




DELIVERABLE D5.3 - EVALUATION REPORT

LIMSI-CNRS

WP5 (T5.3)

	CHIST-ERA	Subproject : WP5 Task : T5.3 Date : August 9, 2016 Page : 2
---	-----------	--

Historique des modifications			
Version	Auteur	Date	Description des modifications
1	Xavier Tannier Patrick Paroubek	May. 22 th 2014	Annotation representation Model
2	Amel Fraisse	Jul. 20 th 2014	Lexical Ressources Evaluation
3	Patrick Paroubek	Jul. 22 th 2014	Evaluation Infrastructure
4	Amel Fraisse	Oct. 06 th 2014	Expert Annotation Guidelines French v1
5	Patrick Paroubek	Nov. 06 th 2014	State of the Art on Crowdsourcing
6	Patrick Paroubek Lucie Gianola	Fév. 25 th 2016	Expert Annotation Guideline French v2
7	Patrick Paroubek	Mar. 31 th 2016	Expert Annotation Guideline German v1
8	Patrick Paroubek Cyril Grouin Thierry Hamon Pierre Zweigenbaum	June. 10 th 2016	Evaluation Campaigns
9	Patrick Paroubek Marie-Anne Moreau (INALCO)	Juil. 22 th 2016	Expert Annotation Guideline German v2


Validation			
Role	Organisation	Name	Date

Table of Content

1	Introduction	4
1.1	Aim	4
1.2	Responsibility	4
2	Affect Lexica	4
2.1	Aim	4
2.2	State of the art	4
2.2.1	Semantics	5
2.2.2	Opinion Mining & Sentiment Analysis	9
2.2.3	Named Entities	10
2.2.4	Machine Translation	11
2.2.5	Information Retrieval/Extraction & Question Answering	13
2.2.6	Language Communication Enhancement/Validation	14
2.2.7	Real Time Dialog System	15
2.2.8	Conclusion	16
2.3	7 language lexica Experiment	16
2.3.1	The HC experiment with CrowdFlower	17
3	Evaluation Campaign	21
3.1	Aim	21
3.2	Data	21
3.3	A model of language data annotation	22
3.3.1	Annotation representation	22
3.4	Applying our model to real evaluations	24
3.4.1	Classification	25
3.4.2	Transduction	29
3.5	Expert Annotation Guidelines	30
3.5.1	Guideline for French	30
3.5.2	Guideline for German	41
3.6	Expert Annotation Software	53
3.7	Evaluation Campaign Deployment	53
3.7.1	French Language Sentiment Analysis Evaluation Campaign	54

Index of Figures

Index of tables

	CHIST-ERA	Subproject : WP5 Task : T5.3 Date : August 9, 2016 Page : 4
---	-----------	--

1 Introduction

1.1 Aim

This document describes the design and execution of experiments for assessing HC effectiveness. It constitutes the first version of the deliverable planned for the end of the project, as such it focuses on experiments design and present only results for what has been achieved so far (T5.1 and part of T5.2). The whole set of evaluation results, including the results of the evaluation campaign (T 5.3) planned to happen during the last year of uComp will be presented in the final devliverable, due for M30.

1.2 Responsibility

The partner responsible the production of this document is LIMSI-CNRS.

2 Affect Lexica

2.1 Aim

The task for *HC-Based Knowledge Resource Evaluation* focuses on the evaluation of the HC-sourced affect lexicons both in terms of improving the recall, precision and F-measure of existing sentiment and opinion mining algorithms ([Gindl et al., 2010] and [Pak et al., 2014]), as well as by using a set of additional games aimed at evaluating specific aspects of these resources.

We plan to assess the interest of replacing/complementing entropy-based data browsing algorithms by/with human guidance. This will help evaluate the effectiveness of the HC paradigm and framework, and address the following questions:


- *How do crowdsourced resources compare to those created manually, automatically or through mechanised labour (in terms of both quantity and quality, which will be measured by the k-value of inter-annotator agreement)?*
- *Is the cost/benefit ratio of this process better than that of other knowledge acquisition approaches?*

Analogous to WP4, T5.2 will primarily focus on the climate change domain and the repository built in T1.5.

2.2 State of the art

Since it was created in 2005 according to [Safire, 2009], the word *crowdsourcing* found rapidly¹ its way into the research community, where people tried from the beginning to assess the benefits and gains they could get from using the power of human computation. Most of the early

¹in 2009 it reached the 1 million hit limit in Google ([Safire, 2009])


	<p>CHIST-ERA</p>	<p>Subproject : WP5 Task : T5.3 Date : August 9, 2016 Page : 5</p>
---	------------------	---

studies reported in the literature mention Amazon Mechanical Turk (AMT), with CrowdFlower appearing a little later. Our survey of previous work is organized along the following broad themes:

- Semantics,
- Opinion Mining & Sentiment Analysis,
- Named Entities,
- Machine Translation,
- Information Retrieval/Extraction & Question Answering,
- Language Communication Enhancement/Validation,
- and Real Time Dialog Systems.

2.2.1 Semantics

In [Rumshisky et al., 2009], the authors report on an experiment to build a fuzzy sense inventory database for a set of polysemous verbs of medium difficulty, using non-expert annotators which were presented with sentences containing one of the verb form and had identify particular meaning. The results were then evaluated against the groupings created by a professional lexicographer (set-matching F-score of 0.93), each sentence being annotated by 5 AMT annotators. On average 1 minute was required to process 10 sentences and the amount paid to the annotator was 0.03 USD. The total sum spent during the first experiment was 10\$. Later, [Rumshisky, 2011] conclude that clustering the 350 concordance lines into sense related groups would yield only reliable results for about 140 concordance lines and [Rumshisky et al., 2012] report that one of the major issue is to attract high-quality annotators on a service like AMT to perform complex linguistic tasks. If the use of best practice guidelines helps, it does not suppresses the need to run preliminary experiments to calibrate the task parameters and interface. Another quality improvement in the result can be obtained by comparing AMT worker results using worker-quality weighted majority votes. However, while writing about the 2009 experiment [Rumshisky et al., 2012] report that “*the experiments run with the same parameters today do not lead to either fast completion or quality annotation.*”. The authors found they could get an important increase of results quality by restricting the location of the annotators to the USA, but the task took longer to complete and costed more. Factors that attract good annotators in large numbers are the pay rate (which can varied at that time from a few cents to 20 \$ per task), the height of the task in the task search space proposed to annotators and the apparent simplicity of the task and the clarity of the task description, as well as the standing of the image of the task proposer in the AMT annotators specific social media which depends on the rapidity of the payment and speed of answering AMT annotators’ questions. For instance [Rumshisky et al., 2012] report that in the US local restricted prototype experiments, the task failed to complete when only 0.01 \$ were offered to provide five judgments and the location of the tasks was very low in the task list because not much data was proposed for

	<p>CHIST-ERA</p>	<p>Subproject : WP5 Task : T5.3 Date : August 9, 2016 Page : 6</p>
---	------------------	---

annotation. Keeping the in US requirement, the best strategy that the authors found was to pay 3 cents for 10 judgments in one task, allowing a maximum worker error rate of 27%, with at least a 85% previous task approval rate on at least 200 tasks, and offering bonuses to the top workers. They obtained an F-score of 0.72 with a kappa of 0.69. The differences observed by the authors between the 2009 and 2012 experiments show that the AMT market place is evolving rapidly and that the validity of best practice guidelines needs to be checked frequently.


In [Munro et al., 2010], the authors investigate the use of AMT for a range of linguistic experiments from semantics to psycholinguistics dealing with:

1. verbs semantics,
2. segmentation of audio speech stream,
3. language models,
4. speech grammaticality,
5. thematic role,
6. methapor brain processing,
7. and reading attentiveness.

For instance they report between crowdsourced and laboratory condition kappa values of 0.9 for the verb semantics task and of 0.759 for the language model task. Following their experimental results, they conclude that crowdsourcing provides the means to run systematic, large-scale judgment studies at a lower cost and much more easily than when doing them under laboratory conditions.

The experiment with AMT presented in [Negri et al., 2011] addresses sentence modification and textual entailment annotation in a multilingual context involving English, Italian and German. From a set of aligned sentences in the three languages, modifications are done in a monolingual set-up (English) by paraphrasing and rephrasing with information addition or deletion. Modified sentences are then translated after having been annotated with entailment information with respect to the source sentence. The process of multilingual entailment is then simplified by having entailment processing done in a monolingual set-up. For quality control, only the annotation that gathered the agreement of 4 out of 5 annotators were retained and gold standard data test were run. Six types of tasks were submitted to the AMT annotators:

1. paraphrasing,
2. grammaticallity checking,
3. bidirectionnal entailment,

	<p style="text-align: center;">CHIST-ERA</p>	<p style="text-align: right;">Subproject : WP5 Task : T5.3 Date : August 9, 2016 Page : 7</p>
---	--	---

4. rewriting while adding information,
5. rewriting while removing information,
6. unidirectional entailment,

for a total of 1,620 pairs of sentences added to the corpus, while 721 were discarded, which required 22 days and 11 hours of AMT work for a cost of 435.71 \$. Including the work of the expert to prepare the gold data and manage the pipeline, the overall operation was a success and proved the usability and efficiency of crowdsourcing for building the multilingual entailment corpus.


Because of language variability, rule based semantic processing requires solving a large number of small inference problems, like for instance knowing that if “*X work as Y*” then it also means that “*X was hired as Y*”. But the evaluation of such rules is problematic because no reference dataset exists and extrinsic evaluation in the context of an applicative task will not necessarily produce evaluation of the rules since it is difficult to assess to which cause the application performance measured must be imputed to. So [Zeichner et al., 2012] experimented using crowdsourcing (CrowdFlower) to perform inference rule evaluation. To assess a particular rule application, one must answer 3 questions:

1. Is the left-hand side of the rule meaningful?
2. Is the right-hand side of the rule meaningful?
3. If yes was to answer to both previous questions, does the entailment holds?

There were 2 types of cascading tasks submitted to the annotators:

1. to appreciate the rule relevance, which corresponds to the first two previous questions
2. and to judge the entailment.


Depending on the answer produced by a task from the first type, the annotators working on the second type of task has either to validate the entailment or to ascribe the identified non-relevance of the rule to an erroneous answer provided for one the three original evaluation questions. Four inference algorithms were tested on a database of predicates extracted from ClueWeb09 web crawl², where each extraction comprises a predicate and two arguments, providing four datasets. For each 5,000 extraction were sampled and for each dataset four rules common to all datasets were extracted which resulted in 20,000 rule applications, out of which 10,443 were discarded due to low CrowdFlower annotators confidence, a further 1,281 were flagged as meaningless left-hand side applications and another 1,012 as meaningless right-hand side applications. Out of the remaining 8,264. rule applications that were passed on to the second type of task (entailment assessment), 5,555 were judged with a confidence high enough, 2,445 with positive entailment and 3,108 with negative entailment. In the experiment a total

	<p style="text-align: center;">CHIST-ERA</p>	<p style="text-align: right;">Subproject : WP5 Task : T5.3 Date : August 9, 2016 Page : 8</p>
---	--	---

of 6,567 application rules were annotated for a total cost of 1,000 \$. Note that the second type of task required specific experimentation from the authors with the annotators in order to calibrate the communication of the task in order to improve the kappa performance. The experiment was considered a success by the authors.

Statistical Modality Tagging was investigated by [Prabhakaran et al., 2012] Modality can be defined as a quantification over possible worlds or a speaker attitudes with respect to a proposition. Acquiring training data for building an automatic modality tagger is a difficult problem. In a pilot study, the author obtained and ran the modality tagger described in [Baker et al., 2010] on the English side of the Urdu-English LDC language pack. Using AMT, they estimated the precision of this type of approach at around 60%. They posted on AMT a set of randomly selected sentences (1997) that the tagger had labeled as not having the Want modality. Each sentence was checked by 3 annotators to decide whether it had the Want modality or not. Using majority rules on the annotations, 95 (4.76%) of the sentences were marked as validated Another set of 1993 sentences annotated by the tagger as not having the Want modality were posted on AMT, out of which 1238 were validated by the annotators. Hence, the authors decided to apply a simple tagger as a first pass, with positive examples subsequently hand-annotated through AMT. The simple tagger used a word spotting approach with a set of trigger words like “try”, “plan”, “aim”, “wish”, “want” etc. In the corpus, the number of sentences for each modality was limited to 50 for each trigger word in order to preserve linguistic variability. The AMT annotators were asked to check that the modality was not present in the sentence, otherwise they had to highlight the target of the modality. Each sentence was annotated by 3 persons. Only the output produced by adult annotator with an approval rating above 95% were considered if they had completed at least 50 tasks. They were paid 0.10 \$ for each set of ten sentences. Only the annotation which gathered the approval of at least 2 annotators out of 3 on the modality and target were kept. Out of the resulting 1,008 examples, 674 had 2 annotators agreeing while 334 collected unanimous agreement. This work proved that it is possible to combine a high-recall simple tagger with crowdsourcing annotations to produce training data for a modality tagging.


[Lafourcade and Fort, 2014] propose to use the paradigm of GWAP (Game With A Purpose) to build lists of semantically-related terms, for instance needed to to deploy parental control systems on Internet. At the heart of the system is a GWAP that is used to build a semantic network where players collect credits when they provide answers similar to other players, with higher rewards for original contributions. The game played involve lexico-semantic relations like is-a, hyponym, characteristic, location, agent, patient, etc. In their paper the authors propose an algorithm to exploit the resulting semantic network to deploy a filtering service. The network produced for the oldest game on French contained in October 2013 approximately 300,000 terms, including 15,000 to 20,000 word usages and more than 6 million relations. The main game has been played more than 1.3 million times by more than 3,500 registered players. The experiment demonstrated the possibility to build and dynamically update a semantic network through crowd sourcing with a GWAP interface.

	<p style="text-align: center;">CHIST-ERA</p>	<p style="text-align: right;">Subproject : WP5 Task : T5.3 Date : August 9, 2016 Page : 9</p>
---	--	---

2.2.2 Opinion Mining & Sentiment Analysis

Also from 2009, the work of [Hsueh et al., 2009] report on the use of AMT to classify text snippets extracted from political blogs, first according to the the political candidate they concern and the polarity they express (Positive, Negative, Both, or Neutral) and second, according to whether they support or oppose the political candidate they are related to. The measured agreement among three expert annotator on the relevance to a particular candidate was 77.8% while overall agreement on the four-classes sentiment annotation was 70.4%. For the second classification step (support/oppose/neutral) the author report an agreement of 76.8%. In comparison, each snippet of a set of 100, was marked by 5 AMT annotators taken from a group of 25 annotators selected on the basis of an approval rate higher than 95% and paid 0.04 \$ per annotation. Each snippet required on average 40s to be completed and the lower overall agreement measured on all four-class sentiment task was 35.3%. On the second classification subtask (support/oppose/neutral) they achieved a value of 47.2% of agreement, but they nevertheless managed to reach a relatively good level of agreement on the simpler tasks of determining whether a snippet is relevant to a particular political candidate (81.0%), whether the snippet is subjective or not (81.8%) and whether the snippet is positive versus negative (61.9%). But there exists a group of annotators that produces more noisy annotations than the other, judging against annotations produced by a majority vote, 20% of the AMT annotators have a noise level that exceeds 60% which disagree in 70% of the cases with the result of majority. The more the text snippet is ambiguous the lower is the agreement. In this paper 3 quality measures were found to be useful for selecting annotations: the noise level of annotator, the inherent ambiguity of the class labels and the informativeness of the annotated data.

A fine grained annotation task identifying word expressing a sentiment about a particular in-sentence target was experimented by [Sayeed et al., 2011] with CrowdFlower. This task is in a simpler form (no identification of text spans required by the annotators), the task that will be used in the uComp evaluation campaign. The annotators had to classify as POSITIVE, NEGATIVE or NONE the preselected opinion words depending on their relation with a preselected target word located in the same sentence. The authors used the same techniques as [Hsueh et al., 2009] for discarding noisy annotations. The experiment had 200 tasks paid 0.04 \$ per task, with three different annotators performing each task, for a total cost 60 \$ and a time span of 24 hours to complete the job. In addition, 30 tasks were used to define a gold standard which served to identify unreliable annotators (with less than 65% accuracy). This gold standard produced 117 words annotated as NONE, 35 as POSITIVE and 17 as NEGATIVE. Aggregation of the information produced by the annotators was done by majority voting (agreement above 50%) at a word level. There were 155 words with a majority consensus was reached (>50%). The authors of the paper determined 48 to have a particular opinion weight (POSITIVE or NEGATIVE). Only 22 annotators passed the Crowdfower quality control. Removing unreliable annotators based on the gold standard test had a remarkable effect on the F-measure and kappa values. the best kappa measure (0.65) was achieved when the 7 worse ranked annotators had been dropped. But highest precision (0.85) and accuracy (0.88) were

	CHIST-ERA	Subproject : WP5 Task : T5.3 Date : August 9, 2016 Page : 10
---	-----------	---


achieved after dropping the 10 worse annotators, note that in that case the kappa dropped to 0.54, because when going over 80% limit, performance start to decrease if more annotators were dropped because they were not numerous enough anymore for the voting mechanism to have a smoothing effect. This experiment showed that fine grained annotation is possible with crowdsourcing but requires extensive quality control procedures.

2.2.3 Named Entities

CrowdFlower is mentioned along AMT in [Finin et al., 2010], which describes an experiment about the annotation of Named Entities in Twitter (person, organization, location, or none) and the collection of judgements on the quality of “word clouds“ for semantics and sentiment representation. The dataset was split into tasks containing 4 previously unlabeled tweets and 1 previously labeled tweet. On AMT, 251 tasks were submitted, each to be completed twice for an overall duration of 15 hours to complete the whole set of tasks at a total cost 27.61 \$, which corresponds to 0.0275 \$ per tweet. There were 42 AMT annotators mostly from the US and India and some from Australia. Most annotators performed a single task and one annotator did most of them. Inter-annotator agreement was checked using an algorithm akin to Google’s PageRank ([Page et al., 1999]). The effectiveness of AMT annotators was judged inferior the to the one of the expert annotators but it was possible nevertheless to achieve the same results at a lower costs by by carefully combining annotations. The CrowdFlower experiment involved 30 tweets and each task had 3 tweets for a price of 0.05 \$ to be done in 30s and representing a total cost of 2.19 \$ (overhead included), and a rough pay of 2 \$ per hour per annotator. The cloud comparison experiment was run on AMT, a task consisting in deciding which of the two word cloud presented to the annotator describes best the query topic. After selection with an average accuracy rating of at least 0.75 % on 7 questions, there were 8 AMT annotators selected and they did achieved a performance level of 61% of accuracy against gold data. The authors found particularly helpful the extra functionalities for managing the task and validating the annotators work provided by CrowdFlower over those of AMT.

[Sayeed et al., 2010] deploy crowdsourcing tasks to evaluate name entity recognition algorithms. They use AMT to assess the performance of an algorithm that identifies names of persons and organizations (ENAMEX NIST ACE-standard). For each task, each annotator was paid 0.05 \$, summing up to 150.00 \$ for 3 annotators. It took more than an estimated two person weeks to complete the work. They showed that crowdsourcing can provide reliable results and provide simple means of verifying algorithm performances, in a context where the aim is to reduce the rate of false positives.

The authors of [Higgins et al., 2010] describe how they used AMT to collect Arabic nicknames for completing exiting Named Entity lexicons. In addition, they experimented the effect that increasing the pay rate had on taks completion speed. On average, a pay of 0.03 \$ per task yielded 9.8 names per day, increasing the pay to 0.05 \$ made the number of collected names up

	CHIST-ERA	Subproject : WP5 Task : T5.3 Date : August 9, 2016 Page : 11
---	-----------	---

to 25 names per day and to reach the value of 100 names per day we had to go up to 0.25 \$ per task. In another experiment they increased the price from 0.01 \$ per task to 0.05 \$, of course increasing the price shortened the collection time, but surprisingly enough, the publication of the higher paying task had also an accelerating effect on the lower paying task.

2.2.4 Machine Translation


Translation from Urdu to English was the theme of the AMT experiment reported in [Zaidan and Callison-Burch, 2011], the price paid was 0.10 \$ per translation (approx. 0.005 \$ per word). Four translations for each sentence were collected from different translators and a set of AMT task was created to cross validate the translations by other AMT annotators (ranking four translations at a time from best to worse or postediting). The task were paid on a basis of 0.10 \$ for translating a sentence and 0.25 \$ to edit a set of ten sentences, and 0.06 \$ to rank a set of four translations. The overall costs were:

- translation cost: 716.80 \$,
- editing cost: 447.50 \$,
- Ranking cost: 134.40 \$.

Adding Amazon's 10% fee resulted in a total less than 1,500 \$ to more than 7,000 translations produce around 17,000 edited translations and rank 35,000 labels (since each ranking task involved judging 16 translations, in groups of four). Including the cost of the professional translation reference would add 1,000 \$ to the total cost which now amounts to 2,500 \$. The participation of the AMT translators was 52 for translation (138 sentences on average), 320 for editing (56 sentence on average) and 245 for ranking 245 Turkers (averaging 9.1 ranking task each, or 146 rank labels). The authors investigated cost reduction by eliminating the need for professional translation and decreasing the amount of edited translations. The first measure produced a significant drop of quality (BLEU score of 34.86), while the second measure greatly reduced the cost but managed to maintain good performance (BLEU score of 38.67). Then, the best strategy for translation seems to be

- for each source sentence produce several translations,
- rank the multiple translation,
- edit only the top ranked translations.

The work of [Hu et al., 2011] does not make use of AMT or CrowdFlower but is nevertheless interesting as it deploys two crowds of translators in a setup comparable with this two collaborative infrastructures. The experiment combined machine translation with human computation using two crowds of monolingual source (Haitian Creole) and target (English) speakers, for the WMT 2011 Haitian Creole to English translation task. The result showed that the combined approach translated 38% of the sentences well compared to Google Translate's 25%. The 4

	<p>CHIST-ERA</p>	<p>Subproject : WP5 Task : T5.3 Date : August 9, 2016 Page : 12</p>
---	------------------	--

Haitian Creole speakers were recruited from Haiti and did not speak English while the 26 English speakers were for 6 of them paid UMD undergraduates while the other 21 were volunteers of various background. Over 13 days, the total effort was 15 hours for Creole and 29 hours for English. The Haitian Creole sentence was first automatically translated into English and presented to the English speakers who could take one of the following actions:


1. mark a phrase in the candidate as an error,
2. suggest a new translation candidate translation,
3. change the ranking of the candidate translation.

New translation candidates were back translated into Haitian Creole and along with spans marked as errors which were projected back to identify the corresponding spans in the source sentence by means of word alignment. In turn, the Haitian Creole speakers could:

1. Rephrase the entire source sentence,
2. explain spans marked as errors,
3. basing their action on the back translation, change the ranking of the candidate translation.

Source speakers could document error spans either by rephrasing, either by annotating the spans with images or Web links (e.g. Wikipedia). The process was asynchronous for participants from both locations and the voting based best translation could be extracted at anytime.

[Post et al., 2012] applied AMT to the building of a collection of parallel corpora between English and six languages from the Indian subcontinent, low-resourced and under-studied which are difficult form machine translation: Bengali, Hindi, Malayalam, Tamil, Telugu, and Urdu. The source documents in English were the top-100 most viewed wikipedia pages for from each language. Since the authors were not proefficient in all the six Indian languages, the decided to appreciate the quality of the AMT translator by comparing their production with the lexical translation of the original sentence by means of bilingual dictionaries bootstrapped for each language by means of another set of AMT tasks. In this other set of disctionary tasks, AMT translators had single words or very short phrases to translate while the validation was done with the wikipedia page titles whose translation can be assumed to be found in Wikipedia following the cross-lingual links. In the translation work, subsequent sentences from the original text were grouped by 10 to provide an AMT task, and translators had to provide a free-form translation. They were paid 0.70 \$ per task. Decision to accept or reject the translation was done manually, by checking various factors like comparison to a lexical translation obtaine with the dictionaries developped for the project,the percentage of empty translations, the amount of time that the translator required to complete the task, his geographic location (self reported and identified through is IP address) and by measuring the distances between the various translation of the same source excerpt. Malayalam provided the highest through- put, generating half a million

	CHIST-ERA	Subproject : WP5 Task : T5.3 Date : August 9, 2016 Page : 13
---	-----------	---


words in just under a week. Some authors assess the cost of professional translation from Tamil to English at 0.30 \$ per word while the cost measured in our experiment was less than 0.01 \$ per word. Because no professional translation was available to the authors they could not use direct performance measures like BLEU to assess the variation of quality of the translations. Nevertheless, they designed another task in which original sentences were displayed along with four of its translations to another set of AMT translators for ranking the best. Each task was performed by 5 translators. Approximately 65% of the sentences had five votes cast on just 1 or 2 translations while 95% of the sentences had all the votes attributed to 1 to 3 sentences. This suggests that differences in translation quality existed but also that the translators took their did their assesment job seriously enough to report on the differences.

2.2.5 Information Retrieval/Extraction & Question Answering

The study of [Grady and Lease, 2010] is about human factor parametrization for a crowdsourcing tasks of relevance judgement in information retrieval. The author investigated, on AMT, the impact that the following four parameters had on the cost, time, and accuracy of the assessments:

1. providing the annotator with only a title for a query versus a detailed description,
2. different wording for the task title (specialized, i.e. “binary relevance judgement”, versus layman, i.e. yes/no),
3. the amount paid per task (0.01 \$ versus 0.02 \$),
4. and a bonus: (0 versus 0.02 \$).

Assessment were done with document from the TREC TIPSTER collection of news articles (using gold standard data enabled easy computation of AMR annotators accuracy). Five batch evaluations were done, for each of four topics, five documents were assessed and 10 assessments were collected for each document. A total of 200 tasks were submitted to AMT for each batch, resulting in 1000 tasks for the five batches together. The length of the documents was between 162 words and 2129 words (including HTML tags and single-character tokens). For each task, the annotator had to perform a single binary relevance judgment linking a query and a document. There were 149 annotators who participated, some of them to the 5 batches. In batches 2 and 3 one variable was modified with respect to batch 1, while in Batches 4 and 5 it was against batch 3 configuration that one variable was modified. In batch 5, 23 bonuses were given for a total cost of 0.46 \$. Statistical significance was measured via a twotailed unpaired t-test. The only significant outcomes observed were increase in comment length and number of comments for higher-paying or bonus batches. The highest accuracy, 70.5% was reached with batch 3, which used a title query and a simple yes/no response. The use of description query did not entailed an accurary improvement. The fastest task completion, 72s, was measured for batch 4 while the average time completion per task across the five batches was 63s. The number of unique annotators per batch varied between 64 to 72 for batches 1 to 4 but fell to 38 for batch 5, probably because of the bonus incentive, annotators in this last batch tended to complete

	<p>CHIST-ERA</p>	<p>Subproject : WP5 Task : T5.3 Date : August 9, 2016 Page : 14</p>
---	------------------	--


more tasks better accuracy (3.37 documents correctly annotated as compared to 2.10-2.20 for batches 1 to 3 and 1.85 for batch 4). But bonus attribution requires expert human supervision and the question remains of the gain versus expert time spent for the operation to become beneficial. From the words of the authors, this experiment provided largely inconclusive results.

The following work was realized with the help of a web application game to collect crowd-sourced input. After qualitatively examining how humans perform incremental classification [Boyd-Graber et al., 2012] have shown in their article how crowdsourced knowledge of a human’s incremental classification process improves state-of-the-art rapacious classification. Then they built a Bayesian models that embedded in a Markov decision process to replicate the improved classification process and develop new hierar- chical models combining local and thematic content to better capture the underlying content. The corpus used was made of 37,225 quiz bowl questions with 25,498 distinct labels from 121 tournaments (between 1999 and 2010). The authors created a web application to simulates the experience of playing quiz bowl where text is incrementally revealed until the user decides to answer. The answer is judged with a string matching algorihtm. More than 7000 questions were answered in the first day, and over 43000 questions were answered in two weeks by 461 users.

2.2.6 Language Communication Enhancement/Validation

It is to palliate the lack of common reference corpus for evaluating grammatical error detection that [Madnani et al., 2011] decided to use crowdsourcing. For their experiment, the authors studied with AMT the presence of extraneous preposition in a corpus of students writing for a test of English as a foregin language. In this experiment 75 sentences were used as gold standard built by 3 experts and the remaining 923 sentences were annotated by 20 annotators located in the USA, whithin one day. Using 3 annotators per judgement with a majority vote yields an agreement with any one of annotator of 0.87 on average which corresponds to a kappa of 0.76. The extraneous preposition annotation costed 325 \$. Further experiments were done with a gold standard of 20 sentences obtained with CrowdFlower, in particular to propose new evaluation measures, derived from precision and recall by weighting the evaluation of annotations items depending on the proportion of agreement for this item by the annotators. The authors found out that the weighted measures are more stable and contrary to regular precision and recall they display less a tendency to overestimate the performance of the system under evaluation. Similarly, for comparing two systems they propose to replace precision-agreement estimation by a kappa-agreement measure

Augmented and Alternative Communication (AAC) devices enable users with communi- cation disabilities to participate in everyday conversations. For designing such devices, the elaboration of language models representing as best as possible the style of the users’ intended communications is essential. Since collection of geniune AAC material for designing such device is quite difficult, the authors of [Vertanen and Kristensson, 2011] decided to use crowdsourcing

	CHIST-ERA	Subproject : WP5 Task : T5.3 Date : August 9, 2016 Page : 15
---	-----------	---

to create a corpus of fictional AAC messages. In their paper, they show that the messages they produced a better model of AAC communication than the material generally used so far from telephone conversations or newswire text. Two types of tasks were proposed:

1. the AMT workers tried to imagine how he would used an AAC device to communicate if he was subject to a communication impairment,
2. AMT annotators were asked to assess the plausibility of communications produced by the previous type of task.

Task of the second type gathered more easily AMT participants than tasks of the first type. The average task completion time was shorter (24s on average) for the second type of task, as opposed to an average of 36s for the tasks of the first type. During the experiment, the pay for the the second type of task was reduced from 0.04 \$ (the price paid for both types of tasks) to 0.02 \$ without loss of performance. For comparisons the authors trained language model with material of different origins (Wikipedia discussions, Usenet, Switchboard, newspapers, Twitter...) and compared the performance with models trained on the crowdsourced data. Compared to a model trained only on Switchboard, their best model reduced perplexity by 60-82% relative on three AAC-like test sets, which represents a potential keystroke saving of 5-11% on a predictive keyboard interface.

2.2.7 Real Time Dialog System

The work of [Bessho et al., 2012] addresses the creation of utterance/reply pairs for Japanese dialog system design. Here the wisdom of the crowd is not provided by a crowdsourcing framework but by Twitter. For each user input, the system will extract from the utterance-pair database the pair for which the tweet is most similar to the input part of the pair and the system response is provided by the output part of the pair. The utterance pair database was built using a corpus of 1.2 million utterance-pairs from Twitter which were written in Japanese, contemporary, and had a in-reply-to field. The author propose to integrate into the dialog system a "real-time crowdsourcing" functionality to handle the cases where the system cannot provide an adequate answer (similarity distance is below a certain threshold between processed utterance an input parts of the utterance pairs stored in the database). The original user input is recast into a tweet from the dialog system chatbot and if a crowd member responds before a certain delay the crowd answer is used as reply by the system. For evaluation, 90 user input examples were selected and 20 utterance-pairs were extracted from the database retrieved from Twitter for each per user input, totaling to 1,800 of triples (user input and utterance pair). Thirty subjects evaluated naturalness and versatility of the responses (600 triples each). Various scoring functions were investigated by means of a ROC curve representation (true positives versus false positives). The area under the curve (AUC) was used to measure the classifiers performance. A random classifier has an AUC of 0.5, and ideal classifier has an AUC of 1.0. The scoring function selected was the one with the best performance, here an AUC of 0.803.

2.2.8 Conclusion

The conclusion that we draw from this survey is that crowdsourcing or similar approaches like collaborative ones mediated by Internet or GWAPs ([Lafourcade and Fort, 2014]) offer the possibility to develop language resources for which human language processing functionality is required at a cost much lower than with classical means of production, for various kind of tasks, in particular task like semantics or translation. This is so because the people employed in this kind of infrastructure are recruited in large numbers, potentially from all over the planet and do not need to have a particular expertise except of being proefficient in (most often) one or several languages. If the number can compensate the quality of the individual annotations by cross-validation, the measure that a task proposer has to deploy to ensure a minimum of quality require either gold standard data in sufficient number or the involvement of an expert, in addition to designing specific procedures to prevent cheating. Despite the extra cost incurred by the measures deployed to ensure quality of the resource produced by crowdsourcing, if the feasibility study has been properly done the resulting cost is much lower than with traditional means. But crowdsourcing infrastructure evolve rapidly and crowds of participants adapts rapidly to new tasks which means that quality measure need to be revised frequently.


There is a question that is rarely raised in the litterature, nevertheless addressed by [Fort et al., 2014] and also the following presentations available on the Web [Lease, 2013a], [Lease, 2013b], [Larson, 2013], it is the question of ethics about the minimal return pay somebody working full time in a crowdsourcing infrastructure is able to achieve. As a corollary, one cannot but consider the question that the new infrastructures for crowdsourcing raises about social tax evasion.

2.3 7 language lexica Experiment

The first part of Deliverable D5.2 consists in the 7 language lexica extracted using as baseline the pointwise mutual information (PMI) measure ([Manning and Schütze, 2002]) collected using climate change linguistic markers obtained by machine translation from the set of patterns developed for English, German and French [Scharl, 2014].

D5.2 V1# of entries	D5.2 V2	Language
12,137	0	en
9,394	0	es
6,129	0	it
5,606	0	pt
5,294	9,930	fr
4,336	0	de
760	0	ru

Table 13: Number of entries for the 7 languages in first version of D5.2

	<p style="text-align: center;">CHIST-ERA</p>	<p style="text-align: right;">Subproject : WP5 Task : T5.3 Date : August 9, 2016 Page : 17</p>
---	--	--

Since the amount of data was too small, a second corpus extraction was done with a wider source set (less focused on institutional sources), initially on French only, which resulted in a 7,000 Tweets corpus, out of which a 9,930 entry lexicon was extracted using PMI, then further lemmatized by alignment with the content of the LEFF lexicon [Sagot, 2010].

2.3.1 The HC experiment with CrowdFlower

Concerning the comparison between automatic extraction of lexica and HC computations, there is an ongoing experiment running on Crowdflower with the French lexicon as input. The parametrization of the experiment is as follows:

- Each annotation task consists in a set of 8 lexical entries (units in Crowdflower terminology, note that the number of units per task advised by Crowdflower was 5) for which the annotator must provide answers to two questions:
 1. Is the term related to an expression of opinion/sentiment/emotion (yes/no)?
 2. Which among the following 20 categories ([Fraisie and Paroubek, 2013]) of opinion/sentiment/emotion is the most appropriate to describe the meaning of the term?
- the price set of the completion of a task is 0.07 €(not that the price advised by Crowdflower was 0.10€),
- the total cost of the experiment is 200 €
- the experiment was started on Monday, July 21th 2014,

The experiment was submitted directly to Crowdflower, we did not test the GATE API yet. Although the task is not yet completed, we already have gained a useful information concerning the potential use of services like Crowdflower with respect to difficulty to provide a task description for relatively complex classification schemes, since the amount of information that one can provide to the *taskers* is relatively limited.

2.3.1.1 Job design

As described in Figure 1 data consists of 9939 terms.

As shown un Figure 2 unit consists in the 4 questions:

- This term may be used to express opinion, sentiment or emotion ?
- This term may be a trigger of an opinion, sentiment or emotion ?
- What is the connotation of this term ?
- what type of opinion, sentiment or emotion does this term express ?

Unit ID	State	Judgments	Agreement	scratchage
506076166	Golden	423		indignation
506076189	Judgable	0		rationaliser
506076195	Judgable	1		ignorance
506076196	Judgable	0		artificieux
506076212	Judgable	0		implacable
506076221	Judgable	0		désserter
506076222	Judgable	0		agressive
506076231	Judgable	0		givrure
506076232	Judgable	0		crêpage
506076233	Judgable	0		réembarcation
506076234	Judgable	0		pas-de-porte
506076235	Finalized	4	0.875	polychète
506076236	Judgable	4		polyviser
506076237	Finalized	5	0.55	au mieux
506076238	Finalized	5	0.9	cartouchière
506076239	Judgable	1		désindemniser
506076240	Judgable	1		dualisation
506076241	Judgable	0		guiper
506076242	Judgable	0		napoléon
506076243	Judgable	0		esquintant
506076244	Judgable	0		caviarder
506076245	Judgable	0		héliciculteur
506076246	Finalized	4	0.9375	talonnière
506076247	Finalized	5	0.85	retour
506076248	Finalized	5	0.6	criminalisation

Page 1 of 398 | Displaying units 1 - 25 of 9939

Figure 1: Data sample

2.3.1.2 Quality management

Concerning the quality control, we defined one test question (Figure 3). However, in order to have as many contributors as possible, we selected the *performance level 1* (the lowest level proposed by CrowdFlower) for French (Figure ??). For the CrowdFlower job, we put 10 units per task that paid 0.06 € (Figure ??) for a task to be completed within an expiration delay of 30 minutes.

2.3.1.3 Results

The job was running for 3 months and as shown in Figure 4, only 9.1% of the job was completed. This results can be explained first by the complexity of the task. In fact the fine-grained classification task is more complicate than a simple binary classification task and this is true even if the contributors are native speakers. The second raison, for this low score is the payment of the task. In fact, contributors are less motivated when the task is underpaid (0.06 € instead of 0.10 € that was recommended by CrowdFlower).

In the Figure 5, each bar represents a contributor and the number of judgments they have submitted for this job. Contributors who have a low trust score and have submitted a significantly larger amount of judgments than other contributors are likely scammers. So, based on

ignorance

Ce terme peut-il être utilisé pour exprimer une opinion, un sentiment ou une émotion ?

Oui
 Non

Ce terme peut-il déclencher une opinion, un sentiment ou une émotion ?

Oui
 Non

Ce terme a-t-il une connotation

Positive
 Negative
 Neutre

Quelle est l'opinion, le sentiment ou l'émotion que ce terme peut exprimer ou déclencher ?

AMOUR
 APPAISEMENT
 DERANGEMENT
 ENNUI
 PLAISR
 DEPLAISIR
 TRISTESSE
 PEUR
 COLÈRE
 SURPRISE POSITIVE
 SURPRISE NEGATIVE
 SATISFACTION
 INSATISFACTION
 VALORISATION
 DEVALORISATION
 ACCORD
 DESACCORD
 DÈGOÛT
 AUCUNE (TERME NEUTRE)

Figure 2: Unit description

Questions Settings Stats

Filter judgments: All Quiz mode Work mode

No filters

1 total Test Question

ID	Judgments	% Missed	% Contested	Last updated	Enabled	Actions
506076166	423	<div style="width: 100%; height: 10px; background-color: #ccc;"></div>	<div style="width: 100%; height: 10px; background-color: #ccc;"></div>	3 months ago by fraisse@lmsi.fr	<input checked="" type="checkbox"/>	Show Details

Figure 3: Test question

the gold data set, their work will be rejected. Only the top 100 contributors are displayed in this graph.

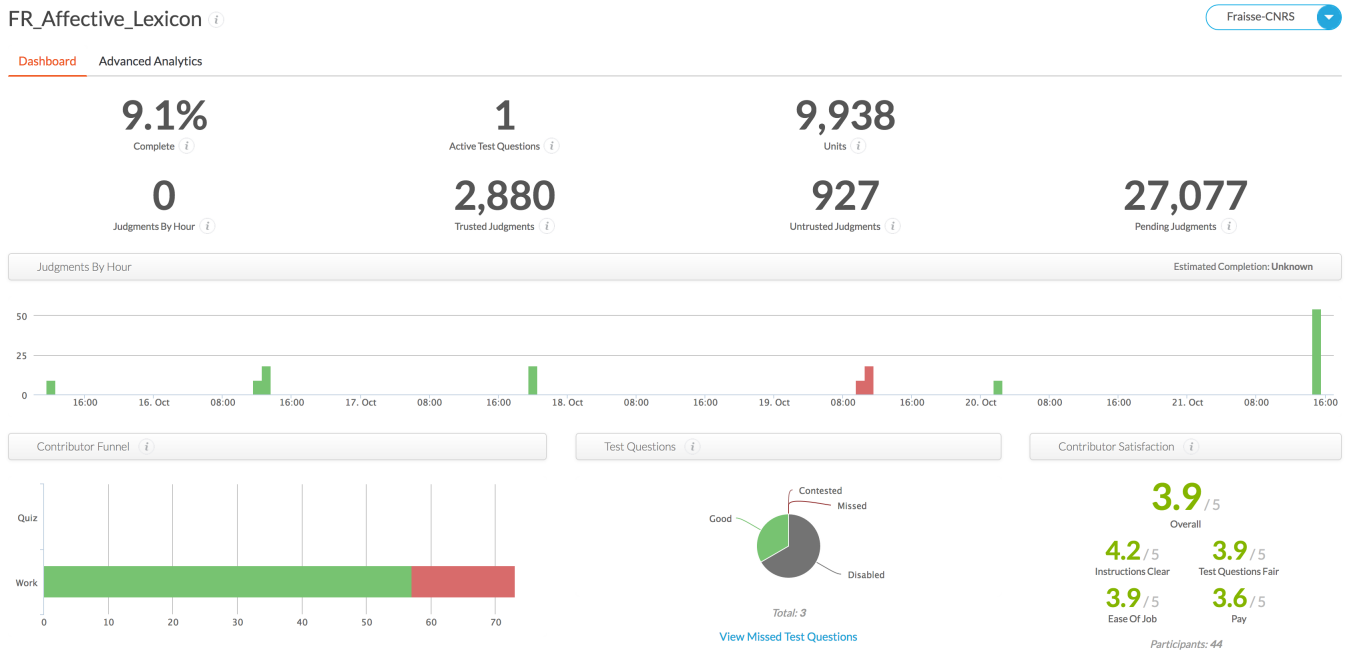


Figure 4: Dashboard of the Workflow job concerning the validation of the french affective lexicon

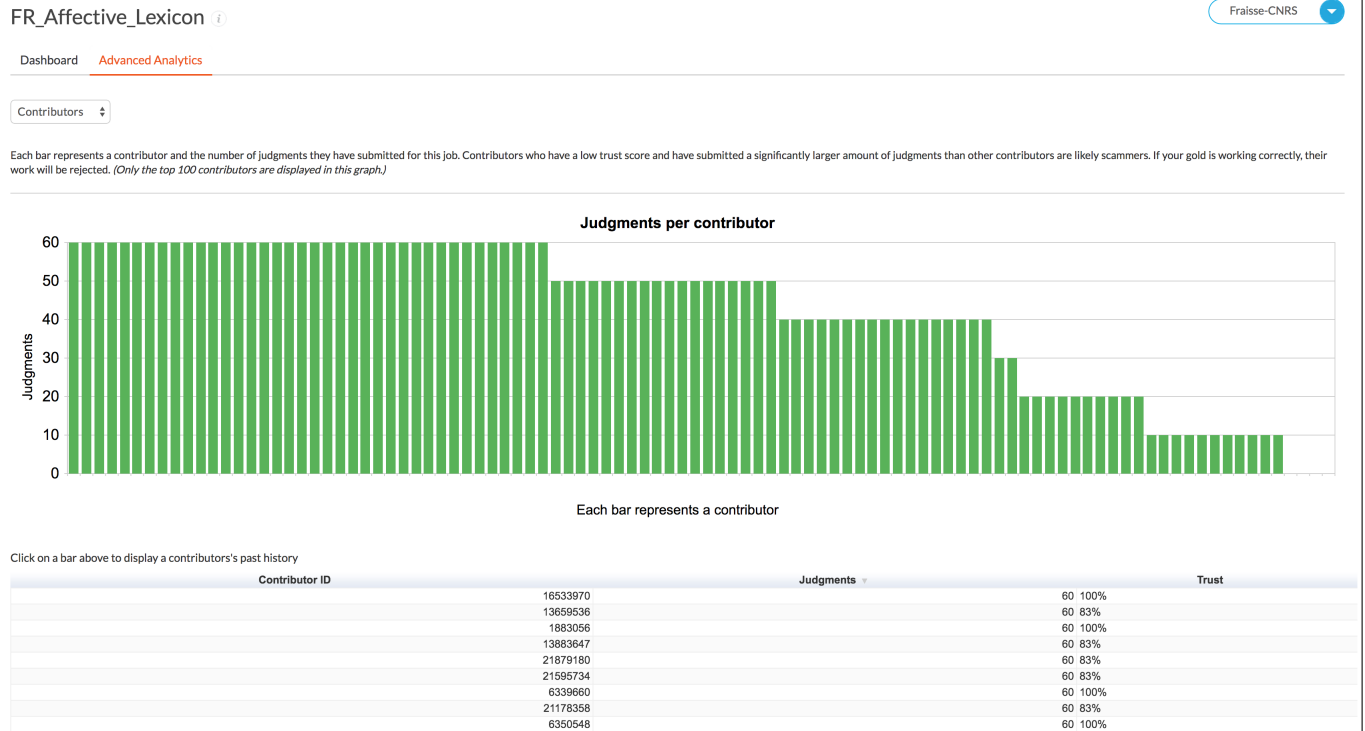


Figure 5: Judgments per contributor


3 Evaluation Campaign

3.1 Aim

T5.3 aims at organizing an open evaluation campaign on affective document analysis (identifying source, target and affect expressions) with a corpus-based quantitative black-box evaluation methodology in the climate change domain, for French and German. The HC-based workflow developed in WP2 will be used first to build and validate the gold standard, second to propose input resources to the participants (ontology elements of WP4, affect lexicons, etc.). Thus a participant will be able to either use their own resources, or build on the proposed input resources to improve their system (or both, in which case we will be able to assess the impact of the resource choice on system performance). The campaign will also be an occasion to test replacing a static comparison against a gold standard by a dynamic assessment of the participants' data through HC.

3.2 Data

A public call for tender has been published in February 2014 by LIMSI-CNRS (PUMA Nbr. 43826) for providing reference annotations on microblog textual data (Tweets), for an amount of 4,200,000 signs (or 30,000 messages) equally spread between French and German (15,000 of

	<p>CHIST-ERA</p>	<p>Subproject : WP5 Task : T5.3 Date : August 9, 2016 Page : 22</p>
---	------------------	--

each). ELDA, the Evaluations and Language resources Distribution Agency (located 9, rue des Cordelières, 75013 Paris) was the only institution to make an offer and was selected.

So far, a the French subcorpus was extracted by MODUL from the data collected with the *Media Watch on Climate Change* demonstrator. A query with a list of regular expressions [Scharl, 2014] resulted in 3,433 tweets, 112 YouTube videos, 77 Facebook messages and 11 Google+ postings for the period 26 March to 26 May 2014. The first exploration of the corpus showed a relatively high proportion of retweets or purely informative tweets. LIMSIS using the same set of regular expression queries did a new extraction from Twitter (July 2014) on French which resulted in a corpus with a relatively higher proportion of subjective messages, containing 7,000 posts. Preliminary tests for the deploying the annotation framework have been done in collaboration with ELDA in July 2014 and production of the annotation of the French reference corpus is planned to start in full during August 2014.

3.3 A model of language data annotation

Comparing the HC framework against the classical annotation approach for producing evaluation resources calls for some automatic support for what concerns the mapping from the classical linguistic annotation scheme, designed for expert annotators, onto a smaller grain annotation procedure suitable for crowdsourcing or GWAPS users. To this end we first provide a task generic annotation formalisation which we will use in a second step to draw the specification of a data converter that will automatically prepare a corpus for crowd annotation from a corpus prepared for classical annotation, allowing the task designer to specify relevant parameters.

3.3.1 Annotation representation

In what follows, we call “control task”, the information processing task for which we want to assess the performance of some computer system, for instance if we look at Information Retrieval (IR), given a set of documents and a query, the control task consists in identifying which documents are relevant with respect to the given query [Cleverdon, 1960]. In fact, for any objective evaluation task, the systems under test are asked to output a symbolic representation in function of the objects and relations that they have found to be present in the input representation (the test data). The control task may be limited to identifying only the objects present in the test set, *e.g.* POS tagging [Paroubek, 2007] for which the output representation is made of the word boundaries and their class label. Sometimes the boundaries of the objects present in the test set are given and the systems need only to identify the class they belong to (*e.g.* IR or Word Sense Disambiguation [Edmonds and Kilgarriff, 2002]). On the other hand, some control tasks are much more complex and require identify objects, primitive relations holding between objects and also higher level relations holding between primitive relations, like in parsing [de la Clergerie et al., 2008], anaphoric resolution [Vilain et al., 1995] or image recognition [Unnikrishnan et al., 2007]. For tasks like machine translation, we are in general only interested in the final result of the transformation of the objects and relations that the system under test has identified in the input data.

In the most general case, a control task can be seen as a process that links together test data units resulting from a segmentation process and annotation symbols possibly organized in several layers. Assuming that the test data is the result of a segmentation process of an input medium (character stream, speech signal, pixel array, etc.) represented by $S = \{s_i/0 \leq i \leq N \in \mathbb{N}\}$, and the set of annotation labels by A , the m layers of relations graphs R resulting from the annotation process can be expressed as follows²:

$$\begin{aligned}
 R &= \bigcup_{j=1}^m R_j, \quad m \in \mathbb{N} \\
 R_1 &= \bigcup_{k=1}^q \{r_l/l \in \mathbb{N}, r_l \subset \mathcal{P}(S^k \times A)\} \\
 R_i &= \bigcup_{k=1}^u \{r \subset \mathcal{P}((S \cup R_{x_1}) \times (S \cup R_{x_2}) \cdots \times (S \cup R_{x_k}) \times A), 1 \leq x_k < i\}
 \end{aligned} \tag{1}$$

R_1 represents the first layer of annotation of the test data (i.e. the set of all relations of any arity between initial segmentation units) and the R_i represent the successive layers of annotations that may reference annotations from any previously existing layer down to the initial data segments themselves (see Figure 6).

Note that with respect to the *annotation graph* model from LDC [Bird and Liberman, 2000], which directly link annotations to events from the various linear input streams that they decorate, we encode in our model the (potentially recursive) structure of the annotations. Such information is important in our opinion when comparing annotation schemes to take into account their relative structural complexity. Our representation is very much like the one proposed in [Roth and Sammons, 2008] except that we adopt a slightly more general point of view by abstracting any implementation detail like annotations identifiers, types, constituents etc. to simply a kind of relation.

(S (NP-SBJ I) (VP consider (S (NP-SBJ Kris) (NP-PRD a fool))))

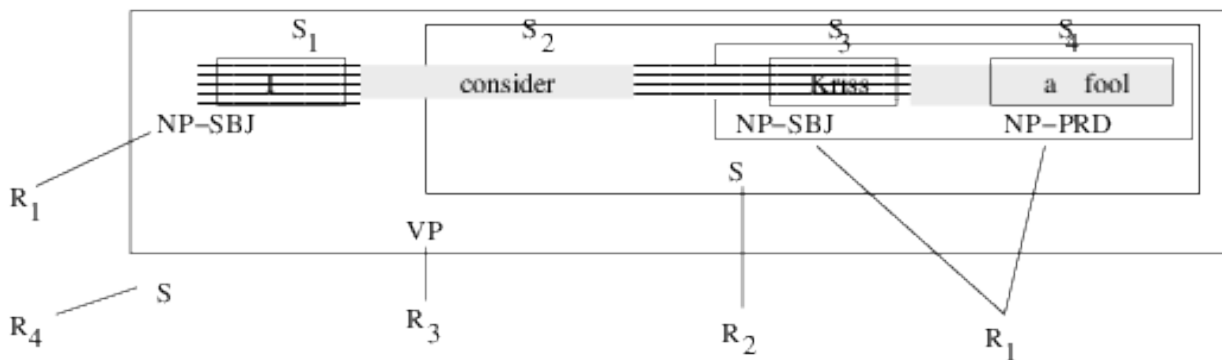


Figure 6: S_i and R_i for a PennTreebank syntactic annotation sample.

In the next section, we give examples of the instantiation of S , R_i and A for a selection of well known evaluation campaigns.

²In formula 1, $\mathcal{P}(x)$ is the set of all subsets of x .

	CHIST-ERA	Subproject : WP5 Task : T5.3 Date : August 9, 2016 Page : 24
---	-----------	---

In the literature, computer traces resulting from performing the control task are called *hypothesis* (annotations) and the human ones: *gold standard* or *reference* (annotations). The objective evaluation result is obtained by comparing the human and computer traces produced in response to the test data, *i.e.* comparing the set of relations identified by the computer (H) with the one identified by humans (G). When the evaluation is quantitative, the result is obtained by computing some *measure* defined over the two sets of relations. Here it is important to point out that the result of evaluation, which is a measure $R \times R \rightarrow \mathbb{R}$, is a function of T the test data, whose role in linking reference and hypothesis data is essential for the computation of the evaluation result. Even more so, when the system under test uses relations R' which have different semantics from the reference ones (but nevertheless mappable to them), or when it modifies the input data because of noise, data corruption or specific normalization, or when it uses a segmentation function S' different from the reference one. With N the “noise” function, the hypothesis is then better described as :

$$\begin{aligned}
 H &= \bigcup_{j=1}^n H_j, n \in \mathbb{N} \\
 H_1 &= \bigcup_{k=1}^q \{S' : T \rightarrow \mathcal{P}(T), S' \neq S, A' \neq A, N : T \rightarrow T, r_l \subset \mathcal{P}((S'(N(T)))^k \times A')\} \quad (2) \\
 H_i &= \bigcup_{k=1}^u \{r \subset \mathcal{P}((S'(N(T)) \cup H_{x_1}) \cdots \times (S'(N(T)) \cup H_{x_k}) \times A'), x_k < i\}
 \end{aligned}$$

In addition to the mapping μ from relation annotation labels A' to A , which in general is provided by the participating system, one must then be able to find a “reasonable” mapping M between $S(T)$ and $S'(N(T))$ to be able to compute an evaluation result. By “reasonable”, we mean a mapping that maximizes the global similarity between the reference and hypothesis versions of the annotated material with respect to a particular similarity function σ . This is what is done for instance when one uses dynamic programming to find the mapping that minimizes the edit distance between two slightly different versions of the same text to compare their POS tag annotations [Paroubek et al., 1998].

$$M = \underset{m}{\operatorname{argmax}} \sum_{h,g} \sigma(m(h), g), \quad (3)$$

$$m \in \mathcal{P}(S'(N(T))) \rightarrow \mathcal{P}(S(T)), \sigma : \mathcal{P}(S(T)) \times \mathcal{P}(S(T)) \rightarrow [0, 1]$$

Most of the time, an evaluation campaign will define several measures in conjunction and use the vector space corresponding to the measurement tuples to represent the performance of each system as a point in the n -dimensional Euclidean space. Their relative position is then characterized by their distance, justifying the use of the term *metric* instead of measure.

3.4 Applying our model to real evaluations

In this section, we use the model proposed in section ?? to represent well known evaluation protocols from different domains of natural language processing [Paroubek et al., 2007].

3.4.1 Classification

The first type of applications that we identified is the general domain of classification and relation extraction. In this kind of task, the purpose is to segment a data set in order to highlight parts of this set that belong to specific classes (predefined or not), and possibly to provide relations existing between these segments.

We take the examples of information retrieval, named entity recognition, temporal annotation and parsing.

Information retrieval Classical information retrieval aims at finding full documents that are relevant to a given query Q . The document collection C is the input data (the retrieval unit is the entire document level). Here we consider a simplified instance of the general model presented in the previous section, in a sense that the segmentation of the test data into units to be annotated is provided, it is made of the documents themselves, see Figure 7.

This is a classification task, since the aim is to produce a partition of the collection, between relevant and non-relevant documents, with respect to the query. In practice the evaluation data contains several queries, but since in general they are considered independent of each other, the evaluation resolves to a series of single query evaluation (see Figure 7).

In other words, variables introduced in Section ?? are instantiated in the following way:

- T : the set of documents
- $S(T)$: the structuration corresponds to the existing document boundaries,
 $\forall t \in T, S(t) = \{t\}$
- A : a set of two labels: *relevant for the query* or *not relevant for the query*
- $R = R_1$: a singleton made of one unary relation that tag the relevance of each documents.

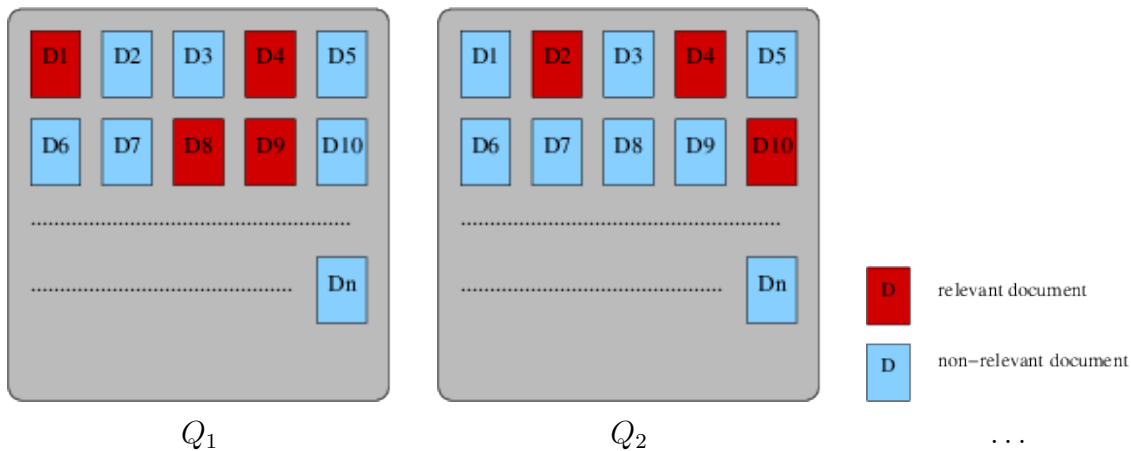


Figure 7: Classification of the set of documents among two parts (relevant and non relevant) for each query: for Q_1 , D1, D4, D8 and D9 are relevant; for Q_2 , D2, D4 and D10 are relevant.

Named entity recognition The following steps can be identified concerning named entity recognition:

1. Identification: finding which data units of the test set need to be annotated.
2. Categorization: finding the appropriate relation to annotate a data unit from the test set, *e.g.* tagging word sequences with labels for locations, persons, organisations, etc.

A normalization step can be added, as for example at the Temporal Expression Recognition and Normalization (TERN) Task of EVALITA [Magnini et al., 2008], where temporal expression should be associated with a universal representation of the expression. All NE types can be concerned by this normalization, for example person names, since they exhibit often many variations in their realization: “Barack Obama”, “B. Obama”, “President Obama”, “Barack H. Obama”.

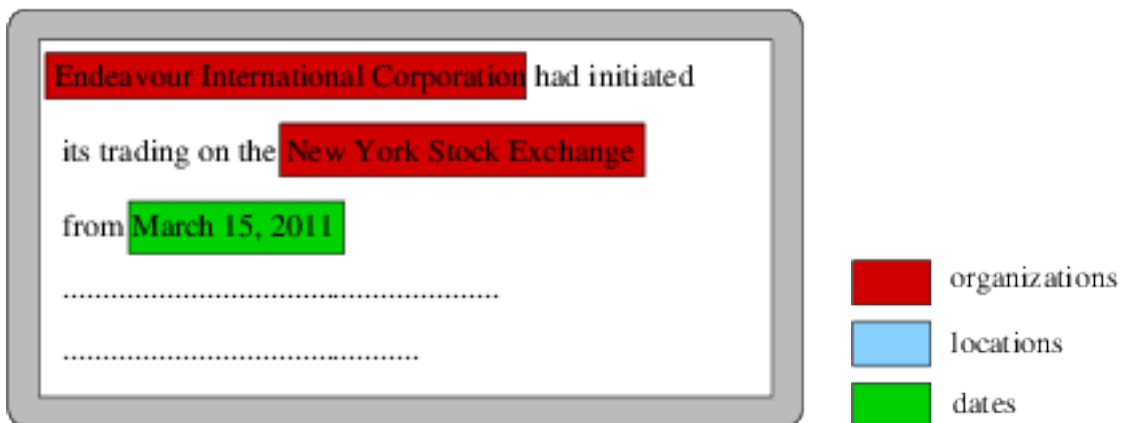


Figure 8: Classification of the sets of characters considering the named entity types (here, organizations, locations, dates, none).

Note that for named entity recognition, the segmentation function of the test data into elementary units is generally not provided by the evaluation organizers. This is not the case for the following example: TempEval.

- T : a document, seen as a stream of words or characters
- $S(T)$: the segmentation of NE types, at character or word level
- A : the set of NE class labels
- $R = R_1$: a singleton holding the unary relation linking the NE to its class label.

Temporal annotation Temporal annotation as defined by TempEval evaluation campaign [Verhagen et al.] consists in the following: given a set of test texts for which sentence boundaries are annotated, as well as all temporal expressions and events in texts, the control task goal is to link events to other events, or events to time expressions (see Figure 10).

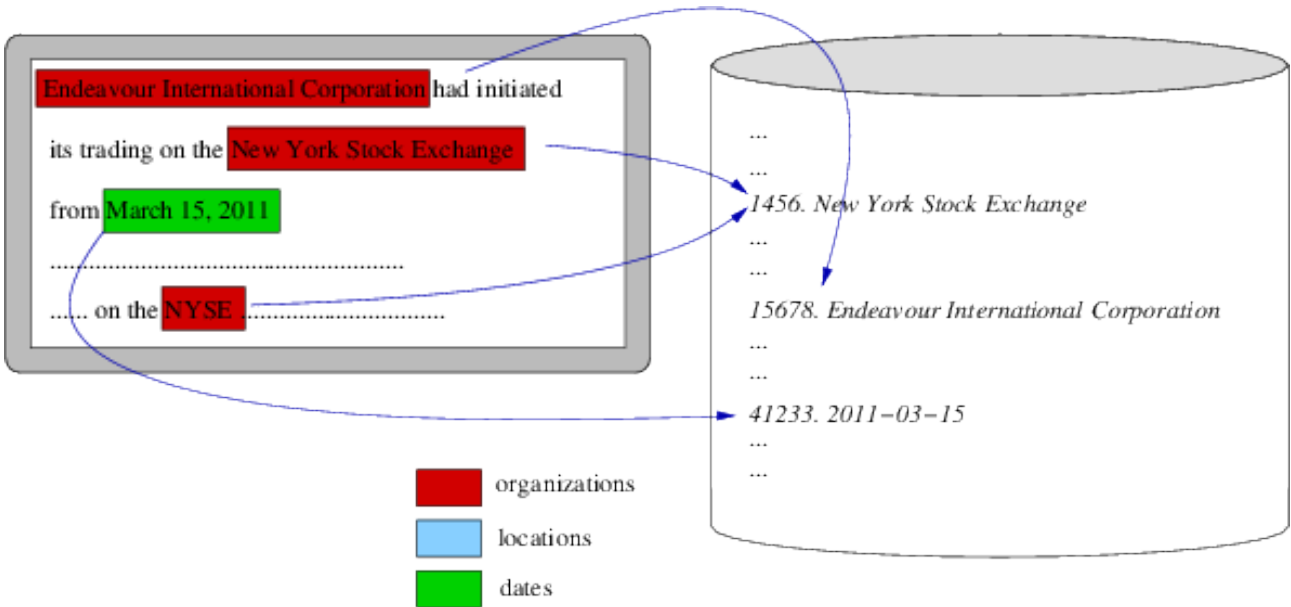


Figure 9: Classification of the sets of characters considering the named entity types and relations to normalized entities in a separate knowledge base.

- T : a document, seen as a stream of words or characters.
- $S(T)$: the segmentation into temporal expressions, signals, and events.
- A : the set of temporal expression signal and event class labels, as well as temporal relations labels.
- $R = R_1$: 1/ the relation that links a temporal expression, signal or event to its class label, *e.g.* *kidnapped* is an event.
2/ plus all the labeled time relations between the temporal elements *e.g.* *kidnapped* is **before** *rescued*.

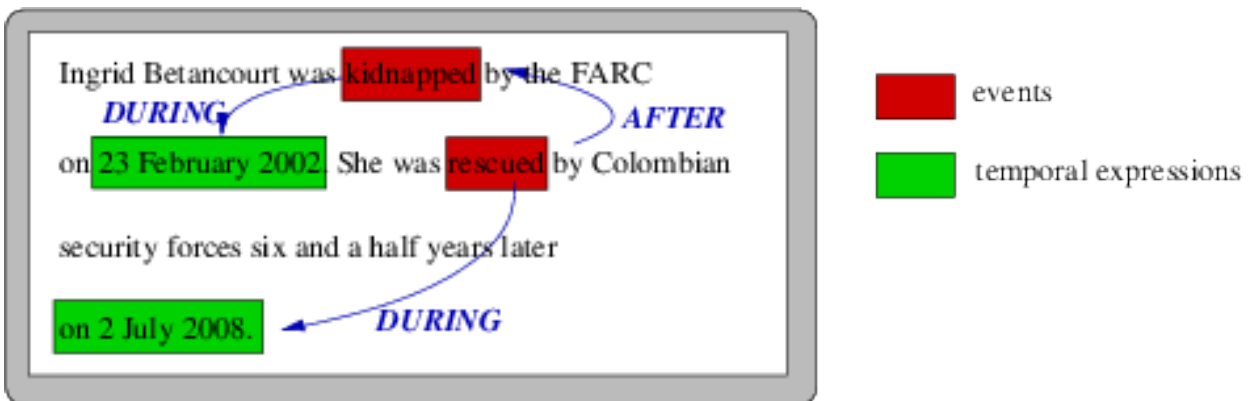


Figure 10: Temporal annotation.

Parsing The aim of automatic parsing is to provide a complete/partial structural analysis of a sentence expressed in terms of:

- chunks, sequences of words with some syntactic meaning,
- constituents, sequences of words which function as a single units within a hierarchical structure,
- dependencies, relations linking a particular word (the head) and one of its dependents,
- links, relation between pairs of words without necessarily referring to a tree hierarchy,
- grammatical relations, i.e. relation/head/dependent tuples [Watson et al., 2005],
- derivation/derived tree [Schmitz and Le Roux, 2008] describing the construction of the syntactic parse tree ,
- etc.

Since theories and annotation schemes are quite numerous and diverse in parsing, we present here only a few annotation schemes which have been used for evaluation: the PennTreebank [Marcus et al., 1993] for constituent analysis of English and PASSAGE [Vilnat et al., 2010] for chunks and grammatical relations in French. The PennTreebank example (see Figure 6 on page 23) is the first example of annotation scheme in this article which exhibits both relations between annotated elements (words) and their class label (e.g. the relation between NP-SBJ and “T”), as well as relations between annotations themselves (e.g. the toplevel relation between S and the constituents NP-SBJ and VP).

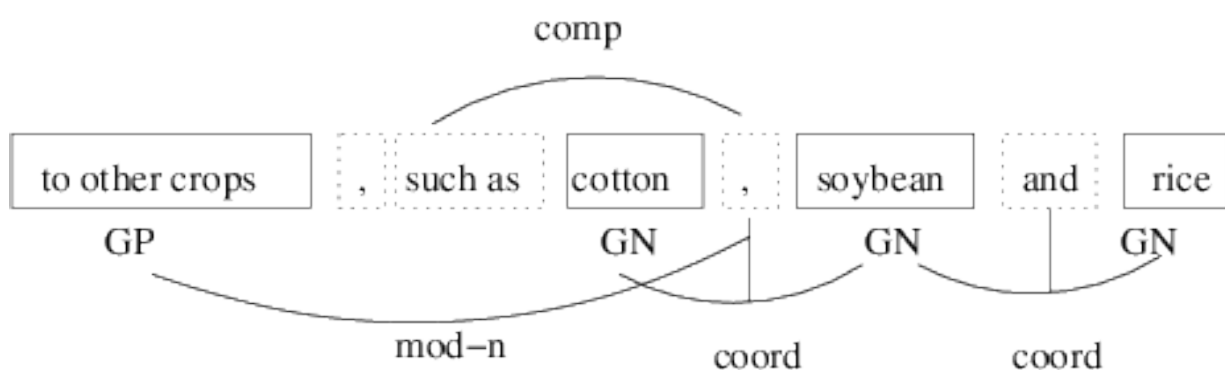



Figure 11: Example of PASSAGE annotation

The Penn Treebank constituent annotation model:

	CHIST-ERA	Subproject : WP5 Task : T5.3 Date : August 9, 2016 Page : 29
---	-----------	---

- T : a documents seen as a stream of characters
- $S(T)$: the segmentation into words
- A : the set of constituent labels
- R_1 : relations between words and their deepest layer of constituent label
- R_i : relations between words and constituent labels, or between constituent labels of deeper levels
- $R : \bigcup_{j=1}^m R_j, m \in \mathbb{N}$

The PASSAGE annotation scheme has only one layer of non-recursive syntactic chunks and grammatical relations defined between words and/or chunks (cf Figure 11). Element of comparison between the passage annotation scheme and PARC, SD and GR, three other syntactic annotation schemes used for English parsing evaluation are provided in [Paroubek et al., 2009].

The PASSAGE annotation model:

- T : a documents seen as a stream of tokens
- $S(T)$: the segmentation into words
- A : the set of chunk and relation labels
- R_1 : the relations between words and their chunk labels, or relations for which at least one argument is word (e.g. coordinating relation for whose coordinating conjunction argument is always a single word not included in any chunk, see Figure 11)
- R_2 : the relations linking chunks only
- $R : R_1 \cup R_2$


3.4.2 Transduction

Lastly, a very different type of applications is the set of applications producing an output that is not an enrichment (or annotation) of an existing test set, but a new object obtained by transformation from or in response to another object. Examples of transduction applications are: machine translation, speech synthesis, automatic summarization, language generation [Koller et al., 2010] or machine dialogue.

For all these examples, T is a document seen as a sequence of characters to be either translated, synthesized, summarized, etc. $S(T)$ is the existing segmentation into language units, while A is the result of the operation: the translation of a language unit into the target language, the synthesis of a language unit into sound generation instructions, etc. R is the set of links between language units in the test set and elements from A .

- T : a documents seen as a stream of tokens
- $S(T)$: the segmentation into transduction source units
- A : the corresponding transduction target units labels
- $R = R_1$: relations between source and target units

Note that we can consider multilingual alignment tasks [Chiao et al., 2006] to be degenerate

	<p>CHIST-ERA</p>	<p>Subproject : WP5 Task : T5.3 Date : August 9, 2016 Page : 30</p>
---	------------------	--

cases of transduction task, where the target labels are provided as input data and the systems under test need only to identify the relation between source and target units.

3.5 Expert Annotation Guidelines

Opinion, Sentiment and Emotion annotations are applied on microblogs text message (maximum length of 140 characters). The annotations scheme was inspired from the PASSAGE evaluation campaign of syntactic parser of French for the structure to which a specific semantics for opinion, sentiment and emotion annotation were added. With the previous formalism, the annotation structure is as follows :

T : a documents seen as a list of tokens

$S(T)$: the segmentation into words

A : the set of chunk and relation labels

R_1 : the relations between words and their chunk (group) labels, or relations for which at least one argument is word

R_2 : the relations linking chunks only

R : $R_1 \cup R_2$

3.5.1 Guideline for French

Note that annotation guidelines are written in French, but an English version is in preparation.

Guide d'annotation d'opinion/sentiment/émotions pour le français

Projet uComp

Patrick Paroubek (pap@limsi.fr)

Amel Fraisse (amel.fraisse@univ-lille3.fr)

Lucie Gianola (gianola@limsi.fr)

LIMSI-CNRS

v. 5.3 / 2016 03 31

Introduction

Les annotations d'opinion, de sentiment ou d'émotion concernent le français et l'allemand sur des énoncés de moins de 140 caractères.

L'objectif est de représenter la sémantique des expressions de sentiment, d'opinion et d'émotion exprimée au niveau d'un énoncé court. Pour annoter une séquence de mots, l'annotateur ne devra prendre en compte que les informations explicitées dans le co-texte (les mots proches de la séquence annotée, par exemple présents dans la phrase précédente) et non le contexte. Il devra par exemple ignorer les méta-données ou des connaissances spécifiques au domaine, à la culture, à l'histoire des protagonistes auxquels il est fait référence et ne considérer que les informations liées à la sémantique du message, autrement dit, la signification associée à la séquence de mots qui reste vraie en toutes circonstances.

L'annotateur devra identifier les expressions d'opinion, qu'elles soient explicites, par exemple «*Je déteste la pluie*», ou implicite. Dans ce dernier cas, il devra se limiter aux présupposés liés au sens des mots présents dans le co-texte et choisir l'interprétation la plus générale. Bien entendu, il est toujours possible que des restrictions ou des modifications de sens présentes dans le co-texte viennent modifier cette interprétation.

Par exemple dans «*Le fournisseur d'énergie X va construire une centrale hydro-électrique dans la région Y.*», la définition sémantique de «*construire*» implique nécessairement un investissement du fournisseur d'énergie qui réalise ou commande la construction. Cet investissement peut être de nature très variable : temps, énergie, matériaux etc. On peut donc en déduire que le résultat de toute action de construction est nécessairement valorisée par l'entité qui la réalise ou la commande. Comme exemple de prise en compte du co-texte pour modifier le sens, nous pouvons citer les marqueurs de négation, qui inversent la polarité des expressions.

L'expression d'opinion, sentiment ou émotion (OSEE) associée a priori à un terme ou une séquence de mots peut aussi être neutre. Par exemple dans le cas de «*détruire*», la polarité de l'expression «*A détruit B*», dépend de la situation décrite, c'est-à-dire de la polarité associée aux entités auxquelles les expressions A et B se réfèrent. Cette polarité est déterminée en fonction de la sémantique de A et de B en appliquant des règles qui prennent en compte

- la relation des entités auxquelles les expressions A et B se réfèrent par rapport à la situation décrite dans l'énoncé. On pourra ici s'inspirer de la notion de rôle thématique (par ex. agent, patient,...)
- l'orientation de l'agent humain, posée a priori par rapport à la situation, en fonction d'un modèle «*Belief, Desire, Intention*» (Croyance, Désir, Intention) instancié au plus haut niveau d'abstraction à partir des principes et contraintes de survie de l'agent dans son environnement, par exemple avec le besoin d'assurer son approvisionnement en énergie, sa sécurité, son confort et sa reproduction. De manière générale, toute situation ou entité qui va dans le sens de l'augmentation du potentiel de survie et de confort de l'agent sera considérée positive (attraction), et réciproquement, négative si elle s'y oppose (répulsion).

Par exemple, l'énoncé «*ce nouveau médicament détruit les cellules cancéreuses*» est clairement positif, puisque l'entité objet B se réfère ici à une maladie, mais l'énoncé «*les pesticides détruisent aussi les animaux indispensables à la vie du jardin*» est négatif, puisque si le sens de «*détruire*» n'a pas changé, l'objet B sur lequel s'applique l'action est connoté dans le sens positif de la préservation de la vie du jardin.

Les annotations sont de deux types :

- les groupes (non récursifs)
- les relations

Les groupes et les relations présentent une certaine redondance, par exemple le type d'une relation va contraindre le type de groupes qui peut servir de point de départ à cette relation. Notez que cette redondance sera utile lors de l'annotation manuelle par la vérification de cohérence qu'elle apporte.

Les groupes identifient dans le texte les mots ou groupes de mots (empans de texte) faisant référence aux entités impliquées dans la situation décrite par l'énoncé, donc l'entité porteuse de l'opinion, sentiment ou émotion (SOURCE) et l'entité objet de cette expression (CIBLE); ainsi que l'expression elle-même, les éventuels marqueurs de négation (NEGATION) et modificateurs d'intensité (MODIFIEUR) associés.

Types de groupe	Contenu
Expression d'Opinion/Sentiment/Émotion (OSEE)	séquence de mots considérée comme une expression de subjectivité
SOURCE	séquence de mots faisant référence à l'entité qui exprime cette subjectivité
CIBLE	séquence de mots faisant référence à l'entité sur laquelle porte cette subjectivité
MODIFIEUR	marqueurs d'intensité de l'expression de subjectivité
NEGATION	marqueurs de négation
DESTINATAIRE	séquence de mots faisant référence à l'entité à qui est destinée l'expression de subjectivité

N.B. : pour le moment nous n'annoterons pas les émoticônes. Table 1. Les différents types de groupe.

Pour les groupes, nous aurons deux stratégies complémentaires concernant l'identification de l'empan de texte associé à un groupe :

- d'une part une **stratégie maximaliste pour les groupes de type SOURCE, CIBLE, ou DESTINATAIRE** consistant à sélectionner la séquence de mots la plus étendue, incluant au besoin, les modificateurs, appositions, subordonnées relatives, voire conjonctions de modificateurs de toutes sortes y compris de relatives, jusqu'au discours rapporté ou aux citations.
- d'autre part une **stratégie minimaliste pour les OSEE, MODIFIEUR et NEGATION**, qui se concentrera sur les mots sémantiquement porteurs du sens, à l'exclusion des modificateurs, apposés compléments ou subordonnées relatives.

Notez qu'il peut exister des groupes qui ne sont reliés par aucune relation.

Les relations servent à qualifier les liens entre les groupes identifiés dans le texte pour faire référence aux entités impliquées dans la situation décrite par l'énoncé. Elles permettent de relier la SOURCE à l'OSEE par le biais de la relation **DIT** et l'expression à la CIBLE par le biais de la relation **SUR**. À ces deux relations s'ajoutent des relations reliant les mots modifiant l'expression d'opinion, de sentiment ou d'émotion, en distinguant d'une part les marqueurs de négation (**NEG**) et les modificateurs d'intensité (**MOD**). On annotera aussi, lorsqu'elle existe, la référence au destinataire (**DESTINATAIRE**) de l'expression d'opinion, lorsqu'il est explicitement mentionné dans le texte, qu'il s'agisse d'un individu ou d'un ensemble d'individus.

Type de relations	Fonction
DIT	relie la SOURCE à l'OSEE
SUR	relie l'OSEE à la CIBLE, ou aux CIBLES intermédiaires (transitivité de la relation SUR)
MOD	relie le MODIFIEUR à l'OSEE
NEG	relie la NÉGATION à l'OSEE ou aux CIBLES intermédiaires
RECEPTEUR	relie l'OSEE au DESTINATAIRE (ou SOURCE ou CIBLE ou OSEE, lorsque le RECEPTEUR a une double fonction, on annotera la fonction concernant l'opinion, le sentiment ou l'émotion)

Table 2. Les différents types de relation.

Les différents types de groupes et leurs délimitations

SOURCE

La source est constituée du groupe de mots qui représente l'entité à laquelle est attribuée l'expression d'opinion, sentiment ou émotion (OSEE).

S'il n'y a pas de mention explicite de la source, on supposera que la personne qui publie le message (auteur du microblog) est la source et on ne fera aucune annotation particulière.

On regroupe sous l'annotation SOURCE, la mention explicite de la source la plus large possible, incluant ses modificateurs, multiples apposés, relatives, voire conjonctions de modificateurs de toutes sortes, y compris de relatives.

Ex. 1 La SOURCE de l'OSEE est le groupe nominal: «*Plus de 200 villes*». On considère que ces 200 villes valorisent («*renforcent*») leur action contre le changement climatique, la séquence de mots y faisant référence est donc annotée comme SOURCE de l'OSEE.

SOURCE	VALORISATION	CIBLE
Plus de 200 villes	renforcent	leur résilience face au changement climatique
G1101	G1102	G1153
G1101 DIT G1102		
	G1102 SUR G1153	

Ex. 2 Absence de source explicite dans cet énoncé, car l'origine du lancement est inconnue (pays, organisation, société etc.)

CIBLÉ		VALORISATION					
Un engin propulsé par l'énergie solaire	lancé	dans	l'	espace	en	2016	
G956	G1108	F10	F11	F12	F13	F14	

G1108 SUR G956

Ex. 3 Une SOURCE de l'OSEE qui inclue une subordonnée relative: «les écologistes que nous sommes»

SOURCE		DESACCORD		MODIFICATEUR		CIBLÉ	
(2/2)	les écologistes que nous sommes	combattont	sans relâche	tout ce qui sera contraire aux valeurs écol.	#potec	#juncker	
F2	G1101	G1106	G1109	G1107	F19	F20	

G1101 DIT G1106

G1109 MOD G1106

G1106 SUR G1107

Remarque 2

Dans l'exemple 3, nous avons identifié les mots «sans relâche» comme un MODIFICATEUR de l'OSEE, car l'expression accentue ici l'intensité de l'OSEE.

Les Expressions d'Opinion de Sentiment ou d'Émotion (OSEE)

Une expression d'Opinion de Sentiment ou d'Émotion (OSEE) est constituée de l'empan de texte dont la sémantique exprime l'OSEE. Elle peut être portée par [un verbe](#), [un nom](#), [un adjectif](#) ou bien encore [un adverbe](#), voire [une préposition](#)

OSEE verbale

Dans l'empan de texte associé à l'OSEE, on regroupe la séquence minimale de mots dont la sémantique exprime l'OSEE, à l'exclusion de ses modificateurs, compléments, subordonnées relatives etc. Il faut noter que l'exemple 4 comporte une expression secondaire de désaccord (opposition introduite par «face au»), mais on privilégie l'expression de valorisation, à cause du verbe «renforcer» qui indique une action de valorisation (investissement d'effort pour préserver l'existant).

Ex. 4 L'OSEE est le groupe verbal : «renforcent»

SOURCE		VALORISATION		CIBLÉ	
Plus de 200 villes	renforcent	leur résilience face au changement climatique			
G1101	G1102	G153			

G1101 DIT G1102

G1102 SUR G153

Dans l'exemple 5, l'OSEE est le participe passé : «épargnés». Elle n'inclut pas le complément d'objet indirect «par les #éoliennes» car il n'est pas indispensable à l'interprétation sémantique de l'OSEE, la définition du verbe «épargner», quel que soit l'objet sur lequel porte l'action d'épargner, implique une intention de préservation, donc une valorisation de l'objet.

Ex. 5 L'OSEE est le participe passé : «épargnés» sans son COI

DESTINATAIRE		CIBLÉ						SATISFACTION				
Coucou	@valeurs	voici	une	photo	d'un de ces coins de paradis français	encore	épargnés	par	les	#éoliennes	Bisous	!
F2	G1170	F4	F5	F6	G834	F15	G833	F17	F18	F19	F20	F21

G833 RECEPTEUR F3

G833 SUR G834

OSEE nominale

Dans l'exemple 6, l'OSEE est le nom «champion». L'OSEE inclut aussi le groupe prépositionnel «de l'efficacité» qui modifie le nom «champion», car le terme «efficacité» est aussi un marqueur de valorisation. Cependant on exclut le modifieur «énergétique», car sa sémantique n'exprime pas obligatoirement un sentiment de valorisation.

Ex. 6 L'OSEE est le groupe nominal : «champion de l'efficacité énergétique»

CIBLÉ		VALORISATION	
Les clusters HPC sous Nvidia	champions de l'efficacité	énergétique	
G1133	G1132	F11	

G1132 SUR G1133

OSEE ajective

Dans l'exemple 7, l'OSEE est composée de l'adjectif épithète antéposé «beau» et de l'adjectif postposé «festif».

Ex. 7 L'OSEE sont les adjectifs : «beau» et «festif».

VALORISATION		CIBLÉ		PLAISIR						
Un	beau	marché public	festif	dédié	aux	produits	écologiques	et	aux	entreprises...
F2	G209	G210	G1155	F7	F8	F9	F10	F11	F12	F13

G209 SUR G210

G1155 SUR G210

Dans l'exemple 8, l'OSEE est l'adjectif épithète antéposé «grand». Notez que les termes «Science» et «Vérité» sont considérés a priori comme marqueurs de valorisation, le premier car l'accumulation de savoir est bénéfique a priori pour un agent, l'augmentation de ses connaissances augmentant sa probabilité de survie. De même, la notion de vérité est valorisante, car elle est définie comme l'accord d'une représentation avec la réalité. De plus, les deux concepts ont des antonymes a priori fortement connotés négativement : ignorance, méconnaissance pour le premier et mensonge, fausseté pour le second.

Ex. 8 L'OSEE est l'adjectif : «grand».

VALORISATION		CIBLÉ		VALORISATION			VALORISATION						
c'	est	l'	anniversaire	d'	un	grand	Homme... un Homme qui consacre sa vie à la	Science	,	à	la	Vérité	,
F5	F6	F7	F8	F9	F10	G311	G1160	G1159	F22	F23	F24	G1158	F26

G311 SUR G1160

G1159 SUR G1160

G1158 SUR G1160

OSEE adverbiale

Dans l'exemple 9, l'OSEE est l'adverbe «basement». Notez «sans», annoté comme marque de négation car il modifie le nom «calcul» qui est lui-même modifié par l'adjectif «politique», qui est lui-même modifié par l'adverbe «basement», qui est l'OSEE (ici exprimant le mépris). Cette chaîne complexe de relations SUR relie de manière indirecte l'expression d'opinion «basement» à ce qu'elle qualifie, ici l'action du ministère, d'où l'annotation du nom «ce ministère» et du verbe «agit» avec son groupe prépositionnel modifieur «Avec #RoyalSégolène».

Ex. 9 L'OSEE est l'adverbe «basement».

MODIFICATEUR		VALORISATION		CIBLÉ		DESACCORD		CIBLÉ	
Quelle	bonne	idée	a	eu	#duflot	de refuser	l'écologie.		
G1161	G457	G458	F5	F6	G461	G459	G460		

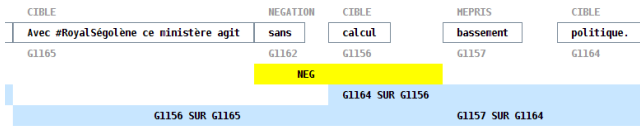
G457 SUR G458

G461 DIT G459

G459 SUR G460

G1161 MOD G457

G458 SUR G459



OSEE prépositionnelle

Dans l'exemple 10, l'OSEE est la préposition «Contre».

Ex. 10 L'OSEE est la préposition «Contre».



Ex. 11 L'OSEE est la préposition : «VS» (versus, c'est-à-dire «contre»).

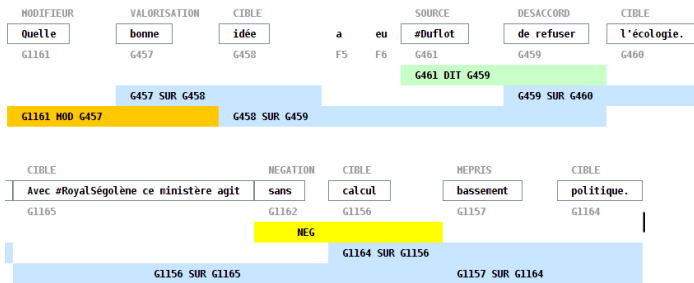
Notez que l'on n'annote qu'une seule direction d'opposition (celle du sens de lecture), de «Écologie» vers «Économie», alors que la relation est considérée comme symétrique.



CIBLE

La CIBLE regroupe les mots qui font référence à ce sur quoi porte l'expression d'opinion de sentiment ou d'émotion, en considérant la mention la plus large possible dans le texte, incluant ses modificateurs, multiples apposés, relatives, voire conjonctions de modificateurs de toutes sortes y compris de relatives (comme pour la SOURCE). Dans l'exemple 12, une chaîne complexe de relations SUR relie les cibles de manière indirecte pour aller de l'expression d'opinion «bassement» à ce qu'elle qualifie, ici l'action du ministère, d'où l'annotation du nom «ce ministère» et du verbe «agit» avec son groupe prépositionnel modifieur «Avec #RoyalSégolène» comme cible finale.

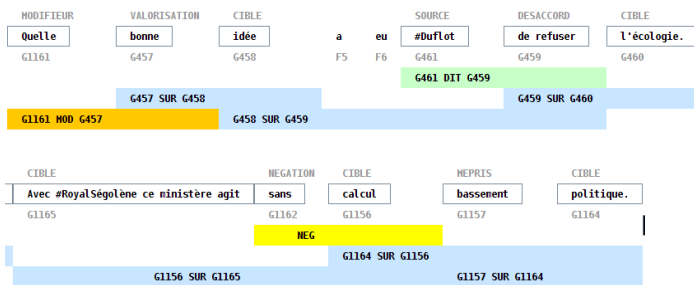
Ex. 12 L'OSEE est l'adverbe : «bassement».



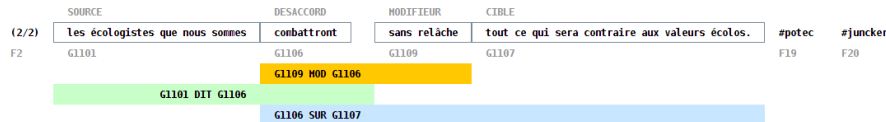
MODIFIEUR

Seront annotés comme MODIFIEUR les termes qualifiant une intensité d'expression d'opinion, de sentiment ou d'émotion. Dans l'exemple suivant le déterminant exclamatif «Quelle» est considéré comme un intensifieur.

Ex. 13 Le MODIFIEUR est le déterminant exclamatif «Quelle».



Ex. 14 Le MODIFIEUR est le groupe nominal «sans relâche».

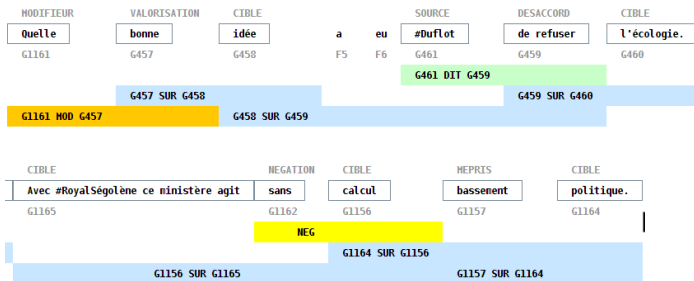


NEGATION

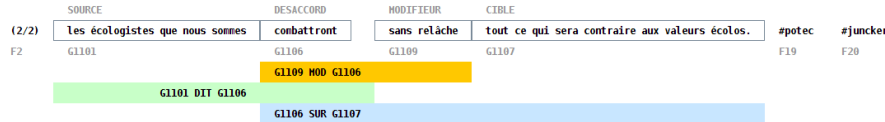
Les termes indiquant la négation ne sont annotés que lorsqu'ils concernent une OSEE ou une CIBLE intermédiaire comme c'est le cas ici. Les négations indiquent alors soit une absence du sentiment exprimé par le groupe pointé par la relation de NEGATION, soit le sentiment opposé à celui exprimé par le groupe pointé par la négation.

Ex. 15 La NEGATION : «sans» marque l'absence du sentiment associé au nom «calcul».

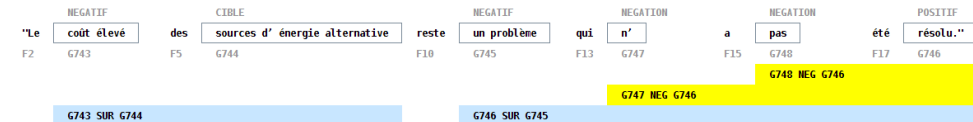
Notez qu'ici, le nom «calcul» n'est pas indiqué comme faisant partie d'une OSEE, le terme porteur de l'OSEE est «bassement», qui propage le sentiment par une cascade de relations SUR jusqu'au nom «calcul» par le biais de l'adjectif qualificatif «politique».



Ex. 16 Un exemple où l'on n'annote pas de négation pour le mot «sans» car cette négation ne porte pas directement sur une OSEE.



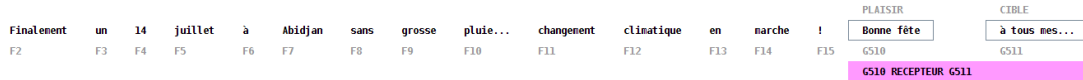
Ex. 17 Un exemple de NEGATION où le sentiment de l'OSEE est inversé par la NEGATION. Sur ce point, il faut remarquer que l'on annote la valeur originale du sentiment pour l'OSEE (ici positif). Notez aussi que les deux particules de négation «n» et «pas» sont annotées NEGATION.



DESTINATAIRE

Les termes qui indiquent explicitement à qui l'on s'adresse sont rassemblés sous l'étiquette DESTINATAIRE.

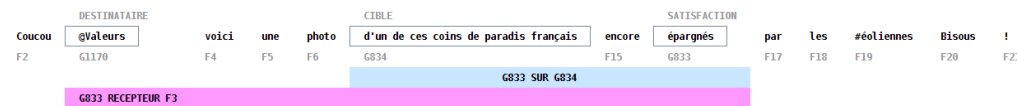
Ex. 18 Le DESTINATAIRE est «à tous mes...». Notez que ni «14 juillet» ni «pluie» ne sont annotés comme des OSEE car cela supposerait utiliser des connaissances d'ordre pragmatique.



Ex. 19 Le DESTINATAIRE est ici interpellé dans un discours rapporté dont le titre est: «Monsieur Junker, serez-vous l'homme..?» et qui contient lui-même une INSTRUCTION/DEMANDE adressée à M. Junker (c'est pourquoi le titre n'est pas regroupé dans un groupe CIBLE mais est éclaté entre un DESTINATAIRE et une INSTRUCTION/DEMANDE reliés par une relation RECEPTEUR). Le destinataire est aussi rappelé comme destinataire initial de l'objet valorisé, l'apposition «discours de Philippe Lambert, président du groupe écologiste» qui est la CIBLE de l'OSEE «beau». Notez que cette CIBLE est devenue de ce fait éligible pour être origine d'une relation RECEPTEUR.



Ex. 20 Le DESTINATAIRE est indiqué par son nom d'utilisateur Twitter : «@Valeurs».



Les classes sémantiques affectives fines

- 1.6.1 Annotation en classes sémantiques affectives fines
 - 1.6.1.1 DÉPLAISIR
 - 1.6.1.2 DÉRANGEMENT
 - 1.6.1.3 MÉPRIS
 - 1.6.1.4 SURPRISE (NÉGATIVE)
 - 1.6.1.5 PEUR
 - 1.6.1.6 COLÈRE
 - 1.6.1.7 ENNUI
 - 1.6.1.8 TRISTESSE
 - 1.6.1.9 PLAISIR
 - 1.6.1.10 APAISEMENT
 - 1.6.1.11 AMOUR
 - 1.6.1.12 SURPRISE (POSITIVE)
 - 1.6.1.13 SATISFACTION
 - 1.6.1.14 INSATISFACTION
 - 1.6.1.15 ACCORD
 - 1.6.1.16 VALORISATION
 - 1.6.1.17 DÉSACCORD
 - 1.6.1.18 DEVALORISATION
 - 1.6.1.19 INSTRUCTION/DEMANDE
- 1.6.2 Annotation en classes génériques de polarité
 - 1.6.2.1 NÉGATIF
 - 1.6.2.2 POSITIF

Nous répertorions les différentes classes sémantiques affectives en trois grandes catégories:

1. Opinions (intellectif): cette catégorie contient deux classes sémantiques positives : ACCORD et VALORISATION et deux classes sémantiques négatives : DÉSACCORD et DEVALORISATION.
ex. Je suis d'accord avec toi.
2. Sentiments (affectif-intellectif): cette catégorie contient une classe sémantique positive : SATISFACTION et une classe sémantique négative : INSATISFACTION.
ex. Je suis satisfait de ce travail.
3. Émotions (affectif): cette catégorie contient quatre classes sémantiques positives : PLAISIR, APAISEMENT, AMOUR et SURPRISE (POSITIVE) et huit classes sémantiques négatives : DÉPLAISIR, DÉRANGEMENT, MÉPRIS, TRISTESSE, PEUR, COLÈRE, ENNUI et SURPRISE (NÉGATIVE).
ex. Je suis triste

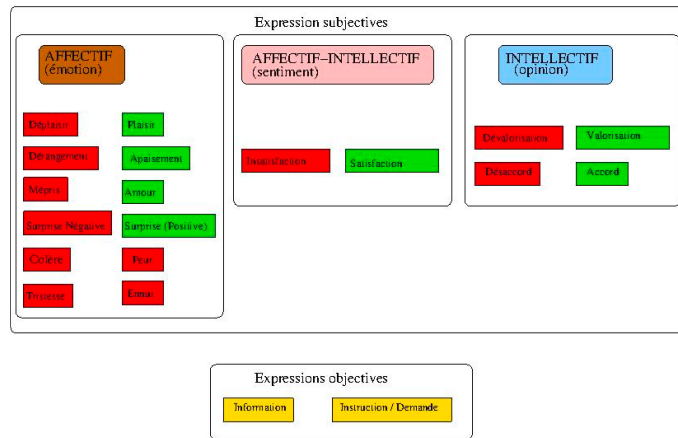


Fig 1. Les grandes catégories de classe sémantiques.

Dans les sections qui suivent, nous donnons pour chaque classe sa définition et les caractéristiques qui permettent de la distinguer des autres classes. Nous exprimerons les définitions en fonction de trois concepts qui décrivent l'état mental de la SOURCE qui ressent les opinions/sentiments/émotions :

1. **Perceptions, croyances et connaissances** sur le monde, mémoire des événements.
2. **Désirs** qui sont les souhaits qui résultent des pulsions instinctives de survie ou des préférences subjectives.
3. **Intentions** qui sont les actions que l'on se prépare à réaliser dans un avenir plus ou moins proche.

La détermination de l'OSEE ressentie par la SOURCE dépendra du point de vue que la SOURCE entretient sur les **événements** du monde, qui sont **localisés dans le temps** par rapport au **présent de la SOURCE**. Les différentes sortes d'OSEE seront distinguées les unes des autres en fonction de :

1. des **Désirs** de la SOURCE, considérés comme constants dans le temps,
2. des **Croyances** de la SOURCE qui évoluent en fonction des événements,
3. des **Intentions** de la SOURCE. Notez que la TRISTESSE et le PLAISIR nécessitent de distinguer des actions génériques un type d'actions spécifiques, dites de «signalement» de l'état interne de la SOURCE à destination de ses congénères (par exemple pleurer pour la TRISTESSE et rire pour le PLAISIR).
4. de la **localisation temporelle** des événements tels que perçus par la SOURCE, c'est à dire dans le **passé, le présent ou le futur du ressenti de l'OSEE**.

ÉTIQUETTE GÉNÉRIQUE DE LA CLASSE SEMANTIQUE	TYPE	CLASSES SPÉCIFIQUES DE LA CLASSE SEMANTIQUE
1 DÉPLAISIR	e-	déplaisir / déception
2 DÉRANGEMENT	e-	dérangement / embarras
3 MÉPRIS	e-	mépris / dédain / dégoût / haine
4 SURPRISE NÉGATIVE	e-	surprise négative / étonnement négatif
5 PEUR	e-	peur / terreur / inquiétude / anxiété
6 COLÈRE	e-	colère / rage / agacement / exaspération / énervement / impatience
7 ENNUI	e-	ennui / torpeur
8 TRISTESSE	e-	tristesse / chagrin / souffrance / désespoir / résignation
9 PLAISIR	e+	plaisir / divertissement / joie / euphorie / bonheur / extase
10 APAISEMENT	e+	apaisement / soulagement / reconnaissance / pardon / sérénité
11 AMOUR	e+	amour / tendresse / affection / dévouement / passion / envie / désir
12 SURPRISE (POSITIVE)	e+	surprise (positive) / étonnement positif
13 SATISFACTION	s+	satisfaction / contentement / fierté
14 INSATISFACTION	s-	insatisfaction / mécontentement
15 ACCORD	o+	accord / compréhension / approbation
16 VALORISATION	o+	valorisation / intérêt / appréciation
17 DÉSACCORD	o-	désaccord / désapprobation
18 DÉVALORISATION	o-	dévalorisation / désintérêt / dépréciation
19 INSTRUCTION/DEMANDE	i	instruction / demande / requête / ordre / conseil / recommandation
20 INFORMATION	i	information

Tab.1- Tableau-1 récapitulatif des différentes classes sémantiques d'opinion/sentiment/émotion, e-=émotion négative, e+, s=sentiment, o=opinion, i=information, +=valeur positive, -=valeur négative.

OSEE	TYPE	type de cause	intentions, actions, connaissances
1 DÉPLAISIR	e-	événement présent non-désiré	pas d'intention de remédiation
2 DÉRANGEMENT	e-	événement présent non-désiré	intention de remédiation
3 MÉPRIS	e-	Existence d'un Désir (répulsion)	-
4 SURPRISE NÉGATIVE	e-	événement présent non-désiré	absence de connaissance antérieure de l'événement
5 PEUR	e-	absence de remédiation pour un événement non-désiré	connaissance antérieure d'un événement similaire
6 COLÈRE	e-	événement présent ou passé non-désiré	action de (pseudo) remédiation
7 ENNUI	e-	absence d'événement désiré	-
8 TRISTESSE	e-	événement présent ou passé non-désiré	action de signalement
9 PLAISIR	e+	événement présent désiré	action de signalement
10 APAISEMENT	e+	2 événement consécutifs, l'un non-désiré, l'autre annulant les effets du précédent	-
11 AMOUR	e+	Existence d'un Désir (attraction)	-
12 SURPRISE (POSITIVE)	e+	événement présent (désiré)	absence de connaissance antérieure de l'événement
13 SATISFACTION	s+	Croyance	connaissance de la réalisation du désir
14 INSATISFACTION	s-	Croyance	absence de connaissance de la réalisation du désir
15 ACCORD	o+	similarité d'opinion de 2 SOURCES sur une CIBLE commune	Connaissance de l'opinion de l'autre SOURCE sur la CIBLE
16 VALORISATION	o+	accord des Connaissances sur la CIBLE les Désirs, Croyances ou Intentions	-
17 DÉSACCORD	o-	disimilarité d'opinion de 2 SOURCES sur une CIBLE commune	Connaissance de l'opinion de l'autre SOURCE sur la CIBLE
18 DÉVALORISATION	o-	désaccord des Connaissances sur la CIBLE les Désirs, Croyances ou Intentions	-
19 INSTRUCTION//DEMANDE	i	réponse ou action attendue	absence d'OSEE
20 INFORMATION	i	besoin de communication	absence d'OSEE

Tab.2- Tableau-2 récapitulatif des caractéristiques discriminantes pour les différentes classes sémantiques d'opinion/sentiment/émotion, selon la **cause** et les **intentions, actions, connaissances**.

1.6.1.1 DÉPLAISIR

Définition : Émotion suscitée par un événement passé ou présent qui est non-désiré par la SOURCE mais sans que la SOURCE ait l'intention d'y remédier.
Classes spécifiques : déplaisir / déception

Ex. 21 L'OSEE indiquant le DÉPLAISIR est le groupe verbal «passe mal».

#abadbuzz	RT	:	Un	lion	en	cage	,	un	singe	et	des	serpents	:	CIBLE	DEPLAISIR						
F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15	G240	G241	F22	F23	F24	F25		
														une soirée cirque qui passe mal		au	Stamp	#Waterloo	#Gaia		
														G241 SUR G240							

1.6.1.2 DÉRANGEMENT

Définition : Émotion suscitée par un évènement non-désiré par la SOURCE engendrant une Intention d'action de remédiation.
Classes spécifiques : dérangement / embarras

Ex. 22 L'OSEE indiquant le DÉRANGEMENT est le nom «Nuisance».

CIBLE	DERANGEMENT	MODIFICATEUR		SOURCE
Éoliennes	Nuisances	graves	pour	la santé.
G827	G824	G825	F5	F6
	G825 MOD F3			
G824 SUR G827				
	G826 DIT G824			
	témoignages du monde entier http://leventourne.org/eoliennes-nuisances-graves-pour-la-sante-temoignages-du-monde-entier/ ...			
				G826

#anti	#éolien	#Eoliennes	#Santé	#Témoignage
F14	F15	F16	F17	F18

1.6.1.3 MÉPRIS

Définition : Émotion suscitée par une connaissance sur une CIBLE en désaccord avec les désirs de la SOURCE.
Classes spécifiques : mépris / dédain / dégoût / haine

Ex. 23 L'OSEE indiquant le MÉPRIS est «gâchis».

CIBLE	MEPRIS	MODIFICATEUR	CIBLE	INSTRUCTION_DEMANDE
Les appareils connectés en mode veille	gâchis	planétaire.	Le lobby des fournisseurs d'énergie	serait innocent ?
G106	F8	G105	G111	G109
		G111 MOD G105		
G105 SUR G106		G108 SUR G109		

Ex. 24 L'OSEE indiquant le MÉPRIS est «haine».

MEPRIS	DESACCORD	SOURCE	DESTINATAIRE	DESTINATAIRE	DESTINATAIRE	DESTINATAIRE	DESTINATAIRE
Denis_Baupin	Toute incitation à la haine	est étrangère aux valeurs des écologistes	@eelv	@Loiselet	@DinahZeghir	@AlbertZennou	@ECOLodep
F2	F3	F4	F5	F6	F7	F8	F9
				G822 SUR G401			
				G823 DIT G822			
				G822 RECEPTEUR G815			
				G822 RECEPTEUR G816			
				G822 RECEPTEUR G817			
				G822 RECEPTEUR G819			

1.6.1.4 SURPRISE NÉGATIVE

Définition : Émotion suscitée par la découverte d'un évènement non-désiré et non-attendu par la SOURCE.
Classes spécifiques : surprise négative / étonnement négatif

Ex. 25 L'OSEE indiquant la surprise négative est «Mauvaise nouvelle».

SURPRISE_NEGATIVE	SOURCE	ACCORD	CIBLE
Mauvaise nouvelle	le ministre Henry a	accordé	le permis unique à Spe Luminus pour 5 éoliennes sur Abée et Ramelot
G828	F4	G1242	F8
		G1242 DIT G1241	
G828 SUR G1241		G1241 SUR G1243	
			Encore!!
			!
			#éolien

1.6.1.5 PEUR

Définition : Émotion suscitée par la (possible) réalisation d'un évènement non-désiré par la SOURCE et pour lequel la SOURCE ne connaît pas d'action de remédiation.
Classes spécifiques : Peur / terreur / inquiétude / anxiété

Ex. 26 L'OSEE indiquant la peur est «en danger».

CIBLE	MODIFICATEUR	PEUR
La sardine, l'un des poissons	les plus	en danger
G941	G942	G943
	G942 MOD G943	
G943 SUR G941		
		en Méditerranée

1.6.1.6 COLÈRE

Définition : Émotion suscitée par un évènement non-désiré engendrant une réaction chez la SOURCE qui engendre une Croyance nouvelle donnant à la SOURCE l'impression qu'une action de remédiation existe (impression de reprise du contrôle de la situation).
Classes spécifiques : colère / rage / agacement / exaspération / énervement / impatience

Ex. 27 L'OSEE indiquant la COLÈRE est le terme lui-même «colère».

MODIFICATEUR	COLERE	SOURCE
certaine	colère	de la #société civile au Bourget
G2	G1	G1034
G2 MOD G1		
G1034 DIT G1		

1.6.1.7 ENNUI

Définition : Émotion suscitée par l'absence d'évènement désiré par la source.
Classes spécifiques : ennui / torpeur

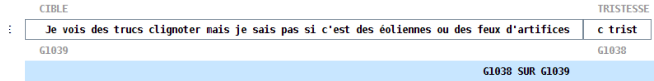
Ex. 28 L'OSEE indiquant l'ennui est le groupe verbal «je m'ennuie».

SOURCE	ENNUI	
je	m'ennuie	
G1037	G1036	
	DIT	

1.6.1.8 TRISTESSE

Définition : Émotion suscitée par un événement passé ou présent en désaccord avec les Désirs de la SOURCE et pour lequel il n'existe pas d'action de remédiation connue de la SOURCE. Cette connaissance suscite de la part de la SOURCE une intention de signaler son désarroi aux autres par un changement d'attitude (par ex. pleurs, posture de faible énergie, etc.).
Classes spécifiques : tristesse / chagrin / souffrance / désespoir / résignation

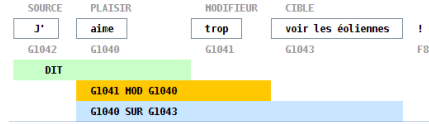
Ex. 29 L'OSEE indiquant la tristesse est le groupe verbal «c trist».



1.6.1.9 PLAISIR

--> **Définition** : Émotion suscitée par un événement (présent ou future) désiré par la SOURCE et pour lequel la SOURCE va avoir l'intention de signaler sa joie aux autres par un changement d'attitude (par ex. sourire, posture énergique, etc.).
Classes spécifiques : plaisir / divertissement / joie / euphorie / bonheur / extase

Ex. 30 L'OSEE indiquant le PLAISIR est le verbe «aime». Il s'agit sans doute ici de sarcasme, mais l'absence de marqueur explicite d'ironie résulte dans une annotation de PLAISIR plutôt que l'inverse.



1.6.1.10 APAISEMENT

Définition : Émotion suscitée par la connaissance par la SOURCE de deux événements consécutifs, le premier en désaccord avec ses Désirs et le second annulant les effets du précédent.
Classes spécifiques : apaisement / soulagement / reconnaissance / pardon / sérénité

Attention : Lorsque les deux événements sont mentionnés dans l'énoncé, chacun sera identifié comme une CIBLE d'un relation SUR partant de l'OSEE d'APAISEMENT qui aura donc deux cibles.

Notez que l'APAISEMENT implique la SATISFACTION mais ne l'inclut pas pour autant.

Ex. 31 L'OSEE indiquant l'APAISEMENT le groupe verbal «s'en sort».



1.6.1.11 AMOUR

Définition : Émotion suscitée par l'attraction empathique de la SOURCE pour la CIBLE.
Classes spécifiques : amour / tendresse / affection / dévouement / passion / envie / désir

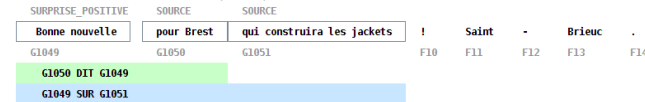
Ex. 32 L'OSEE indiquant l'AMOUR est ici le terme lui-même.



1.6.1.12 SURPRISE (POSITIVE)

Définition : Émotion suscitée par la découverte d'un événement (désiré) non-attendu par la SOURCE. On classera aussi comme positives, les surprises qui ne sont pas en rapport avec aucun désir de la SOURCE.
Classes spécifiques : amour / tendresse / affection / dévouement / passion / envie / désir

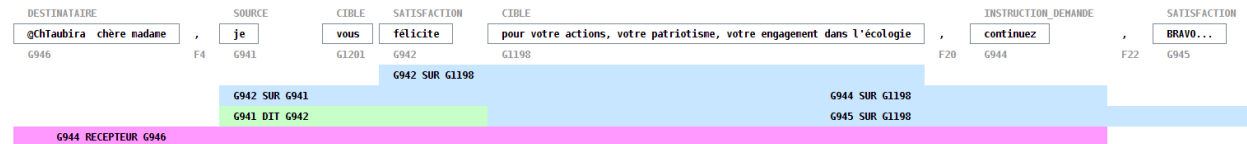
Ex. 33 La SURPRISE POSITIVE est exprimée par «Bonne nouvelle».



1.6.1.13 SATISFACTION

Définition : Sentiment suscité par la Connaissance de la réalisation d'un Désir.
Classes spécifiques : satisfaction / contentement / fierté

Ex. 34 Ici plusieurs OSEE expriment la SATISFACTION : le verbe «félicite» et le nom «BRAVO».



1.6.1.14 INSATISFACTION

de remédiation--> **Définition** : Sentiment suscité par la Connaissance de l'impossibilité de réaliser un Désir.
Classes spécifiques : insatisfaction / mécontentement

Ex. 35 L'OSEE indiquant l'INSATISFACTION est le groupe verbal «avait tout raté».



1.6.1.15 ACCORD

Définition : La SOURCE a connaissance d'un autre SOURCE qui entretient une opinion similaire à la sienne sur la CIBLE.

Classes spécifiques : accord / compréhension / approbation

Ex. 36 L'OSEE indiquant l'accord est le verbe «forment», dans le sens de créer ensemble quelque chose.

SOURCE	ACCORD	CIBLE									
Groen et Ecolo	forment	un groupe commun au Sénat	:	Les	partis	écologistes	francophone	et	néerlandopho...		
G1217	G1215	G1219		F11	F12	F13	F14	F15	F16	F17	
G1215 DIT G1217											
G1215 SUR G1219											

1.6.1.16 VALORISATION

C'est l'émotion duale de l'AMOUR mais appliquée aux inanimés concrets ou abstraits.

Définition : les connaissances de la SOURCE sur la CIBLE sont en accord avec ses Croyances, Désirs ou Intentions.

Classes spécifiques : valorisation / intérêt / appréciation

Ex. 37 Plusieurs OSEE indiquent la valorisation : «héros» et «super».

CIBLE	PEUR	CIBLE		VALORISATION		VALORISATION	CIBLE						
#tortueuth espèce	menacée	"Vu du ciel" #Cabon,	les	héros	de	la nature"	super	doc	@Yann_A_B	sur	@France0tv	#ecc	
G226	G225	G229	F10	G1395	F12	F13	F14	G227	G228	F17	F18	F19	F20
G225 SUR G226		G227 SUR G229											
G227 SUR G228													

1.6.1.17 DÉSACCORD

Définition : La SOURCE a connaissance d'une autre SOURCE qui entretient une opinion différente de la sienne sur la CIBLE.

Classes spécifiques : désaccord / désapprobation

Ex. 38 L'OSEE indiquant le désaccord est la plus évidente et la plus courte possible : «non». Notez qu'ici «non» n'est pas annoté comme une négation.

DESACCORD	CIBLE
Non	aux éoliennes de la ferme du Torpt à Tourville et St -Meslin
G1225	G1226
G1225 SUR G1226	

1.6.1.18 DÉVALORISATION

uniquement aux entités animées.--> **Définition** : les connaissances de la SOURCE sur la CIBLE sont en désaccord avec ses Croyances, Désirs ou Intentions.

Classes spécifiques : dévalorisation / désintérêt / dépréciation

Ex. 39 Ce message comporte de nombreuses expressions de dévalorisation. Il faut remarquer que la première, «saloperies», est à la fois une expression de DÉVALORISATION et une cible pour les autres expressions de DÉVALORISATION. Notez également que «dévalorisant» a deux cibles, la première entité du message ainsi que le nom «foncier» (à défaut de pouvoir exprimer une hiérarchisation entre les deux, elles seront situées sur le même plan). Remarquez aussi la NEGATION qui inverse la polarité d'une expression de VALORISATION. Enfin, la première OSEE, «saloperies», est cible de la seconde, «promues».

DEVALORISATION	VALORISATION	SOURCE	DEVALORISATION	DEVALORISATION	CIBLE	NEGATION	VALORISATION						
Les saloperies	promues	par Royal	c'bruyant, laid	dévalorisant	pour le foncier	et mène pas	efficace						
2	G1229	G1236	G1234	F7	G1235	F12	G1230	G1231	F17	F18	F19	G1232	G1233
G1234 DIT G1236													
G1236 SUR G1229		G1230 SUR G1231											
G1235 SUR F3													
G1230 SUR G1229													
G1233 SUR G1229													
										G1232 NEG G1233			

1.6.1.19 INSTRUCTION/DEMANDE

Définition : OSEE exprimant un ordre, une demande, une question ou une recommandation.

Classes spécifiques : instruction / demande / requête / ordre / conseil / recommandation

Ex. 40 Il s'agit d'une question.

CIBLE	MEPRIS	MODIFIEUR	CIBLE	INSTRUCTION_DEMANDE
Les appareils connectés en mode veille	gâchis	planétaire.	Le lobby des fournisseurs d'énergie	serait innocent ?
G106	F8	G105	G111	G109
G105 SUR G106		G108 SUR G109		

Ex. 41 Il s'agit d'une requête ou recommandation adressée à une personne.

DESTINATAIRE	SOURCE	CIBLE	SATISFACTION	CIBLE	INSTRUCTION_DEMANDE	SATISFACTION				
@ChTaubira chère nadane	je	vous	félicite	pour votre actions, votre patriotisme, votre engagement dans l'écologie	continuez	BRAVO...				
G946	F4	G941	G1201	G942	G1198	F20	G944	F22	G945	
G942 SUR G941		G942 SUR G1198								
G941 DIT G942		G944 SUR G1198								
G945 SUR G1198										
G944 RECEPTEUR G946										

Ex. 42 Il s'agit d'une requête ou recommandation aux lecteurs en général.

C'	est	clair	INSTRUCTION_DEMANDE	CIBLE	
F2	F3	F4	F5	F12	
G1227 SUR G1228					

1.6.2 Annotation en classes génériques de polarité

Dans le cas où il est difficile d'identifier la catégorie sémantique de l'expression d'opinion/sentiment/émotion, on indiquera uniquement la polarité de l'expression (NÉGATIF ou POSITIF).

1.6.2.1 NÉGATIF

On annotera NÉGATIF les expressions d'opinion/sentiment/émotion qui ont une polarité négative et dont il est difficile d'identifier la classe sémantique exacte.

Ex. 43 L'expression «maudit» dans cet exemple sera annoté NÉGATIF car le contexte ne permet pas de savoir si la source de cette expression (l'auteur du tweet) ressent de la colère ou de la tristesse (mais les deux cas sont négatifs).

CIBLE	NEGATIF	MODIFIEUR	NEGATIF								
Ivoire	Haudit	Kenya	le	pire	nassacre	de	rhinocéros	depuis	1988	#Quebec	
G1239	G1238	F4	F5	F6	G1240	G1237	F9	F10	F11	F12	F13
SUR		G1240 MOD G1237									

1.6.2.2 POSITIF

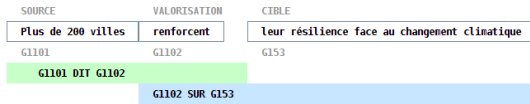
On annotera POSITIF les expressions d'opinion/sentiment/émotion qui ont une polarité positive et dont il est difficile d'identifier avec certitude la classe sémantique exacte.

Les relations

DIT

La relation DIT met **toujours** en rapport une SOURCE avec une OSEE.

Ex. 44 Mise en relation de la SOURCE «Plus de 200 villes» avec l'OSEE «renforcement» par le biais de la relation DIT.

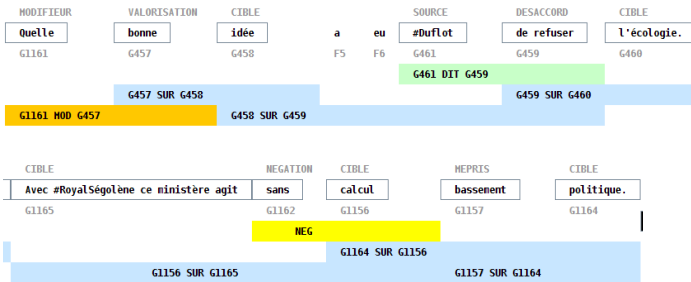


SUR

La relation SUR permet de mettre en rapport une OSEE avec une CIBLE ou bien de propager la relation par le biais de CIBLEs intermédiaires. Dans ce dernier cas, il peut y avoir une chaîne, de relations SUR (portée par des fonctions syntaxiques de modification), partant d'une OSEE pour aboutir à la CIBLE à laquelle fait effectivement référence l'OSEE initiale.

L'exemple 45 comporte deux chaînes complexes de relations SUR. La première part de «bonne» en passant par «idée», pour aboutir à «de refuser». La seconde, partant de l'OSEE «bassement», aboutit à la cible effective, l'action du ministère («Avec #RoyalSégolène ce ministère agit»), et passe par les CIBLEs intermédiaires que sont l'adjectif «politique» et le nom «calcul». Ces deux termes sont devenus éligibles pour être des SOURCES de relation SUR, car ils sont eux-mêmes CIBLEs de relation SUR. Notez que cette propriété, que l'on pourrait appeler transitivité, ne concerne que la relation SUR et pas la relation DIT, qui relie toujours directement une SOURCE à une OSEE.

Ex. 45 Mise en relation de l'OSEE «bonne» avec la CIBLE «idée» par la relation SUR. Notez la présence de deux chaînes de relations SUR, l'une allant de «bonne» à «de refuser», l'autre allant de «bassement» à «calcul».

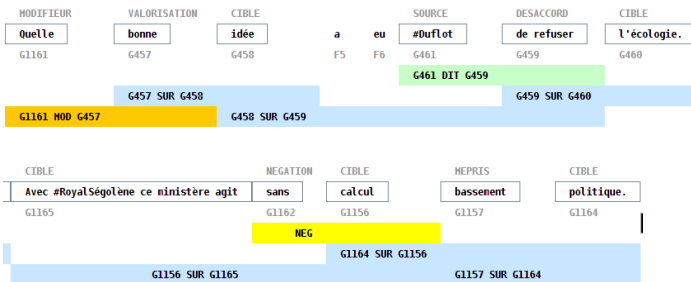


MOD

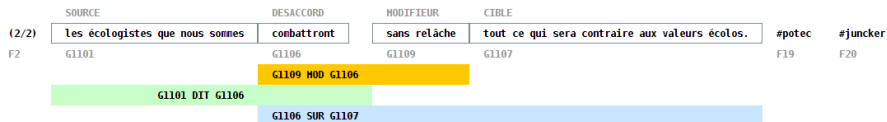
La relation MOD permet de mettre en rapport les termes (groupe MODIFIEUR) qualifiant une intensité d'OSEE. Elle relie **toujours** un MODIFIEUR à une OSEE.

Dans l'exemple 46, le déterminant exclamatif «Quelle» est annoté comme départ d'une relation MOD car il qualifie l'intensité de l'adjectif «bonne». Notez que ce déterminant n'est pas interprété en tant que marqueur de surprise, même si l'on peut lui associer cette émotion, car la surprise concerne l'intensité du déterminé, elle est donc considérée ici comme secondaire.

Ex. 46 La relation MOD va du déterminant exclamatif «Quelle» à l'adjectif «bonne» qu'il qualifie.



Ex. 47 La relation MOD relie le groupe nominal «sans relâche» au verbe «combattront»

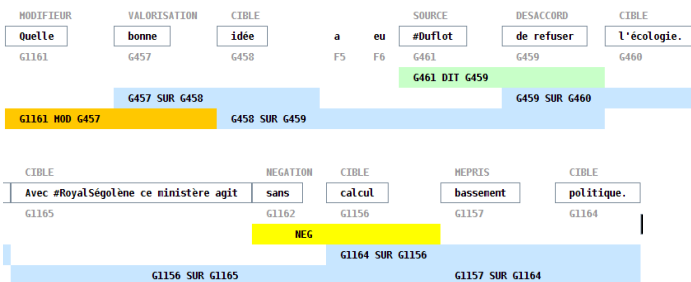


NEG

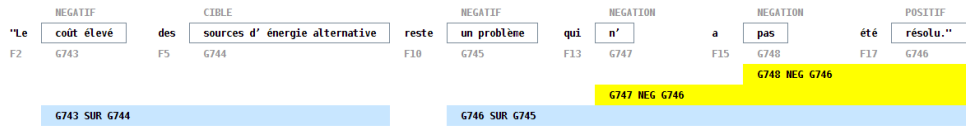
La relation NEG permet de mettre en rapport les éventuels marqueurs de négation avec l'OSEE dont ils modifient la sémantique. Elle part d'un groupe NEGATION pour aller vers une OSEE ou bien une CIBLE intermédiaire.

Dans le cas où la destination de la relation NEG est une CIBLE, nous serons en présence d'une chaîne de relations SUR partant d'une OSEE. Dans l'exemple 48, le nom «calcul» est une des CIBLES de la chaîne partant de l'OSEE «bassement» pour aboutir à la cible effective (l'action du ministère : «Avec #RoyalSégolène ce ministère agit») en passant par les CIBLEs intermédiaires «politique» et «calcul».

Ex. 48 La relation NEG relie «sans» à la CIBLE «calcul», dégné par la chaîne de relations SUR partant de «bassement».



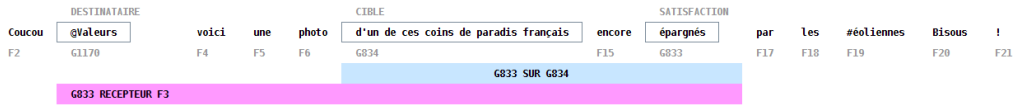
Ex. 49 Un exemple de relations NEG portées par les deux particules de négation «n» et «pas».



RÉCEPTEUR


La relation RÉCEPTEUR va de l'OSEE ou d'une CIBLE au DESTINATAIRE de l'expression de subjectivité.

Ex. 50 La relation RÉCEPTEUR relie l'OSEE «épargnés» au nom d'utilisateur Twitter : «@Valeurs».



Ex. 51 Le DESTINATAIRE est ici interpellé dans un discours rapporté dont le titre est: «Monsieur Junker, serez-vous l'homme..?» et qui contient lui-même une INSTRUCTION/DEMANDE adressée à M. Junker (c'est pourquoi le titre n'est pas regroupé dans un groupe CIBLE mais est éclaté entre un DESTINATAIRE et une INSTRUCTION/DEMANDE reliés par une relation RÉCEPTEUR). Le destinataire est aussi rappelé comme destinataire initial de l'objet valorisé, l'apposition «discours de Philippe Lambert, président du groupe écologiste» qui est la CIBLE de l'OSEE «beau». Notez que cette CIBLE est devenue de ce fait éligible pour être origine d'une relation RÉCEPTEUR.



	CHIST-ERA	Subproject : WP5 Task : T5.3 Date : August 9, 2016 Page : 41
---	-----------	---

3.5.2 Guideline for German

Note that annotation guidelines are written in French since they were adapted from the guidelines for French, an English version is in preparation.

Guide d'annotations d'opinion/sentiment/émotions pour le projet uComp

Patrick Paroubek (pap@limsi.fr)

LIMSI-CNRS

Marie-Anne Moreau

INALCO

v. 5.4 / 20160722

Introduction

Les annotations d'opinion, de sentiment ou d'émotion concernent le français et l'allemand sur des énoncés de moins de 140 caractères.

L'objectif est de représenter la sémantique des expressions de sentiment, d'opinion et d'émotion exprimée au niveau d'un énoncé court. Pour annoter une séquence de mots, l'annotateur ne devra prendre en compte que les informations explicites dans le co-texte (les mots proches de la séquence annotée, par exemple présents dans la phrase précédente) et non le contexte. Il devra par exemple ignorer les méta-données ou des connaissances spécifiques au domaine, à la culture, à l'histoire des protagonistes auxquels il est fait référence et ne considérer que les informations liées à la sémantique du message, autrement dit, la signification associée à la séquence de mots qui reste vraie en toutes circonstances.

L'annotateur devra identifier les expressions d'opinion, qu'elles soient explicites, par exemple «*Je déteste la pluie*», ou implicite. Dans ce dernier cas, il devra se limiter aux présupposés liés au sens des mots présents dans le co-texte et choisir l'interprétation la plus générale. Bien entendu, il est toujours possible que des restrictions ou des modifications de sens présentes dans le co-texte viennent modifier cette interprétation.

Par exemple dans «*Le fournisseur d'énergie X va construire une centrale hydro-électrique dans la région Y.*», la définition sémantique de «*construire*» implique nécessairement un investissement du fournisseur d'énergie qui réalise ou commande la construction. Cet investissement peut être de nature très variable : temps, énergie, matériaux etc. On peut donc en déduire que le résultat de toute action de construction est nécessairement valorisée par l'entité qui la réalise ou la commande. Comme exemple de prise en compte du co-texte pour modifier le sens, nous pouvons citer les marqueurs de négation, qui inversent la polarité des expressions.

L'expression d'opinion, sentiment ou émotion (OSEE) associée a priori à un terme ou une séquence de mots peut aussi être neutre. Par exemple dans le cas de «*détruire*», la polarité de l'expression «*A détruit B*», dépend de la situation décrite, c'est-à-dire de la polarité associée aux entités auxquelles les expressions A et B se réfèrent. Cette polarité est déterminée en fonction de la sémantique de A et de B en appliquant des règles qui prennent en compte

- la relation des entités auxquelles les expressions A et B se réfèrent par rapport à la situation décrite dans l'énoncé. On pourra ici s'inspirer de la notion de rôle thématique (par ex. agent, patient,...),
- l'orientation de l'agent humain, posée a priori par rapport à la situation, en fonction d'un modèle «*Belief, Desire, Intention*» (Croyance, Désir, Intention) instancié au plus haut niveau d'abstraction à partir des principes et contraintes de survie de l'agent dans son environnement, par exemple avec le besoin d'assurer son approvisionnement en énergie, sa sécurité, son confort et sa reproduction. De manière générale, toute situation ou entité qui va dans le sens de l'augmentation du potentiel de survie et de confort de l'agent sera considérée positive (attraction), et réciproquement, négative si elle s'y oppose (répulsion).

Par exemple, l'énoncé «*ce nouveau médicament détruit les cellules cancéreuses*» est clairement positif, puisque l'entité objet B se réfère ici à une maladie, mais l'énoncé «*les pesticides détruisent aussi les animaux indispensables à la vie du jardin*» est négatif, puisque si le sens de «*détruire*» n'a pas changé, l'objet B sur lequel s'applique l'action est connoté dans le sens positif de la préservation de la vie du jardin.

Les annotations sont de deux types :

- les groupes (non récursifs)
- les relations

Les groupes et les relations présentent une certaine redondance, par exemple le type d'une relation va contraindre le type de groupes qui peut servir de point de départ à cette relation. Notez que cette redondance sera utile lors de l'annotation manuelle par la vérification de cohérence qu'elle apporte.

Les groupes identifient dans le texte les mots ou groupes de mots (empan de texte) faisant référence aux entités impliquées dans la situation décrite par l'énoncé, donc l'entité porteuse de l'opinion, sentiment ou émotion (SOURCE) et l'entité objet de cette expression (CIBLE); ainsi que l'expression elle-même, les éventuels marqueurs de négation (NEGATION) et modificateurs d'intensité (MODIFIEUR) associés.

Types de groupe	Contenu
Expression d'Opinion/Sentiment/Emotion (OSEE)	séquence de mots considérée comme une expression de subjectivité
SOURCE	séquence de mots faisant référence à l'entité qui exprime cette subjectivité
CIBLE	séquence de mots faisant référence à l'entité sur laquelle porte cette subjectivité
MODIFIEUR	marqueurs d'intensité de l'expression de subjectivité
NEGATION	marqueurs de négation
DESTINATAIRE	séquence de mots faisant référence à l'entité à qui est destinée l'expression de subjectivité

N.B. : pour le moment nous n'annoterons pas les émoticônes. *Table 1. Les différents types de groupe.*

Pour les groupes, nous aurons deux stratégies complémentaires concernant l'identification de l'empan de texte associé à un groupe :

- d'une part une **stratégie maximaliste pour les groupes de type SOURCE, CIBLE, ou DESTINATAIRE** consistant à sélectionner la séquence de mots la plus étendue, incluant au besoin, les modificateurs, appositions, subordinées relatives, voire conjonctions de modificateurs de toutes sortes y compris de relatives, jusqu'au discours rapporté ou aux citations.
- d'autre part une **stratégie minimaliste pour les OSEE, MODIFIEUR et NEGATION**, qui se concentrera sur les mots sémantiquement porteurs du sens, à l'exclusion des modificateurs, apposés compléments ou subordinées relatives.

Notez qu'il peut exister des groupes qui ne sont reliés par aucune relation.

Les relations servent à qualifier les liens entre les groupes identifiés dans le texte pour faire référence aux entités impliquées dans la situation décrite par l'énoncé. Elles permettent de relier la SOURCE à l'OSEE par le biais de la relation DIT et l'expression à la CIBLE par le biais de la relation SUR. A ces deux relations s'ajoutent des relations reliant les mots modifiant l'expression d'opinion, de sentiment ou d'émotion, en distinguant d'une part les marqueurs de négation (NEG) et les modificateurs d'intensité (MOD). On annotera aussi, lorsqu'elle existe, la référence au destinataire (DESTINATAIRE) de l'expression d'opinion, lorsqu'il est explicitement mentionné dans le texte, qu'il s'agisse d'un individu ou d'un ensemble d'individus.

Type de relations	Fonction
DIT	relie la SOURCE à l'OSEE
SUR	relie l'OSEE à la CIBLE, ou aux CIBLES intermédiaires (<i>transitivité de la relation SUR</i>)
MOD	relie le MODIFIEUR à l'OSEE
NEG	relie la NEGATION à l'OSEE ou aux CIBLES intermédiaires
RECEPTEUR	relie l'OSEE au DESTINATAIRE (ou SOURCE ou CIBLE ou OSEE, lorsque le RECEPTEUR a une double fonction, on annotera la fonction concernant l'opinion, le sentiment ou l'émotion)

Table 2. Les différents types de relation.

Les différents types de groupes et leurs délimitations

SOURCE

La source est constituée du groupe de mots qui représente l'entité à laquelle est attribuée l'expression d'opinion, sentiment ou émotion (OSEE).

S'il n'y a pas de mention explicite de la source, on supposera que la personne qui publie le message (auteur du microblog) est la source et on ne fera aucune annotation particulière.

On regroupe sous l'annotation SOURCE, la mention explicite de la source la plus large possible, incluant ses modificateurs, multiples apposés, relatives, voire conjonctions de modificateurs de toutes sortes, y compris de relatives.

Ex. 1 La SOURCE de l'OSEE est le groupe nominal: «*17 meiner "Freunde"*».

VALORISATION	SOURCE	VALORISATION	CIBLE	INSTRUCTION_DEMANDE	HEPRIS
OSE_GLOBALE	17 meiner "Freunde"	"gefallt"	Burger King.	Heisst das , dass	die da essen gehen? #Ekel
660	659	657	658	F8 F9 F10 F11	G783 G782
	659 DIT 657		657 SUR 658		
				6782 SUR 6783	

Ex. 2 Absence de source explicite dans cet énoncé.

VALORISATION	MODIFIEUR	VALORISATION	CIBLE	HEPRIS	
OSE_GLOBALE	Mobiltelefon.	Wie	gut	, dass man dadurch nicht unterwegs erreichbar ist. #hass	
662	F2	G769	G61	F5 G263	G770
	6769 MOD 661		661 SUR 6263		

Ex. 3 Une SOURCE de l'OSEE qui inclue une subordinée relative: «*Alle die bei uns wegen Hochwasser anfragten*»

SOURCE	APPAISEMENT	CIBLE	NEGATION	NEGATIF
Alle die bei uns wegen Hochwasser anfragen	können	beruhigt	sein	Halle wird nicht in Fluten versinken
G21	F9	G22	F11 F12	G23 F14 G24 F16 F17 G25 F1
G21 DIT G22		G24 NEG G25		
		G25 DIT G23		

Les Expressions d'Opinion de Sentiment ou d'Émotion (OSEE)

Une expression d'Opinion de Sentiment ou d'Émotion (OSEE) est constituée de l'empan de texte dont la sémantique exprime l'OSEE. Elle peut être portée par [un verbe](#), [un nom](#), [un adjectif](#), ou bien encore [un adverbe](#), voire [une préposition](#)

OSEE verbale

Dans l'empan de texte associé à l'OSEE, on regroupe la séquence minimale de mots dont la sémantique exprime l'OSEE, à l'exclusion de ses modificateurs, compléments, subordinées relatives etc.

Ex. 4 L'OSEE est le verbe : « gefällt »

DEPLAISIR	CIBLE	PLAISIR	SOURCE	NEGATION	PEUR
OSE_GLOBALE	Meine Katze beobachtet nicht, das	gefällt	mir	nicht	#angst
G63	G288	G14	G289	G15	G757
		G289 DIT G14		id=E9	
		G15 NEG G14			
G14 SUR G288					

Dans l'exemple 5, l'OSEE est le participe passé : « gefickt ».

Ex. 5 L'OSEE est le participe passé : « gefickt » sans son COI

DEVALORISATION	CIBLE	DEVALORISATION	HEPRIS	MODIFIEUR
OSE_GLOBALE	Freundschaften werden	gefickt	Hass	überall
G40	G273	F3 G38	G39	G771
G38 SUR G273				
G771 MOD G39				
#nonopoly	HEPRIS	HEPRIS		
F13	#hass	#hate...		
	G772	G773	id=E18	

OSEE nominale

Dans l'exemple 6, l'OSEE est le nom « SonderMÜLL ». Notez que la portée de l'OSEE exclue son modificateur « Verbaler » et que le modificateur d'intensité « ohne Grenzen » est lui étiqueté comme MODIFIEUR et supporte une relation MOD.

Ex. 6 L'OSEE est le nom : « SonderMÜLL »

DEVALORISATION	MODIFIEUR	CIBLE
OSE_GLOBALE	ohne Grenzen	SonderMÜLL
F1	F2	F3
G27 MOD G26		
G26 SUR G28		
ihren Antseid nachzukommen, siehe NSA.		
F18	F19	F20 F21 F22 F23
id=E301		

OSEE ajectivale

Dans l'exemple 7, l'OSEE est l'adjectif épithète antéposé « schlechtes ».

Ex. 7 L'OSEE sont les adjectifs : « schlechtes ».

DEVALORISATION	INSTRUCTION_DEMANDE	DEVALORISATION	CIBLE	CIBLE	DEVALORISATION	INSTRUCTION_DEMANDE
OSE_GLOBALE	Ist es ein	schlechtes	Omen	einen Porzellan #Schornsteinfeger ein Bein	abzubrechen	?
G400	G780	G399	G398	G779	G778	G781
G399 SUR G398						
G780 SUR G399						
G781 SUR G399						
G778 SUR G779						
PEUR	#Angst	#Aberglaube	#glueckauf			
G774	F15	F16	id=E184			

Dans l'exemple 8, l'OSEE est l'adjectif épithète antéposé « kleinen ».

Ex. 8 L'OSEE est l'adjectif : « kleinen ».

VALORISATION	DEVALORISATION	CIBLE
OSE_GLOBALE	Machtspiele	kleinen
F1	F2	F4
G29 SUR G30		
G30 DIT F3		
Tornados über Syrien https://t.co/v20sKE7KAX		
F14	F15	F16 F17
id=E302		

OSEE adverbiale

Dans l'exemple 9, l'OSEE est l'adverbe « gut ». Notez qu'en allemand la plupart des adjectifs peuvent être employés comme adverbes (ils sont alors invariable). Remarquez aussi, l'emploi du groupe MODIFIEUR et de la relation MOD pour relier la particule « aus » au verbe « sieht ». On remarquera aussi la forme verbale « geht herunter » qui est ici un modificateur d'intensité.

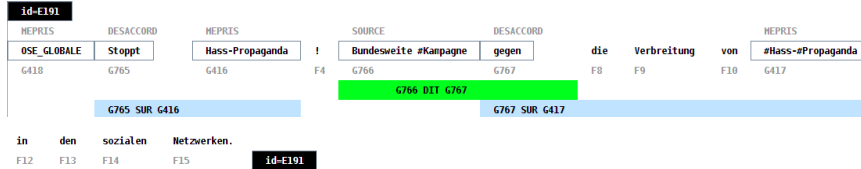
Ex. 9 L'OSEE est l'adverbe « gut ».

CIBLE	VALORISATION	PEUR	MODIFIEUR	
OSE_GLOBALE	Die Vorhersage ist	gut	Lawinengefahr geht herunter	
F1	F4	G32	F6 F7	
G32 SUR G33				
G35 MOD G34				
CIBLE	VALORISATION	CIBLE		
es sieht	gut	aus für unseren Skiride durch die Stubaier...		
G37	G36	G38	F16 F17	F18 F19 F20 F21
G36 SUR G37				
G38 MOD G37				
id=E303				

OSEE prépositionnelle

Dans l'exemple 10, l'OSEE est la préposition « gegen ».

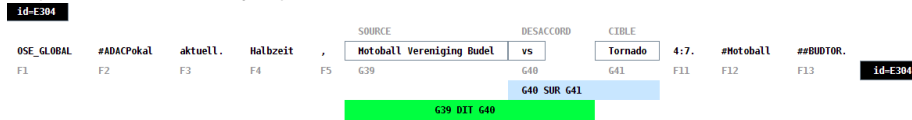
Ex. 10 L'OSEE est la préposition «gegen».



Dans l'exemple 11 Dans le cas de Notez que la préposition : «VS» (versus, c'est-à-dire «contre») porte une seule relation SUR selon une seule direction d'opposition (celle du sens de lecture), alors que la relation est considérée comme symétrique.

Ex. 11 L'OSEE est la préposition : «VS» (versus, c'est-à-dire «contre»).

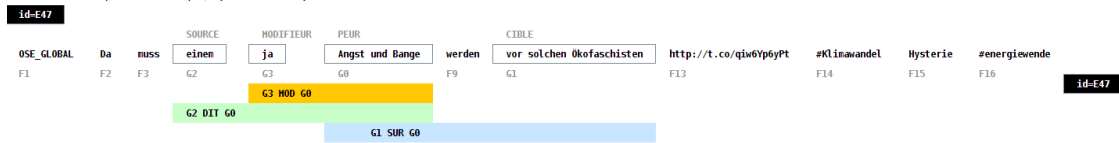
Notez que l'on n'annote qu'une seule direction d'opposition (celle du sens de lecture), alors que la relation est considérée comme symétrique.



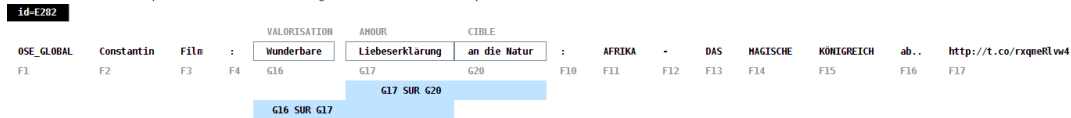
CIBLE

La CIBLE regroupe les mots qui font référence à ce sur quoi porte l'expression d'opinion de sentiment ou d'émotion, en considérant la mention la plus large possible dans le texte, incluant ses modificateurs, multiples apposés, relatives, voire conjonctions de modificateurs de toutes sortes y compris de relatives (comme pour la SOURCE).

Ex. 11 bis Un exemple de cible simple, « vor solchen Ökofaschisten »

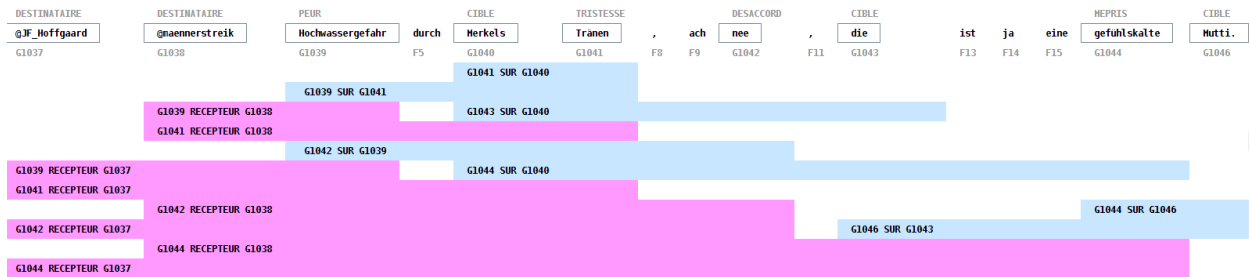


Ex. 11 ter Deux cibles sont présentes ici : « Liebeserklärung » et « an die Natur » dont la première est elle-même une OSE.



L'exemple 12 comporte une chaînes complexes de relations SUR en cascade. La première part de «geföhkalte» en passant par «Mutti», puis par le pronom relatif «die» pour aboutir à «Merkels». Ici la cible initiale «Mutti» est devenue éligibles pour être une SOURCEs de relation SUR, car elle est CIBLES de relation SUR (venant de «geföhkalte»). Notez que cette propriété, que l'on pourrait appeler transitivité, ne concerne que la relation SUR et pas la relation DIT, qui relie toujours directement une SOURCE à une OSEE. Les cascades de relation SUR sont portées en générales par les relations syntaxique de modifieur ou d'attribut, voire par le lien unissant un pronoms relatifs à son référent. Notez aussi ici dans cet exemple, le fait que l'OSEE est portée par une négation «nee», qui n'est pas annotée comme un group NEGATION mais comme une OSEE DESACCORD.

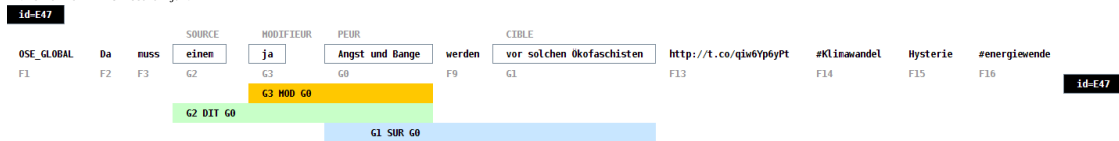
Ex. 12



MODIFIEUR

Seront annotés comme MODIFIEUR les termes qualifiant une intensité d'expression d'opinion, de sentiment ou d'émotion. Dans l'exemple suivant le déterminant exclamatif «Quelle» est considéré comme un intensifieur.

Ex. 13 Le MODIFIEUR est ici «ja».



Ex. 14 Le MODIFIEUR est le groupe nominal «ohne Grenzen».



NEGATION

Les termes indiquant la négation ne sont annotés que lorsqu'ils concernent une OSEE ou une CIBLE intermédiaire comme c'est le cas ici. Les négations indiquent alors soit une absence du sentiment exprimé par le groupe pointé par la relation de NEGATION, soit le sentiment opposé à celui exprimé par le groupe pointé par la négation.

Ex. 15 Un exemple de relations portée par «kein».

id-E175

OSE_GLOBALE	#Soziologie	Hartz	IV	'Eine #Gesellschaft der	#Angst	kann	kein	'bestes	Programm'	sein'
F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11

G760 SUR G764 G761 NEG G763 F13 SUR G762

Ex. 16 Un exemple où l'on n'annote pas de négation pour le mot « ohne » car cette négation ne porte pas directement sur une OSEE.

id-E201

OSE_GLOBALE	#annewill	Verbaler	Sondermüll	ohne Grenzen	aka #Merkel	!	Die	hat	sich	noch	nie	dafür	interessiert
F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14

G27 MOD G26 G26 SUR G28

ihren Antseid nachzukommen , siehe NSA. id-E301

Ex. 17 Un exemple de NEGATION où le sentiment de l'OSEE est inversé par la NEGATION. Sur ce point, il faut remarquer que l'on annote la valeur originale du sentiment pour l'OSEE (ici PLAISIR). Notez aussi que les deux particules de négation «n'» et «pas» sont annotées NEGATION.

id-E9

DEPLAISIR	CIBLE	PLAISIR	SOURCE	NEGATION	PEUR
OSE_GLOBALE	Meine Katze beobachtet mich, das	gefällt	mir	nicht	#angst
G63	G288	G14	G289	G15	G757

G289 DIT G14 G15 NEG G14

G14 SUR G288

Ex. 17 bis Un exemple de négation qui n'est pas une, la « ne » ici annule un fait hypothétique et exprime dans cet énoncé le désaccord.

DESTINATAIRE	DESTINATAIRE	PEUR	CIBLE	TRISTESSE	DESACCORD	CIBLE	MEPRIS	CIBLE							
@J_Hoffgaard	#anmenstreik	Hochwassergefahr	durch	Merkels	Tranen	,	ach	nee	,	die	ist	ja	eine	gefühlskalte	Mutti.
G1037	G1038	G1039	F5	G1040	G1041	F8	F9	G1042	F11	G1043	F13	F14	F15	G1044	G1046

G1039 SUR G1041 G1041 SUR G1040

G1039 RECEPTEUR G1038 G1043 SUR G1040

G1041 RECEPTEUR G1038 G1042 SUR G1039

G1039 RECEPTEUR G1037 G1044 SUR G1040

G1041 RECEPTEUR G1037 G1042 RECEPTEUR G1038 G1044 SUR G1046

G1042 RECEPTEUR G1037 G1044 RECEPTEUR G1038 G1046 SUR G1043

G1044 RECEPTEUR G1037

DESTINATAIRE

Les termes qui indiquent explicitement à qui l'on s'adresse sont rassemblés sous l'étiquette DESTINATAIRE.

Ex. 18 La DESTINATAIRE est l'identifiant twitter suivi du pronom «@Dagi_Bi Du». Comme le destinataire est aussi CIBLE d'une OSE il pourrait aussi recevoir une relation SUR, mais dans ce cas on considère que le relation RECEPTEUR subsume la relation SUR.

id-E305

OSE_GLOBALE	DESTINATAIRE	DEVALORISATION	PEUR	keiner	will	:D.
F1	F2	F3	F4	F5	F6	F7

G43 RECEPTEUR G42

Ex. 19 La DESTINATAIRE est ici interpellée par l'auteur du tweet qui fait une INSTRUCTION/DEMANDE. Notez que le destinataire reçoit une relation RECEPTEUR provenant d'une INSTRUCTION/DEMANDE.

id-E306

OSE_GLOBALE	Also	neine	VALORISATION	DESTINATAIRE	INSTRUCTION_DEMANDE
F1	F2	F3	F4	F5	F6

G45 SUR G44

G47 RECEPTEUR G44

entweder	alle	NEGATIF	Tornados	in	die	Luft	oder	neue...
F14	F15	F16	F17	F18	F19	F20	F21	F22

id-E306

Les classes sémantiques affectives fines

- 1.6.1 Annotation en classes sémantiques affectives fines
 - 1.6.1.1 DEPLAISIR
 - 1.6.1.2 DERANGEMENT
 - 1.6.1.3 MÉPRIS
 - 1.6.1.4 SURPRISE (NÉGATIVE)
 - 1.6.1.5 PEUR
 - 1.6.1.6 COLÈRE
 - 1.6.1.7 ENNUI
 - 1.6.1.8 TRISTESSE
 - 1.6.1.9 PLAISIR
 - 1.6.1.10 APAISEMENT
 - 1.6.1.11 AMOUR
 - 1.6.1.12 SURPRISE (POSITIVE)
 - 1.6.1.13 SATISFACTION
 - 1.6.1.14 INSATISFACTION
 - 1.6.1.15 ACCORD
 - 1.6.1.16 VALORISATION
 - 1.6.1.17 DESACCORD
 - 1.6.1.18 DEVALORISATION
 - 1.6.1.19 INSTRUCTION/DEMANDE
- 1.6.2 Annotation en classes génériques de polarité
 - 1.6.2.1 NÉGATIF
 - 1.6.2.2 POSITIF

Nous répertorions les différentes classes sémantiques affectives en trois grandes catégories:

- Opinions (intellectif): cette catégorie contient deux classes sémantiques positives : ACCORD et VALORISATION et deux classes sémantiques négatives : DÉACCORD et DÉVALORISATION. ex. Je suis d'accord avec toi.
- Sentiments (affectif-intellectif) : cette catégorie contient une classe sémantique positive : SATISFACTION et une classe sémantique négative : INSATISFACTION. ex. Je suis satisfait de ce travail.
- Émotions (affectif) : cette catégorie contient quatre classes sémantiques positives : PLAISIR, APAISEMENT, AMOUR et SURPRISE (POSITIVE) et huit classes sémantiques négatives : DÉPLAISIR, DERANGEMENT, MÉPRIS, TRISTESSE, PEUR, COLÈRE, ENNUI et SURPRISE (NÉGATIVE). ex. Je suis triste

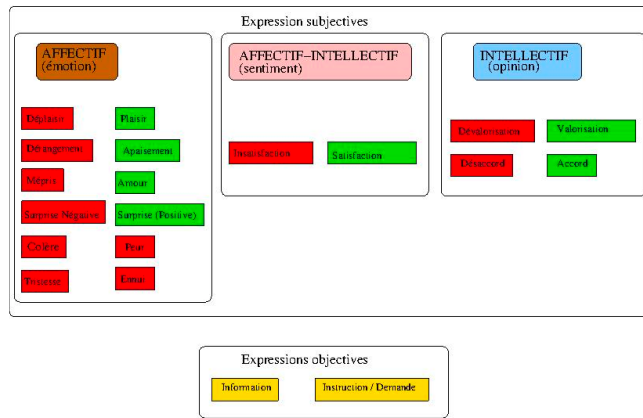


Fig 1. Les grandes catégories de classe sémantiques.

Dans les sections qui suivent, nous donnons pour chaque classe sa définition et les caractéristiques qui permettent de la distinguer des autres classes. Nous exprimerons les définitions en fonction de trois concepts qui décrivent l'état mental de la SOURCE qui ressent les opinions/sentiments/émotions :

1. **Perceptions, croyances et connaissances** sur le monde, mémoire des événements.
2. **Désirs** qui sont les souhaits qui résultent des pulsions instinctives de survie ou des préférences subjectives.
3. **Intentions** qui sont les actions que l'on se prépare à réaliser dans un avenir plus ou moins proche.

La détermination de l'OSEE ressentie par la SOURCE dépendra du point de vue que la SOURCE entretient sur les **événements** du monde, qui sont **localisés dan le temps** par rapport au **présent de la SOURCE**. Les différentes sortes d'OSEE seront distinguées les unes des autres en fonction de :

1. des **Désirs** de la SOURCE, considérés comme constants dans le temps,
2. des **Croyances** de la SOURCE qui évoluent en fonction des événements,
3. des **Intentions** de la SOURCE. Notez que la TRISTESSE et le PLAISIR nécessitent de distinguer des actions génériques un type d'actions spécifiques, dites de «signallement» de l'état interne de la SOURCE à destination de ses congénères (par exemple pleurer pour la TRISTESSE et rire pour le PLAISIR).
4. de la **localisation temporelle** des événements tels que perçus par la SOURCE, c'est à dire dans le **passé, le présent ou le futur du ressenti de l'OSEE**.

ÉTIQUETTE GÉNÉRIQUE DE LA CLASSE SÉMANTIQUE	TYPE	CLASSES SPÉCIFIQUES DE LA CLASSE SÉMANTIQUE
1 DÉPLAISIR	e-	déplaisir / déception
2 DÉRANGEMENT	e-	dérangement / embarras
3 MÉPRIS	e-	mépris / dédain / dégoût / haine
4 SURPRISE NEGATIVE	e-	surprise négative / étonnement négatif
5 PEUR	e-	peur / terreur / inquiétude / anxiété
6 COLÈRE	e-	colère / rage / agacement / exaspération / énervement / impatience
7 ENNUI	e-	ennui / torpeur
8 TRISTESSE	e-	tristesse / chagrin / souffrance / désespoir / résignation
9 PLAISIR	e+	plaisir / divertissement / joie / euphorie / bonheur / extase
10 APAISEMENT	e+	apaisement / soulagement / reconnaissance / pardon / sérénité
11 AMOUR	e+	amour / tendresse / affection / dévouement / passion / envie / désir
12 SURPRISE (POSITIVE)	e+	surprise (positive) / étonnement positif
13 SATISFACTION	s+	satisfaction / contentement / fierté
14 INSATISFACTION	s-	insatisfaction / mécontentement
15 ACCORD	o+	accord / compréhension / approbation
16 VALORISATION	o+	valorisation / intérêt / appréciation
17 DÉSACCORD	o-	désaccord / désapprobation
18 DÉVALORISATION	o-	dévalorisation / désintérêt / dépréciation
19 INSTRUCTION/DEMANDE	i	instruction / demande / requête / ordre / conseil / recommandation
20 INFORMATION	i	information

Tab.1- Tableau-1 récapitulatif des différentes classes sémantiques d'opinion/sentiment/émotion, e-=émotion négative, e+, s=sentiment, o=opinion, i=information, +=valeur positive, -=valeur négative.

OSEE	TYPE	type de cause	intentions, actions, connaissances
1 DÉPLAISIR	e-	événement présent non-désiré	pas d'intention de remédiation
2 DÉRANGEMENT	e-	événement présent non-désiré	intention de remédiation
3 MÉPRIS	e-	Existence d'un Désir (répulsion)	-
4 SURPRISE NEGATIVE	e-	événement présent non-désiré	absence de connaissance antérieure de l'événement
5 PEUR	e-	absence de remédiation pour un événement non-désiré	connaissance antérieure d'un événement similaire
6 COLÈRE	e-	événement présent ou passé non-désiré	action de (pseudo) remédiation
7 ENNUI	e-	absence d'événement désiré	-
8 TRISTESSE	e-	événement présent ou passé non-désiré	action de signallement
9 PLAISIR	e+	événement présent désiré	action de signallement
10 APAISEMENT	e+	2 événements consécutifs, l'un non-désiré, l'autre annulant les effets du précédent	-
11 AMOUR	e+	Existence d'un Désir (attraction)	-
12 SURPRISE (POSITIVE)	e+	événement présent (désiré)	absence de connaissance antérieure de l'événement
13 SATISFACTION	s+	Croyance	connaissance de la réalisation du désir
14 INSATISFACTION	s-	Croyance	absence de connaissance de la réalisation du désir
15 ACCORD	o+	similarité d'opinion de 2 SOURCES sur une CIBLE commune	Connaissance de l'opinion de l'autre SOURCE sur la CIBLE
16 VALORISATION	o+	accord des Connaissances sur la CIBLE les Désirs, Croyances ou Intentions	-
17 DÉSACCORD	o-	disimilarité d'opinion de 2 SOURCES sur une CIBLE commune	Connaissance de l'opinion de l'autre SOURCE sur la CIBLE
18 DÉVALORISATION	o-	désaccord des Connaissances sur la CIBLE les Désirs, Croyances ou Intentions	-
19 INSTRUCTION/DEMANDE	i	réponse ou action attendue	absence d'OSEE
20 INFORMATION	i	besoin de communication	absence d'OSEE

Tab.2- Tableau-2 récapitulatif des caractéristiques discriminantes pour les différentes classes sémantiques d'opinion/sentiment/émotion, selon la **cause** et les **intentions, actions, connaissances**.

1.6.1.1 DÉPLAISIR

Définition : Émotion suscitée par un événement passé ou présent qui est non-désiré par la SOURCE mais sans que la SOURCE ait l'intention d'y remédier.

Classes spécifiques : déplaisir / déception

Ex. 21 L'OSEE indiquant le DÉPLAISIR est l'adjectif «unangenehm».

id=E315

DESTINATAIRE	SOURCE	MODIFICATEUR	DEPLAISIR	SOURCE						
05E_GL08AL	gjnfringerle	diesen	Menschen	das Leben hier	so	unangenehm	wie möglich	zu	gestalten	-
F1	G94	F3	F4	G79	G82	G78	G83	F12	F13	F14
					G82 MOD G78		G83 MOD G78			
									G78 SUR G79	
										F9 RECEPTEUR 694

VALORISATION
 dies nennt nan dann erhaltung der kulturellen Werte...
 F15 F16 F17 F18 F19 F20 F21 G81

id-E315

Ex. 21 bis L'OSEE indiquant le DÉPLAISIR est l'adjectif «enttäuscht».

id-E316

SOURCE MODIFICATEUR DEPLAISIR CIBLE SOURCE DEPLAISIR
 OSE_GLOBALE Branche ist tief enttäuscht von EEG-Referentenwurf - Vertreter der Wind-und Bioenergie-Branche kritisieren ...
 F1 G86 F3 G85 G84 G87 F8 G88 G89 F15
 G85 MOD G84
 G86 DIT G84
 G84 SUR G87
 G88 DIT G89

1.6.1.2 DÉRANGEMENT

Définition : Émotion suscitée par un évènement non-désiré par la SOURCE engendrant une Intention d'action de remédiation.

Classes spécifiques : dérangement / embarras

Ex. 22 L'OSEE indiquant le DÉRANGEMENT est le nom «Nuisance».

CIBLE DERANGEMENT MODIFICATEUR SOURCE
 Éoliennes Nuisances graves pour la santé. témoignages du monde entier http://leventtourne.org/eoliennes-nuisances-graves-pour-la-sante-tenoignages-du-monde-entier/ ...
 G827 G824 G825 F5 F6 F7 G826
 G825 MOD F3
 G824 SUR G827
 G826 DIT G824

#anti #éolien #Éoliennes #Santé #Ténoignage
 F14 F15 F16 F17 F18

1.6.1.3 MÉPRIS

Définition : Émotion suscitée par une connaissance sur une CIBLE humaine en désaccord avec les désirs et/ou belief de la SOURCE.

Classes spécifiques : mépris / dédain / dégoût (pour une CIBLE humaine) / haine

Ex. 23 L'OSEE indiquant le MÉPRIS est «Hass-Propaganda».

id-E191

MÉPRIS DESACCORD MÉPRIS SOURCE DESACCORD MÉPRIS
 OSE_GLOBALE Stopp Hass-Propaganda i Bundesweite #Kampagne gegen die Verbreitung von #Hass-#Propaganda
 G418 G765 G416 F4 G766 G767 F8 F9 F10 G417
 G765 SUR G416
 G766 DIT G767
 G767 SUR G417

in den sozialen Netzwerken.

F12 F13 F14 F15 id-E191

1.6.1.4 SURPRISE NÉGATIVE

Définition : Émotion suscitée par la découverte d'un évènement non-désiré et non-attendu par la SOURCE.

Classes spécifiques : surprise négative / étonnement négatif

Ex. 25 L'OSEE indiquant la surprise négative est le nom «Unfall».

id-E317

SOURCE MODIFICATEUR SURPRISE_NEGATIVE
 OSE_GLOBALE "In jeden Atonkraftwerk ist ein schwerer Unfall möglich" -
 F1 G92 F5 F6 G90 G91 F9 F11
 G90 MOD G91
 G91 SUR G92
 G91 DIT G91
 SOURCE
 möglich" - Studie des @bundnaturschutz https://t.co/I2xrW0uShw https://t.co/k8a5HZLc0q.
 F9 F10 G93 id-E317

1.6.1.5 PEUR

Définition : Émotion suscitée par la (possible) réalisation d'un évènement non-désiré par la SOURCE et pour lequel la SOURCE ne connaît pas d'action de remédiation.

Classes spécifiques : Peur / terreur / inquiétude / anxiété

Ex. 26 L'OSEE indiquant la peur est «Angst und Bange».

id-E47

SOURCE MODIFICATEUR PEUR CIBLE
 OSE_GLOBALE Da muss einem ja Angst und Bange werden vor solchen Okofaschisten http://t.co/qiv6Yp6yPt #Klimawandel Hysterie #energievende
 F1 F2 F3 G2 G3 G0 F9 G1 F13 F14 F15 F16 id-E47
 G3 MOD G0
 G2 DIT G0
 G1 SUR G0

Ex. 26 bis L'OSEE indiquant la peur est «Sorge».

id-E186

PEUR PEUR SOURCE CIBLE DEVALORISATION
 OSE_GLOBALE Sorge von Sozialphobikern scheint unbegründet, dass ihre Freundschaften oberflächlich sind.
 G403 G401 G402 F5 F6 F7 F8 G405 G404 F12
 G402 DIT G401
 G404 SUR G405

1.6.1.6 COLÈRE

Définition : Émotion suscitée par un évènement non-désiré engendrant une réaction chez la SOURCE qui engendre une Croyance nouvelle donnant à la SOURCE l'impression qu'une action de remédiation existe (impression de reprise du contrôle de la situation).

Classes spécifiques : colère / rage / agacement / exaspération / énervement / impatience

Ex. 27 L'OSEE indiquant la COLÈRE est le terme «Ärger».

SOURCE		COLERE		CIBLE		PEUR		CIBLE	
OSE_GLOBALE	Spiegel	online	berichtet	Ärger	über Großwildjägerin	"Giraffen	sind	gefährliche	Tiere"
F1	G9	F3	F4	F5	G4	F9	F10	F11	G7
				G4 SUR G5					
				G9 DIT G4		G7 SUR G8			

1.6.1.7 ENNUI

Définition : Émotion suscitée par l'absence d'événement désiré par la source.
Classes spécifiques : ennui / torpeur

Ex. 28 L'OSEE indiquant l'ennui est l'adjectif « langweilig ».

CIBLE		ENNUI		NEGATIF	
OSE_GLOBALE	Geo	ist	Langweilig.	Und	draußen
F1	G49	F3	G48	F5	F6
		G48 SUR G49			
				id=E307	

1.6.1.8 TRISTESSE

Définition : Émotion suscitée par un événement passé ou présent en désaccord avec les Désirs de la SOURCE et pour lequel il n'existe pas d'action de remédiation connue de la SOURCE. Cette connaissance suscite de la part de la SOURCE une intention de signaler son désarroi aux autres par un changement d'attitude (par ex. pleurs, posture de faible énergie, etc.).
Classes spécifiques : tristesse / chagrin / souffrance / désespoir / résignation

Ex. 29 L'OSEE indiquant la tristesse est l'adjectif « traurig » avec son émoticône « :-(- ».

SOURCE		INSTRUCTION DEMANDE		TRISTESSE	
OSE_GLOBALE	"AfD-Parteiprogramm : Mehr Polizei, mehr Waffen, mehr Schadstoffe"	Was ist bloß los mit unseren Nachbarn	,	traurig :	https://t.co/08514Y0060.
F1	G62	G63	F19	G61	F22
		G61 SUR G62			

1.6.1.9 PLAISIR

--> **Définition :** Émotion suscitée par un événement (présent ou future) désiré par la SOURCE et pour lequel la SOURCE va avoir l'intention de signaler sa joie aux autres par un changement d'attitude (par ex. sourire, posture énergique, etc.).
Classes spécifiques : plaisir / divertissement / joie / euphorie / bonheur / extase

Ex. 30 L'OSEE indiquant le PLAISIR est le nom « Forfreude ».

SOURCE		MODIFICATEUR		PEUR		CIBLE	
OSE_GLOBALE	Irgendwie	hab	ich	ja	schon	ein wenig	#Angst
F1	G9	F2	F3	F5	F6	G755	G7
		G291				G290	
		G755 MOD G7				G7 SUR G290	
		G291 DIT G7				G8 SUR G290	
		G291 DIT G8					
CIBLE		MODIFICATEUR		PEUR		CIBLE	
OSE_GLOBALE	aber	die	Vorfreude	ist	stärker	#berlinmarathon15	#GoFor42
F14	F15	G8	F17	G756	F19	F20	
		G756 MOD G8					
						id=E4	

Ex. 30 bis L'OSEE indiquant le PLAISIR est le verbe « gefällt » avec une négation.

SOURCE		MODIFICATEUR		PEUR	
OSE_GLOBALE	Meine Katze beobachtet mich, das	gefällt	nir	nicht	#angst
F1	G63	G288	G14	G289	G15
		G289 DIT G14			
		G15 NEG G14			
		G14 SUR G288			
				id=E9	

1.6.1.10 APAISEMENT

Définition : Émotion suscitée par la connaissance par la SOURCE de deux événements consécutifs, le premier en désaccord avec ses Désirs et le second annulant les effets du précédent.
Classes spécifiques : apaisement / soulagement / reconnaissance / pardon / sérénité

Attention : Lorsque les deux événements sont mentionnés dans l'énoncé, chacun sera identifié comme une CIBLE d'un relation SUR partant de l'OSEE d'APAISEMENT qui aura donc deux cibles.

Notez que l'APAISEMENT implique la SATISFACTION mais ne l'inclut pas pour autant.

Ex. 31 L'OSEE indiquant l'APAISEMENT le participe passé « beruhigt ».

SOURCE		APPAPSEMENT		CIBLE		NEGATION		NEGATIF	
OSE_GLOBALE	Alle die bei uns wegen Hochwasser anfragen	können	beruhigt	sein	,	Halle	wird	nicht	in
F1	G21	F9	G22	F11	F12	G23	F14	G24	F16
		G22 DIT G22				G24 NEG G25			
						G25 DIT G23			

1.6.1.11 AMOUR

Définition : Émotion suscitée par l'attraction empathique de la SOURCE pour la CIBLE qui est humaine ou animale.
Classes spécifiques : amour / tendresse / affection / dévouement / passion / envie / désir

Ex. 32 L'OSEE indiquant l'AMOUR est ici le terme lui-même « Liebe ».

SOURCE		AMOUR		SOURCE		CIBLE	
OSE_GLOBALE	Die Meinung	:	Obanas	Liebe	fur die Umwelt	-	Thüringer
F1	F2	F3	F4	G13	G14	G15	F10
		G13 DIT G14		G14 SUR G15			
						id=E95	
SOURCE		AMOUR		CIBLE		NEGATIF	
OSE_GLOBALE	#Umweltdaten2015	:	Die deutsche	Liebe	zu mehr #PS	treibt	den
F1	F2	F3	G12	G10	G11	F10	F11
		G10 SUR G11					
		G12 DIT G10					

Ex. 32 bis L'OSEE indiquant l'AMOUR est ici le terme « Liebeserklärung ».

id=E282

OSE_GLOBAL Constantin Film : **Wunderbare** **Liebeserklärung** **an die Natur** : AFRIKA - DAS MAGISCHE KÖNIGREICH ab. http://t.co/rxqerLw4

F1 F2 F3 F4 F5 F6 F7 F8 F9 F10 F11 F12 F13 F14 F15 F16 F17

VALORISATION AMOUR CIBLE

G16 G17 G20

G16 SUR G17 G17 SUR G20

1.6.1.12 SURPRISE (POSITIVE)

Définition : Émotion suscitée par la découverte d'un événement (désiré) non-attendu par la SOURCE. On classera aussi comme positives, les surprises qui ne sont pas en rapport avec aucun désir de la SOURCE.

Classes spécifiques : amour / tendresse / affection / dévouement / passion / envie / désir

Ex. 33 La SURPRISE POSITIVE est exprimée par « Gute Nachrichten »

id=E9

OSE_GLOBAL **Gute Nachrichten** **für @kosmonautfstvl** **das Gelände an Stausee wird nun Jahr für Jahr festivalsauglicher** #StadtratC.

F1 F2 F3 F4 F5 F6 F7 F8 F9 F10 F11 F12 F13 F14 F15 F16 F17

SURPRISE_POSITIVE SOURCE CIBLE

G58 G59 G60

G59 DIT G58 G58 SUR G60

Ex. 33 bis La SURPRISE POSITIVE est exprimée par « erstaunlichen », classé ici comme marqueur de surprise (plutôt neutre, donc annoté comme surprise positive).

id=E8

OSE_GLOBAL **An BER** **hilft** **nun sogar der Klimawandel.** **Alle** **Verantwortlichen** halten...

F1 F2 F3 F4 F5 F6 F7 F8 F9 F10 F11 F12 F13 F14 F15 F16 F17

SOURCE ACCORD SOURCE CIBLE INSATISFACTION

G51 G52 G53 G56 G57 F11

G51 DIT G52 G52 SUR G53 G57 SUR G56

mit teils **erstaunlichen** **Argumenten** fest.

F12 F13 F14 F15 F16 F17

SURPRISE_POSITIVE CIBLE

G54 G55

G54 SUR G55

1.6.1.13 SATISFACTION

Définition : Sentiment suscité par la Connaissance de la réalisation d'un Désir.

Classes spécifiques : satisfaction / contentement / fierté

Ex. 34 Ici L'OSEE exprimant la SATISFACTION est: le verbe « erfreut ».

id=E312

OSE_GLOBAL **"Die prognostizierte Klimaerwärmung scheint sich zu bewahrheiten. Dies** **erfreut** **uns** **naturlich, da** **wir**

F1 F2 F3 F4 F5 F6 F7 F8 F9 F10 F11 F12 F13 F14 F15 F16 F17

CIBLE SATISFACTION SOURCE SOURCE

G70 G72 G71 G69

G71 DIT G72 G72 SUR G70 G69 DIT G67

dadurch immer **bessere** **Weine** **produzieren"**

F14 F15 F16 F17

VALORISATION CIBLE

G67 G68

G67 SUR G68

1.6.1.14 INSATISFACTION

de remédiation--> **Définition** : Sentiment suscité par la Connaissance de l'impossibilité de réaliser un Désir.

Classes spécifiques : insatisfaction / mécontentement

Ex. 35 L'OSEE indiquant l'INSATISFACTION est le groupe verbal « unzufrieden ».

id=E311

OSE_GLOBAL **Aufschrei** : **Umweltschützer** **sind** **unzufrieden** : **Wolfratshausen – Das geplante Kraftwerk und Baumfällungen sorgen** ...

F1 F2 F3 F4 F5 F6 F7 F8 F9 F10 F11 F12 F13 F14 F15 F16 F17

SOURCE INSATISFACTION CIBLE

G65 G64 G66

G65 DIT G64 G64 SUR G66

1.6.1.15 ACCORD

Définition : La SOURCE a connaissance d'un autre SOURCE qui entretient une opinion similaire à la sienne sur la CIBLE.

Classes spécifiques : accord / compréhension / approbation

Ex. 36 L'OSEE indiquant l'accord est le nom « einigung ».

id=E313

OSE_GLOBAL **Keine** **Einigung** : **Koalition** **verschiebt** **Votum** **über** **Fracking-Gesetz** **https://t.co/AoTsva00bb** **via** **faznet.**

F1 F2 F3 F4 F5 F6 F7 F8 F9 F10 F11 F12 F13 F14 F15 F16 F17

NEGATION ACCORD

G74 G73 G74 G73

F2 NEG. F3

Ex. 36 bis L'OSEE indiquant l'accord est le nom « vereinbaren ».

id=E314

OSE_GLOBAL **Stadt und Landkreis Würzburg** **vereinbaren** **engere Kooperation in Klimaschutz** **https://t.co/R8DJ6I1515.**

F1 F2 F3 F4 F5 F6 F7 F8 F9 F10 F11 F12 F13 F14 F15 F16 F17

SOURCE ACCORD CIBLE

G75 G76 G77

G75 DIT G76 G76 SUR G77

1.6.1.16 VALORISATION

C'est l'émotion duale de l'AMOUR mais appliquée aux inanimés concrets ou abstraits.

Définition : les connaissances de la SOURCE sur la CIBLE (matérielle inanimée ou abstraite) sont en accord avec ses Croyances, Désirs ou Intentions.

Classes spécifiques : valorisation / intérêt / appréciation

Ex. 37 L'OSEE indiquant la valorisation est: « Wunderbar ».

id=E282

OSE_GLOBAL Constantin Film : **Wunderbare** **Liebeserklärung** **an die Natur** : AFRIKA - DAS MAGISCHE KÖNIGREICH ab. http://t.co/rxqerLw4

F1 F2 F3 F4 F5 F6 F7 F8 F9 F10 F11 F12 F13 F14 F15 F16 F17

VALORISATION AMOUR CIBLE

G16 G17 G20

G16 SUR G17 G17 SUR G20

1.6.1.17 DÉSACCORD

Définition : La SOURCE a connaissance d'au moins une autre SOURCE (qui peut être implicite dans l'énoncé, mais néanmoins déduite/impliquée de/par ce dernier), qui entretient une opinion différente de la sienne sur la CIBLE. Notez que toutes les personnes en désaccord sont des SOURCES et que la CIBLE (l'objet du désaccord) peut être absent de l'énoncé).

Classes spécifiques : désaccord / désapprobation

Ex. 38 Ici il y a deux expressions de DESACCORD, le verbe à l'impératif « Stoppt » et la préposition « gegen ». Notez que lorsqu'une instruction se superpose à une OSE nous annoterons uniquement l'OSEE, comme c'est le cas ici avec le verbe à l'impératif « Stoppt » qui est annoté DESACCORD.

id-E191

HEPRIS	DESACCORD	HEPRIS	SOURCE	DESACCORD	HEPRIS
OSE_GLOBALE	Stoppt	Hass-Propaganda	Bundesweite #Kampagne	gegen	#Hass-#Propaganda
G418	G765	G416	F4 G766	G767	F8 F9 F10 G417

G765 SUR G416 G766 DIT G767 G767 SUR G417

in den sozialen Netzwerken.

F12 F13 F14 F15 **id-E191**

Ex. 38 bis Un exemple de DESACCORD porté par une négation (« ne ») qui ici annule un fait hypothétique et exprime dans cet énoncé le desaccord.

DESTINATAIRE	DESTINATAIRE	PEUR	CIBLE	TRISTESSE	DESACCORD	CIBLE	ist	ja	eine	HEPRIS	CIBLE
JF_Hoffgaard	onaemerstreik	Hochwassergefahr	MerkeIs	Tranen	nee	die				gefühlskalte	Mutti.
G1037	G1038	G1039	F5 G1040	G1041	F8 F9 G1042	F11 G1043	F13	F14	F15	G1044	G1046

G1039 SUR G1041 G1041 SUR G1040

G1039 RECEPTEUR G1038 G1043 SUR G1040

G1041 RECEPTEUR G1038

G1042 SUR G1039 G1044 SUR G1040

G1039 RECEPTEUR G1037 G1044 SUR G1046

G1041 RECEPTEUR G1037

G1042 RECEPTEUR G1038

G1044 RECEPTEUR G1038

G1044 SUR G1043 G1044 SUR G1046

G1044 RECEPTEUR G1037

1.6.1.18 DÉVALORISATION

uniquement aux entités animées.--> **Définition** : les connaissances de la SOURCE sur la CIBLE matérielle (inanimée) ou abstraite sont en désaccord avec ses Croyances, Désirs ou Intentions.
Classes spécifiques : dévalorisation / désintérêt / dépréciation / dégoût (pour une CIBLE matérielle, e.g. nourriture)

Ex. 39 L'expression de dévalorisation est ici portée par l'adjectif « oberflächlich »

id-E186

PEUR	PEUR	SOURCE	CIBLE	DEVALORISATION
OSE_GLOBALE	Sorge	von Sozialphobikern	ihre Freundschaften	oberflächlich
G403	G401	G402	F5 F6 F7 F8 G405	G404

G402 DIT G401 G404 SUR G405

1.6.1.19 INSTRUCTION/DEMANDE

Définition : OSEE exprimant un ordre, une demande, une question ou une recommandation.
Classes spécifiques : instruction / demande / requête / ordre / conseil / recommandation

Ex. 40 Il s'agit d'une question.

id-E121

INSTRUCTION_DEMANDE	INSTRUCTION_DEMANDE	PEUR	INSTRUCTION_DEMANDE	INSTRUCTION_DEMANDE	DESTINATAIRE	CIBLE
OSE_GLOBALE	Was ist	#Angst	eigentlich ?	Brauchst	du	die
G223	G221	G222	G220	G775	G776	G777

G221 SUR G222 G220 SUR G222 G775 RECEPTEUR G775

G775 SUR G777

id-E121

Ex. 41 Il s'agit d'une requête ou recommandation adressée à une personne.

id-E306

OSE_GLOBALE	Also	neine	VALORISATION	DESTINATAIRE	INSTRUCTION_DEMANDE
F1	F2	F3	Liebe	Frau Ministerin	dass müssen Sie nu befehlen
G45	G45	G44	G44	G47	F13

G45 SUR G44 G47 RECEPTEUR G44

entweder	alle	NEGATIF	Tornados	in	die	Luft	oder	neue...
F14	F15	G46	F17	F18	F19	F20	F21	F21

id-E306

Ex. 42 Il s'agit d'une requête ou recommandation aux lecteurs en général.

id-E13

COLERE	#Mut	#Glaube	INSTRUCTION_DEMANDE	DESTINATAIRE	NEGATION	COLERE	COLERE	INSTRUCTION_DEMANDE
OSE_GLOBALE	#Mut	#Glaube	Geben	Sie sich	nicht	der	Wut	oder
G29	F2	F3	G280	G281	G282	F9	G27	F11

G280 RECEPTEUR G281

G282 NEG G280

G280 SUR G27

G280 SUR G28

G759 MOD G280

Notez que lorsqu'une instruction se superpose à une OSE nous annoterons uniquement l'OSEE, comme c'est le cas ici avec le verbe à l'impératif « Stoppt » qui est annoté DESACCORD.
 Ex. 40 bis Exemple d'INSTRUCTION/DEMANDE non annotée comme telle (verbe « Stoppt »).

id-E191

HEPRIS	DESACCORD	HEPRIS	SOURCE	DESACCORD	HEPRIS
OSE_GLOBALE	Stoppt	Hass-Propaganda	Bundesweite #Kampagne	gegen	#Hass-#Propaganda
G418	G765	G416	F4 G766	G767	F8 F9 F10 G417

G765 SUR G416 G766 DIT G767 G767 SUR G417

in den sozialen Netzwerken.

F12 F13 F14 F15 **id-E191**

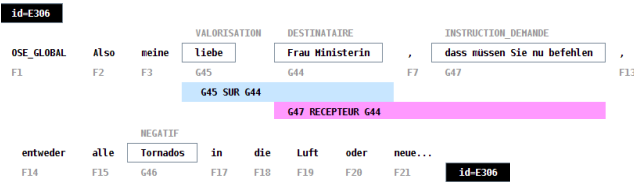
1.6.2 Annotation en classes génériques de polarité

Dans le cas où il est difficile d'identifier la catégorie sémantique de l'expression d'opinion/sentiment/émotion, on indiquera uniquement la polarité de l'expression (NÉGATIF ou POSITIF).

1.6.2.1 NÉGATIF

On annotera NÉGATIF les expressions d'opinion/sentiment/émotion qui ont une polarité négative et dont il est difficile d'identifier la classe sémantique exacte.

Ex. 43 Le nom « Tornado » est connoté négativement sans que le contexte puisse préciser s'il s'agit plus de PEUR que de DÉRANGEMENT



1.6.2.2 POSITIF

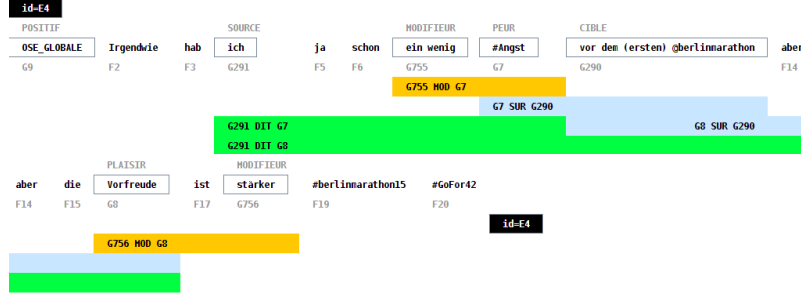
On annotera POSITIF les expressions d'opinion/sentiment/émotion qui ont une polarité positive et dont il est difficile d'identifier avec certitude la classe sémantique exacte.

Les relations

DIT

La relation DIT met **toujours** en rapport une SOURCE avec une OSEE.

Ex. 44 Ici la source est référencée par le pronom «ich» et porte deux relations DIT vers deux émotions antagonistes.



SUR

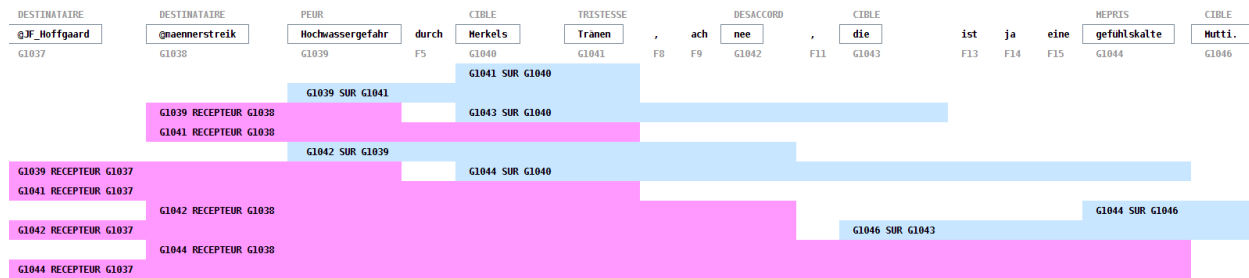
La relation SUR permet de mettre en rapport une OSEE avec une CIBLE ou bien de propager la relation par le biais de CIBLEs intermédiaires. Dans ce dernier cas, il peut y avoir une chaîne, de relations SUR (portée par des fonctions syntaxiques de modification), partant d'une OSEE pour aboutir à la CIBLE à laquelle fait effectivement référence l'OSEE initiale.

Ex. 44 bis Deux relations SUR, dont la cible de la seconde est un group verbal.



L'exemple 45 comporte une chaînes complexes de relations SUR en cascade. La première part de «geföhkalte» en passant par «Mutti», puis par le pronom relatif «die» pour aboutir à «Merkels». Ici la cible initiale «Mutti» est devenue éligibles pour être une SOURCES de relation SUR, car elle est CIBLES de relation SUR (venant de «geföhkalte»). Notez que cette propriété, que l'on pourrait appeler transivité, ne concerne que la relation SUR et pas la relation DIT, qui relie toujours directement une SOURCE à une OSEE. Les cascades de relation SUR sont portées en générales par les relations syntaxique de modifieur ou d'attribut, voire par le lien unissant un pronom relatifs à son référent. Notez aussi ici dans cet exemple, le fait que l'OSEE est portée par une négation «nee», qui n'est pas annotée comme un group NEGATION mais comme une OSEE DESACCORD.

Ex. 45 Une chaînes complexes de relations SUR en cascade. La première part de «geföhkalte» en passant par «Mutti», puis par le pronom relatif «die» pour aboutir à «Merkels». Ici la cible initiale «Mutti» est devenue éligibles pour être une SOURCES de relation SUR, car elle est CIBLES de relation SUR (venant de «geföhkalte»).



MOD

La relation MOD permet de mettre en rapport les termes (groupe MODIFIEUR) qualifiant une intensité d'OSEE, elle relie donc un MODIFIEUR à une OSEE; dans le cas de l'allemand elle est aussi utilisée pour relier une particule en fin de phrase à son verbe, la particule est alors contenue dans un groupe de même nature que le verbe.

Dans l'exemple 46, la particule «hin» (contenu dans un groupe INSTRUCTION/DEMANDE comme le verbe qu'elle modifie) est reliée au verbe «Geben» par une relation MOD.

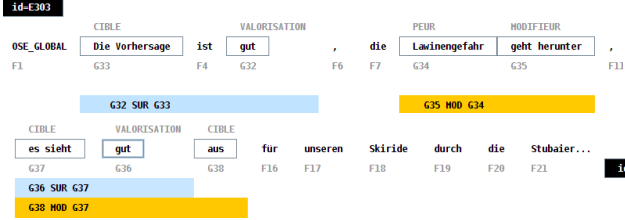
Ex. 46 La relation MOD va de la particule «hin» au verbe «Geben» qu'elle modifie.



Ex. 47 Nous avons ici deux relations MOD, portées respectivement par les modifieurs «ein wenig» et «stärker»



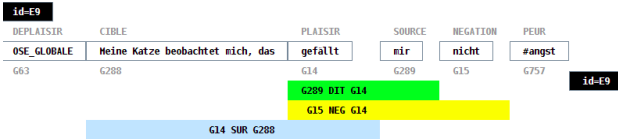
Ex. 47 bis Dans ce exemple, le modifieur est une expression verbale « geht herunter ». Remarquez aussi, l'emploi du groupe MODIFICATEUR et de la relation MOD pour relier la particule « aus » au verbe « sieht ».



NEG

La relation NEG permet de mettre en rapport les éventuels marqueurs de négation avec l'OSEE dont ils modifient la sémantique. Elle part d'un groupe NEGATION pour aller vers une OSEE ou bien une CIBLE intermédiaire.

Ex. 48 La relation NEG relie « nicht » au verbe « gefällt ».



Ex. 49 Un exemple de relations portée par « kein ».



RÉCEPTEUR

La relation RÉCEPTEUR va de l'OSEE ou d'une CIBLE au DESTINATAIRE de l'expression de subjectivité ou l'expression d'une INSTRUCTION/DEMANDE.

Ex. 18 Le DESTINATAIRE est l'identifiant twitter suivi du pronom «@Dagi_Bi Du». Comme le destinataire est aussi CIBLE d'une OSE il pourrait aussi recevoir une relation SUR, mais dans ce cas on considère que le relation RECEPTEUR subsume la relation SUR.



Ex. 19 La DESTINATAIRE est ici interpellée par l'auteur du tweet qui fait une INSTRUCTION/DEMANDE. Notez que le destinataire est ici recepneur



3.6 Expert Annotation Software

A specific annotation stand-alone annotation interface with the accompanying text importer has been developed by customizing the Pasta annotation interface for syntactic annotation developed in the PASSAGE project ([Vilnat et al., 2010]). The text importer is written in C++ and converts raw text into xml suitable to be imported in Pasta. The Pasta interface is written in Java.

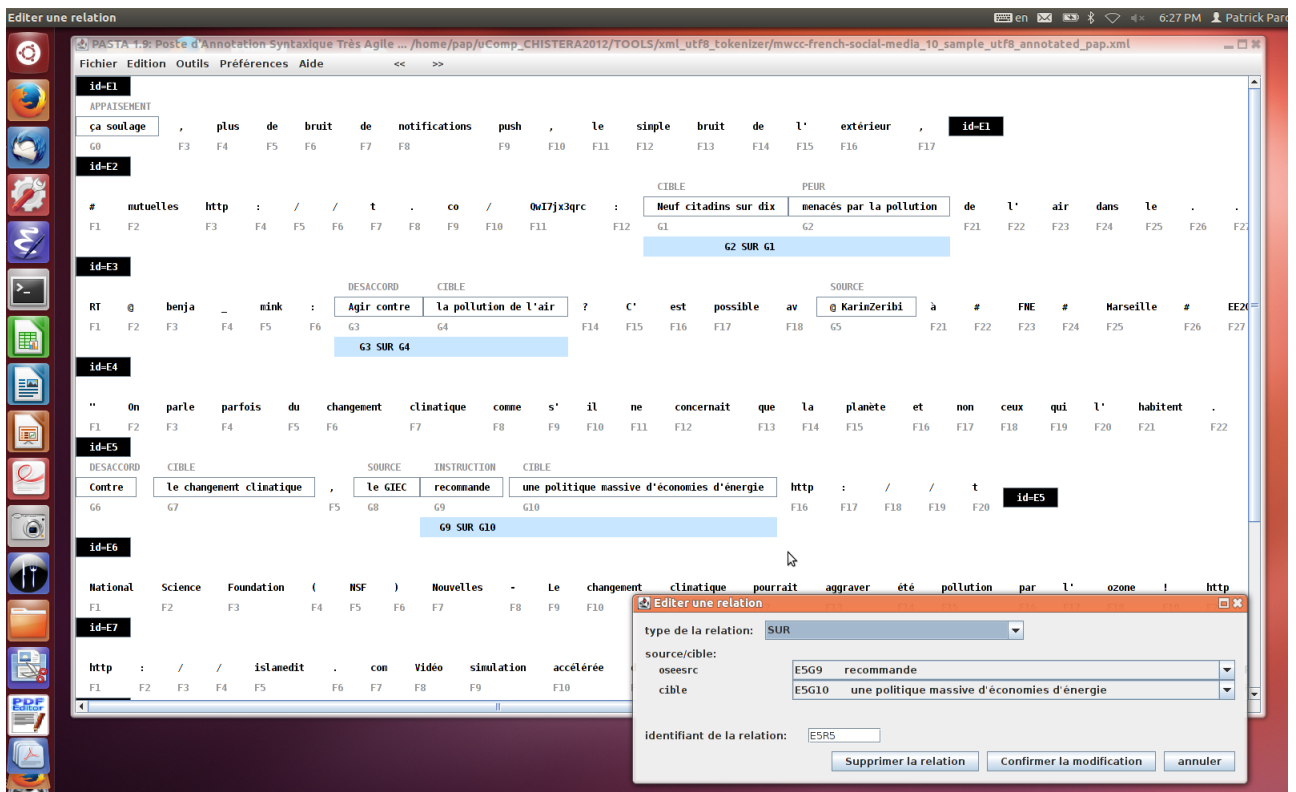



Figure 12: Pasta-uComp annotation interface.

3.7 Evaluation Campaign Deployment

Two evaluation campaigns for sentiment analysis were planned in uComp as test field for HC computations, one on the French language and the other on the German language. Because of the rescheduling of some of the uComp tasks, only the evaluation campaign on French took place within the course of the project. For the evaluation campaign on German, everything is ready for deployment:

- the guideline have been adapted to German, after having been revised following the results of the French evaluation campaign,
- a gold standard corpus of German Tweet on climate change has been hand annotated by experts (15k tweets). Up to know it has only been disclosed to uComp partners, but it

	<p>CHIST-ERA</p>	<p>Subproject : WP5 Task : T5.3 Date : August 9, 2016 Page : 54</p>
---	------------------	--

will be made available for everybody during the German evaluation campaign due start before the end of the year.

- a scientific committee for the evaluation campaign is being organized, already Diana Maynard (U. Scheffeld), Marie-Anne Moreau (INALCO/Paris) and Patrick Paroubek (LIMSI-CNRS) have agreed to be part of it.
- the tentative calendar for the German evaluation is the following: announcement will be done in september 2016, the training and development phase will take place in the fall, evaluation is scheduled to happen early in 2017 and the closing workshop will happen between spring and summer 2017.

3.7.1 French Language Sentiment Analysis Evaluation Campaign

Introduction

The evaluation campaign on French language was organized in collaboration with the organizers of the series of DEFT challenges on French text mining as the 2015 event. In the past, the DEFT challenge series had already addressed opinion and sentiment analysis for their shared tasks at two occasions:


- **DEFT 2007** [Grouin et al., 2007, Grouin et al., 2009b]
 - The corpus was made of parliamentary debates retranscriptions, with texts of relatively refined register with long sentences.
 - The task was to determine the opinion expressed in the message.
- **DEFT 2009** [Grouin et al., 2009a]
 - The corpus was made of parliamentary debates and newspaper articles (chronicles, analyses and debates, surveys and studies on particular issues).
 - One task was about identifying the relative objectivity/subjectivity of an article.
 - For the other task the participants had to identify opinion bearing expressions
Note that for this campaign the gold standard had been built by combining all the output of the participating systems.

The corpus

The corpus was made of French tweets about climate change. There were 15 000 messages from Twitter, essentially holding short sentences, with a relatively high frequency of acronyms specific to Internet social media (*MDR*, *STP*, *klk1*, *kan*, *dla*). Depending on the author, the language register was more or less high, with the presence of emoticons or smileys in addition to the keywords specific to twitter like, *hashtags* as thematic marker (*#Irak*, *#NoControlDay*) or as an indication of sarcasm (*#FeedTheTroll*) as well as some user identifiers (*@CNRS*).

The 2015 DEFT/uComp task

There were four tracks spread across 3 main tasks described as follows:

	<p>CHIST-ERA</p>	<p>Subproject : WP5 Task : T5.3 Date : August 9, 2016 Page : 55</p>
---	------------------	--

- **Task 1 – Identifying the Message Polarity**

- *positive (+)*
- *neutral or mixed (=)*
messages with a neutral polarity (objective content) or holding a mix of both positive and negative polarity roughly counterbalancing each other.
- *negative (-)*

- **Task 2.1 – Identifying the Generic Global Opinion, Sentiment or Emotion Class of the messages**

- *opinion, sentiment, emotion, information (objective content)*

- **Task 2.2 – Identifying the Specific Global Opinion, Sentiment or Emotion Class of the messages**

- *agreement, love, appeasement, anger, displeasure, upset, disagreement, devalorisation, boredom, dissatisfaction, contempt, fear, pleasure, satisfaction, negative surprise, positive surprise, sadness, valorisation*

- **Task 3 – Identification of Source, Target and Expression of Opinion, Sentiment or Emotion at word level.**

- Fine grained identification of word groups making the the source, the target and the sentiment expression in a message their precise category labelling (specific opinion, sentiment or emotion class) and identification of their relation links (matching the source and target with the corresponding expression).

The annotations for the three first tracks were built automatically from the annotations for the finer information grain track: Task 3 (Identification of Source, Target and Expression of Opinion, Sentiment or Emotion at word level) which had been hand annotated by experts according to the annotation guidelines presented previously for French in section 3.5.1. In the annotation protocol there was a double annotation for 500 messages with kappa estimation of inter-annotator agreement [Artstein and Poesio, 2008]. For some sentiment classes it was particularly difficult to reach an agreement among annotators, in particular for *satisfaction*, *devalorisation* or *anger*. The following tables provides the details of the annotation distribution per category, task and subcorpus.

Task	Category	Training	Test
T1	+	2464 (31,08%)	1057 (31,28%)
	-	1894 (23,89%)	804 (23,79%)
	=	3571 (45,04%)	1518 (44,92%)
T2.1	Emotion	826 (12,23%)	351 (10,39%)
	Information	3571 (52,87%)	1518 (44,92%)
	Opinion	2275 (33,68%)	973 (28,80%)
	Sentiment	82 (1,21%)	537 (15,89%)
T2.2	Accord	154 (4,84%)	67 (4,92%)
	Love	8 (0,25%)	4 (0,29%)
	Apeasement	9 (0,28%)	5 (0,37%)
	Anger	210 (6,60%)	87 (6,39%)
	Displeasure	47 (1,48%)	21 (1,54%)
	Dérangement	13 (0,41%)	6 (0,44%)
	Désaccord	216 (6,79%)	92 (6,76%)
	Dévalorisation	401 (12,60%)	170 (12,49%)
	Ennui	4 (0,13%)	2 (0,15%)
	Insatisfaction	9 (0,28%)	5 (0,37%)
	Mépris	176 (5,53%)	75 (5,51%)
	Peur	274 (8,61%)	114 (8,38%)
	Plaisir	35 (1,10%)	15 (1,10%)
	Satisfaction	73 (2,29%)	32 (2,35%)
	Surprise_négative	10 (0,31%)	4 (0,29%)
	Surprise_positive	4 (0,13%)	2 (0,15%)
	Tristesse	36 (1,13%)	16 (1,18%)
Valorisation	1504 (47,25%)	644 (47,32%)	

Table 32: Details of the annotation distribution per category, task and subcorpus.

Campaign schedule

- 16 February 2015 : participant registration opened and training data were made available.
 - registration period from 16 February to 1st of May, most of the participants joined in March.
 - 23 teams registered :
 - * deux from industry (Proxem et Synapse Développement)
 - * a non-French academic team (TU Moldova)
- the tests took place from the 4th of May to the 10th of Mai 2015 :

- each participant had a 3 day window to complete its runs that he could choose at will during this week
- runs were submitted by 12 teams
(there was a maximum of 3 runs per team per task)
 - * Task 1 : 27 runs from 12 teams
 - * Task 2.1 : 24 runs from 9 teams
 - * Task 2.2 : 21 runs from 7 teams
 - * Task 3 : no participant
- from the 14th of May to the 18th of May 2015 : results were communicated to the participants with the gold standard data for adjudication.

Résultats

For the tasks 1, 2.1 and 2.2 the evaluation measure was the macro-precision (formula 4)

$$\text{Macro-précision} = \frac{\sum_{i=1}^n \left(\frac{\text{vrais positifs}(i)}{\text{vrais positifs}(i) + \text{faux positifs}(i)} \right)}{n} \quad (4)$$

For the Task 1 the results observed had a mean value of 0,582 ; a median value of 0,693 ; with a standard deviation of 0,238.

Team	Runs			Ranking
LIF (team 3)	0,736	0,722	0,688	1
INaLCO (team 2)	0,692	0,711	0,734	2
LIRMM (team 17)	0,732	0,725	0,733	3
Synapse (team 19)	0,701			4
Proxem (team 15)	0,699			5 ex-aequo
IRISA (team 14)	0,699	0,672	0,658	5 ex-aequo
LIMSI (team 23)	0,687	0,688		7
LINA / Dictanova (team 6)	0,655	0,676		8
IRIT / LIMSI (team 10)	0,577	0,578	0,580	9
TU Moldova (team 22)	0,559	0,547		10
LINA-Dimeco (team 25)	0,000	0,000	0,136	11
(team 4)	0,041			12

Figure 13: Task 1 Evaluation results per team.

For the Task 2.1 the results observed had a mean value of 0,514 ; a median value of 0,217 ; and a standard deviation of 0,029.

Team	Runs			Ranking
LIRMM (team 17)	0,613	0,563	0,552	1
INaLCO (team 2)	0,572	0,562	0,575	2
IRISA (team 14)	0,572	0,478	0,502	3
LIF (team 3)	0,558	0,560	0,535	4
LINA / Dictanova (team 6)	0,508	0,514		5
TU Moldova (team 22)	0,383	0,382		6
IRIT / LIMSI (team 10)	0,269	0,332	0,332	7
LINA-Dimeco (team 25)	0,000	0,000	0,097	8
(team 4)	0,029	0,029		9

Table 36: Task 2.1 Evaluation results per team.

For the Task 2.2 the results observed had a mean value of 0,180 ; a median value of 0,200 ; and a standard deviation of 0,152.


Team	Runs			Ranking
LIF (team 3)	0,347	0,327	0,327	1
INaLCO (team 2)	0,337	0,292	0,304	2
IRISA (team 14)	0,325	0,258	0,316	3
TU Moldova (team 22)	0,226	0,175		4
LIRMM (team 17)	0,037	0,174	0,007	5
LINA / Dictanova (team 6)	0,028	0,027		6
(team 4)	0,002	0,002		7
LINA-Dimeco (team 25)	0,000	0,000	0,000	8

Table 37: Task 2.2 Evaluation results per team.

The next table present the performance gain obtained by combining the runs of the participants with a majority vote weighted by their performance measure (the ROVER procedure introduced by [Fiscus, 1997]), according to the following schemes. Comparison for each task between the best participant performance and the performance of the ROVER was computed as follows:


- max = best participant result
- Rover = weighted vote combination of participant runs
- Rover - max = difference between the best performance and the ROVER
- Rover4best = ROVER restricted to the best 4 runs
- Rover4best - max = difference between the best four participants ROVER and the best performance.

Task	max	Rover	Rover - max	Rover4best	Rover4best - max
Task 1	0,736	0,765	0,029	0,760	0.024
Task 2.1	0,613	0,589	-0,024	0,607	-0,006
Task 2.2	0,347	0,330	-0,016	0,346	-0,001


	<p style="text-align: center;">CHIST-ERA</p>	<p style="text-align: right;">Subproject : WP5 Task : T5.3 Date : August 9, 2016 Page : 60</p>
---	--	--

References


- [Artstein and Poesio, 2008] Artstein, R. and Poesio, M. (2008). Inter-coder agreement for computational linguistics. *Computational Linguistics*, 34(4):555–96.
- [Baker et al., 2010] Baker, K., Bloodgood, M., Dorr, B. J., Filardo, N. W., Levin, L. S., and Piatko, C. D. (2010). A modality lexicon and its use in automatic tagging. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, Valletta, Malta. European Language Resources Association (ELRA).
- [Bessho et al., 2012] Bessho, F., Harada, T., and Kuniyoshi, Y. (2012). Dialog system using real-time crowdsourcing and twitter large-scale corpus. In *Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 227–231, Seoul, South Korea. Association for Computational Linguistics.
- [Bird and Liberman, 2000] Bird, S. and Liberman, M. (2000). A Formal Framework for Linguistic Annotation. *Speech Communication*, 33:23–60.
- [Boyd-Graber et al., 2012] Boyd-Graber, J., Satinoff, B., He, H., and Daume III, H. (2012). Besting the quiz master: Crowdsourcing incremental classification games. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 1290–1301, Jeju Island, Korea. Association for Computational Linguistics.
- [Chiao et al., 2006] Chiao, Y.-C., Kraif, O., Laurent, D., Minh, T., Nguyen, H., Semmar, N., Stuck, F., Véronis, J., and Zaghouani, W. (2006). Evaluation of multilingual text alignment systems: the ARCADE II project. In *Proceedings of the Fifth conference on International Language Resources and Evaluation (LREC'06)*, pages 1975–1979, Genoa, Italy.
- [Cleverdon, 1960] Cleverdon, C. (1960). The ASLIB Cranfield research project on the comparative efficiency of indexing systems. *ASLIB Proceedings*, 12:421–431. ISSN: 0001-253X / DOI: 10.1108/eb049778.
- [Edmonds and Kilgarriff, 2002] Edmonds, P. and Kilgarriff, A. (2002). Introduction to the special issue on evaluating word sense disambiguation systems. *Nat. Lang. Eng.*, 8:279–291.
- [Finin et al., 2010] Finin, T., Murnane, W., Karandikar, A., Keller, N., Martineau, J., and Dredze, M. (2010). Annotating named entities in twitter data with crowdsourcing. In *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk*, pages 80–88, Los Angeles. Association for Computational Linguistics.
- [Fiscus, 1997] Fiscus, J. G. (1997). A post-processing system to yield reduced word error rates: recognizer output voting error reduction (rover). In *In proceedings of the IEEE Workshop on Automatic Speech Recognition and Understanding*, pages 347–357, Santa Barbara, CA.

	<p style="text-align: center;">CHIST-ERA</p>	<p style="text-align: right;">Subproject : WP5 Task : T5.3 Date : August 9, 2016 Page : 61</p>
---	--	--


- [Fort et al., 2014] Fort, K., Adda, G., Sagot, B., Mariani, J., and Couillaud, A. (2014). Crowdsourcing for Language Resource Development: Criticisms About Amazon Mechanical Turk Overpowering Use. In Vetulani, Zygmunt and Mariani, Joseph, editor, *Human Language Technology Challenges for Computer Science and Linguistics*, pages 303–314. Springer International Publishing.
- [Fraisie and Paroubek, 2013] Fraisie, A. and Paroubek, P. (2013). uComp Deliverable D5.1 - Requirements of Affective Knowledge Extraction. Technical report, LIMSI-CNRS.
- [Gindl et al., 2010] Gindl, S., Weichselbraun, A., , and Scharl, A. (2010). Cross-domain contextualization of sentiment lexicons. In *Proceedings of the Eur. Conf. on Artif. Intelligence (ECAI-2010)*, pages 771–776. IOS Press.
- [Grady and Lease, 2010] Grady, C. and Lease, M. (2010). Crowdsourcing document relevance assessment with mechanical turk. In *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon’s Mechanical Turk*, pages 172–179, Los Angeles. Association for Computational Linguistics.
- [Grouin et al., 2009a] Grouin, C., Arnulphy, B., Berthelin, J.-B., El Ayari, S., Garcia-Fernandez, A., Grappy, A., Hurault-Plantet, M., Paroubek, P., Robba, I., and Zweigenbaum, P. (2009a). Présentation de l’édition 2009 du défi fouille de textes (deft’09). In *Actes de DEFT*, pages 35–50, Paris, France.
- [Grouin et al., 2007] Grouin, C., Berthelin, J.-B., El Ayari, S., Heitz, T., Hurault-Plantet, M., Jardino, M., Khalis, Z., and Lastes, M. (2007). Présentation de DEFT’07. In *Actes de DEFT*, Grenoble, France. AFIA.
- [Grouin et al., 2009b] Grouin, C., Hurault-Plantet, M., Paroubek, P., and Berthelin, J.-B. (2009b). DEFT’07 : une campagne d’évaluation en fouille d’opinion. *Revue des Nouvelles Technologies de l’Information*, RNTI E-17:1–24.
- [Higgins et al., 2010] Higgins, C., McGrath, E., and Moretto, L. (2010). Mturk crowdsourcing: A viable method for rapid discovery of arabic nicknames? In *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon’s Mechanical Turk*, pages 89–92, Los Angeles. Association for Computational Linguistics.
- [Hsueh et al., 2009] Hsueh, P.-Y., Melville, P., and Sindhvani, V. (2009). Data quality from crowdsourcing: A study of annotation selection criteria. In *Proceedings of the NAACL HLT 2009 Workshop on Active Learning for Natural Language Processing*, pages 27–35, Boulder, Colorado. Association for Computational Linguistics.
- [Hu et al., 2011] Hu, C., Resnik, P., Kronrod, Y., Eidelman, V., Buzek, O., and Bederson, B. B. (2011). The value of monolingual crowdsourcing in a real-world translation scenario: Simulation using haitian creole emergency sms messages. In *Proceedings of the Sixth Workshop on Statistical Machine Translation*, pages 399–404, Edinburgh, Scotland. Association for Computational Linguistics.

	CHIST-ERA	Subproject : WP5 Task : T5.3 Date : August 9, 2016 Page : 62
---	-----------	---


- [Koller et al., 2010] Koller, A., Striegnitz, K., Gargett, A., Byron, D., Cassell, J., Dale, R., Moore, J., and Oberlander, J. (2010). Report on the Second NLG Challenge on Generating Instructions in Virtual Environments (GIVE-2). In *Proceedings of the International Natural Language Generation Conference (INLG)*, Dublin.
- [Lafourcade and Fort, 2014] Lafourcade, M. and Fort, K. (2014). Propa-l: a semantic filtering service from a lexical network created using games with a purpose. In Chair), N. C. C., Choukri, K., Declerck, T., Loftsson, H., Maegaard, B., Mariani, J., Moreno, A., Odijk, J., and Piperidis, S., editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland. European Language Resources Association (ELRA).
- [Larson, 2013] Larson, M. (2013). Responsible crowdsourcing. http://homepage.tudelft.nl/q22t4/lib/EthicsOfCrowdsourcing_v2.pdf. Delft University of Technology, seminar on "Crowdsourcing: From Theory to Practice and Long-Term Perspectives", Dagstuhl, September 1-4, 2013.
- [Lease, 2013a] Lease, M. (2013a). Crowdsourcing & ethics: a few thoughts and refences. <http://fr.slideshare.net/mattlease/crowdsourcing-ethics-a-few>. Extracts and addendums from an earlier talk, for those interested in ethics and related issues in regard to crowdsourcing, particularly research uses. Slides updated Sept. 2, 2013.
- [Lease, 2013b] Lease, M. (2013b). Crowdsourcing for information retrieval: From statistics to ethics presentation transcript. <http://fr.slideshare.net/mattlease/lease-statisticsethics>. Revised October 27, 2013. Talk at UC Berkeley (October 21, 2013), Syracuse University (October 28, 2013).
- [Madnani et al., 2011] Madnani, N., Chodorow, M., Tetreault, J., and Rozovskaya, A. (2011). They can help: Using crowdsourcing to improve the evaluation of grammatical error detection systems. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 508–513, Portland, Oregon, USA. Association for Computational Linguistics.
- [Magnini et al., 2008] Magnini, B., Cappelli, A., Tamburini, F., Bosco, C., Mazzei, A., Lombardo, V., Bertagna, F., Calzolari, N., Toral, A., Lenzi, V. B., Sprugnoli, R., and Speranza, M. (2008). Evaluation of Natural Language Tools for Italian: EVALITA 2007. In Chair), N. C. C., Choukri, K., Maegaard, B., Mariani, J., Odjik, J., Piperidis, S., and Tapias, D., editors, *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco. European Language Resources Association (ELRA). <http://www.lrec-conf.org/proceedings/lrec2008/>.
- [Manning and Schütze, 2002] Manning, C. D. and Schütze, H. (2002). *Foundation of Statistical Natural Language Processing*. Massachusetts institute of Technology Press, 5th edition.
- [Marcus et al., 1993] Marcus, M., Santorini, B., and Marcinkiewicz, M. (1993). Building a large annotated corpus of English: the Penn Treebank. *Computational Linguistics*, 19.

	CHIST-ERA	Subproject : WP5 Task : T5.3 Date : August 9, 2016 Page : 63
---	-----------	---

- [de la Clergerie et al., 2008] de la Clergerie, E., Hamon, O., Mostefa, D., Ayache, C., Paroubek, P., and Vilnat, A. (2008). PASSAGE: from French Parser Evaluation to Large Sized Treebank. In ELRA, editor, *In proceedings of the sixth international conference on Language Resources and Evaluation (LREC)*, Marrakech, Morocco.
- [Munro et al., 2010] Munro, R., Bethard, S., Kuperman, V., Lai, V. T., Melnick, R., Potts, C., Schnoebelen, T., and Tily, H. (2010). Crowdsourcing and language studies: the new generation of linguistic data. In *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon’s Mechanical Turk*, pages 122–130, Los Angeles. Association for Computational Linguistics.
- [Negri et al., 2011] Negri, M., Bentivogli, L., Mehdad, Y., Giampiccolo, D., and Marchetti, A. (2011). Divide and conquer: Crowdsourcing the creation of cross-lingual textual entailment corpora. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 670–679, Edinburgh, Scotland, UK. Association for Computational Linguistics.
- [Page et al., 1999] Page, L., Brin, S., Motwani, R., and Winograd, T. (1999). The pagerank citation ranking: Bringing order to the web. Technical Report 1999-66, Stanford InfoLab. Previous number = SIDL-WP-1999-0120.
- [Pak et al., 2014] Pak, A., Paroubek, P., Fraisse, A., and Francopoulo, G. (2014). *Human Language Technology Challenges for Computer Science and Linguistics*, chapter Normalization of Term Weighting Scheme for Sentiment Analysis. Springer International Publishing Switzerland. *In print*.
- [Paroubek, 2007] Paroubek, P. (2007). *Evaluation of Text and Speech Systems*, volume 36 of *Text, Speech and Language Technology*, chapter Evaluating Part Of Speech Tagging and Parsing, pages 97–116. Kluwer Academic Publisher. ISBN-10: 1-4020-5815-2, ISBN-13: 978-1-4020-5815-8.
- [Paroubek et al., 2007] Paroubek, P., Chaudiron, S., and Hirschman, L. (2007). Principles of Evaluation in Natural Language Processing. *Traitement Automatique des Langues (TAL)*, 48(1):7–31.
- [Paroubek et al., 1998] Paroubek, P., Lecomte, J., Adda, G., Mariani, J., and Rajman, M. (1998). The GRACE French Part-Of-Speech Tagging Evaluation Task. In *Proceedings of the First International Conference on Language Resources and Evaluation (LREC)*, pages 433–441, Granada, Spain. ELDA.
- [Paroubek et al., 2009] Paroubek, P., de la Clergerie, E., Loiseau, S., Vilnat, A., and Francopoulo, G. (2009). The PASSAGE Syntactic Representation. In *Proceedings of the 7th International Workshop on Treebanks and Linguistic Theories*, pages 91–102, Gröningen. Netherlands Graduate Schools of Linguistics (LOT).

	<p style="text-align: center;">CHIST-ERA</p>	<p style="text-align: right;">Subproject : WP5 Task : T5.3 Date : August 9, 2016 Page : 64</p>
---	--	--

- [Post et al., 2012] Post, M., Callison-Burch, C., and Osborne, M. (2012). Constructing parallel corpora for six indian languages via crowdsourcing. In *Proceedings of the Seventh Workshop on Statistical Machine Translation*, pages 401–409, Montréal, Canada. Association for Computational Linguistics.
- [Prabhakaran et al., 2012] Prabhakaran, V., Bloodgood, M., Diab, M., Dorr, B., Levin, L., Pitatko, C. D., Rambow, O., and Van Durme, B. (2012). Statistical modality tagging from rule-based annotations and crowdsourcing. In *Proceedings of the Workshop on Extra-Propositional Aspects of Meaning in Computational Linguistics*, pages 57–64, Jeju, Republic of Korea. Association for Computational Linguistics.
- [Roth and Sammons, 2008] Roth, D. and Sammons, M. (2008). A Unified Representation and Inference Paradigm for Natural Language Processing. Technical Report UIUCDCS-R-2008-2969, UIUC Computer Science Department.
- [Rumshisky, 2011] Rumshisky, A. (2011). Crowdsourcing word sense definition. In *Proceedings of the 5th Linguistic Annotation Workshop*, pages 74–81, Portland, Oregon, USA. Association for Computational Linguistics.
- [Rumshisky et al., 2012] Rumshisky, A., Botchan, N., Kushkuley, S., and Pustejovsky, J. (2012). Word sense inventories by non-experts. In Chair), N. C. C., Choukri, K., Declerck, T., Doğan, M. U., Maegaard, B., Mariani, J., Moreno, A., Odijk, J., and Piperidis, S., editors, *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC’12)*, Istanbul, Turkey. European Language Resources Association (ELRA).
- [Rumshisky et al., 2009] Rumshisky, A., Moszkowicz, J., , and Verhagen, M. (2009). The holy grail of sense definition: Creating a sense-disambiguated corpus from scratch. In *In Proceedings of 5th International Conference on Generative Approaches to the Lexicon*, Pisa, Italy.
- [Safire, 2009] Safire, W. (2009). On language. *New York Times Magazine*.
- [Sagot, 2010] Sagot, B. (2010). The Lefff, a freely available and large-coverage morphological and syntactic lexicon for French. In *7th international conference on Language Resources and Evaluation (LREC 2010)*, Valletta, Malta.
- [Sayeed et al., 2011] Sayeed, A., Rusk, B., Petrov, M., Nguyen, H., Meyer, T., and Weinberg, A. (2011). Crowdsourcing syntactic relatedness judgements for opinion mining in the study of information technology adoption. In *Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, pages 69–77, Portland, OR, USA. Association for Computational Linguistics.
- [Sayeed et al., 2010] Sayeed, A. B., Meyer, T. J., Nguyen, H. C., Buzek, O., and Weinberg, A. (2010). Crowdsourcing the evaluation of a domain-adapted named entity recognition system. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 345–348, Los Angeles, California. Association for Computational Linguistics.

	<p style="text-align: center;">CHIST-ERA</p>	<p style="text-align: right;">Subproject : WP5 Task : T5.3 Date : August 9, 2016 Page : 65</p>
---	--	--

- [Scharl, 2014] Scharl, A. (2014). Domain-specific content repository. Technical report, MODUL University.
- [Schmitz and Le Roux, 2008] Schmitz, S. and Le Roux, J. (2008). Feature Unification in TAG Derivation Trees. In Gardent, C. and Sarkar, A., editors, *TAG+9*, pages pages 141–148, Tübingen, Allemagne. 12 pages, 4 figures.
- [Unnikrishnan et al., 2007] Unnikrishnan, R., Pantofaru, C., and Hebert, M. (2007). Toward Objective Evaluation of Image Segmentation Algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(6):929–944.
- [Verhagen et al.,] Verhagen, M., Gaizauskas, R., Schilder, F., Hepple, M., Katz, G., and Pustejovsky, J. SemEval-2007 - 15: TempEval Temporal Relation Identification.
- [Vertanen and Kristensson, 2011] Vertanen, K. and Kristensson, P. O. (2011). The imagination of crowds: Conversational aac language modeling using crowdsourcing and large data sources. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 700–711, Edinburgh, Scotland, UK. Association for Computational Linguistics.
- [Vilain et al., 1995] Vilain, M., Burger, J., Aberdeen, J., Connolly, D., and Hirschman, L. (1995). A model-theoretic coreference scoring scheme. In *MUC6 '95: Proceedings of the 6th conference on Message understanding*, pages 45–52, Columbia, Maryland, USA. ACL.
- [Vilnat et al., 2010] Vilnat, A., Paroubek, P., de la Clergerie, E. V., Francopoulo, G., and Guénot, M.-L. (2010). PASSAGE Syntactic Representation: a Minimal Common Ground for Evaluation. In Chair), N. C. C., Choukri, K., Maegaard, B., Mariani, J., Odijk, J., Piperidis, S., Rosner, M., and Tapias, D., editors, *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, Valletta, Malta. European Language Resources Association (ELRA).
- [Watson et al., 2005] Watson, R., Carroll, J., and Briscoe, T. (2005). Efficient extraction of grammatical relations. In *Proceedings of the Ninth International Workshop on Parsing Technologies (IWPT)*, pages 160–170, Vancouver. Association for Computational Linguistics.
- [Zaidan and Callison-Burch, 2011] Zaidan, O. F. and Callison-Burch, C. (2011). Crowdsourcing translation: Professional quality from non-professionals. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 1220–1229, Portland, Oregon, USA. Association for Computational Linguistics.
- [Zeichner et al., 2012] Zeichner, N., Berant, J., and Dagan, I. (2012). Crowdsourcing inference-rule evaluation. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 156–160, Jeju Island, Korea. Association for Computational Linguistics.