratings for specific products.
- 2- Extracting short noun phrases.
- 3- Classifying noun phrases into predefined product aspect.
- 4- Constructing adjective scales.

In the following, we describe the 4 steps.

3.1 Collection of Data

The data on which our analysis is based is feedback provided by users of different cosmetic products. The feedback is in the form of adjectives pertaining to different aspects of the perfume (packaging, fragrance, how long the perfume lasts and the satisfaction-to-price ratio). From the Beauté-test webpage, we extract 7544 reviews of 48 perfumes.

The metadata for each review focuses on the star ratings of the perfume and on the text which is structured as follows: strengths, weaknesses, and general reviews. In this we follow the increasing tendency of websites to separate ratings for the distinct aspects of products belonging to different domains (for instance, <http://www.ciao.fr/> that deals with High-Tech, Family, Sports and Hobbies, Vehicles, etc.).

3.2 Extraction of Short Noun Phrases

We next extract short noun phrases from the raw review text by applying syntactic patterns such as the ones listed below using Holmes' parser¹. The phrases are n-grams, where n ranges from one to four:

- 4-grams : NC + ADV + ADV + ADJ / ADV + ADV + P + NC
ex: flacon vraiment très joli (*really nice bottle*), pas trop de tenue (*not too persistent*)
- 3-grams : ADV + ADJ + NC / NC + ADV + ADJ / ADV + ADV + ADJ
ex: très bon parfum (*very good perfume*), pas trop cher (*not too expensive*)
- 2-grams : ADJ + NC / NC + ADJ / ADV + ADJ
ex: bonne odeur (*good smell*), pas beau (*not pretty*)
- 1-gram : ADJ
ex: excellent (*excellent*), écœurant (*disgusting*)

Furthermore, we developed a rule to normalize patterns. For example we transform “NC + être (*be*) + ADJ” into “NC + ADJ” :

- NC + est (*is*) + ADV* + ADJ => NC + ADV* + ADJ

Thus, for the sentence “l’odeur est envoûtante” (*the smell is captivating*), we will extract « odeur envoûtante » (*captivating smell*).

Such a patterns are the most frequent short noun phrases involving adjectives in our dataset. Focusing on these patterns we ensure a low noise in our final resource.

Since modifiers are important to our investigation, our approach favors longer patterns that can be applied. Thus, for the phrase “il a une très bonne odeur” (*it has a very good smell*), “très bonne odeur” (*very good smell*) is a target phrase rather than “bonne odeur” (*good smell*).

Taking into account negation is crucial for this task. Indeed, the score associated to “bon” (*good*) might be different to the score associated to “bon” when it is modified by “pas” (*not*). Holmes' does not tag negation:

we capture negation through negative polarity items tagged as adverbs (such as “pas”) which are stored in our resource. Each user has the possibility to represent his or her opinion in increments of half-stars. We transformed the 5 star ratings attached to each aspect of the product into a score between -1 and 1. We defined a table of equivalence between the number of selected stars and the polarity. For instance 5 stars = +1, and 2.5 stars = 0.

Our final data set consists of 4,775 structured reviews, each of which contains at least one n-gram, with a stored morphosyntactic annotation. An example of a structured review is given in Figure 1. In total we extracted more than 30,000 n-grams.

```
<unit id="891">
  < strengths >
    <phrase>
      <NC>flacon</NC><ADJ>magnifique</ADJ
    >
    </phrase>
  </ strengths >
  < weaknesses />
  < comment />
  < packaging >1</ packaging >
  <fragrance>0.5</fragrance>
  < persistence >0.5</ persistence >
  < price >0.25</ price >
  < global >0.5</ global >
</unit>
```

Figure 1: Example of a structured review.

3.3 Classification of noun phrases into predefined product aspect

The objective of this step is to associate each n-gram with the corresponding product aspect and its related star score. For example, we want to associate “odeur envoûtante” (*captivating smell*) with the “fragrance” product aspect, and “joli flacon” (*pretty bottle*) with “packaging”. Noun phrases extracted from the “commentaire” (*comment*) field are attached to the global note.

For this purpose, we define lexical fields for each product aspect², which are then compared with n-gram lemmas. As we dealt only with four product aspects in this experiment, lexical fields were created manually. For future scaling, we foresee to use a French resource such as Wolf (Sagot and Fiser, 2008) or JeuxDeMots (Lafourcade, 2007).

In this way, we associate each n-gram with a product aspect (including an “empty” one) and then with its score. The following are example triples:

- “envoûtant”, “fragrance”, 1
- “joli”, “packaging”, 0.5
- “assez cher”, “price”, 0.25

² For instance : fragrance = {fragrance, parfum, senteur, odeur, sillage, note, effluve, arôme, bouquet, empreinte, ...}

¹ <http://www.ho2s.com/>

The important point is that an adjective can be attached to several product aspects. Such adjectives are thus associated to different polarity scores according to the related product aspect. For instance “joli parfum” (*pretty smell*) and “joli packaging” (*pretty packaging*) generate two different scores for the adjective “joli” (*pretty*).

3.4 Merging scores

At this step, as the reviews are totally subjective, we can obtain different scores for a given noun phrase. For instance, “très bonne tenue” (*very good persistence*) appears 120 times in the reviews with scores between 0 and 1. Then, when the noun phrases are exactly the same, we decided to merge the scores into their arithmetic mean. Otherwise, the number of occurrences for each noun phrase is stored in the resource: this is a good clue to estimate the confidence on the score.

As an example, in Figure 2 we present the entry “odeur extra” (*fantastic smell*) that obtains a score of 1 based on 4 occurrences.

```
<unit id="435">
  <post>425</post>
  <noun>odeur</noun>
  <adjective>extra</adjective>
  <modifier />
  <concept>fragrance</concept>
  <score>1.0</score>
  <occ>4</occ>
  <domain>cosmetics</domain>
</unit>
```

Figure 2: Example of an entry in our resource.

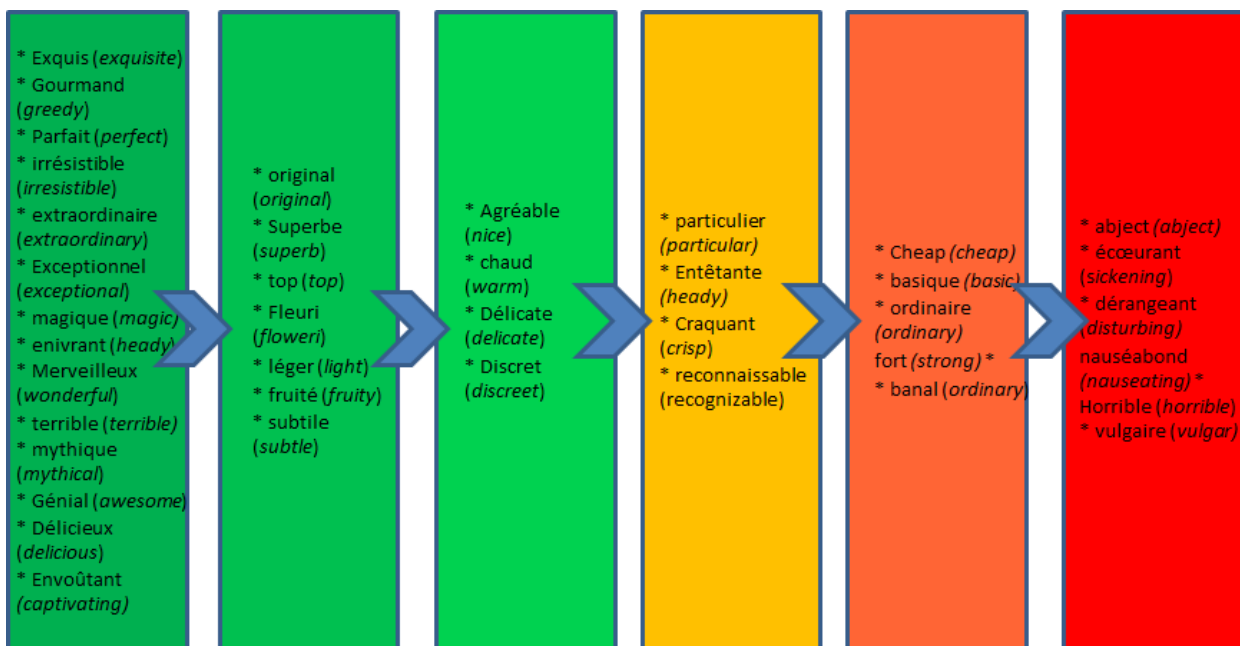


Figure 3: Generated scale of adjectives for the “smell” aspect. The “*” indicates the position of the common noun “odeur” (*smell*) with regard to the adjective (preceding or following “odeur”)

3.5 Generating intensity scales

Having transformed the data in the way described above, we can now generate scales reflecting the relative strength of specific adjectives associated with a shared attribute. An example is given in Figure 3, which shows 39 adjectives describing « odeur » (*smell*) grouped into six categories of intensity based on their scores (we manually defined the threshold between each category).

4. Our Resource

The resource consists of 3449 different entries, that is, 3449 different “modifier, noun, adjective” triples (cf. Table 1). We identified between 35 and 346 different adjectives for each product aspect. 570 adjectives are not assignable to any product aspect: we include them in the “global” rating. All in all, 553 different adjectives have been extracted from 7544 reviews of perfumes.

Product aspects	Nb. of different entries in the resource	Nb. of different adjectives
Fragrance	952	346
Persistence	93	56
Packaging	344	148
Price	98	35
Global	1962	570
Total	3449	553

Table 1: Number of entries and adjectives in the resource according to each product aspect.

We asked whether the adjectives in our dataset are truly product aspect specific and if their intensity can vary depending on the context. We tried to answer this

question by examining the use of the adjectives with each aspect of the product. We computed the number of specific adjectives related to each product aspect (cf. Table 2). Out of the “global” rating, the table shows that between 16% and 41% of the adjectives are specific to one aspect of the perfume. For instance, 41% of adjectives from our resource are specifically used to describe a fragrance, 59% are not. Such a results suggests the possibility to use our resource for other product aspects. Next, we consider the question in terms of scalability. For instance, intuitively, “cheap” should give different scores depending on the aspect it refers to (e.g. “cheaper price” is towards the positive end of scale, while “cheap perfume” is towards the negative end of the scale).

In our corpus, only 4 adjectives are used to describe the 5 aspects: “correct” (*correct*), “excellent” (*excellent*), “bon” (*good*), “sympa” (*nice*). In table 3, we compare their score according to each aspect. Note that we focus on adjectives without modifiers in order to avoid a bias in the scores.

The results show that some adjectives are not aspect dependent in terms of scalability. This is the case of “bon” which obtains the same score whatever the product aspect. This indicates a popular consensus and a less subjective adjective than expected. By contrast, adjectives such as “correct” (*correct*) or “sympa” (*nice*) seem to be more aspect specific. For instance, it appears that “sympa” obtains very different scores when it qualifies a fragrance (0.25) or a packaging (0.7).

Finally, we show that even if adjectives are not frequently specific to any product aspects in terms of usage, they might be considered as aspect specific in terms of scalability.

Product aspects	Nb. of Specific adjectives	Examples of specific adjectives
Frag.	142 (41%)	équilibré (<i>well-proportioned</i>), citronné (<i>lemony</i>), chaud (<i>warm</i>)
Persis.	9 (16%)	Exemplaire (<i>exemplary</i>), sensationnel (<i>sensational</i>), honorable (<i>commendable</i>)
Pack.	46 (31%)	Innovant (<i>innovative</i>), rectangulaire (<i>rectangular</i>), astucieux (<i>clever</i>)
Price	8 (23%)	Imbattable (<i>invincible</i>), élevé (<i>high</i>), mini (<i>mini</i>)
Global	332 (58%)	Intime (<i>intimate</i>), indépendant (<i>independent</i>), industriel (<i>industrial</i>)

Table 2: Number of specific adjectives by product aspect.

Evaluating our resource is a very difficult task because to our knowledge no independently derived and agreed-scale of adjectives exists. Nevertheless, as a first

evaluation, we ordered the adjectives “correct”, “sympa”, “bon”, “excellent” according to their score in our resource. Figure 4 represents the scores obtained for each adjective by product aspect, for cases with at least three occurrences. The results show that globally (see the columns for “average”), in our corpus, “correct” is on the bottom of the scale, above there is “sympa”, then “bon”, and, at the top of the scale, “excellent”.

Such a result corresponds to what we would intuitively expect from human judgments.

Asp./Adj.	correct	sympa	bon	excellent
Frag.	0 (3)	0.25 (8)	0.7 (67)	0.9 (31)
Persis.	0.3 (5)	1 (1)	0.7 (179)	1 (23)
Pack.	0 (2)	0.7 (21)	0.7 (13)	1 (2)
Price	0.5 (15)	1 (1)	0.7 (3)	0.75 (2)
Global	0.4 (18)	0.5 (10)	0.6 (80)	0.8 (38)

Table 3: Scores obtained for each adjective by aspect and their number of occurrences (in brackets).

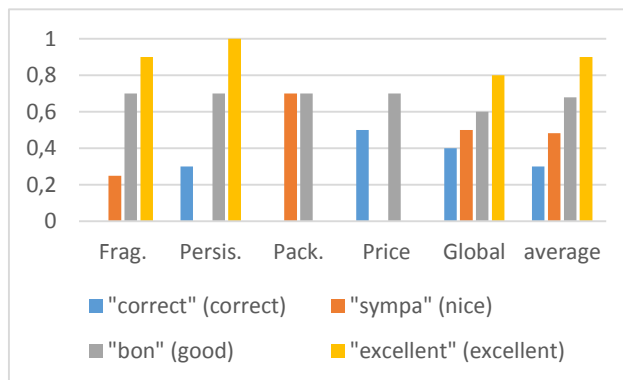


Figure 4: Scores obtained for each adjective by product aspect considering at least 3 occurrences; the average of the scores is presented on the right side of the graph.

5. Conclusion and Future Work

Our work could be applied in several ways. Opinion mining could be performed in a fine-grained way, differentiating several degrees of evaluative adjectives. For instance, from a raw text, we can extract relevant phrases to be attached to the corresponding aspect, giving a polarity (see Figure 5).

Based on these polarity scores, a global opinion can be computed, aggregating scores according to preferences of the user (fragrance over packaging, for instance). Future work should take into account additional linguistic features, such as intensifying and weakening modifiers and negation that will improve the results of the opinion mining tasks.

« très beau flacon mais odeur écoeurante (trop sucrée !!) et de courte durée. En plus il est vraiment cher !! C'est de la pure arnaque !!! »

" Beautiful bottle but sickening smell (too sweet !!) and of short duration. Furthermore it 's really expensive !! This is pure scam !!! "



Figure 5: Example of opinion mining from a raw text, using our resource.

6. Bibliographical References

- Baccianella, S., Esuli, A., & Sebastiani, F. (2010). SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. *LREC*, (pp. 2200-2204).
- Fellbaum, C. (1998). *WordNet: An Electronic Lexical Database*. Cambridge: MIT Press.
- Ifrim, G., & Weikum, G. (2010). The bag-of-opinions method for review rating prediction from sparse text patterns. *Proceedings of the 23rd International Conference on Computational Linguistics*, (pp. 913-921).
- Lafourcade, M. (2007). Making people play for Lexical Acquisition with the JeuxDeMots prototype. In *SNLP'07: 7th international symposium on natural language processing*, p. 7.
- Liu, J., & Seneff, S. (2009). Review sentiment scoring via a parse-and-paraphrase paradigm. *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, (pp. 161-169).
- Lopez, C., Segond, F., Hondermarck, O., Curtoni, P., & Dini, L. (2014). Recognition of Products and Brandnames. Application to the Cosmetic domain. *Proceedings of the Ninth International Conference on Language Resources and Evaluation Conference*, (pp. 2559-2564). Reykjavik, Islande.

- Miller, G. A. (1995). WordNet: A Lexical Database for English. *Communications of the ACM*, 38, pp. 39-41.
- Sagot, B., & Fišer, D. (2008). Building a free French wordnet from multilingual resources. In *OntoLex*.
- Sheinman, V., & Tokunaga, T. (2009). Adjscales: Visualizing differences between adjectives for language learners. *IEICE TRANSACTIONS on Information and Systems*, 92(8), pp. 1542-1550.
- Sheinman, V., Fellbaum, C., Julien, I., Schulam, P., & Tokunaga, T. (2013). Large, huge or gigantic? Identifying and encoding intensity relations among adjectives in WordNet. *Language resources and evaluation*, 47(3), pp. 797-816.