French Wikipedia Talk Pages: Profiling and Conflict Detection

Lydia-Mai Ho-Dac, Veronika Laippala, Céline Poudat and Ludovic Tanguy

University of Toulouse – CLLE-ERSS – University of Turku – TIAS – University of Nice Sophia Antipolis

September 2016
Plan

1. Detecting Conflict in French Wikipedia Talk Pages

2. WikiTalk Corpus

3. Classification of "Conflicting" vs. "Neutral" Discussions

4. A Bottom-Up Approach to Talk Page Profiling

5. Conclusion
Conflict in Wikipedia Talk pages

Wikipedia talk pages
- Discussion forum associated with each article where Wikipedian can discuss the ongoing writing process with other Wikipedian
- "The user discussions on the article Talk pages might shed light on this issue and give an insight into the otherwise hidden processes of collaboration that, until now, could only be analyzed via interviews or group observations in experimental settings." [Ferschke et al., 2012].

Conflict in this context
- Antagonistic edits of the article structure and content
- Disagreements which may turn to conflicts when the editing and discussion processes are deadlocked [Poudat et al., 2016]

Example: https://fr.wikipedia.org/wiki/Psychanalyse
Objectives

Global Objectives

- Linguistic description of the "online discussion genre" starting from Wikipedia talk pages (relatively well-written in contrast with more popular fora)[Ho-Dac and Laippala, 2015]
- Defining efficient discriminating features for Web classification
- Linguistic description of conflict in interaction

More precisely

- Identifying relevant features for selecting suitable discussions for conflict analysis i.e. discussions "conducive" to conflict
Methodology

Presented today

1. WikiTalk Corpus building with a special attention of available meta-data
2. Linguistic description of the WikiTalk corpus using NLP techniques:
   - automatic classification
   - multidimensional analysis

In progress

4. Definition of efficient discriminating features
5. Selection of discussions "conducive" to conflict
6. Conflict annotation in the selected discussions
7. Machine learning for conflict detection in online discussion
Plan

1. Detecting Conflict in French Wikipedia Talk Pages
2. WikiTalk Corpus
3. Classification of "Conflicting" vs. "Neutral" Discussions
4. A Bottom-Up Approach to Talk Page Profiling
5. Conclusion
WikiTalk Corpus building

1. Talk pages extraction from the official dump
global backup frwiki-20150512-pages-meta-current#.xml.bz2 available on http://dumps.wikimedia.org/frwiki/20150512/)

2. Talk pages selection acc. to number of words and posts

3. Syntactic parsing with Talismane [Urieli, 2013]

3,487,480 /<title>Discussion/ on the global backup

<table>
<thead>
<tr>
<th></th>
<th>User Talk pages</th>
<th>Article Talk pages</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1,990,927</td>
<td>1,496,553</td>
</tr>
<tr>
<td>Exact number</td>
<td>57%</td>
<td>43%</td>
</tr>
<tr>
<td>Redirections</td>
<td>116,432</td>
<td></td>
</tr>
<tr>
<td>Empty talk pages (&lt; 2 words)</td>
<td>1,013,791</td>
<td>68%</td>
</tr>
<tr>
<td>Selected talk pages</td>
<td>366,326</td>
<td>24%</td>
</tr>
</tbody>
</table>
Corpus Building – Document structure

Document structure encoding acc. to TEI-P5

- threads marked up as <div>
- threads topic: <head>
- posts: <post who="user" when="timestamp" interactionLevel="#">
WikiTalk Corpus Overview – soon available under CC license

<table>
<thead>
<tr>
<th>talk pages</th>
<th>threads</th>
<th>posts</th>
<th>words</th>
</tr>
</thead>
<tbody>
<tr>
<td>366,326</td>
<td>1,024,351</td>
<td>3,022,240</td>
<td>159,578,279</td>
</tr>
</tbody>
</table>

- 202,856 (55%) single post talks
- 181,503 (50%) under 53 words talks
- Some extremely long talks (up to 1,143 posts and 148,968 words)
- 150,603 (41%) monologues
- 40,413 (10%) talks involving between 8 and 228 different writers
- 80% anonymous posts
Corpus Building – Meta-Data (1)

Meta-Data associated with each talk pages

- "discipline" i.e. associated thematic portals
- "avancement" i.e. article’s quality scale based on Wikipedian assessments
- "conflictness" i.e. information manually inserted by Wikipedians via the template {{keep calm}} which adds a dedicated banner at the top of the talk page.

Please be calm and civil when you make comments or when you present evidence, and avoid personal attacks. Please be patient as we work toward resolution of any issues in a peaceful, respectful manner.

- "talk type": a specific characteristic of the French Wikipedia for allowing several talk pages per article for discussing especially about neutrality, quality, etc.
Corpus Building – Meta-Data (2)

Meta-Data extracted from other resources

- meta-data extracted from the associated article
  - portal sections e.g. History, Art, Sport, etc. (up to 7 sections associated with a same article). 11 sections
- banners (not yet included)

Affaire Bogdanoff

existence of a "parallel" talk page dedicated to neutrality, article quality, etc. (see "talk types")

Autres discussions [liste]

Suppression - Neutralité - Droit d'auteur - Article de qualité - Bon article - Lumière sur - À faire - Archives

Autres discussions [liste]

Suppression - Neutralité - Droit d'auteur - Article de qualité - Bon article - Lumière sur - À faire - Archives
Plan

1. Detecting Conflict in French Wikipedia Talk Pages
2. WikiTalk Corpus
3. Classification of "Conflicting" vs. "Neutral" Discussions
4. A Bottom-Up Approach to Talk Page Profiling
5. Conclusion
Data-driven analysis of conflicts in WikiTalks: automatic classification

Using a classifier for both the automatic detection of conflicts and their analysis

- Linear classifier VowPal Wabbit [Agarwal et al., 2011]
- Supervised Machine Learning: Based on a set of training data, the classifier learns to detect the predefined classes and describe their most typical features
- Automatic classification already tested for classifying Wikipedia articles vs. talk pages (F1 = 0.94) and talk pages vs. eHealth Fora (F1 = 1)

Need for training data with a conflict class

- No available annotated corpus
- Available meta-data that may help for predefining a "conflict class"
Classification of "Conflicting" vs. "Neutral" Discussions

An experimental classification of "conflicting" vs. "neutral" discussions

Training data: a distinction based on available meta-data

<table>
<thead>
<tr>
<th>Conflict discussions</th>
<th>Neutral discussions</th>
</tr>
</thead>
<tbody>
<tr>
<td>More than 100 words in the talk page</td>
<td></td>
</tr>
<tr>
<td>From the header talk page:</td>
<td></td>
</tr>
<tr>
<td><em>Keep calm</em></td>
<td><em>A-class</em></td>
</tr>
<tr>
<td>Existence of a parallel talk page about:</td>
<td></td>
</tr>
<tr>
<td>article neutrality</td>
<td>article A-ranking</td>
</tr>
<tr>
<td>&quot;Talk Type&quot;:</td>
<td></td>
</tr>
<tr>
<td>neutrality talk</td>
<td>na</td>
</tr>
<tr>
<td>2,028 pages</td>
<td>4,569 pages</td>
</tr>
<tr>
<td>11 M words</td>
<td>8.8 M words</td>
</tr>
</tbody>
</table>
Classification task: categorizing talk pages, threads or posts?

Three available levels of granularity

- entire talk page: 6,597 units
- threads: 46,690 units
- posts: 194,289 units

But

- meta-data only available on the talk page level
- conflict seems to be more a matter of threads than of entire talk pages

Adhoc assumption

If a talk page is a priori categorized as conflicting, then the threads or posts it contains are also categorized as conflicting.
Classification features: bag-of-lemmas or syntactic information?

Two options

- A standard option: Lexical information [Scott et al., 2006]
  - Problem: words reflect more the text topic than the text genre
- A more complex option: Syntactic information [Laippala et al., 2015]

Figure: A delexicalized syntactic bi-arc describing a clitic+verb+conjunction as in the clause ‘I find that’.
Classification of "Conflicting" vs. "Neutral" Discussions

Results (training on 2/3, test on 1/3)

<table>
<thead>
<tr>
<th>features</th>
<th>threads P</th>
<th>threads R</th>
<th>threads F1</th>
<th>posts P</th>
<th>posts R</th>
<th>posts F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>lemmas</td>
<td>0.84</td>
<td>0.60</td>
<td>0.72</td>
<td>0.79</td>
<td>0.69</td>
<td>0.74</td>
</tr>
<tr>
<td>bi-arcs</td>
<td>0.55</td>
<td>0.48</td>
<td>0.52</td>
<td>0.63</td>
<td>0.59</td>
<td>0.61</td>
</tr>
<tr>
<td>units</td>
<td>46,690</td>
<td></td>
<td></td>
<td>194,289</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table: Comparison of lexical vs. syntactic approaches for the automatic classification of conflict threads and posts.

- Syntactic approach lost
- (Adhoc) "conflict" posts are better classified than threads with the both approaches
- F1 is relatively low (79% of the detected posts are correctly classified, 69% are detected)
Lexical specificities of the conflicts

- Analysis of 40 most significant lemmas
- Two kinds: words referring to the writing process and those referring to article topics
  - style, to hope, respect, version, way of writing, restructuring, reformulation, neutralisation, clumsy, uncoherent, respect, mistake, controversy, debate, ok
  - rwanda, dictatorship, mandarin, quebec, islam, buddhism

Questionning

- the features used for classification
- our *a priori* classification of conflict and our Adhoc assumption
Plan

1. Detecting Conflict in French Wikipedia Talk Pages
2. WikiTalk Corpus
3. Classification of "Conflicting" vs. "Neutral" Discussions
4. A Bottom-Up Approach to Talk Page Profiling
5. Conclusion
Profiling talk pages

- Exploring Rich Linguistic Features usually associated with interactional and rhetorical structures
- Mining talk pages and threads for discovering classes (without \textit{a priori})
- Identifying relevant classes for conflict analysis
Data mining method

Data mining tool
- R package FactoMineR dedicated to multivariate exploratory data analysis
- Principal Components Analysis (PCA) on talk pages and threads

Rich Linguistic Features
- **Global**: general quantitative characteristics of texts (talk pages and threads) e.g. number of words, number of contributors, presence of a "keep calm" banner;
- **Thema**: portal sections of the associated article
- **Interact**: the frequency per texts of a wide range of interaction and politeness cues e.g. social deixis, marks of (dis)agreement, etc.
- **DiscRel**: the frequency per texts of connectives for each discourse relations as defined in the *LexConn*[Roze et al., 2012].
## Global features

### Information extracted from the talk page itself

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>logNnMots</td>
<td>number of words</td>
</tr>
<tr>
<td>nbFils</td>
<td>number of threads*</td>
</tr>
<tr>
<td>nbPosts</td>
<td>number of posts</td>
</tr>
<tr>
<td>profMax</td>
<td>&quot;interactional depth&quot;</td>
</tr>
<tr>
<td>nbContributeurs</td>
<td>number of different participants</td>
</tr>
<tr>
<td>nbAnonymes</td>
<td>number of anonymous posts</td>
</tr>
<tr>
<td>X.anonymes</td>
<td>% of anonymous posts</td>
</tr>
<tr>
<td>nbBots</td>
<td>number of posts written by bots</td>
</tr>
<tr>
<td>X.bots</td>
<td>% of posts written by bots</td>
</tr>
<tr>
<td>AdQ</td>
<td>&quot;1&quot; if the talk page is linked to a A-class article</td>
</tr>
<tr>
<td>polemique</td>
<td>&quot;1&quot; if the talk page has the banner &quot;keep calm&quot;</td>
</tr>
</tbody>
</table>
Thema features

Wikipedia section of the associated article

- 11 Wikipedia sections: art, geography, history, leisure, medicine, politics, religion, sciences, society, sport, technology
- Some articles are simultaneously in 7 sections!
- *Geography* is the most frequent section (119,359 talk pages)
- 11 features binarized (e.g. geography = 1/0)
- The same feature for talk pages and threads
# Interact features

## 11 features automatically identified with simple regular expressions

<table>
<thead>
<tr>
<th>Feature</th>
<th>Pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>Politeness</td>
<td>thanks, hello, goodbye, hi, sincerely, cheers, please, would you, etc.</td>
</tr>
<tr>
<td>Agreement</td>
<td>OK, agree, yes, no, actually, etc.</td>
</tr>
<tr>
<td>Question</td>
<td>?</td>
</tr>
<tr>
<td>Je</td>
<td>1st singular person pronouns + personally</td>
</tr>
<tr>
<td>Tu</td>
<td>2nd sing. pers. pronouns, informal &quot;you&quot;</td>
</tr>
<tr>
<td>Vous</td>
<td>2nd plur. pers. and formal &quot;you&quot; pronouns</td>
</tr>
<tr>
<td>Nous</td>
<td>1st plur. pers. pronouns</td>
</tr>
<tr>
<td>On</td>
<td>Informal &quot;We&quot;</td>
</tr>
<tr>
<td>WP</td>
<td>Wikipedia or WP</td>
</tr>
<tr>
<td>pour</td>
<td>Sentence-initial For or I’m for</td>
</tr>
<tr>
<td>contre</td>
<td>Sentence-initial Against or I’m against</td>
</tr>
</tbody>
</table>
DiscRel features

22 discourse relations with the number of identified connectives as value

- discourse relations as defined in the Lexconn [Roze et al., 2012]: "a French lexicon of 328 discourse connectives, collected with their syntactic categories and the discourse relations they convey"
- When a connective is polysemious, all possible relations are considered
- alternation; background; commentary; concession; condition; consequence; continuation; contrast; detachment; elaboration; evidence; explanation; flashback; goal; narration; opposition; parallel; rephrasing; result; summary; temporality; unknown relation
ACP parameters

- Considering only discussions with more than 100 words
- Only Interact and DiscRel features are taken into account (normalized on the number of words)
- The other features are just indicated (in blue) for permitting a global overview
Results: ACP on linguistic features

- 5 dimensions that explain around 30% of the total variance
- A first dimension simply related to the number of words (the more words, the more features)
- A second dimension that differs acc. to the unit taken into account (thread or talk page)
- For **threads**, it opposes:
  - Dimension 2+: more *I*, informal *we* (*on*) and discourse relations expressing **contrast**
  - Dimension 2-: agreement cues, formal *you* and discourse relations expressing alternation, consequence, goal and temporal relations
Results: ACP on linguistic features

- 5 dimensions that explain around 30% of the total variance
- A first dimension simply related to the number of words (the more words, the more features)
- A second dimension that differs acc. to the unit taken into account (thread or talk page)
- For talk pages, it opposes:
  - Dimension 2+: more discourse relations expressing *contrast*, background/narration and causality
  - Dimension 2-: politeness cues, formal you and we and discourse relations expressing concession, condition and temporal relations
Results: ACP on linguistic features for talk pages

- Dimension 1: the more words the more features
- Dimension 2+: discourse relations expressing contrast, background/narration and causality
- Dimension 2-: politeness cues, formal you and we and discourse relations expressing concession, condition and temporal relations
Difficulties to go from these results to examples we may interpret

"few politeness cues, formal you and more discourse relations expressing contrast"

- Few politeness cues because few words or no real interaction (one post per thread)
- Potentially only one formal you (perhaps included in a specific locution as "s’il vous plait" please)
- 17 connectives associated with contrast in the LexConn including the two very polysemous "but" and "while"

Example:
Plan

1. Detecting Conflict in French Wikipedia Talk Pages
2. WikiTalk Corpus
3. Classification of "Conflicting" vs. "Neutral" Discussions
4. A Bottom-Up Approach to Talk Page Profiling
5. Conclusion

Conclusion
Conclusion

- CMC genres are complex objects that challenge our traditional methods
- CMC genres require different levels of investigation
- Data mining techniques may give us some leads but...

Qualitative analyses and manual annotation are crucial

- A first annotation on two talk pages shows that only 50% of threads in two *a priori* conflict talk pages are conflict threads
- We must improve features that describe the thread level as for example by looking at the headings, the first post of the section and the context (ex: https://fr.wikipedia.org/wiki/Discussion:Psychanalyse/arch1#choqu.C3.A9)
- A lot of work in progress....


