

Analysis of translation suggestions on Reverso translation engines: initial findings

Daniel Déchelotte

Softissimo, Inc.

2 July 2010



www.faust-fp7.eu

Analysis of translation suggestions on Reverso translation engines: initial findings

Introduction

Millions of users rely on the www.reverso.net site to translate their texts in more than 30 translation directions. After the translation, they now have the possibility to suggest a different translation, a feature that has been first enabled in March 2008. To date (June 2010), Softissimo accumulated more than 100,000 translation suggestions. The challenge is now to automatically extract useful information from those suggestions, by identifying problematic expressions and constructs, and populating lists of new entries to add to our models.

More specifically, the objective of this early study is to compile a list of criteria to select useful feedbacks. Two slightly different objectives will actually be distinguished:

1. Objective 1: The first list will be conservative and only filter out “obviously useless” feedbacks (user mistakes, malevolence, etc). The key is here to include all potentially useful suggestions, even if that means including “too many” of them.
2. Objective 2: The second list will strive to select the cleanest feedbacks, i.e. the feedbacks that are presumably the easiest to exploit. Such a list could be engine-specific, as the notion of being “easy to exploit” is tied to each engine’s technology. For the moment, however, this objective will try to keep all “generally useful feedbacks”, with as little bias as possible towards one particular engine.

Method

All suggestions ever stored from the launch of the feedback interface have been collected in a flat text file, for the two translation directions between French and English. First, some key figures describing the data were computed. Then, the following process was repeated: manual annotation of small sets of feedbacks, analysis of the key “trends” and design of additional filtering criteria. This led to the recommended set of filters.

Data

In the remainder of the document, we call “triple” the context of a translation suggestion, which is defined as the triple (source text, automatic translation, suggested translation).

Translation suggestions for the English-French direction

The following amount of suggestions is available:

- 13586 suggestions in 2008 (launch date: March 2008)
- 15191 suggestions in 2009
- 3869 suggestions in 2010, as of April 30th
- Hence a total of 32645 suggestions

Figure 1 presents the monthly evolution of the number of suggestions, as well as their breakdown in short, middle and long source texts. An annual periodicity can be detected, with a decreased traffic during summer. Also visible is the recent change in the UI, in early 2010, to no longer display

Analysis of translation suggestions on Reverso translation engines: initial findings

“suggestion” button for short source texts (it is now planned to reverse that decision). Overall, the amount of suggestions is fairly flat across time.

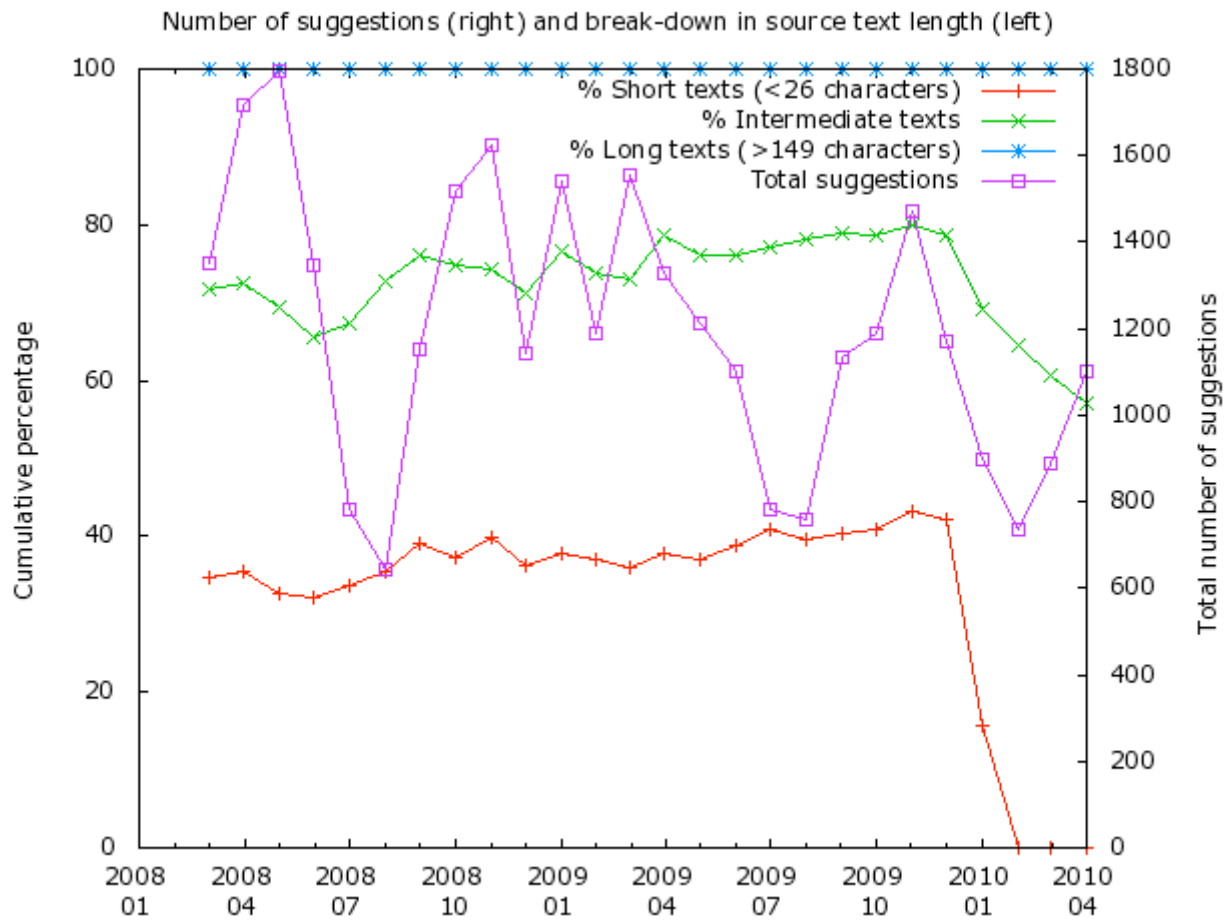


Figure 1: Translation suggestions for English-French

Translation suggestions for the French-English direction

The following amount of suggestions is available:

- 11160 suggestions in 2008
- 11515 suggestions in 2009
- 2922 suggestions in 2010, as of April 27th
- Hence a total of 25597 suggestions

Figure 2 shows similar patterns as Figure 1, especially regarding the crude annual periodicity and the drop in suggestions for short texts.

Analysis of translation suggestions on Reverso translation engines: initial findings

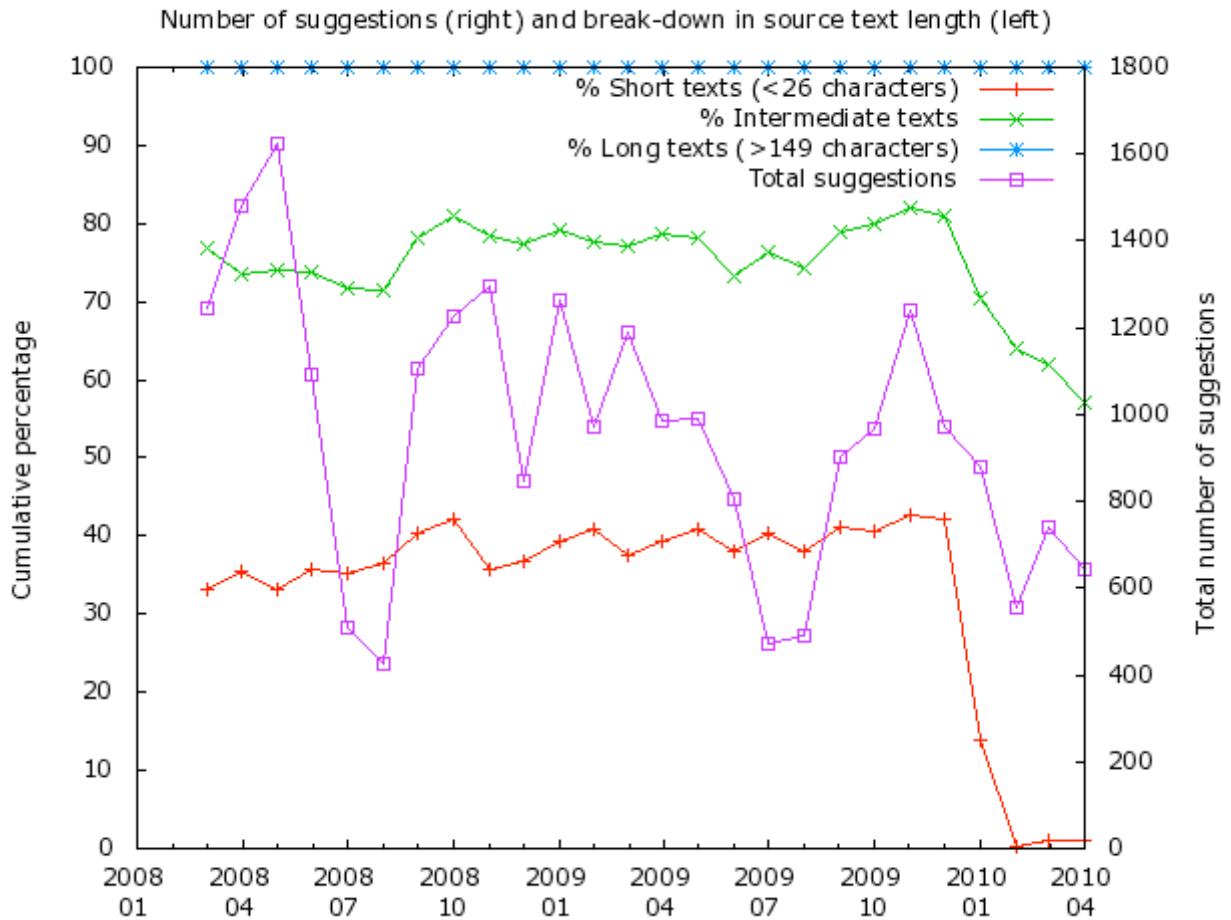


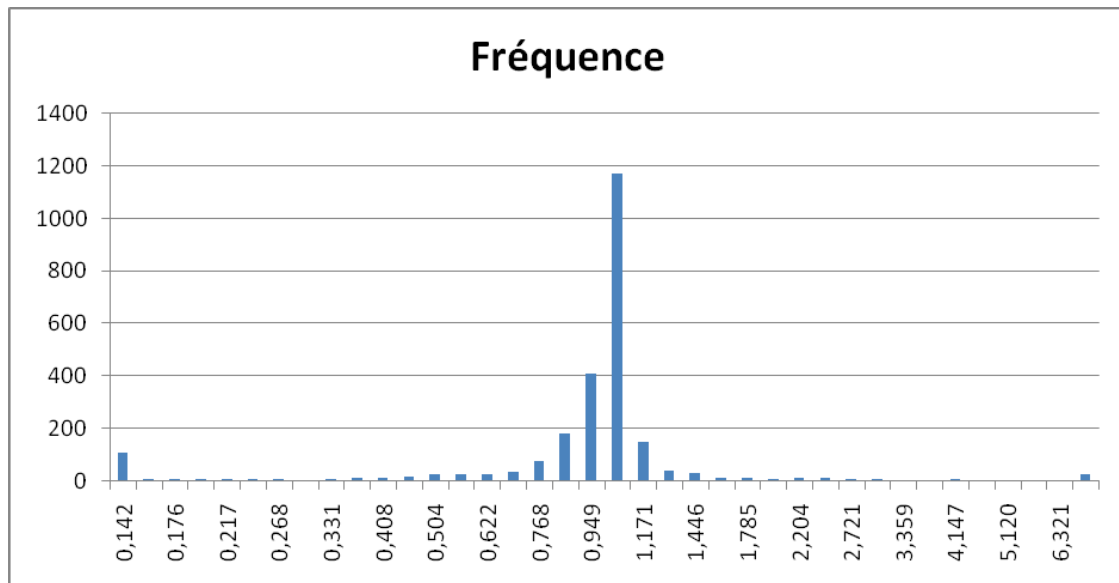
Figure 2: Translation suggestions for French-English

Investigating a criterion on the length of the suggestion text

The rationale behind a criterion on the length of the suggestion text is that a valid suggestion should probably be of similar length as the original, automatic translation.

Below is a histogram plot of the ratio of the length (in characters) of the suggested translation over the length of the automatic one, for the French-English direction.

Analysis of translation suggestions on Reverso translation engines: initial findings



The central bin covers ratios between 0.95 and 1.05. It gathers almost half of all suggestions. 1% of all suggestions are completely empty.

A criterion filtering out extreme values of this ratio could be used as part of the Objective 1 (eliminating the obviously wrong feedbacks). In examined examples, a ratio below 0.45 was always associated to a useless suggestion (either a mistake or a malicious entry). Moreover, no potentially interesting suggestion had a ratio above 1.6. A few limit cases:

- The triple (« comite supérieur de la formation et de la recherche stratégiques », “Comoth superior(higher education) of the strategic training(formation) and the search(research)”, “High council for strategic trainig and research”) has a ratio of 0.49. This is mainly due to the presence of alternatives (in brackets) in the initial translation.
- The triple (« ils etait obliger d'aller a l'école », “They was to oblige to go has the school”, “they had to go to school”) has a ratio of 0.62.
- The triple (« Je travaille ici depuis deux ans », “I work here for two years”, “I have been working here for two years”) has a ratio of 1.52.

Filtering at 0.45 gets rid of a bit less than 8% of all entries, and filtering at 1.6 discards 3.5% of them. It is possible to improve the lower filtering by stripping any content between parentheses in the automatic translation before computing the length ratio. The lower bound can thus be raised to 0.6.

Although the two thresholds of 0.6 and 1.6 were set by observing suggestions for the French-English translation direction, a cursory check on English-French suggestions validated the easy solution of reusing those thresholds unchanged for that direction as well.

One might even be tempted to claim that those thresholds are reasonable as long as the target language uses latin characters. Attention must be paid, though, at the fact that some translation engines might tend to produce rather verbose translations, while others might produce rather short texts.

Analysis of translation suggestions on Reverso translation engines: initial findings

Analysis of a set of 100 translation suggestions (French-English direction)

A first set of 100 suggestions were categorized according to the following classes:

- Improved translation
- Perfect translation (a subset of improved translations)
- Degraded translation
- Unchanged translation
- Mixed: part-improved and part-degraded translation, or other hard to tell cases
- Malevolence or user mistake

Class	#	Comments, examples
Improved	47	<ul style="list-style-type: none">• All errors are not necessarily corrected (Auto : He(It) is still in the head of this organization, Sugg: He is still at the head of (in charge of) this organization)• Includes selection of alternatives (between brackets) (see above, or when the automatic translation of « ils ne comprennent pas » contains “understand(include)” and the suggestions only keeps “understand”).• Includes corrections of errors due to errors in the source text (Src : fenetre (missing accented letter), Auto : fenetre, Sugg : window)
Perfect (subset of improved)	13	Example: Src: Dans les pays les plus démunis en Afrique ; Auto : In countries the most deprived in Africa ; Sugg : In Africa's most deprived countries
Degraded	6	Example : Auto : Oh! You already went in France several times!; Sugg: Oh! You already went in French several times!
Unchanged	11	No changes, or space-changes only
Mixed	9	Various examples where the user visibly did his best, but the suggestion contains new errors: Auto : ecoles of ingenieur ; Sugg : ecoles of ingénieur Auto: being capable of; Sugg: being able of Auto : Saw you I speak too very English, Sugg: You see I speak very well English
Malevolence, user mistake	16	Suggestion contains insults, unintelligible or unrelated text, etc

Analysis of translation suggestions on Reverso translation engines: initial findings

Based on this analysis, it appears that suggestions are mainly produced by well-intentioned users, although not necessarily experts. We were pleasantly surprised to observe that malicious entries and mistakes were in limited number and that most of them could be eliminated by simple heuristics: meaningless sequences of characters, translation suggestion way too short or too long, detecting common profanities, for instance. However, most of the remaining suggestions correct errors that are themselves due to errors in the source text: those suggestions are valid and yet harder to exploit.

The following case occurs frequently. A user types a French text and misses one accent (“fenetre” instead of “fenêtre”, for instance). The unaccented word is unknown to the translation engine, which reproduces it as is. That error is easily spotted by the user, who corrects the translation with “window”.

This observation led to another analysis, geared towards errors in the source text.

Analysis of 50 additional suggestions: errors in source text

A fresh set of 50 triples (source, automatic translation and suggested translation) have been analyzed. The first part of the analysis is to detect and categorize errors in the source texts. Those errors are described in the table below; effects on the translation are detailed further down.

Type of problem	Number of source texts with that problem	Examples
Punctuation	28	Missing final period, or other missing punctuation (comma, period, dash, hyphen, apostrophe, ...)
Space	8	Missing space between a punctuation and a word nearby, between two words or between a figure and its unit. Superfluous spaces are ignored.
Case	25	Lowercased first word or proper noun, uppercased words, etc.
Spelling	12	All kinds of expected and surprising spelling errors (excluding grammar errors below)
Grammar	11	Agreement errors, mostly
Miscellaneous	10	Fragments, texts in foreign languages, abbreviations, insults, ...

Based on this manual detection of problems in the source text, 22 triples (out of 50) are found to contain no serious problems (punctuation-, space- and case-related errors being ignored) and are further examined. Their respective translation suggestions are classified according to their quality:

Analysis of translation suggestions on Reverso translation engines: initial findings

Class (evaluation of the suggestion)	#	Comments, examples
Improved	8	For instance, “un-translation” of some terms (achatvip.com wrongly translated into purchase vip.com), “as indicated to the telephone” corrected into “as indicated on the phone”, ...
Perfect (subset of improved)	4	For instance, correction of a bad translation due to a missing hyphen in « pouvez vous... ». A good one: Src: “Nous vous invitons à venir découvrir avec nous notre coin de paradis”; Auto: “We invite you to come to discover with us our piece of heaven.”; Sugg :” Come share with us our piece of heaven.”
Degraded	5	Example: Auto: “I did not know how to help them”; Sugg: “I did not how help them”
Unchanged	3	No changes, or space-changes only
Mixed	1	Src: “nous vous prions de nous communiquer votre réponse”; Auto: “We you prions to communicate us your answer”; Sugg: “we ask you kindly togie as your answer” Note that somme suggestions annotated as “improvements” are not clear-cut and could be counted in this category
Malevolence, user mistake	5	

A great portion of the suggestions that were previously deemed “useful” actually pertain to source texts containing errors. That’s why restricting to error-free source texts leads to a significantly lower “usefulness rate” (8 out of 22, themselves out of 50, as opposed to the 47 out of 100 as reported in the first set of 100 suggestions).

Note that some perfect suggestions are actually so idiomatic and specific that they will be hard to exploit (like changing “We invite you to come to discover with us our piece of heaven” into “Come share with us our piece of heaven”).

Designing filters for Objectives 1 and 2

A potential feature: spell checking of source text

Running a spell checker on the source text, i.e. the original translation request, could constitute a great filter for Objective 2 for two reasons:

- By restricting the analysis to (presumably) error-free text, the provided feedback is expected to be easier to exploit;
- Being able to correctly spell the source language may be viewed as an indicator of the “linguistic expertise” of the user. In other words, for Objective 2, we will dismiss the feedback of users that can’t even spell correctly in the source language.

Analysis of translation suggestions on Reverso translation engines: initial findings

The performance of the spell checker was quickly assessed on a new set of 60 source texts. No errors were detected in 21 of them, although 6 of them actually contained grammar errors. At least one spelling error was detected in the other 39 source texts, and all of them did have errors (although not necessarily the ones detected by the speller). Therefore, when used as a binary filter, the speller alone had a precision of 71% (15 out of 21) and a recall of 100%.

Another potential feature: language identification

The libtextcat-2.2 package provides a program and a library performing language identification. It is available from <http://software.wise-guys.nl/libtextcat/> under the BSD license. Several sources (web sites, high-profile projects like SpamAssassin) testify of the package's excellent performances, both in terms of classification quality and in terms of execution speed.

Detection relies on comparing the document's "fingerprint" to a set of reference fingerprints computed for numerous languages. The fingerprint is built from character n-grams ranked by decreasing frequency, as described in [Cavnar and Trenkle, 1994]. The software package comes with fingerprints for 69 different languages.

In my tests, I only kept the intersection of the covered languages with the languages for which Softissimo has a system. Those languages are Arabic, Chinese, English, French, German, Hebrew, Italian, Japanese, Portuguese, Russian and Spanish. Dutch was removed because, in my tests, it hindered the detection of English texts. In fact, only French, English and Spanish texts were detected when translating from French to English.

The program is configured not to perform detection on texts shorter than 24 bytes. When used to filter source texts, shorter entries will thus always be considered in the right language, as identified by the user. Similarly, texts that leave the program undecided between several languages (indicating extremely close scores) are generously given a free pass. Only the texts that are identified in a "wrong" language and those classified in the reject class (like, for instance, "xxxxjx uuusususus fjjjjjjjj xxx") are evicted by this filter.

When evaluated on hundreds of sample texts, the above algorithm has been wrong only on two occasions: "Hampton Court Palace visit" and "dominant autosomic transmission" were falsely classified as French.

Proposed filter for Objective 1

The proposed filter performs the conjunction of the following criteria:

1. The source text contains at least one alphabetical character
2. The source text is in the expected language (here, French), or more exactly it is not classified as being in another language, as described in the subsection on language identification
3. The automatic and suggested translations are not identical (case and space insensitive comparison)
4. The ratio of the suggestion over the automatic translation is between 0.45 and 1.6.

Analysis of translation suggestions on Reverso translation engines: initial findings

5. The suggestion cannot be obtained by only deleted words from the automatic translation (insertions or substitutions have to be performed). Deleting letters or selecting a variant (in parentheses) are considered as a substitution and therefore pass this criterion.
6. The suggestion is in the expected language (here, English), or more exactly it is not classified as being in another language, as described in the subsection on language identification

This filter has not been evaluated.

Proposed filter for Objective 2

The proposed filter performs the conjunction of the following criteria:

1. The source text is free from spelling and grammar errors, as detected by the spell checker (corrections related to punctuations and case are ignored)
2. The source text is not entirely uppercased
3. The source text is in the expected language (here, French), or more exactly it is not classified as being in another language, as described in the subsection on language identification
4. The automatic and suggested translations are not identical (case and space insensitive comparison)
5. The ratio of the suggestion over the automatic translation is between 0.6 and 1.6.
6. The suggestion cannot be obtained by only deleting words from the automatic translation (insertions or substitutions have to be performed). Deleting letters or selecting a variant (in parentheses) are considered as a substitution and therefore pass this criterion.
7. The suggestion is in the expected language (here, English), or more exactly it is not classified as being in another language, as described in the subsection on language identification

The first criterion is crucial for the quality of the filtered suggestions and can be explained as follows: the quality of the source text is a good predictor of the suggestion's quality. The second criterion can be viewed as an extension of this principle. The third and fourth criteria go without saying and the fifth one is based on results mentioned earlier in this study. The sixth criterion relies on the observation that suggestions that only deleted words were unusable for us (Source: « Le français à annoncer sur la chaîne de télévision Canal+, le 25 avril 2006, qu'il prendra sa retraite à l'issue de la Coupe du monde, en Allemagne. ». Auto: "French to announce *on the television channel Channel(Canal) +*, on April 25th, 2006, that he(it) will retire at the conclusion of the World cup, in Germany.". Sugg: "French to announce on April 25th, 2006, that he(it) will retire at the conclusion of the World cup, in Germany.") or visibly the consequence of a mistake (Source : « Vérifier la tête de votre enfant car il y a dans sa classe des enfants qui ont des pouls ». Auto: "Verify the head of your child because there is in his class of the children who have *pulses*". Sugg: "Verify the head of your child because there is in his class of the children who have")¹.

¹ Interestingly, in both cases, the deleted part was indeed poorly translated. So *some* signal could potentially be extracted from those "suggestions". However, for the time being, those feedbacks will be filtered out.

Analysis of translation suggestions on Reverso translation engines: initial findings

A criterion on the source text length was considered but eventually rejected: the first criterion is more efficient in preserving the right suggestions.

An email address is sometimes provided. This feature has been ignored as well, as it could show the user's goodwill, but tells nothing about its expertise.

For source texts containing errors, we contemplated correcting the errors, translating the corrected source and then processing the new triple (corrected source, automatic translation of the corrected source, suggestion). Indeed, if the new automatic translation is closer to the suggestion than the original automatic translation was, it means that the suggestion corrects the translation of source errors. The suggestion could contain other interesting corrections. This refinement will be examined in a later study.

Blind evaluation of the filter for Objective 2

All suggestions for French-English provided between April 1st and 3rd 2010 were collected, for a total of 57 suggestions. They were fed to the filter for Objective 2 described above. Nine suggestions passed all criteria. They are listed in the table below, along with a short comment.

Source text	Automatic translation	Suggested translation	Comment
j'aimerais devenir hôtesse de l'air	I would like to become a stewardess	I would like to become a air hotess	Correct, but not very interesting.
. Le plus drôle, c'est lorsqu'il vient près de moi lorsque je suis tranquille, avec un jouet dans la gueule. À ce moment, je suis sûre que s'il pouvait parler il dirait « JE VEUX JOUER !»	. The funniest, it is when come I when I am quiet, with a toy in the mouth. At this moment, I am safe(sure) that if he could speak he would say " I WANT TO PLAY! "	. The funniest,it's when I'm quiet, he come with a toy in the mouth. At this moment, I am sure that if he could speak he would say " I WANT TO PLAY! "	Correct. « Je suis sûr que » -> "sure" instead of "safe(sure)" is interesting. Start of sentence hard to exploit.
charles de gaulle est né le 22 novembre 1890	Charles de Gaulle was born on November 22nd, 1890	Charles de Gaulle was born on November 22rd, 1890	Bad.
Hier, j'étais à la mer mais il pleuvait alors je suis rentré chez moi	Yesterday, I was in the sea but it rained then I returned at home	Yesterday, I was at the beach but it rained then I came back home	Correct and interesting.
Demande de documentation. le 19 février 1990. Monsieur, Je cherche des étagères pour livres qui couvriraient tout un mur de mon salon. Pourriez-vous, s'il vous plaît, m'envoyer un catalogue qui présenterait tous les modèles que vous avez en stock ainsi que la liste des prix. En vous remerciant d'avance, je vous prie de croire, Monsieur, à l'expression de mes sentiments les meilleurs.	Demand(Request) of documentation. February 19th, 1990. Sir, I look for shelves for books(pounds) which would cover a whole wall of my lounge(show). You could, please, send me a catalog which would present all the models That you have in stock as well as the price-list. Sincerely Yours The best feelings.	Demand of documentation. February 19th, 1990. Sir, I look for booksshelves which would cover a whole wall of my living room. You could, please, send me a catalog which would present all the models That you have in stock as well as the price-list. Sincerely Yours The best feelings.	« étagères pour livres » -> "bookshelves" instead of "shelves for books(pounds)" and « salon » -> "living room" instead of "lounge(show)" are both interesting.
le climat devient de plus en plus	The climate becomes more and	the climate is becoming colder	Correct and

Analysis of translation suggestions on Reverso translation engines: initial findings

froid	more cold	and colder	interesting.
beaucoup de gens ne veulent pas être contactés n'importe quand.	A lot of people does not want to be contacted whenever.	A lot of people don't want to be contacted at any time.	Correct and very interesting.
Vous êtes invité à participer à notre prochaine séance d'information qui aura lieu le 8 avril 2009 à 16 h 30 dans les locaux du Service des ressources humaines situés au sixième étage. L'ordre du jour portera sur les congés annuels, lesquels doivent être planifiés à l'intérieur de chaque service ainsi que sur vos demandes particulières concernant les vacances de Pâques. De plus, nous profiterons de l'occasion pour discuter de nos nouvelles heures d'ouverture durant les mois de juillet et août.	You are invited to participate in our next information session which will take place on April 8th, 2009 at 4:30 am in the premises of the Service of the human resources situated in the sixth floor. The agenda will concern the annual leaves, which must be planned inside every service as well as on your particular demands(requests) concerning Easter holidays. Furthermore, we shall take the opportunity to discuss our new opening hours during July and August.	You are invited to participate in our next information session which will take place on April 8th, 2009 at 4:30 p.m. in the premises of the department of the human resources situated in the sixth floor. The agenda will concern the annual leaves, which must be planned inside every department as well as on your particular demands(requests) concerning Easter holidays. Furthermore, we shall take the opportunity to discuss our new opening hours during July and August.	Not too interesting (suggests "department" instead of "service" for « service » in « service des ressources humaines »).
Et ne prenez pas s'il vous plaît Vos vêtements de. Je n'apprendrai pas.	And do not set please Your clothes of. I shall not learn.	And do not set please Your clothes of. I will not learn.	Not very interesting (will instead of shall)

All nine entries look "reasonable", in that only it now takes an expert to determine whether, for instance, "will" is a better translation than "shall" in this context, or that "a lot of people" indeed asks a plural verb. Out of the nine feedbacks, manual analysis gives:

- 5 suggestions identify relevant errors, or provide clearly better translations ("I was in the sea/at the beach", « je suis sûre que », "becomes more and more cold/is becoming colder and colder", "a lot of people does/do", "shelves for books(pounds)/bookshelves")
- 3 suggestions suggest plausible but not clearly better translations ("stewardess/air hostess", "service/department of the human resources", "shall/will")
- 1 suggestion is clearly wrong ("22nd/22rd")

Applying the filter to all suggestions

When applied on the 25597 suggestions for the French-English direction, the filter described above for "Objective 2" keeps 8179 suggestions, i.e. approximately 32%.

For the English-French direction, the filter is adapted as follows: spell checking is naturally performed on the suggested translation (in French). Out of all 32645 suggestions, 13861 are kept, or a bit more than 42%.

Analysis of translation suggestions on Reverso translation engines: initial findings

Conclusion

Various analyses of past user feedbacks led to the design of two filters, in order to get rid of the most useless ones (Objective 1) or on the contrary to keep those that are the most likely to be useful in improving the translation engine (Objective 2). On a minimal blind testset of 57 suggestions, an algorithm relying entirely on automatic tools identified 9 potentially useful feedbacks, among which 5 actually pointed to translation errors, or suggested a much improved translation. Even though the filter successfully identified worthy suggestions, some additional work is necessary in order to (i) further filter the suggestions and integrate the feedback in a completely automated manner, and (ii) define *how to present the suggestions* to expert linguists, so that their reviewing work is as efficient as possible.

Future directions

Due to the relative large portion of translation requests containing spelling errors, filtering translation suggestions when the source text is erroneous significantly cuts down the amount of translation suggestions taken into account. Instead, all translation suggestions could be considered when updating the translation models, with entries with fewer errors in the source or target parts “simply” weighting more in the update process than entries with several errors. As noted earlier in this document, another approach would be to automatically correct the original translation request, and translate the corrected text.

Knowing the mother tongue of the user has the potential of being a key feature, although it hasn't been explored in this work. Suggestions might differ in type and quality depending on whether the target language is the mother tongue of the user, or not. That information could be extracted from the profile of a registered user, or could be inferred from the interface language or even from some network properties (browser information, IP address, etc).

Acknowledgments

The author (Daniel Déchelotte) would like to thank Benoît le Ny, John Michaelis and Théo Hoffenberg for ideas and fruitful discussions.

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7-ICT-2009-4) under grant agreement n° 247762.