



UPDATING WATT'S TRANSLATION MEMORY PART 2: EVALUATION OF WATT

Fabrizio Gotti – gottif@iro.umontreal.ca

Guy Lapalme – lapalme@iro.umontreal.ca

29 FEBRUARY 2016

RALI

Recherche Appliquée en
Linguistique Informatique

rali.iro.umontreal.ca

1 Table of contents

1	Table of contents	2
2	Introduction	3
2.1	Context of the study: WATT translation engine.....	3
2.2	New material provided by Environment Canada.....	3
2.3	The goals of this study (<i>Part 2</i>).....	3
3	2016 CAP archive content and resulting translation memory	4
3.1	The new data	4
3.2	Integration into WATT	6
4	Evaluation methodology	7
4.1	Use-case scenario	7
4.2	Goals of the evaluation	7
4.3	Methodology.....	7
4.4	Bringing WATT's translations up to publishing standards	10
5	Evaluation results and analysis.....	11
5.1	Translation quality	11
5.2	The importance of the translation memory	11
5.3	Manually correcting WATT's output.....	12
6	Conclusions.....	13
	Appendix A	14

2 Introduction

2.1 Context of the study: WATT translation engine

Since 2008, RALI has been working on the dissemination of environmental information¹. This project led to the creation of WATT², a prototype dedicated to the translation of public weather warnings issued by Environment Canada.

This prototype has been documented in an article (Gotti, Lapalme, & Langlais, 2014) entitled *Designing a Machine Translation System for Canadian Weather Warnings: a Case Study*, published in the journal *Natural Language Engineering*, in 2014³.

WATT combines a statistical machine translation engine (SMT) and a translation memory (TM). When a source sentence is submitted to the engine, it is first looked up in the TM. If it is present, the corresponding translation is retrieved and proposed. If the sentence is not found, the sentence is sent over to the SMT. The TM contains high-quality sentence pairs, mined from human translations.

2.2 New material provided by Environment Canada

On 6 January 2016, EC sent to RALI an archive of weather warnings published on Datamart from 2013 to January 2016, in Common Alerting Protocol (CAP) format⁴. Since human-produced weather warnings are used to train WATT (briefly described below), RALI naturally wanted to update it with the discussion text available in these new alerts.

This report follows a previous study by the same authors (henceforth called *Part 1*), subtitled “Findings on the CAP data”. The latter was interested in the quality of the text extracted from the CAP archive mentioned above and showed several issues with its quality. We summarize these findings in Section 3.

2.3 The goals of this study (*Part 2*)

While *Part 1* was interested in the nature and the quality of the text and sentence pairs mined from the CAP archive sent by EC, the present study focuses on how this new resource is integrated into WATT and how our tool fares when updated with this new resource.

¹ <http://rali.iro.umontreal.ca/rali/?q=en/EnvironmentalInfo>

² <http://rali.iro.umontreal.ca/EnvironmentalInfo/WarningTranslation.html>

³ <http://rali.iro.umontreal.ca/rali/sites/default/files/publis/WATT-NLE-Final.pdf>

⁴ <http://docs.oasis-open.org/emergency/cap/v1.2/CAP-v1.2-os.html>

3 2016 CAP archive content and resulting translation memory

3.1 The new data

As explained in Part 1, we extracted the sentences in the WW and WO weather warnings (71,699 files) found in the 2016 CAP archive, aligned the sentences⁵ and created a bitext (parallel corpus). Figure 1 (taken from Part 1) illustrates the protocol used, starting from CAP files (in XML format) as input and then converting them into a parallel corpus, or bitext.

Pairs of sentences can be readily derived from this bitext. Each sentence pair consists of an English sentence and its corresponding translation (or vice versa), along with the frequency with which it is found in the bitext. This constitutes the basis for a translation memory (TM). Table 1 below (taken from Part 1) shows an example of this translation memory.

Table 1 – Random excerpt of the translation memory derived from the CAP bitext. The red text highlights a problem in a French translation.

Freq	English sentence	French sentence
13 847	ENVIRONMENT CANADA METEOROLOGISTS WILL UPDATE ALERTS AS REQUIRED .	LES METEOROLOGISTES D ENVIRONNEMENT CANADA METTRONT LES ALERTES A JOUR AU BESOIN .
29	CONDITIONS ARE NO LONGER EXPECTED TO REACH HEAT WARNING CRITERIA .	ON NE PREVOIT PLUS DE CONDITIONS JUSTIFIANT L EMISSION D UN AVERTISSEMENT DE CHALEUR .
1	WINTRY WEATHER EXPECTED OVER HIGHER TERRAIN ON _DAY_ INTO EARLY _DAY_ .	ON PREVOIT DU TEMPS HIVERNAL SUR EN SECTEURS MONTAGNEUX _DAY_ ET TO T _DAY_ .
1	WINTRY WEATHER EXPECTED OVER HIGHER TERRAIN ON _DAY_ INTO EARLY _DAY_ .	ON PREVOIT DU TEMPS HIVERNAL EN TERRAIN MONTAGNEUX _DAY_ ET TOT _D AY_ .

There are 100k (different) sentence pairs in the TM. As explained in Part 1, there are a number of linguistic issues with these sentence pairs, including source text with misspelled words, and dubious translations. Since the quality of this new resource is crucial to WATT, we had to filter out as many unwanted sentences as possible.

Typically, one way of achieving this is to add to the TM the sentence pairs that appear more than n times, where n is found empirically. After examination, we set $n = 2$, and filtered out all other pairs. Out of the original 100k sentence pairs, we only kept 17k by applying this scheme. More elaborate strategies could also be used, like removing sentence pairs with unknown words, but this is not guaranteed to work, since the remaining sentence pairs could very well present undetected translation problems. An additional difficulty with unknown words is that the large number of proper nouns in the warnings renders the use of a dictionary delicate.

⁵ The text alignment is performed with our tool yasa, available online at <http://rali.iro.umontreal.ca/rali/?q=en/yasa>.

Step 1. CAP file description (XML)	
<p><description>Blizzard conditions with poor visibility in snow and blowing snow are expected or occurring.</p> <p>Blizzard conditions over the Churchill and York regions are expected to continue tonight and into Friday morning.</p> <p>Very strong northwest winds of 50 km/h with gusts to 70 km/h have developed early this afternoon bringing blizzard conditions to the Churchill and York regions. Weather conditions will improve Friday morning.</description></p>	<p><description>Il y a ou il y aura du blizzard avec une visibilité mauvaise sous la neige et dans la poudrerie.</p> <p>Blizzard sur les régions de Churchill et de York devraient continuer ce soir, cette nuit et vendredi matin.</p> <p>Des vents très forts du nord-ouest de 50 km/h avec rafales à 70 km/h se sont levés tôt cet après-midi apportant du blizzard sur les régions de Churchill et de York. La situation s'améliorera vendredi matin.</description></p>
Steps 2. and 3. Text extraction and whitespace normalization	
<p>Blizzard conditions with poor visibility in snow and blowing snow are expected or occurring. Blizzard conditions over the Churchill and York regions are expected to continue tonight and into Friday morning. Very strong northwest winds of 50 km/h with gusts to 70 km/h have developed early this afternoon bringing blizzard conditions to the Churchill and York regions. Weather conditions will improve Friday morning.</p>	<p>Il y a ou il y aura du blizzard avec une visibilité mauvaise sous la neige et dans la poudrerie. Blizzard sur les régions de Churchill et de York devraient continuer ce soir, cette nuit et vendredi matin. Des vents très forts du nord-ouest de 50 km/h avec rafales à 70 km/h se sont levés tôt cet après-midi apportant du blizzard sur les régions de Churchill et de York. La situation s'améliorera vendredi matin.</p>
Step 4. Tokenization and serialization	
<p>BLIZZARD CONDITIONS WITH POOR VISIBILITY IN SNOW AND BLOWING SNOW ARE EXPECTED OR OCCURRING . BLIZZARD CONDITIONS OVER THE CHURCHILL AND YORK REGIONS ARE EXPECTED TO CONTINUE TONIGHT AND INTO __DAY__ MORNING . VERY STRONG NORTHWEST WINDS OF __NUM__ KM / H WITH GUSTS TO __NUM__ KM / H HAVE DEVELOPED EARLY THIS AFTERNOON BRINGING BLIZZARD CONDITIONS TO THE CHURCHILL AND YORK REGIONS . WEATHER CONDITIONS WILL IMPROVE __DAY__ MORNING .</p>	<p>IL Y A OU IL Y AURA DU BLIZZARD AVEC UNE VISIBILITE MAUVAISE SOUS LA NEIGE ET DANS LA POUDRERIE . BLIZZARD SUR LES REGIONS DE CHURCHILL ET DE YORK DEVRAIENT CONTINUER CE SOIR , CETTE NUIT ET __DAY__ MATIN . DES VENTS TRES FORTS DU NORD-OUEST DE __NUM__ KM / H AVEC RAFALES A __NUM__ KM / H SE SONT LEVES TOT CET APRES-MIDI APPORTANT DU BLIZZARD SUR LES REGIONS DE CHURCHILL ET DE YORK . LA SITUATION S AMELIORERA __DAY__ MATIN .</p>
Step 5. Sentence alignment and resulting bitext	
<p>BLIZZARD CONDITIONS WITH POOR VISIBILITY IN SNOW AND BLOWING SNOW ARE EXPECTED OR OCCURRING .</p>	<p>IL Y A OU IL Y AURA DU BLIZZARD AVEC UNE VISIBILITE MAUVAISE SOUS LA NEIGE ET DANS LA POUDRERIE .</p>
<p>BLIZZARD CONDITIONS OVER THE CHURCHILL AND YORK REGIONS ARE EXPECTED TO CONTINUE TONIGHT AND INTO __DAY__ MORNING .</p>	<p>BLIZZARD SUR LES REGIONS DE CHURCHILL ET DE YORK DEVRAIENT CONTINUER CE SOIR , CETTE NUIT ET __DAY__ MATIN .</p>
<p>VERY STRONG NORTHWEST WINDS OF __NUM__ KM / H WITH GUSTS TO __NUM__ KM / H HAVE DEVELOPED EARLY THIS AFTERNOON BRINGING BLIZZARD CONDITIONS TO THE CHURCHILL AND YORK REGIONS .</p>	<p>DES VENTS TRES FORTS DU NORD-OUEST DE __NUM__ KM / H AVEC RAFALES A __NUM__ KM / H SE SONT LEVES TOT CET APRES-MIDI APPORTANT DU BLIZZARD SUR LES REGIONS DE CHURCHILL ET DE YORK .</p>
<p>WEATHER CONDITIONS WILL IMPROVE __DAY__ MORNING .</p>	<p>LA SITUATION S AMELIORERA __DAY__ MATIN .</p>

Figure 1 – Example of text extraction pipeline for a single CAP file. Excerpt from file T_WWCN12_C_CWWG_201501082025_BA5AF4B7.cap, with English as the original.

3.2 Integration into WATT

The 17k new sentence pairs were merged with the existing translation memory already used by WATT, which contained 118k sentence pairs. The resulting memory therefore contains 135k sentence pairs, derived from weather warning material ranging from 2000 up to now.

In addition to the update of this resource, we made additional modifications to WATT's code to account for new occurrences in the data, for instance the appearance of social media hashtags (e.g. #MétéoQC). WATT was not designed to handle these (unusual) tokens, but the modifications were straightforward.

The updates to WATT are now deployed on RALI's demo website, available here:

<http://rali.iro.umontreal.ca/EnvironmentalInfo/WarningTranslation.html>

Moreover, RALI continues to translate weather warnings that EC sends to us, in real time, with this updated version. Five years ago, RALI set up a translation server that receives these warnings as they are issued, and the updated version of our engine now handles incoming weather warnings.

4 Evaluation methodology

4.1 Use-case scenario

The evaluation we propose here relies on a specific use-case scenario for WATT:

1. A meteorologist creates a new warning in a source language (English or French).
2. WATT produces a draft translation in the target language (this takes about 10 seconds).
3. The source warning and the draft translation are published simultaneously. The draft translation comes with a notice that informs the user that a human translation will be available shortly.
4. A human translator revises WATT's output and produces a final translation.
5. This final translation replaces the draft previously published.

This proposition draws on WATT's strength, i.e. the possibility to produce an intelligible translation in seconds, while informing the end user that there might be imperfections or lost meaning in the draft translation. It provides first and foremost a way to satisfy the need for a timely dissemination of important information. This scenario is also transparent to the user, and WATT is not used as a replacement for a human translator. The approach is already used by companies with relatively large documentation corpora, e.g. Microsoft's knowledge bases⁶.

4.2 Goals of the evaluation

This evaluation seeks to assess if WATT is ready to be used in the context of the scenario explained above. Specifically, we want to know **how much of the original meaning of a given warning is faithfully transferred in the target language by WATT**. We also want to measure the time taken by WATT to produce these translations.

This faithfulness of the translation to the original, as well as its timely production, will determine whether WATT renders the above use-case scenario reasonably possible.

This evaluation is carried out on the type of weather warnings WATT is designed for, i.e. weather warnings WW and omnibus warnings WO concerning the weather.

4.3 Methodology

We randomly picked 55 weather warnings (42 in English, 13 in French) out of the 71,699 contained in the 2016 CAP archive, so as to perform this evaluation on—relatively—recent warnings. We restricted our sample to warnings containing more than 6 sentences, so as to select “harder” text to translate. Indeed, a cursory examination of the shorter warnings indicates they are mostly stereotypical in nature and readily translatable by looking them up in historical translations (translation memory).

This test corpus is described in Table 2. The complete list of warnings used is listed in Appendix A. All of the French warnings were issued in Montreal.

⁶ See for instance the disclaimer at the beginning of <https://support.microsoft.com/fr-fr/kb/315939>.

Table 2 – Statistics for the test corpus

Statistic	English test corpus	French test corpus
Nb. of warnings	42	13
Avg. nb. of words/warning	116.4	127.8
Avg. nb. of sentences/warning	6.9	6.5

The text description available in these warnings was submitted to the updated version of WATT described in Section 3.2. Prior to this, we had to slightly alter WATT’s translation memory. We had to compensate for the fact that some sentence pairs in this memory were derived from the very test documents submitted to WATT, resulting in an unfair advantage. We corrected for this.

For each bulletin, we collected the original translation, WATT’s translation and (for reference only) the human translation. We created a Microsoft Word document containing these bulletins, so that they could be annotated by us.

We manually assessed each warning translated by WATT to determine how well its original meaning was carried over to its translation. We use a simple scale, from 1 (worst) to 3 (best), arguing that the goals we set for this evaluation would not benefit from a more complex evaluation scheme. More precisely, the ratings we give to each warning are:

- 1 (bad)** The translation process failed to transfer some of the original meaning. This may be because of gibberish on output, or because otherwise linguistically sound sentences fail to convey the original information, or any other reason. In this case, we can say that the draft translation produced by WATT is inappropriate for the use-case scenario described above. A single dubious sentence is enough to rate the complete machine translation for a whole bulletin as 1 (bad).
- 2 (fair)** Most of the meaning is properly carried over from the original sentence, but some subtleties or details are missing or distorted.
- 3 (good)** The draft translation is comparable in quality and content to the reference human translation OR the draft translation has slight grammar problems, but the original meaning is faithfully conveyed and understandable.

This is arguably a subjective process, so we want to provide the reader a sense of our methodology by illustrating it with examples, shown in Table 3. The complete evaluation results are made available to the reader as well.

It is noteworthy that, besides illustrating rating examples, Table 3 also shows problems associated with human interventions. The first case is a misspelled word in the *source* sentence (example 2) and the second case is a dubious *human* translation (example 3). Although it remains anecdotal evidence in this table, these are problems that we have encountered very often (and described extensively) in previous studies, including in Part 1 of the present report.

Table 3 – An example for each translation rating for WATT’s output. For each rating, we give the original, the translation and a short explanation of our decision. We only give examples of French to English translations to alleviate this text. Errors are highlighted in red and are underlined.

1 (bad)	
Source	<i>[6 sentences not shown in this warning]</i> En général, les quantités de pluie attendues seront de 20 à 30 millimètres jusqu'à mercredi soir sauf de 50 à 70 millimètres dans un corridor qui s'étire de Pointe-des-Monts vers la Minganie.
WATT	<i>[6 sentences not shown, properly translated by WATT]</i> General rainfall amounts expected will be 20 to 30 mm through Wednesday evening except 50 to 70 mm in a corridor <u>stretching from Pointe-des-Monts near the Minganie.</u>
Reference translation	Generally, rainfall amounts of 20 to 30 mm are expected through Wednesday evening, except for a corridor stretching from Pointe-des-Monts toward the Minganie where 50 to 70 mm of rain is expected.
Explanation	WATT produces “a corridor stretching from A near B” instead of “corridor stretching from A toward B” (meaning is lost). The rest of the translated sentences are perfect.

2 (fair)	
Source	<i>[5 sentences not shown in this warning]</i> Les vents du nord-est à 80 km/h ou plus et la neige <u>donne</u> de la poudrière généralisée sur la région.
WATT	<i>[5 sentences not shown, properly translated by WATT]</i> <u>Northeast winds to 80 km/h or more</u> and snow <u>will give</u> widespread blowing snow over the area.
Reference translation	Northeasterly winds of 80 km/h or more combined with the snow are causing generalized blowing snow over the region.
Explanation	WATT produces “Northeast winds to 80 km/h or more” instead of “Northeasterly winds of 80 km/h or more”, which is understandable but non-grammatical. Moreover, the French present tense is distorted into the future tense “will give”. The rest of the translated sentences are perfect. Note that the original French contains an agreement error (“donne” should be spelled “donnent”).

3 (good)	
Source	<i>[6 sentences not shown in this warning]</i> Le front chaud qui lui est associé s'étire vers le sud-est jusqu'au centre du Québec et se dirigera vers le nord-est donnant du temps doux.
WATT	<i>[6 sentences not shown, properly translated by WATT]</i> An associated warm front extends southeastward to <u>central</u> Québec and will track northeastward bringing mild temperatures.
Reference translation	The warm front <u>associated to</u> it stretches southeastward to Central Quebec and will track northeastward <u>giving of mild conditions.</u>
Explanation	WATT phrases its translation differently from the human one, but the meaning is the same. A case-restoration problem creates an error on the word “central”, which should be capitalized. The rest of the translated sentences are perfect. Note that the reference (human) translation contains the dubious expression “giving of mild conditions” as well as the Gallicism “associated to” instead of “associated with”.

4.4 Bringing WATT's translations up to publishing standards

As we examined WATT's translation, we used Microsoft Word's revision feature to flag the errors found in WATT's output and to correct those errors so that the resulting text is publication-ready, to the best of our knowledge. A screen capture of this setup is shown in Figure 2. As explained above, the human (reference) translation is also made available to the annotator in order to make sure that his corrections are valid.

This allowed us to produce additional statistics pertaining to the number of edition operations (text insertion, text deletions) necessary to correct the machine output. A timestamp on these operations further allows us to estimate the time required for these correction operations.

<p>Source text</p> <p>La visibilité est ou sera mauvaise sous la neige et dans la poudrerie dans certains secteurs.</p> <p>Les vents du nord-est à 80 km/h ou plus et la neige donne de la poudrerie généralisée sur la région.</p> <p>On peut s'attendre à des conditions routières dangereuses en raison de la visibilité réduite à certains endroits. Adaptez votre conduite aux conditions routières changeantes. Évaluez la possibilité de retarder tout déplacement non essentiel jusqu'à ce que les conditions s'améliorent.</p> <p>Un avis de poudrerie est émis lorsque l'on prévoit que les vents produiront de la poudrerie réduisant la visibilité à 800 mètres ou moins durant au moins 3 heures.</p> <p>Watt translation</p> <p>Poor visibility in snow and blowing snow is expected or occurring in some locations.</p> <p>Northeast winds of up to 80 km/h combined with snow will cause widespread blowing snow over the area.</p> <p>Travel is expected to be hazardous due to reduced visibility in some locations. Be prepared to adjust your driving with changing road conditions. Consider postponing non-essential travel until conditions improve.</p> <p>Weather advisories for blowing snow are issued when winds are expected to create blowing snow giving poor visibility to 800 m or less for at least 3 hours.</p> <p>Human translation</p> <p>Poor visibility in snow and blowing snow is expected or occurring in some locations.</p> <p>Northeasterly winds of 80 km/h or more combined with the snow are causing generalized blowing snow over the region.</p> <p>Travel is expected to be hazardous due to reduced visibility in some locations. Be prepared to adjust your driving with changing road conditions. Consider postponing non-essential travel until conditions improve.</p> <p>Weather Advisories for blowing snow are issued when winds are expected to create blowing snow giving poor visibility to 800 metres or less for at least 3 hours.</p> <p>Additional info</p> <p>Bullname: T_WOCN10_C_CWUL_201412110626_866FFE94</p> <p>Hit ratio: 5/6</p> <p>Translation quality: 2</p>	<p>Fabrizio Gotti Supprimé: to</p> <p>Fabrizio Gotti Supprimé: or more and</p> <p>Fabrizio Gotti Supprimé: give</p>
---	---

Figure 2 – Annotated warning in Microsoft Word. The blue sentences in WATT's translation indicate memory hits.

5 Evaluation results and analysis

5.1 Translation quality

We carried out the evaluation protocol described in Section 4.3 and evaluated the translation quality of WATT-translated warnings. Table 4 shows the results.

Table 4 – Translation quality statistics for WATT’s translation of the test corpus.

% of warnings with a rating of...	English to French (42 warnings)	French to English (13 warnings)	All warnings (55 warnings)
1 (bad)	17%	8%	15%
2 (fair)	24%	38%	27%
3 (good)	59%	54%	58%

Generally speaking, the combined ratings of 2 (fair) and 3 (good) make up 83% of French warnings and 92% of English warnings, when they are produced by WATT. In other words, should the scenario proposed in Section 4.1 be adopted, and draft translations be disseminated pending a human revision, it seems that their meaning would be clear and useful to the end user, for 83% of French draft translations, and 92% of English draft translations. Overall, for all 55 warnings, this corresponds to 85% of WATT’s translations whose meanings convey almost all or all of the original content.

The remainder of the translated warnings (those that fall in the category 1 – bad) would not be readily understandable, and some significant meaning would be lost. In this case, the user would have to wait for the human revision of the translation to access the information (or switch to the other language if he’s bilingual).

The average time taken by WATT to translate an English warning is 10.3 seconds, and 1.5 seconds for a French warning. English warnings are significantly longer to post-process because the corresponding French translation needs to have its diacritics restored (accents, cedillas, etc.). These figures amount to 676 words per minute for English to French and 4990 words per minute for French to English. These times are measured on a standard computer workstation. A more powerful computer infrastructure will outperform this.

5.2 The importance of the translation memory

When we submitted the warnings’ sentences to WATT, we also kept track of which of these sentences were found verbatim in WATT’s translation memory (memory hit) and which ones were not (memory miss). The sentences not found in the TM are translated using WATT’s embedded statistical translation engine.

Obviously, a high hit ratio significantly improves the machine translation quality, because this corresponds to looking up the source sentence in a curated corpus of human translations.

In this evaluation, we found that 58% of the English sentences part of the test corpus were found in the translation memory. This figure rises to 62% for French. These figures are

comparatively very high in the world of machine translation, but are not unexpected. In our 2014 study on WATT, we found a hit ratio of 45%, also very high. This is encouraging, and shows that a simple way of improving WATT is feeding it with as many (good) translation pairs to populate its TM. Conversely, it also shows how quality issues in the human-produced material can contaminate the memory and eventually lead to systematic errors in the machine’s output. These errors were covered in detail in Part 1 of this report. Some of them are even illustrated in Table 3, in the human-produced material (source sentences or human translations containing errors).

The high hit ratio also accounts in part for the very short time WATT takes when translating a warning, since a sentence lookup is very fast.

5.3 Manually correcting WATT’s output

Although the goal of this study is to determine whether WATT’s draft translation can be disseminated (with a disclaimer), one by-product of this evaluation are the annotated draft translations. As explained in Section 4.2, these editing operations (insertions, deletions of text, as illustrated in Figure 2) can be used to approximate the effort needed to bring WATT’s output up to publishing quality. This revision could be done for instance by the professional translator, instead of starting his translation from scratch. In other words, the same draft translation could have two purposes: a temporary warning, pending revision, and the basis the translator would work from.

Table 5 summarizes the statistics for these revision operations. It shows that there are for English to French, on average 7.9 human revisions separating WATT’s draft translation from an acceptable human translation. It should be noted that when the annotator replaces a text excerpt with another, this counts as *two* revisions (one deletion and one insertion). This statistic is 11.2 human revisions for French-to-English translation. The number of words touched by these revisions may paint a more accurate picture of the revision effort: When translating from English, 12.3 words are inserted or deleted on average, while in the other translation direction, 6.5 words are modified.

Timewise, the revision time is less than 2 minutes in both translation directions, which brings the average total duration of the translation process (WATT translation + human revision) to 1 minute 43 seconds from English, and 1 minute 16 seconds for French.

Table 5 – Revision statistics for test warnings translated by WATT.

Revision statistic	English to French (42 warnings)	French to English (13 warnings)
Avg. nb human revisions/warning*	7.9	11.2
Avg. nb revised words/warning	12.3	6.5
WATT translation time	10.3s	1.5s
Human revision time	92.9s	74.8s
Total translation time	103.2s	76.3s

* N.B.: When the annotator replaces a text excerpt with another, this counts as *two* revisions.

6 Conclusions

This two-part study aimed at (1) using 72k newly available weather warnings (issued from 2013 to 2016 inclusively) to improve WATT and at (2) evaluating WATT's performance once updated.

We started by examining the new weather warnings EC sent and we identified a number of issues with the quality of their textual descriptions, ranging from minor formatting problems to more concerning grammar and vocabulary errors present in the source text and in the human translations. We describe this in detail in Part 1.

Despite the errors, we were able to convert the discussions present in the warnings into a translation memory for WATT by filtering out infrequent (and potentially erroneous) sentence pairs, at the cost of eliminating 83% of the available material.

We evaluated this updated version of WATT on a test corpus, with a very specific use-case scenario in mind, i.e. using WATT to swiftly draft temporary translations that could be disseminated along with the original weather warning, as it is issued. WATT's output would indicate that it is a machine translation and that a human revision is pending.

The evaluation results on a benchmark of 55 test warnings show that WATT's performance make it a reasonably suitable tool in this scenario. Overall, 85% of the test warnings translated by WATT conveyed all or almost all of the original warning's meaning. The high quality of the output is in part due to the translation memory curated by RALI, and successfully updated with the new warnings.

In the remaining 15% of warnings, the user would simply have to wait a few minutes for a translation in his language. It should be pointed out that this is already the case presently: the user has to wait for a translation, *for all warnings*. Using WATT would mean that, for 85% of warnings on average, there would not be any waiting period to access the meteorological information.

Since WATT possesses instrumentation features, it is also possible to detect whether the translation engine is struggling with a difficult translation (one that would presumably fall in the 15%) and revert to the scenario currently in place, i.e. withholding the translation and waiting for a human intervention. This simple strategy would limit the number of machine translations not entirely up to par when WATT detects that it is encountering difficulties.

One interesting side-effect of this strategy is the production of a draft translation that could be used entirely or in part by the human translator to bootstrap his work, at his discretion. Our work seems to indicate that less than 2 minutes would be required to sufficiently improve these draft translations so that they can be published.

Appendix A

List of warnings (CAP file) used in the test corpus.

English corpus	French corpus
T_WOCN11_C_CWHX_201402140908_D24678A7	T_WOCN10_C_CWUL_201405070755_56A3A857
T_WOCN11_C_CWHX_2015060800451818	T_WOCN10_C_CWUL_201412110626_866FFE94
T_WOCN11_C_CWTO_201404210235_19EEE8CD	T_WWCN10_C_CWUL_201304250551_B5190DEA
T_WOCN11_C_CWTO_201409101940_8D764CE2	T_WWCN10_C_CWUL_201308131910_A2E375FB
T_WOCN11_C_CWVR_2015120513022727	T_WWCN10_C_CWUL_201401260944_78F1E2D7
T_WOCN12_C_CWTO_201404231506_B672C63C	T_WWCN10_C_CWUL_201401292002_B6DFC01C
T_WOCN12_C_CWWG_2015052921001212	T_WWCN10_C_CWUL_201402142032_8B3ADE5B
T_WOCN13_C_CWWG_201412142022_F89322DE	T_WWCN10_C_CWUL_201402220043_BC6676ED
T_WOCN13_C_CWWG_2015090821282929	T_WWCN10_C_CWUL_201410080818_59127E51
T_WOCN14_C_CWHX_201405280104_A0E3B6CC	T_WWCN10_C_CWUL_201411251155_94F9EACF
T_WOCN16_C_CWHX_201408020237_66165353	T_WWCN10_C_CWUL_201501041608_1FC39B5F
T_WOCN16_C_CWHX_201409131314_9E6735EC	T_WWCN10_C_CWUL_201502012031_5ECCD195
T_WOCN16_C_CWWG_2015052903243535	T_WWCN10_C_CWUL_2015102115391919
T_WOCN16_C_CWWG_2015090421282020	
T_WOCN17_C_CWHX_201409221348_53344A8C	
T_WOCN17_C_CWHX_201411280223_C434A324	
T_WWCN11_C_CWNT_201410310448_889E78E3	
T_WWCN11_C_CWNT_201410311622_95ED5C1E	
T_WWCN11_C_CWTO_201407220114_F81A6180	
T_WWCN11_C_CWTO_201409012056_EB43AD20	
T_WWCN11_C_CWTO_201411201158_90DF2F4B	
T_WWCN11_C_CWVR_201501180017_266E2239	
T_WWCN11_C_CWVR_201501181246_5BBCC89D	
T_WWCN11_C_CWVR_2015082912580909	
T_WWCN11_C_CWVR_2015111219131717	
T_WWCN12_C_CWTO_201309280930_DBF898BC	
T_WWCN13_C_CWNT_201411192040_6490E5C2	
T_WWCN13_C_CWNT_201501091604_EAC17A56	
T_WWCN14_C_CWHX_201302171943_ADCA2BD7	
T_WWCN14_C_CWHX_201401050838_7793D94F	
T_WWCN14_C_CWNT_201502281514_84EC63C2	
T_WWCN14_C_CWNT_2015052314222828	
T_WWCN15_C_CWNT_201409301923_9D5EE11B	
T_WWCN15_C_CWWG_201501301930_AD966DFD	
T_WWCN15_C_CWWG_2015101220063838	
T_WWCN16_C_CWHX_201302080844_90EB1F34	
T_WWCN16_C_CWHX_201408082207_676AF30A	
T_WWCN16_C_CWHX_201408091331_97B64BD9	
T_WWCN16_C_CWHX_201501261912_364689F7	
T_WWCN16_C_CWHX_201502060210_A7D9EEDD	
T_WWCN16_C_CWNT_201310082216_95E17482	
T_WWCN17_C_CWHX_201302051329_8CFA7F93	