



Évaluation du prototype de traducteur d'avertissements

Fabrizio Gotti

Guy Lapalme


14 janvier 2011

WATT: Warning-Alerte Translation-Traduction

Weather Translation Demo

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

lecteur Google



rali Weather Warning Translation Prototype version 2

Source text	Translation
<p>East Vancouver Island 4:22 AM PST Tuesday 04 January 2011 Snowfall warning for East Vancouver Island continued</p> <p>Snow will persist over inland sections of the north and central coasts and will spread to the south coast later today.</p> <p>This is a warning that significant snowfall is expected or occurring in these regions. Listen for updated statements.</p> <p>An approaching Pacific frontal system will sweep down the British Columbia coast today spreading snow to most coastal regions.</p> <p>More snow is forecast for the north and Central Coast - inland sections where 10 to 20 cm of additional accumulations can be expected through Wednesday morning. The snow is expected to change to rain later on Wednesday as warmer air spreads inland.</p> <p>Snow will develop over the south coast later this morning and early this afternoon with widespread amounts of 5 to 15 cm expected by Wednesday morning. Higher amounts are forecast for Howe Sound where up to 20 cm may fall by Wednesday afternoon. In areas near the water, the snow is forecast to slowly change over to rain tonight or early Wednesday as the warmer Pacific air begins to mix down to the surface.</p>	<p>Île de Vancouver Est à 4h22 hnp le mardi 04 janvier 2011 l'avertissement de neige pour l'île de Vancouver Est toujours en vigueur</p> <p>La neige continuera de tomber sur les secteurs de l'Intérieur de la Côte nord et la côte centrale et se propagera sur la Côte sud plus tard aujourd'hui.</p> <p>Ceci est un avertissement indiquant qu'il y aura ou qu'il y a actuellement de fortes chutes de neige dans ces régions. Veuillez surveiller les bulletins météorologiques et leurs mises à jour.</p> <p>Un system frontal du Pacifique balaiera le long de la côte de la Colombie-Britannique aujourd'hui tout en propageant de la neige sur la plupart des secteurs côtiers.</p> <p>On prévoit encore de la neige sur la Côte nord et la côte centrale - secteurs de l'Intérieur ou l'on attend encore de 10 à 20 cm de on peut s'attendre à une jusqu'à mercredi matin. La neige devrait se changer en pluie par la suite mercredi alors que de l'air plus chaud se propagera dans l'Intérieur.</p> <p>Il y aura de la neige sur la Côte sud plus tard ce matin et en début d'après-midi généralisée avec une accumulation de 5 à 15 cm de neige d'ici mercredi matin. Une accumulation plus importante est prévue pour la Baie Howe la ou jusqu'à 20 cm</p>

Translation direction:

Click to get sample source text in: [French](#) or [English](#)

Caractéristiques de WATT

- Développement depuis 2004
- Moteur de traduction statistique Moses
- Entraînement sur un corpus de 40 M de mots (bulletins et avertissements)
- Disponible
 - Sur Linux (ligne de commande Python)
 - Sur le Web (interface graphique)
- Présenté à EC en janvier 2010

Évaluation de WATT par le Bureau de la traduction – été 2010

- 135 bulletins artificiels
- Conclusions très critiques
 - « 65,19 % des échantillons traduits présentent une ou plusieurs erreurs graves[...] »
- Mais méthodologie douteuse
 - Évaluation non à l'aveugle
 - Typologie d'erreur vague
 - Évaluation anglais → français seulement
 - Aucune prise en compte de l'entrée
 - Ne tient pas compte du risque du temps plus long de la traduction manuelle
 - Peu d'information sur le processus d'annotation

Évaluation de WATT par le RALI – *automne 2010*

- Évaluation rationnelle de WATT
 - Protocole état de l'art
 - Agrégation des résultats non complaisante
- Afin de proposer un positionnement de WATT dans la chaîne de traitement d'EC

Évaluation de WATT par le RALI

- 32 bulletins (101 phrases) tirées au hasard par les alertes émises à l'été 2010

Corpus	Nb. de mots	% mots inconnus	Perplexité
Corpus d'évaluation du RALI	1571	3,4 %	27,6
Corpus d'évaluation du BT	12 018	5,1 %	38,1

- 5 annotateurs (non impliqués dans WATT)
- Annotation de paires ⟨source, traduction⟩
 - Pour chaque source: trad. humaine et machine
 - Chaque paire annotée 2 fois: 404 groupes d'annotations à réaliser
- Annotation *à l'aveugle*

7 types d'erreurs à relever

Erreurs sur le sens		
	Majeures	Mineures
Charabia	CHA1 Passage présentant un langage, style incompréhensible ou grossièrement incorrect.	S.O.
Omission	OMIS1 Omission d'une donnée ou d'un élément de sens important pour la compréhension du texte.	OMIS2 Omission sans gravité ne nuisant pas à la compréhension de façon significative.
Faux-sens	FAUX1 Ajout d'une information incorrecte. Sens contraire à l'original.	FAUX2 Insertion d'une ambiguïté gênant un peu la compréhension. Sens différent mais voisin de l'original.
Erreurs linguistiques		
LANG	Non-respect des règles d'orthographe, de grammaire et de syntaxe.	
TYPO	Non-respect des règles de casse, ponctuation et diacritiques (accents, cédilles, etc.) et césure des mots.	

Interface d'annotation

Normal Annot2_a.doc [Mode de compatibilité] - Microsoft Word

Accueil Insertion Mise en page Références Publipostage Révision Affichage Développeur Compléments

Charabia OmissionMajeure OmissionMineure FauxSensMajeur FauxSensMineur Langue Typo

Barres d'outils personnalisées

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18

Paire de phrases 97

Original

Ils devraient traverser le secteur de la minerve - rivière ronge dans la prochaine heure.

Traduction

They should tend to move through the la minerve - ronge river within the next hour.

Commentaire [p17]: OMIS2

Commentaire [p18]: TYPO

Commentaire [p19]: TYPO

Commentaire [p20]: FAUX2

Commentaire [p21]: TYPO

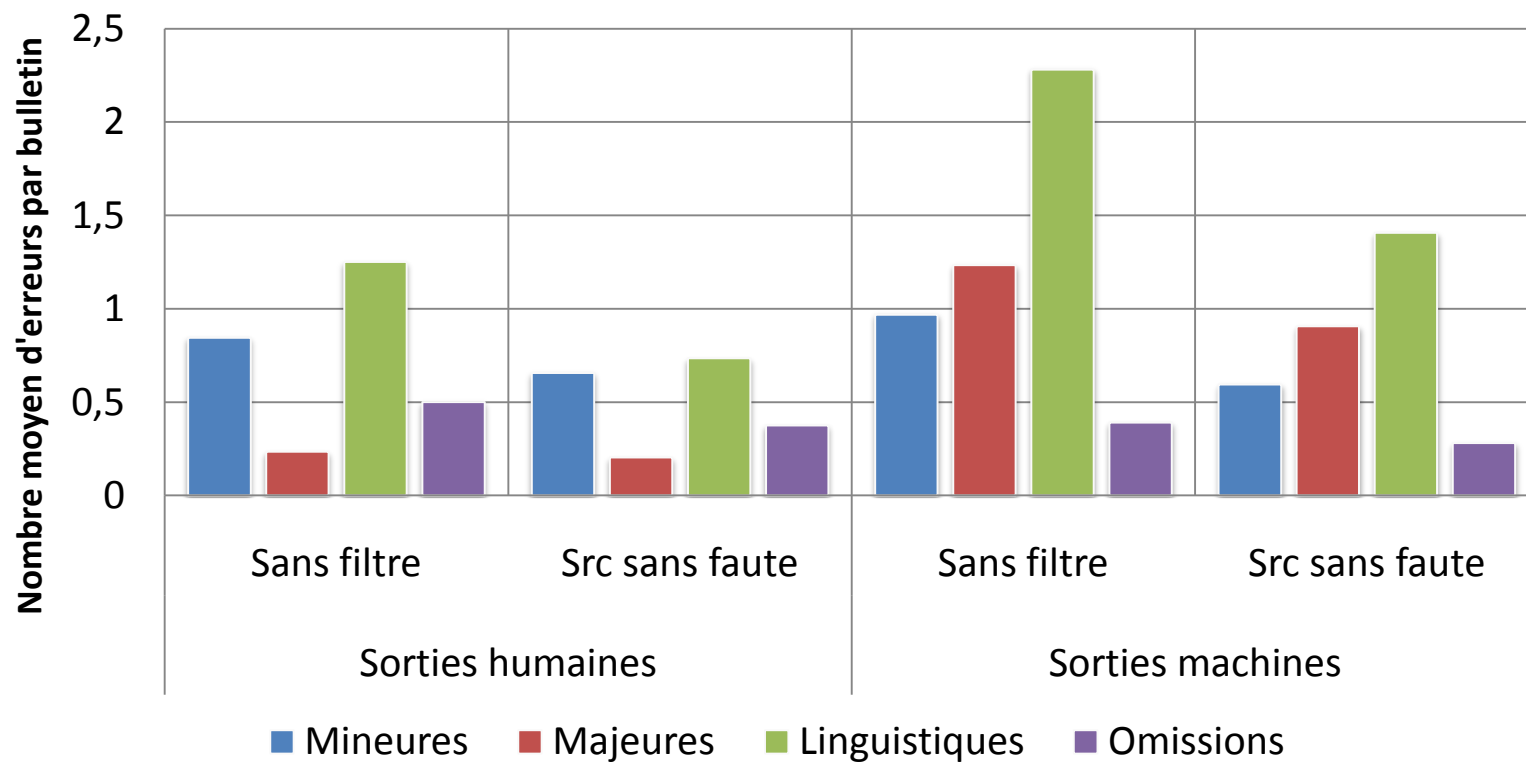
Commentaire [p22]: TYPO

Cette traduction respecte-t-elle l'exactitude du sens et des données de l'original ? Répondre par **oui** ou **non**.

Oui

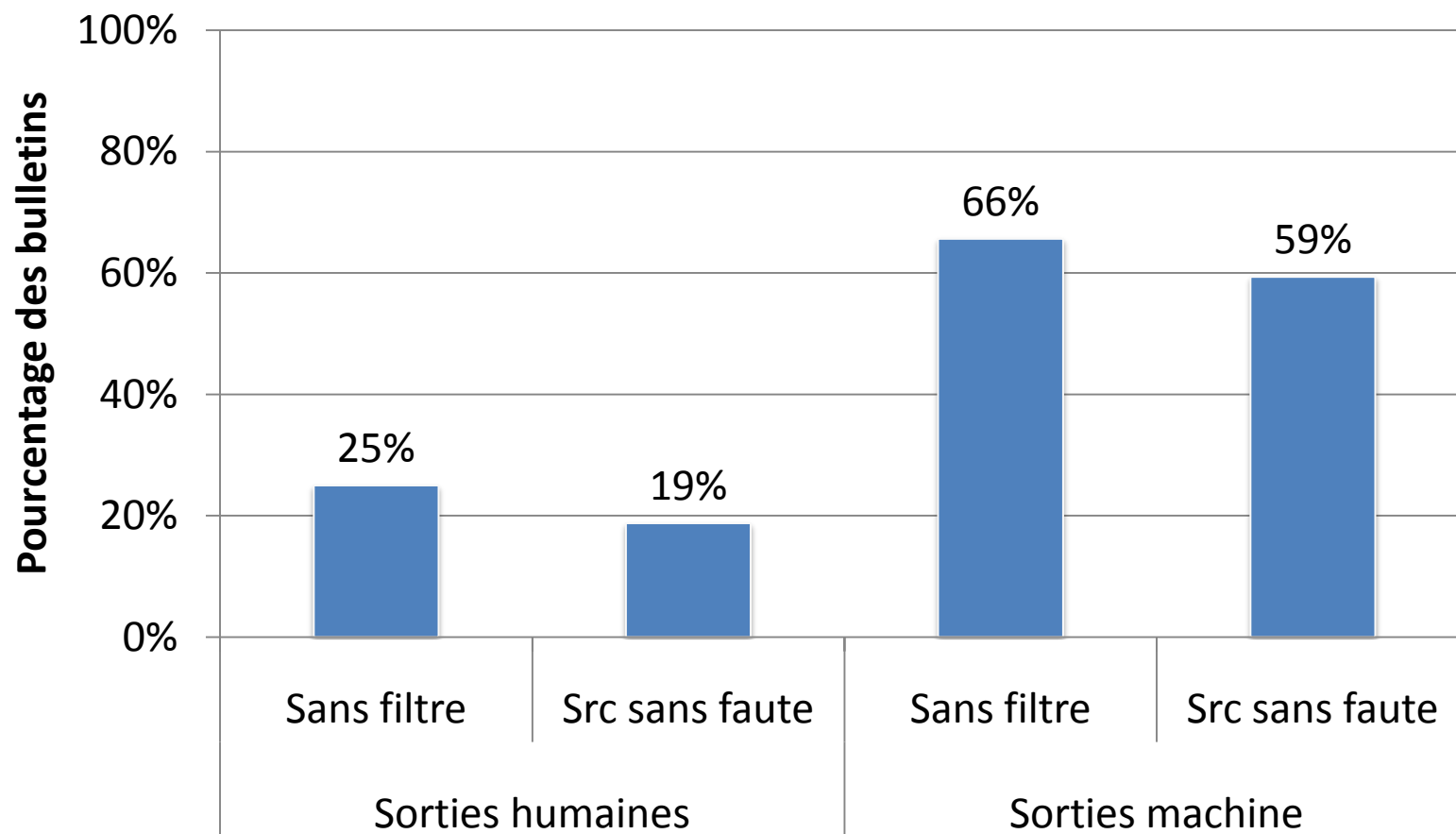
Page : 17 sur 42 Anglais (États Unis) 95 %

Nombre moyen d'erreurs par bulletin traduit, pour différents systèmes de traduction



- Les traductions humaines comportent moins d'erreur
- Erreurs moins fréquentes sur une entrée sans faute

Pourcentage des bulletins traduits contenant au moins une erreur d'exactitude



Erreurs d'exactitude dans la traduction

Traduction humaine

Source

Diminishing northerly winds and clear skies will give ideal conditions tonight for temperatures to fall to the low single digits over Eastern Nova Scotia.

Cible

Des vents du nord diminuant d'intensité et **le ciel** dégagé fourniront les conditions idéales cette nuit pour faire tomber les températures à **1 à 3 degrés** sur l'est de la Nouvelle-Écosse.

Traduction machine

Source

The moisture from the recent rains combined with daytime heating are creating conditions favourable for the development of thunderstorms.

Cible

L'humidité **par les** récentes pluies, combiné au réchauffement diurne **qui tombe** crée des conditions propices à la formation d'orages.

- Annotateurs minutieux, peu complaisants

Erreurs dans la source

Catégorie	Phrase source annotée
Ponctuation	As a result..Patchy frost will form over low lying inland areas where ground temperatures will be near zero.
Nom de lieu mal écrit	Ils devraient traverser le secteur de la minerve - rivière rounge dans la prochaine heure.
Incohérence de notation	About 20 millimetres of rain has already fallen. Another 15 to 25 mm of rainfall is expected in Old Crow before the disturbance begins to weaken this evening.
Orthographe	Weather conditons are no longer favourable to produce funnel clouds.
Casse	attention ...

- 25 % des 101 phrases avec au moins 1 erreur
 - 53 % de ces erreurs: ponctuation
 - 30 % de ces erreurs: noms de lieux

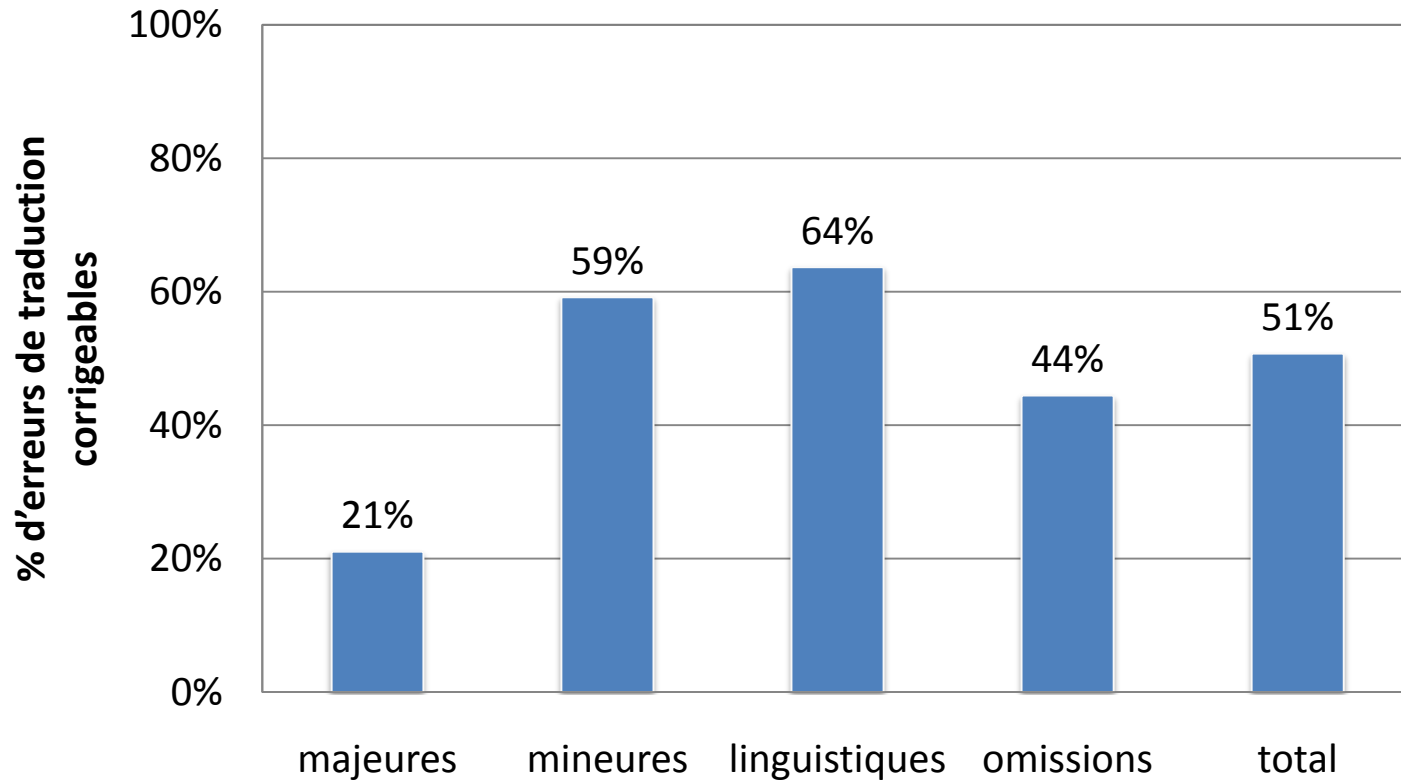
Deux pistes d'amélioration

1. Améliorer le moteur de traduction de WATT
2. Assister le processus de saisie des bulletins

Améliorations du moteur de traduction

- À court terme
 - Tenir compte de tournures particulières
- À moyen terme
 - Accentuation, casse
 - Unités de mesure
- À long terme
 - Fautes de grammaire et d'accord
 - Mots inconnus
 - Anaphores non résolues

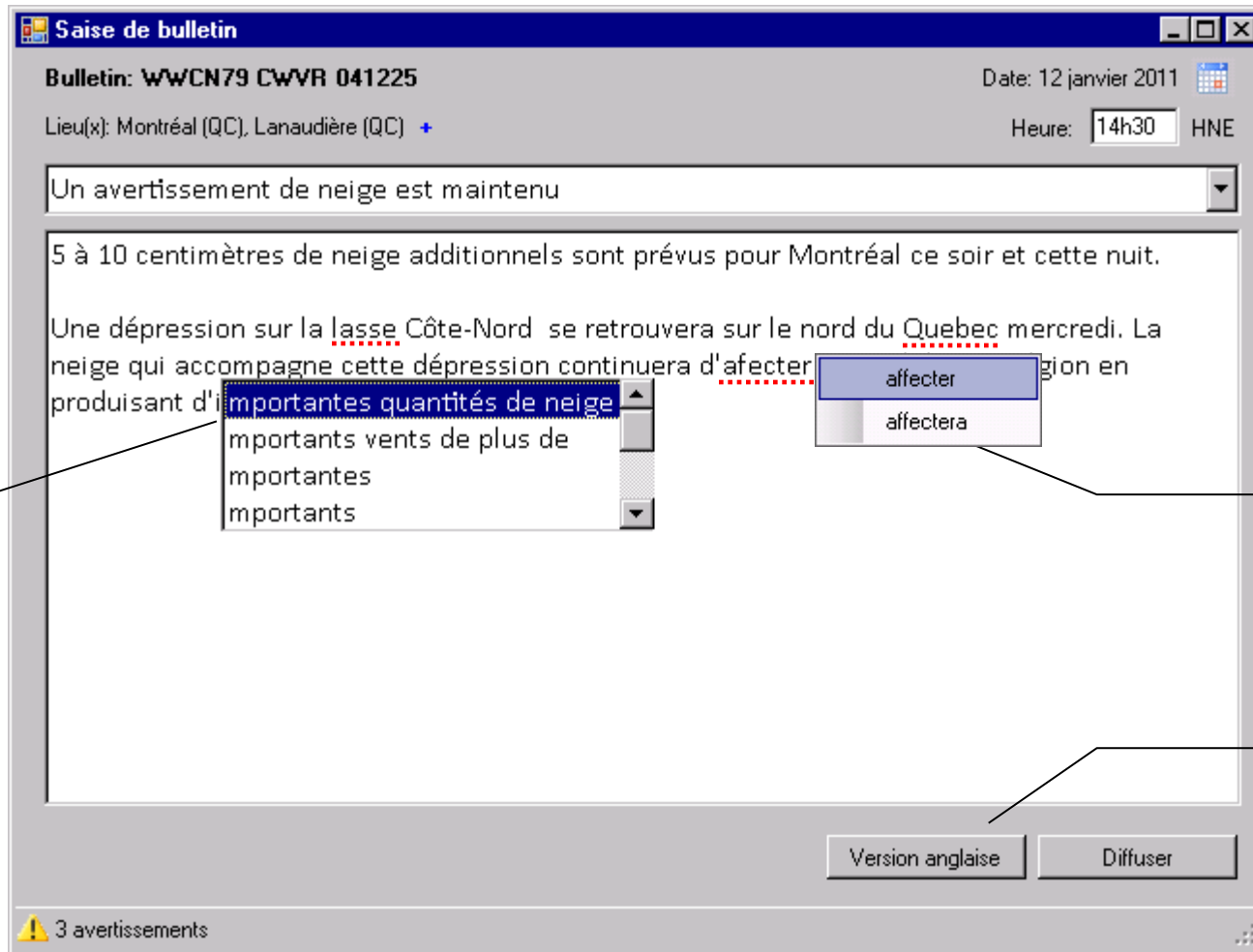
Proportion d'erreurs corrigéables à court et moyen terme, pour différents types d'erreurs



Améliorations lors de la saisie

- Vérification des marqueurs de fin de phrase
- Validation de la casse
- Validation des noms de lieux
- Standardisation des désignations d'unités de mesure (SI)
- Intégration d'un correcteur d'orthographe

Proposition d'interface de saisie

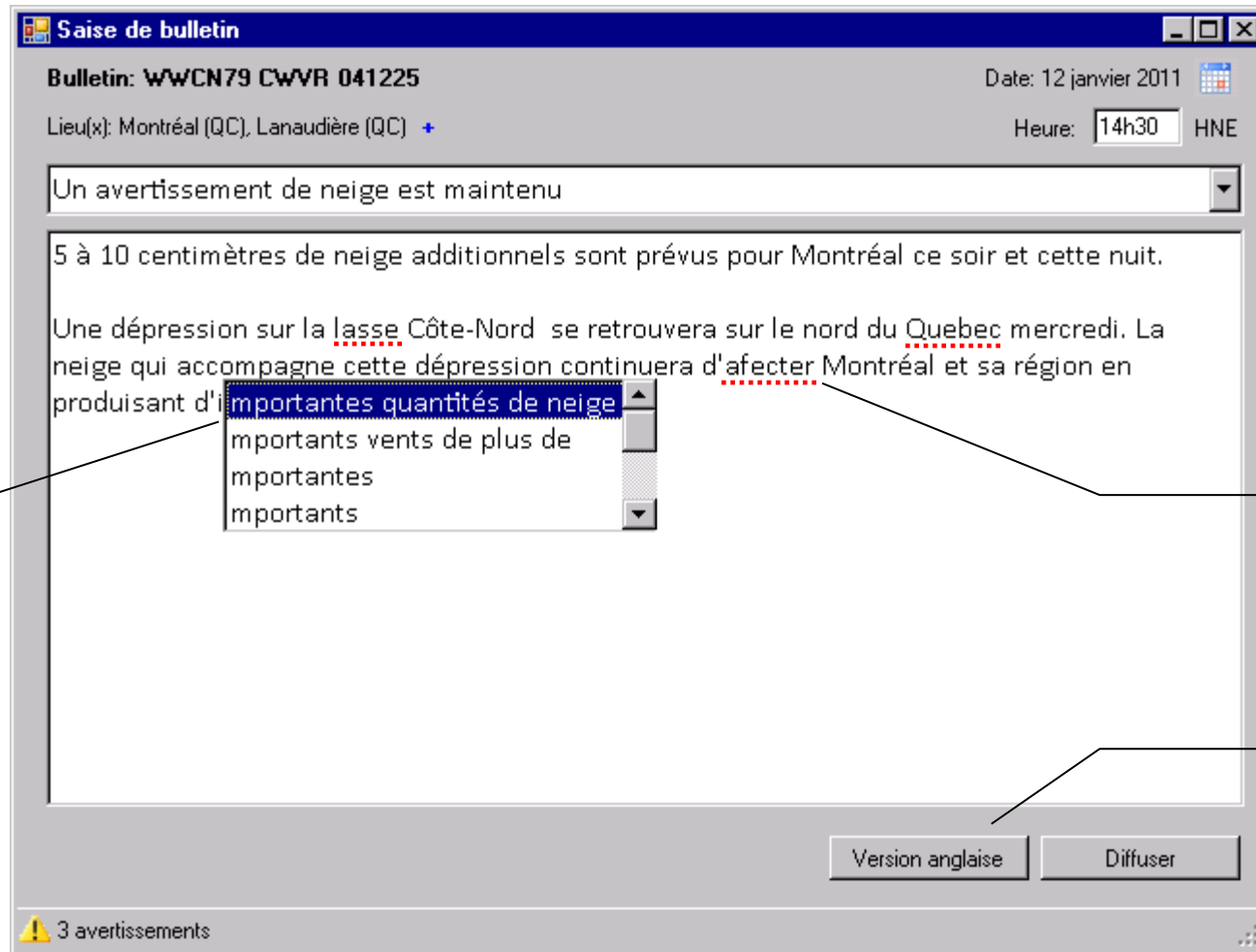


Complétion automatique

Vérification du texte

Traduction

Proposition d'interface de saisie

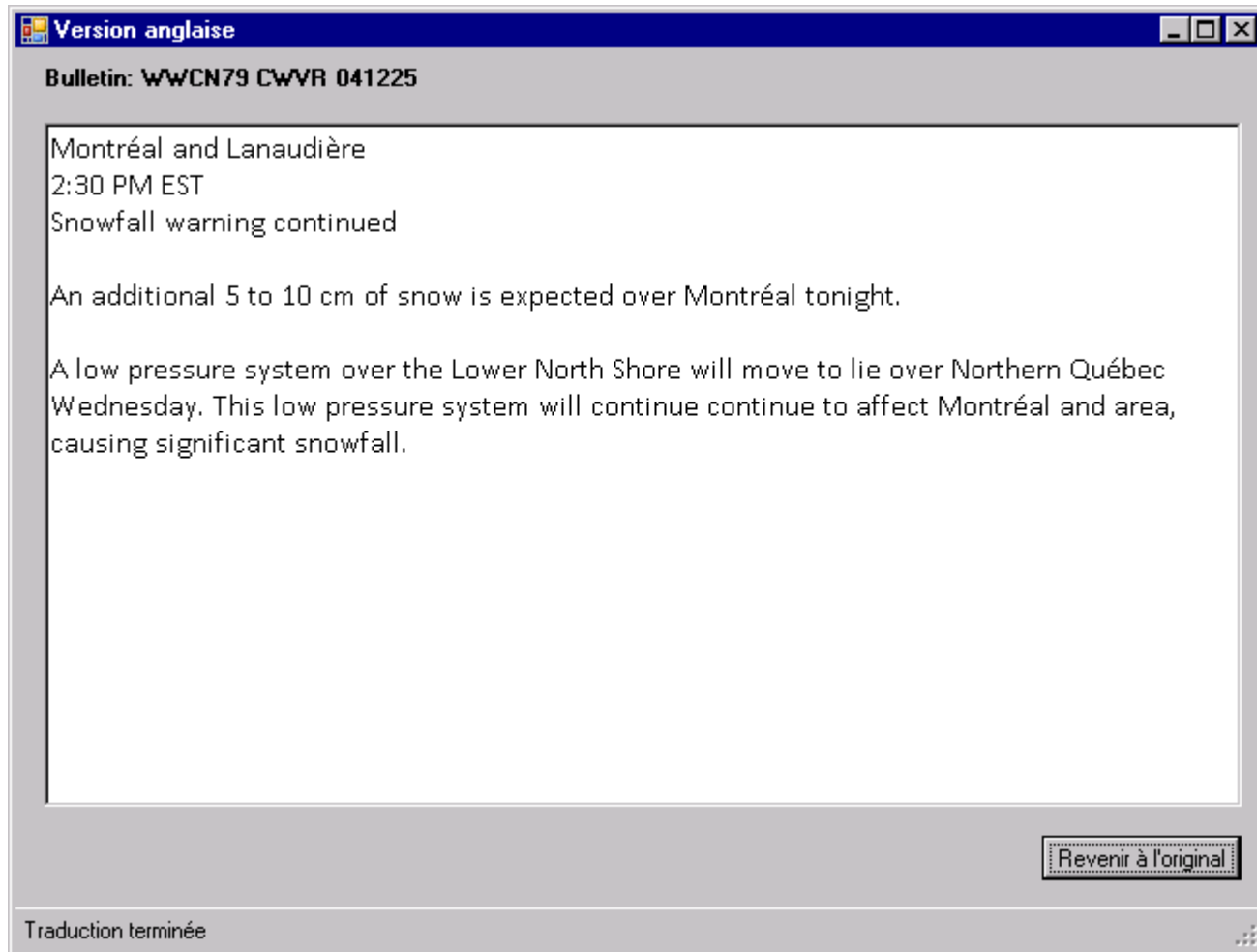


Complétion automatique

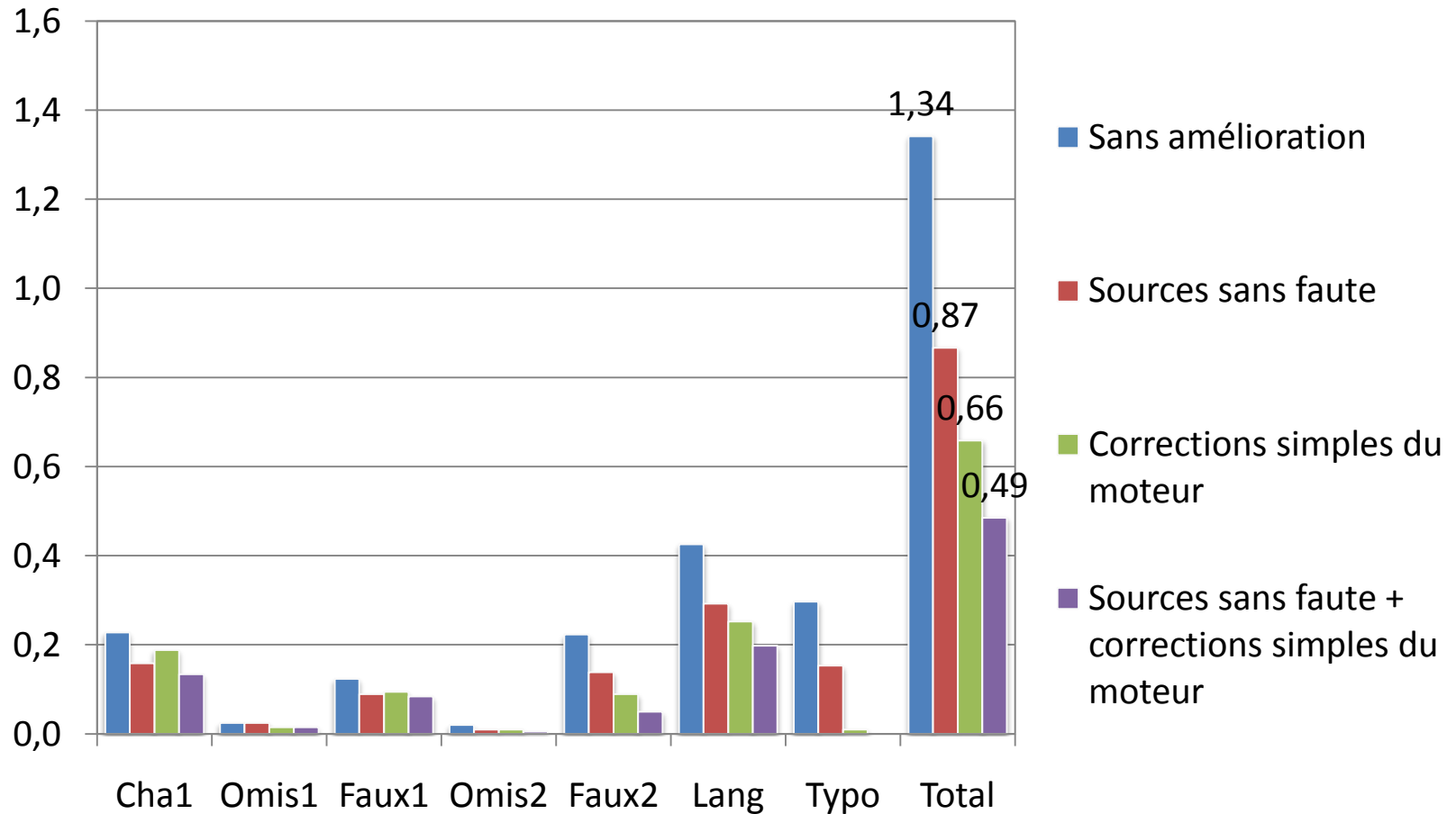
Vérification du texte

Traduction

Proposition d'interface de saisie



Nombre moyen d'erreurs par phrase, pour différentes stratégies d'amélioration



Conclusions

1. WATT semble utilisable comme premier jet de traduction, *avec notice au lecteur*
2. Des améliorations sont envisageables à court terme
3. Il est important de normaliser l'intrant.

➤ **Test à Environnement Canada ?**