Deliverable D4.3 - Benchmarking Application Report.

LIMSI-CNRS

WP4 (T4.3)

Historique des modifications			
Versio	**Auteur**	**Date**	**Description des modifications**
1	Patrick Paroubek	Jul. 18th 2016	Deft 2015 Corpus Lexical coverage Experiment

Validation			
Role	**Organisation**	**Name**	**Date**

Table of Content

Index of Figures

Index of tables

1 Introduction

Setup and results of parameterised benchmarking applications to perform task-based evaluation of uComp ontologies. The partner responsible the production of this document is LIMSI-CNRS. The partner involved is WU.

The aim of this deliverable is to perform application based assesment of the ontologies produced in uComp in order to have an appreciation of the usability of the information they contain in the context of an NLP task, in complement of deliverable *D4.2 uComp Report on Ontology Evalution and Extension*. Condidering, the focus of uComp HC and its use case: *Climate Change*, we made use of the available data, i.e. on the one hand microblog messages from Twitter, in English, French and German (for the last two languages, a corpus of approx. 15,000 Tweets is available in both languages, with fine grained sentiment, opinion and emotion annotations), and on the other hand one ontology produced by uComp as described in D4.2. The Figure 1 give an example of the sentiment annotations in French, more details are availabel in deliverables D5.3 and D5.4.

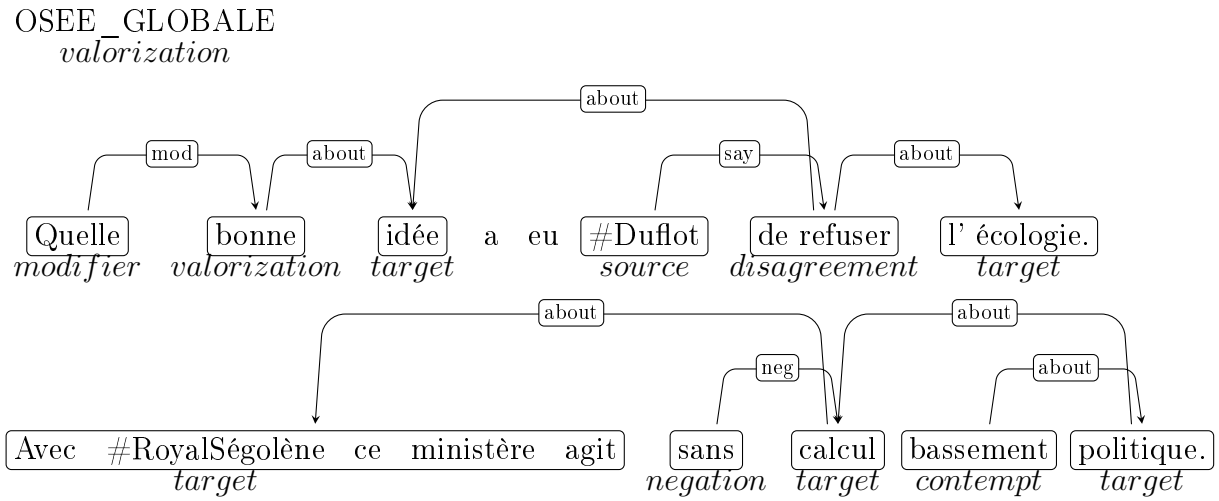

Figure 1: The message is: "*What a good idea (Mrs) #Duflot has had to refuse (the Ministry of) Ecology. With (Mrs) Ségolène Royale this Ministry acts without crassp political scheme.*". The tweet global sentiment expressed (OSEE_GLOBALE) is *valorization*. The relation *neg* and *mod* link the words indicative respectively of a negation ("sans", i.e. *without*) and of an intensity modifier ("Quelle", i.e. *What a...!*) to the sentiment expression they modify, here respectively "calcul" (*scheme*) and "bonne" (*good*). The links *about* connect the sentiment expressions to the objects they qualify, note here that the word "calcul" (*scheme*) a priori not bearing any opinion/sentiment/emotion value receives it through a chain of *about* links from "bassement" (*crassp*). Finally, the relation *say* connects the group of words referring to the sentiment holder to the sentiment expression.

The figure 2 displays the 30 concepts of our ontology as seen with the Gravity RDF browser[1].

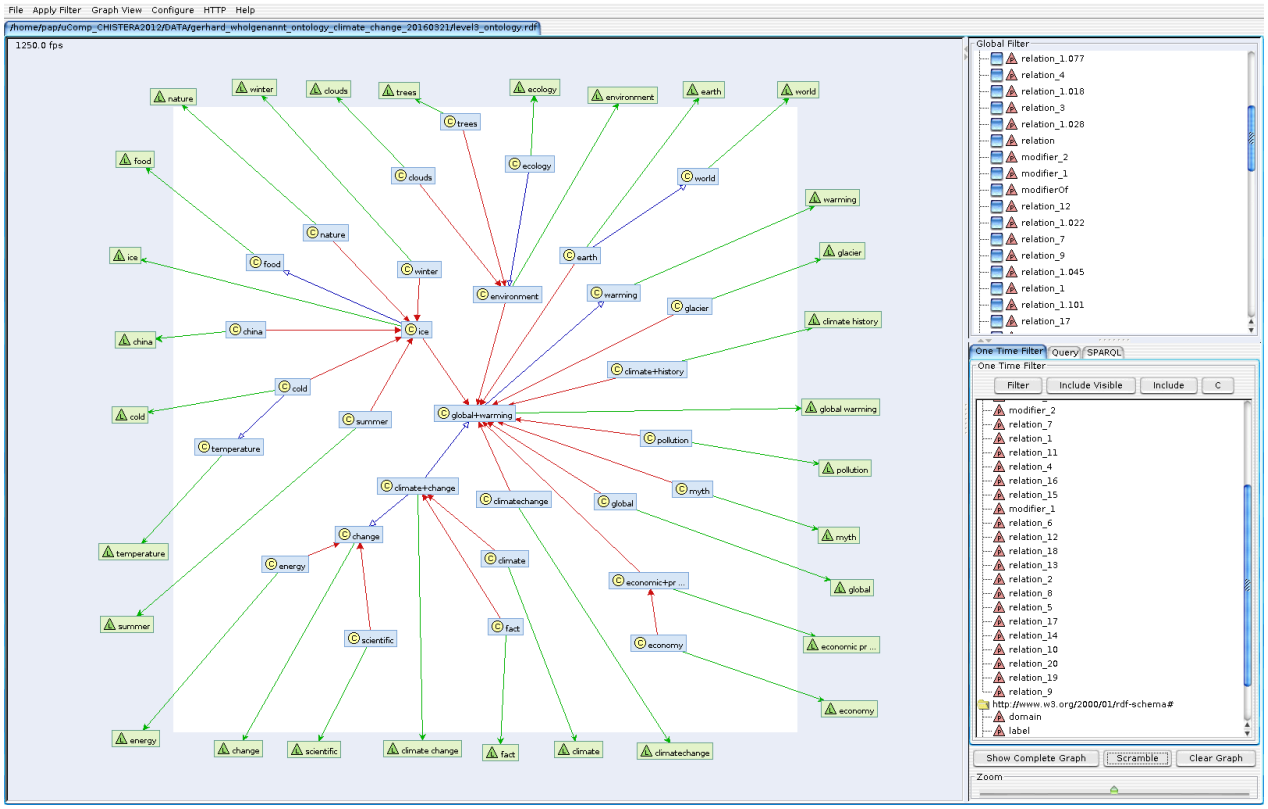

Figure 2: The climate change English ontology used as represented with the Gravity RDF browser; the blue boxes hold concepts and the green boxes literals, the links are blue for SubClassOf and red for other relations.

1.1 Lexical Indexing Experiment

Because of the late availability in the project of the German sentiment gold standard data, the experiments described here concern the English ontology and the French sentiment gold standard. A sample made of the 10 first messages from the corpus is displayed in table 2 We translated the English concept labels as presented in table 3. In this experiment, we used the UNITEX plateform[2] to analyses the presence of the concepts labels in the sentiment corpus. The translated concept labels were encoded into deterministic finite state automata using UNITEX dictionaries and morphological database to cover also their flexional and capitalization variants and to semantically disambiguate term occurence when for instance they are ambiguous between a nouns or a past participle, e.g. “fait”. Note that for this example, the ambiguity results from the translation into French, as in English “fact” is not a verb. Bt the translation process may

[1] `http://semweb.salzburgresearch.at/apps/rdf-gravity/index.html`
[2] `http://www-igm.univ-mlv.fr/~unitex/`

487586134789484544	#festival Décibulles et ses pratiques du #DéveloppementDurable
489717169304109058	@ocarai_slb @envertsetcontre @Val_Moncourtois @F_LeBohellec @phv @psvillejuif Écologie non soluble dans.libéralisme. Verts/UMP non sens.+
489094061408935936	Djibouti : Vers un programme d‘ électrification rurale par énergie solaire _ Afriquinfos
489427879231369218	Transition écologique et énergétique : les contrats de plan Etat-Régions effectifs en 2015 : La mini...
489419483753484289	#BTP #Durable #DD #Ecologie Opération L‘ Esperia , premier habitat collectif Bepos-Effinergie 2013
489607472525475841	#ApFrench #langchat #Anti-gaspillage chez Intermarché : Site Web https://t.co/f8rnb00eit
489754123609202688	La transition écologique et énergétique parmi les priorités des contrats de plan État-Région 2015_2020 _
489796034554789889	#Gard : Le Vidourle sur la voie de l‘ écologie
489802798133809152	Actus dvpt durable : Le projet européen TowerPower contrôle le vieillissement des structures d‘ éoliennes off...
489712240208400385	2nd Forum Trajectoires Développement Durable ! le 2 octobre 2014 , au Nouveau Siècle à Lille

Table 2: The 10 first messages of the corpus.

English	French	English	French
change	changement	food	nourriture
china	chine	glacier	glacier
climate	climat	global	global
climatechange	changementclimatique	global warming	réchauffement climatique
climate change	changement climatique	ice	glace
climate history	histoire du climat		myth mythe
clouds	nuages	nature	nature
cold	froid	pollution	pollution
earth	terre	scientific	scientifique
ecology	écologie	summer	été
economic problems	problèmes économiques	temperature	température
economy	économie	trees	arbres
energy	énergie	warming	réchauffement
environment	environnement	winter	hiver
fact	fait	world	monde

Table 3: Translation of the concept labels from English to French.

also remove ambiguities, e.g. in the case of "change" in English. The three automata are given in figure 3.

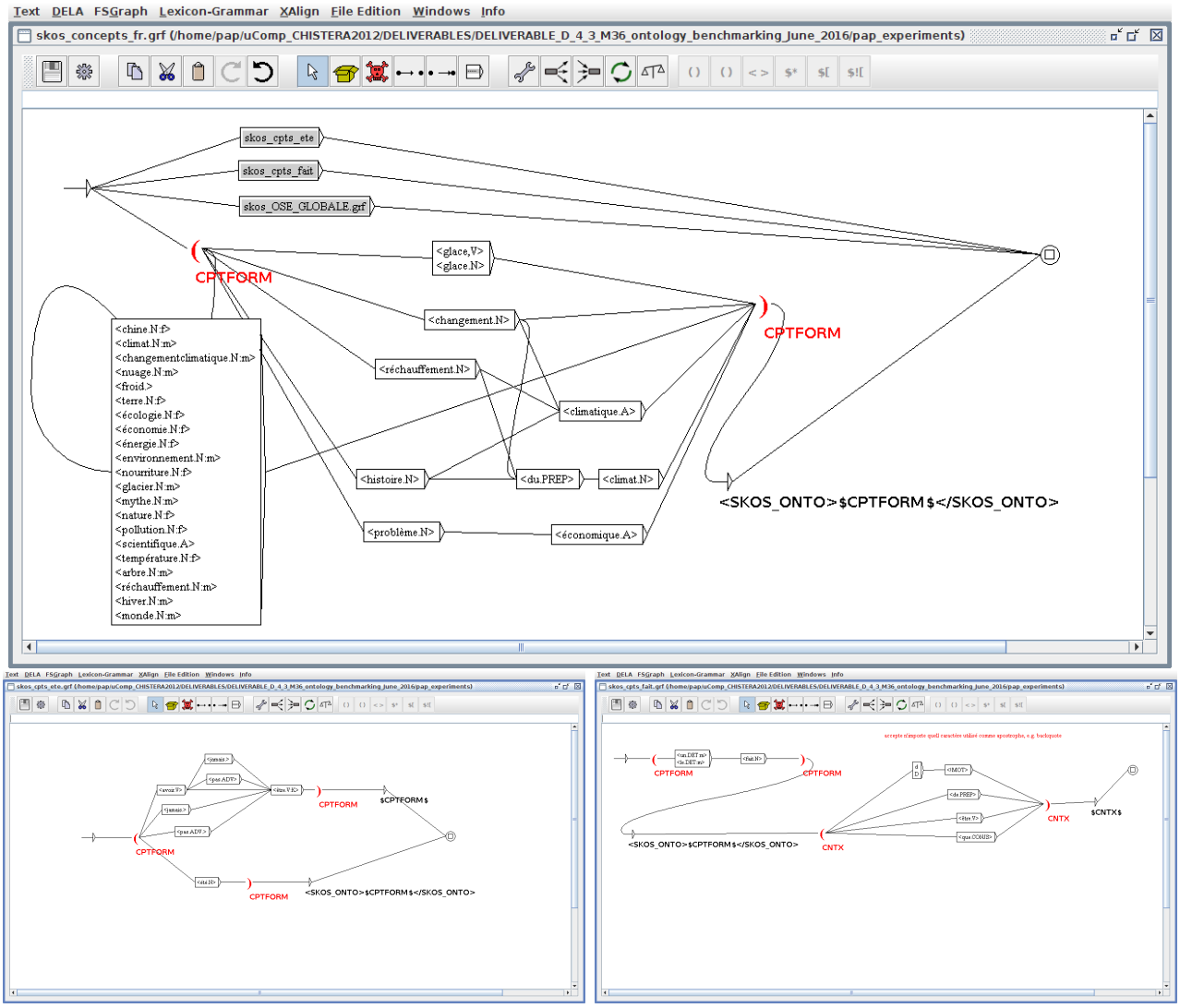

Figure 3: The tree UNITEX automata representing the ontology translated concept labels and their morphological variants (flexion and capitalization).

An excerpt of the concordance resulting from the filtering of the corpus with the concept labels is provided in figure 4

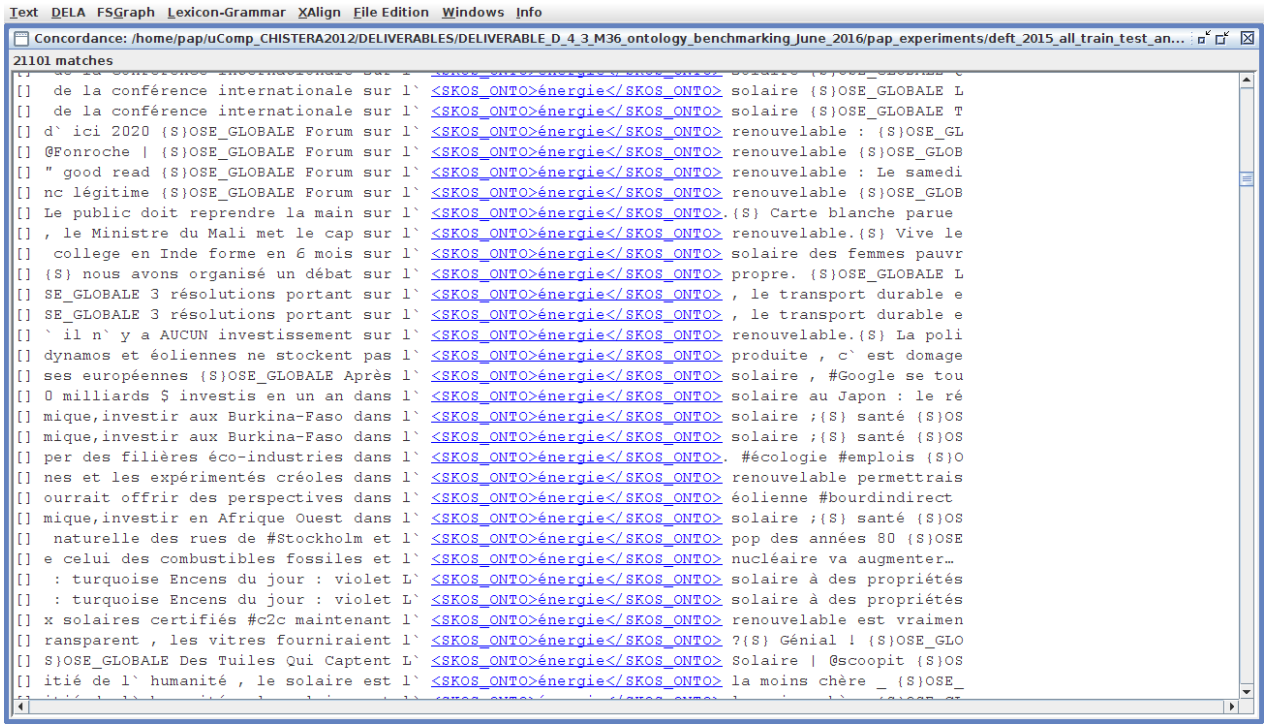

Figure 4: Some concordances resulting from indexing the corpus with the concept labels.

The table 7 gives for each concept label its number of occurences found in the corpus. The label search perfoms a semantic disambiguation based on the context and takes into account the flexional and capitalization variants of the labels. We remark that only three concept labels did not have any occurences in the corpus, *economic problems* (problèmes économiques), *climate history* (histoire du climat) and *climatechange* (changementclimatique). The loss of the last two labels can easily be explained by language and cultural differences between French and English. For the first concept label, although there are 57 occurences of the label "problem" in the corpus, only 10 are followed by an adjective (cf. Table 8), the most fréquent one being "écologique" (*ecologic*) with 6 occurences. Either cultural differences of a particular focus of the corpus may explain the loss of the term "economic problem"; as it seems that in the French corpus the authors of the message are more concerned with ecology. Finally, 90% of the concept labels are found in the corpus, taking into account flexional and capitalization variants and they cover 41% (4476 messages) of the 10,795 tweets.

#occurences	label French translation	English label
1581	énergie	energy
1332	écologie	ecology
525	réchauffement+climatique	global warming
489	changement+climatique	climate change
380	environnement	environment
213	monde	world
154	climat	climate
128	économie	economy
102	nature	nature
76	réchauffement	warming
74	pollution	pollution
60	terre	earth
59	été	summer
44	changement	change
39	scientifique	scientific
38	chine	china
37	global	global
37	arbre	trees
20	température	temperatue
16	froid	cold
10	glace	ice
9	mythe	myth
7	fait	fact
6	hiver	winter
5	nuage	clouds
3	nourriture	food
2	glacier	glacier
0	problèmes économiques	economic problems
0	histoire du climat	climate history
0	changementclimatique	climatechange

Table 7: Nombre d'occurences des labels de concepts de l'ontologie trouvés dans le corpus français.

@McPhyEnergy _ Une	Problème	est, qui "il y a
est confrontée à d'importants	problèmes	écologiques Au premier rang, le
fable qui initie aux	problèmes	écologiques.
sont pas responsables des	problèmes	écologiques." #ecopop #inégalités
@Reporterre les	problèmes+	liés à la croissance démographique
permettrais de résoudre nombreux	problèmes	réunionnais #RUN
climatique, c'est la généralisation de	problèmes	locaux.
L"écologie malgré les	problèmes	environnementaux est traité
l'étalement urbain pose de vrais	problèmes	écologiques.
électrique avait besoin de	problèmes	supplémentaires... #ecologie

Table 8: Concordances for the pattern <problème.N >< A > (all flexional and capitalization variants of the noun "problem" followed by and ajective).