

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

User Talk pages	1,990,927	57%
Article Talk pages	1,496,553	43%
Redirections	116,432	8%
Empty talk pages (< 2 words)	1,013,791	68%
Selected talk pages	366,326	24%

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="#">`

```

</p>
<cp id="6"> Benjac 15 mai 2011 à 17:17 (CEST)</p>
</post>
- <post id="2" who="Celette" bot="no" when="15-05-2011-17:16" interactionLevel="1">
- <cp id="1">
  1) Aucun recul sur l'affaire, même pas une demi-journée.
  </p>
  <cp id="2"> 2) *est une information considérable* : POV.</p>
- <cp id="3">
  3) Vous avez une 4e contributrice pour vous barrer le chemin, moi. Les arguments sont clairs : aucun recul + l'introduction n'est pas un fourre-tout à polémique du moment. On attend de voir comment évolue l'affaire et on avise. Pour ma part, je suis prête à réverter.
  </p>
  <cp id="4"> Cordialement, Celette 15 mai 2011 à 17:16 (CEST)</p>
</post>
- <post id="3" who="Hatonjan" bot="no" when="15-05-2011-17:19" interactionLevel="2">
  <cp id="1">(conflit d'edit)</p>
  <cp id="2">
    Bonjour, avez vous lu RI ? C'est fort quand même vos remarque, à mon premier revert, mon commentaire vous renvoyait vers cette règle ! Et vous avez maintenu votre position malgré les reverts de 3 personnes que vous avez cité, sans rien apport dans votre obstination. Donc, déjà, ne nous accusez pas de discourtoisie. D'ailleurs le dialogue a été débuté par Lomita sur votre page de discussion.
  </p>
  <cp id="3">
    Sur le fond, comme dit dans RI et Wikipédia:Conventions_de_styleBiographies qui sont les deux règles ici, c'est loin, très loin, d'être un fait majeur "pour le moment". Vous confondez une encyclopédie et un site de nouvelles. De plus, l'importance de cette information a déjà été débattu (et toujours en cours de débat) plus haut, et il a déjà été considéré que l'évènement ne justifie pas une partie en lui même au vu des informations à ce jour, alors dans le résumé, surement pas !
  </p>
  <cp id="4">
    Bref, laissez la justice faire son travail, nous sommes peut être à l'aube d'un évènement majeur dans sa carrière, mais pour le moment, rien n'est fait. Hatonjan 15 mai 2011 à 17:19 (CEST)
  </p>
</post>
- <post id="4" who="Lomita" bot="no" when="15-05-2011-17:24" interactionLevel="3">
  <cp id="1">
    Entièrement d'accord avec mes collègues, je n'ai plus rien à rajouter, à part que ce fait divers est déjà traité dans l'article, à la place qui lui revient - --Lomita 15 mai 2011 à 17:24 (CEST)
  </p>
</post>
- <post id="5" who="Suprememangaka" bot="no" when="15-05-2011-17:27" interactionLevel="4">
  <cp id="1">

```


WikiTalk Corpus Overview – soon available under CC license

talk pages	threads	posts	words
366,326	1,024,351	3,022,240	159,578,279

- 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

 **Cet article ne cite pas suffisamment ses sources** (avril 2012).
Si vous disposez d'ouvrages ou d'articles de référence ou si vous connaissez des sites web de qualité traitant du thème abordé ici, merci de compléter l'article en donnant les **références utiles à sa vérifiabilité** et en les liant à la section « Notes et références » (modifier l'article, comment ajouter mes sources ?).

- 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"

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

Training data : a distinction based on available meta-data

Conflict discussions	Neutral discussions
More than 100 words in the talk page	
From the header talk page : <i>Keep calm</i>	<i>A-class</i>
Existence of a parallel talk page about : article neutrality	article <i>A</i> -ranking
"Talk Type" : neutrality talk	na
2,028 pages	4,569 pages
11 M words	8.8 M words

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]

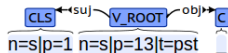


Figure: A delexicalized syntactic bi-arc describing a clitic+verb+conjunction as in the clause 'I find that'.

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

features	threads			posts		
	P	R	F1	P	R	F1
lemmas	0.84	0.60	0.72	0.79	0.69	0.74
bi-arcs	0.55	0.48	0.52	0.63	0.59	0.61
units	46,690			194,289		

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

Questioning

- 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 *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

logNnMots	number of words
nbFils	number of threads*
nbPosts	number of posts
profMax	"interactional depth"
nbContributeurs	number of different participants
nbAnonymes	number of anonymous posts
X.anonymes	% of anonymous posts
nbBots	number of posts written by bots
X.bots	% of posts written by bots
AdQ	"1" if the talk page is linked to a A-class article
polemique	"1" if the talk page has the banner "keep calm"

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

Politeness	<i>thanks, hello, goodbye, hi, sincerely, cheers, please, would you, etc.</i>
Agreement	<i>OK, agree, yes, no, actually, etc.</i>
Question	<i>?</i>
Je	1st singular person pronouns + <i>personally</i>
Tu	2nd sing. pers. pronouns, informal "you"
Vous	2nd plur. pers. and formal "you" pronouns
Nous	1st plur. pers. pronouns
On	Informal "We"
WP	<i>Wikipedia</i> or <i>WP</i>
pour	Sentence-initial <i>For</i> or <i>I'm for</i>
contre	Sentence-initial <i>Against</i> or <i>I'm against</i>

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

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 threads)
- 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 :

https://fr.wikipedia.org/wiki/Discussion:William_Tyndale

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

- 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....



Agarwal, A., Chappelle, O., Dudik, M., and Langford, J. (2011).
A reliable effective terascale linear learning system.
JMLR, 15:1111–1133.



Ferschke, O., Gurevych, I., and Chebotar, Y. (2012).
Behind the article: Recognizing dialog acts in wikipedia talk pages.
In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 777–786. Association for Computational Linguistics.



Ho-Dac, L.-M. and Laippala, V. (2015).
Les discussions wikipedia : un corpus pour caractériser le genre "discussion".
In *International Research Days Social Media and CMC Corpora for the eHumanities*, Rennes, France.



Laippala, V., Kanerva, J., and Ginter, F. (2015).
Syntactic ngrams as kestructures reflecting typical syntactic patterns of corpora in finnish.
Procedia - Social and Behavioral Sciences, 198:233 – 241.



Poudat, C., Vanni, L., and Grabar, N. (2016).
How to explore conflicts in french wikipedia talk pages?
In *JADT*, pages 645–656.



Roze, C., Danlos, L., and Muller, P. (2012).
Lexconn: A french lexicon of discourse connectives.
Discours, 10.



Scott, M., , and Tribble, C. (2006).
Textual Patterns: Key Words and Corpus Analysis in Language Education.
Philadelphia, PA, USA: John Benjamins Publishing Company.



Urieli, A. (2013).
Analyse syntaxique robuste du français : concilier methods syntaxiques et connaissances linguistiques dans l'outil Talismane.
PhD thesis, Université de Toulouse - Jean Jaurès.