Collecting and Evaluating Lexical Polarity with a Game With A Purpose

Mathieu Lafourcade – LIRMM, Montpellier, France Nathalie Le Brun – Imagin@t, Lunel, France Alain Joubert - LIRMM, Montpellier, France





# Polarity ? What ?

 Polarization is defined as a 3-uple of numerical values that are polarities :

#### <Positive, Neutral, Negative>

- Each term has an associated polarization, for example:
  - o joy <high, low, low>
  - sick
     <low, low, high>
  - table
     <low, high, low>
  - car <high, low, high>

mostly positive mostly negative mostly neutral mixed feelings

- Terms are words of any POS, common nouns, named entities, compound words, terms with context...
- Useful in semantic analysis and opinion mining

#### Polarities ? How to collect them ?

- By making people play
- Likelt is a GWAP aimed to build a lexical polarity resource.
- Terms come from the JeuxDeMots lexical network.
  - over 500 000 terms for French
  - connected through 21 million lexico-semantic relations



- over 80 relation types : hyper, syn, semantic roles, etc.
- disambiguated : frigate>boat, frigate>bird
- contextualized : storm [context] insurance

# **Global principles**

 Likelt is a game designed to collect polarization data.



- These data are spread over the lexical network for selecting the next term to play.
- Polarities obtained through likeit are compared to polarities calculated thanks to two other GWAPs.



Emot and JDM are games that require to associate FEELINGS to terms.



#### How does it work?



# A quite simple Game Principles

• The player has to answer either



• to the question

do you like the idea of <a given term> ?

• The user can pass.

#### Game example

barrière

votre réponse précédente était non comme 58% des personnes

#### Est-ce que vous aimez l'idée de

nature

H



e

44.4%

#### Game example



#### Game example: next screen



#### Game example: next screen



#### Zoom on game result



Further to the answer given in the previous screen (barrier), the player immediately can see at the top of the next screen the percentage of players who share his view.

The game thus provides a feedback to the player while being immediately rerun with a new question.

> very efficient in terms of game play

### Example of obtained data

| Term       | Polarity distribution % |    |       | Intensity<br>nb votes |      |    |    |       |   |     |      |   |        |
|------------|-------------------------|----|-------|-----------------------|------|----|----|-------|---|-----|------|---|--------|
| gift       | POS:                    | 82 | NEUT: | 14                    | NEG: | 4  | Nb | votes | : | 280 | norm | : | 232.73 |
| retirement | POS:                    | 48 | NEUT: | 18                    | NEG: | 34 | Nb | votes | : | 303 | norm | : | 190.60 |
| policemen  | POS:                    | 29 | NEUT: | 15                    | NEG: | 56 | Nb | votes | : | 274 | norm | 3 | 177.45 |
| autumn     | POS:                    | 37 | NEUT: | 44                    | NEG: | 18 | Nb | votes | : | 277 | norm | : | 168.34 |

The term *gift* is strongly positively monopolarized, while others show a more heterogeneous distribution.

The norm value is the norm of the vector composed of the values of positive, neutral and negative polarities.

The higher the norm, the more confident we can be in the representativeness of the polarity distribution.

# **Term Selection Algorithm**

- 1. A term T (whose majority polarity is either positive or negative) is **randomly** selected.
  - the proposed term is either T with a probability p of 0.5, either N, a term randomly selected among the **neighbors** of T with a probability 1-p.
- the probability p varies under empirically determined conditions : if the total number of votes for N is under 30 (resp. over 300, over 1000) for N, then p = 0.25 (resp. 0.75, 0. 9).

#### Bootstrapping

The propagation algorithm was initiated by manually assigning a positive polarity to the term *good* (1 *positive* vote) and a negative polarity to the term *bad* (1 *negative* vote).

### Quantitative evaluation (1/6)

many votes

- mid 2008 april 2015
   > 100 millions votes
   > 385 000 polarized terms
- about 75% of the lexical network has polarity information.

| 283 votes | per term              | on average |
|-----------|-----------------------|------------|
| 178 votes | per positive polarity | on av.     |
| 88 votes  | per neutral polarity  | on av.     |
| 83 votes  | per negative polarity | on av.     |
| 120 votes | per polarity          | on av.     |

- Average number of positive votes > neutral + negative
- It is possible that many people unconsciously behave in a "socially correct" way i.e. giving only positive opinions and passing over terms that would generate a negative one.

### Quantitative evaluation (2/6)

 336,461 positive polarities
 (37.1 %)

 373,403 neutral polarities
 (41.1 %)

 198,103 negative polarities
 (21.8 %)

 907,967 polarities
 (100 %)

59,943,107 positive votes (54.9 %) 32,879,876 neutral votes (30.1 %) 16,332,188 negative votes (15 %) 109,155,171 votes (100 %)

- There is a slight predominance of neutral polarity : this is logical because the current vocabulary is mostly neutral.
- The positive polarities are almost twice as high as the negative ones due to different reasons.
  - Socially correct behavior of the players: they tend to express only their positive opinions.
  - Many terms rather perceived as neutral (e.g. Odonata) are often labeled positively.
  - A large proportion of the proposed vocabulary corresponds to named entities or areas usually eliciting approval, such as cinema, gastronomy, art, literature...

#### Quantitative evaluation (3/6)

53,881 polarities < 10 votes 96,301 polarities < 20 votes 222,988 polarities < 40 votes ▶ 468,072 polarities < 80 votes 639,269 polarities < 160 votes 863,043 polarities < 320 votes 907,261 polarities < 640 votes 907,926 polarities < 1280 votes 854,086 polarities  $\ge$  10 votes 811,666 polarities  $\ge$  20 votes 684,979 polarities  $\ge$  40 votes 439,895 polarities  $\ge$  80 votes 268,698 polarities  $\ge$  160 votes 44,924 polarities  $\ge$  320 votes 706 polarities  $\ge$  640 votes 41 polarities  $\ge$  1280 votes

- Distributions of polarities depending upon the number of votes
  - Median value is around 80. Average value (120) is shifted to the right due to the "hub" terms i.e. the very general terms, which are connected to several tens of thousands of words.
  - Such terms are more often proposed than less connected words, and thus rapidly collect a large number of votes.

### Quantitative evaluation (4/6)

| 6,835 te   | erms with positive polarity only             | (1.8 %)  |
|------------|--|----------|
| 13,006 te  | erms with neutral polarity only              | (3.4 %)  |
| 4,388 te   | erms with negative polarity only             | (1.1%)   |
| 167,396 te | erms with positive/neutral polarity only     | (43.4 %) |
| 627 te     | erms with positive/negative polarity only    | (0.2 %)  |
| 31,563 te  | erms with neutral/negative polarity only     | (8.2 %)  |
| 161,752 te | erms with positive/neutral/negative polarity | (42%)    |
| 385567 te  | erms with at least one polarity              | (100 %)  |

The dominant polarities combinations are:

- positive/neutral bipolarity 

   it confirms that people tend to vote either neutral or positive, or more precisely to vote positively even if they are rather indifferent.
- positive/neutral/negative tripolarity → it indicates that many words arouse an opinion shared, although in these polarities distributions there may be a strong dominance of one amongst the three.
- unanimity is rare: only 6.3% shows a single polarity.

### Quantitative evaluation (5/6)

Cumulative number of polarities according to the number of votes (weight).

The median values concerning the negative and neutral polarities (resp. 36 and 70) are significantly lower than the median value for positive polarities (about 200).



Many (in proportion) negative and neutral polarities with a low number of votes, and significantly more positive polarities over 200 votes.

Consistent with the hypothesis that:

People seem more likely to vote positively than negatively or neutrally

#### Quantitative evaluation (6/6)



Distribution of polarities (on the left) and of the log of polarities (on the right) according to their number of votes (weight).

Over approximately 400 votes, negative polarities are more numerous than others.

Due to the presence in the network of very negative "hubs" : hightly connected words for which the vote is almost always negative, as death, illness, accident, cancer...

Ups and downs are a consequence of the structure of the network, the algorithm and the fact that players can pass over, all combined.

# Qualitative evaluation (1/5)

# Difficulty of performing a qualitative assessment due to

- NO golden standard for lexical resource for polarity, to which the data obtained via Likelt could be compared (not for French, not even for English).
- Inability to perform a manual evaluation due to the number of terms (> 500 000).

#### One approach various data cross-analysis

- Within the project JDM, two games allow associations between terms and the feelings they evoke: terms relative to feelings can be proposed openly via a text field in the main game JeuxDeMots, and in a semi-open way (chosen by clic or given through free answer in advanced mode) in Emot game.
- For each term, we get a list of weighted associated feelings.
- Terms of feelings were the first to be reached by the propagation algorithm, so they are polarized like this:

#### arm: *strength* (110)(\*); *protection* (100)(+); *support* (80)(+); *union* (5)(-); *indifference* (4)().

(number)= weight of the relation; (+) positive dominant polarity; (-) négative dominant polarity; () neutral dominant polarity; (\*) absence of a dominant polarity.

### Qualitative evaluation (2/5)



Emot, a game that allows to associate feelings to any term within the network

# Qualitative evaluation (3/5)



JeuxDeMots, an associative game that allows to associate also sentiments to any term within the network

#### Qualitative evaluation (4/5)



### Qualitative evaluation (5/5)

Comparison between polarity inferred and polarity directly established by players through Likelt is done via a *cosine* measure and a measure of the *max* (*max*=1 if both dominant polarities coincide).

| n first terms | Cos average | Max average | Maximal polarities average |
|---------------|-------------|-------------|----------------------------|
| 1 000         | 0.80        | 0.76        | 85.65 %                    |
| 2 000         | 0.83        | 0.79        | 86.55 %                    |
| 3 000         | 0.80        | 0.75        | 87.40 %                    |
| 4 000         | 0.82        | 0.77        | 87.49 %                    |
| 5 000         | 0.83        | 0.79        | 87.63 %                    |

- We calculated the **cos** and **max** values, and ordered the first 5,000 terms by decreasing weights for the feelings relation (thus the most often played for this relation at the first).
- It clearly shows a significant correlation between the polarization defined by Likelt and that induced by the associated feelings.
- The average of the maximal polarities from the game (mpa) can be seen as the maximum rate of agreement reached on average by the general opinion, for the *n* most played words.

## Conclusion

- Our approach and tools used for getting polarity data are relatively recent, and the number of polarized terms represents a significant proportion (>75%) of the entire network.
- It can be assumed that the most interesting common words are those which are the most played in JeuxDeMots, hence the most appropriately linked to other words.

gwaps are

- As our propagation algorithm selects the vast majority of such terms, we may conclude that our approach allows to effectively polarize them.
- Given the results, we reckon we have demonstrated the feasibility, the interest of our project, and broadly undertook to build the corresponding resource.

# Future work

- never ended acquisition Continuing the double approach (polarity inferred from associated) feelings, and polarity directly assigned through the game Likelt) to further expand the already abundant lexical resource of polarity (> 385,000 words with a polarity information as a freely available resource).
  - Extrapolation : different characteristics (size, temperature, weight/baalance, temporality, location...) may be characterized and quantified using crowdsourcing through GWAP. Note that the data generated through these games, that require only knowledge and a good command of language, are of good quality, which justifies this approach.
  - The polarities data are not static but potentially fluctuating, especially in time, and depending on the circumstances. So, going further with the notion of context may be interesting, for example:

DSK [context] IMF and DSK [context] Sofite



#### Thank you for your attention

#### Do you have any questions?

