

# ACCOLÉ: A Collaborative Platform of Error Annotation for Aligned Corpora

**Emmanuelle Esperança-Rodier**

*Univ. Grenoble Alpes, CNRS,  
Grenoble INP\*,  
France*

[Emmanuelle.Esperanca-Rodier@univ-grenoble-alpes.fr](mailto:Emmanuelle.Esperanca-Rodier@univ-grenoble-alpes.fr)

**Francis Brunet-Manquat**

*Univ. Grenoble Alpes, CNRS,  
Grenoble INP\*,  
France*

[Francis.Brunet-Manquat@univ-grenoble-alpes.fr](mailto:Francis.Brunet-Manquat@univ-grenoble-alpes.fr)

**Sophia Eady**

*Univ. Grenoble Alpes, CNRS,  
Grenoble INP\*,  
France*

[Sophia.Eady@univ-grenoble-alpes.fr](mailto:Sophia.Eady@univ-grenoble-alpes.fr)

## Abstract

This article presents a platform, named ACCOLÉ, for the collaborative annotation of translation errors. ACCOLÉ offers a range of services that allow simplified management of corpora and typologies of errors, annotation of effective errors, collaboration during annotation, and finally different kinds of search in corpora. ACCOLÉ allows the annotation of translation errors according to built-in error typologies, Vilar's typology or DQF-MQM or uploaded ones, on several corpora of different texts, translated by different Statistical or Neural Machine Translation systems, as well as processing the annotated corpora created in order to look for typical error models and patterns, related to a specific MT system.

The collaboration feature also gives the possibility to detect any misleading interpretation of an error type among the annotators. ACCOLÉ currently provides 15 corpora, 7 projects of 201,474 words and 18,301 annotations that we will describe. Eventually, we will implement the semi-automatic propagation of found patterns on other corpora to enlarge the scope of linguistic studies, thus providing to the community a wide range of error annotated bilingual parallel corpora.

\* Institute of Engineering University of Grenoble-Alpes

## 1 State of the Art

Translation error analysis is a task that has been developed in recent years in order to improve the machine translation (MT) systems and to qualitatively evaluate these same systems. In addition to the field of MT, translation error analysis is used in particular to improve translation teaching methodologies (Kübler et al, 2016). The integration of MT in translation masters curricula brings a new questioning about the evaluation of MT systems. How do translators feel about different MT systems (Esperança-Rodier *et al.*, 2017)? What tools to use for non-specialists? The answer to these questions requires a linguistic analysis of MT errors in order to be able, firstly, to reveal the typical errors of each type of MT system, secondly, to correlate these typical errors with the post-editing capability of the users, and thirdly to create corpora. Performed manually, the translation error analysis is a very time-consuming task. Research is therefore concerned with the automatic analysis of translation errors (Popović *et al.*, 2006). Manual or automatic, this task requires error analysis criteria, often related to the particular characteristics of the tool used.

The work of Vilar *et al.* (2006), the MeLLANGE project (Castagnoli *et al.*, 2011), Felice and Specia (2012), Wisniewski *et al.* (2013a) and the work of Lommel and Melby (2018) focuses on the definition of typologies of linguistic errors. An error typology consists of a list of errors, categorized according to the corrective actions to be performed to improve the translation. For example, the error type "missing word" indicates that a word of the source segment has not been

translated in the translation hypothesis and therefore does not appear there. In order to improve the translation hypothesis, corrective actions must be taken to translate this "missing word". ACCOLÉ proposes these different typologies of error annotation, but also makes it possible to propose new typologies, better adapted to the correlation of the errors of the systems to the non-correctible errors made by the users. Other work deals with the creation of corpora of error translations and translation error correction, such as Potet *et al.* (2012) and Wisniewski *et al.* (2013b). Other research focuses on the creation of tools for either quality estimation such as Quest ++ (Specia *et al.*, 2015), or quality assessment with Kantan's MT LQR and PET (Aziz *et al.*, 2012). ) or the error analysis such as BLAST (Stymne, 2011), focused on the annotation, Coreference Annotator (Tsoumari and Petasis, 2011) more parallel corpora oriented, YAWAT (Germann, 2008) an alignment tool that allows labelling, and TranslationQ (Steurs and al., 2015) translation and revision oriented.

It is in this last vein that we position ourselves. We want to analyse translation errors on different corpora, in collaboration with people from various backgrounds. Since the task of analysing errors is already tedious, we have created an online platform that can be accessed at <http://lig-accole.imag.fr>. In the rest of this article, we will present our platform ACCOLÉ, before presenting the annotated corpora obtained so far.

## **2 ACCOLÉ, a platform to annotate translation errors**

### **2.1 Purpose**

The initial goal that guided the development of ACCOLÉ is the manual annotation of translation errors according to linguistic criteria. The underlying idea is to be able to provide a user with help in choosing a MT system to use, depending on the context (language and computer skills of the user, knowledge of the domain of the source document to be translated and the task for which s/he needs to translate the source document.) To do this, ACCOLÉ must be able to detect which linguistic phenomena are not handled correctly by the MT system studied. We offer on the same platform a range of services to meet the needs of analysis of translation errors. Thus, the main features of the platform ACCOLÉ are the simplified management of corpora, types of errors, annotators, etc.; annotation of errors; collaboration and/or supervision during the annotation; the search for error models (type of errors at first, later morphosyntactic patterns) in the annotations and finally corpora creation.

Since the task of analysing errors is already tedious, it is important that the people carrying it out have a simple access to the tool as well as the corpus they wish to analyze. This is why the platform is available online on a browser and does not require any specific set up.

The task consists in labelling translation errors by locating manually their occurrences in the target as well as in the source sentence.

### **2.2 Annotation Project Management**

An annotation task is described on the platform thanks to the notion of project. A project is a couple consisting of a corpus and an error typology. Thus, as shown in Figure 1, a corpus can be associated with several error typologies in the form of several annotation projects. The corpus will only be loaded once on the platform.

The project leader will also provide the list of annotators and supervisors. The error typologies are handled by the platform managers.

This notion of project allows not to duplicate corpora.

## Création d'un nouveau projet

[Retour à la liste de projets](#)

TC41

Demonstration at the conference

Corpus

Modèle d'erreurs

Figure 1. Project management.

As presented in table 1, two typologies of errors i.e. Vilar *et al.* (2006) and DQF-MQM (Lommel and Melby, 2018) are available on ACCOLÉ. Fifteen French-English corpora, journalistic news, technical documents, patents, extracts from BTEC (Basic Travel Expression Corpus) climate documents are part of the nineteen existing projects.

Projects	Error Typology	
	Vilar	DQF-MQM
Citi1	Citi1-Vilar/Demonstration	
Citi2	Citi2-Vilar	
News 156 Moses	News 156 Moses - Vilar	
News 156 Google	News 156 Google - Vilar	
News 2000	News 2000 - Vilar	
BTEC1-LIGMoses	BTEC1-LIGMoses - Vilar	BTEC1-LIGMoses – DQF-MQM
BTEC1-NMTLIG	BTEC1-NMTLIG - Vilar	BTEC1-NMTLIG – DQF-MQM
WIPOTranslateN	Hawaii	Brad Pitt
WIPOTranslateS	Madagascar	George Clooney
CANADA_eTranslation	Nouvelle Calédonie	Jude Law
CANADA_LegacyMT@EC	New Zealand	Robert
Doc1_Fujitsu_180118 (monolingue)	Altica test 1	
Doc2_Fujitsu_20180109 (monolingue)	Altica test 2	
<b>TOTAL</b>	<b>13</b>	<b>6</b>
		<b>19</b>

Table 1: Project composition

### 2.3 Error Typology Management

As well as permitting of upload new corpora, the platform offers the possibility to enter other error typologies.

As illustrated in figure 2, an error type will consist of a name, a category (optional), a subcategory (optional) and a code (keyboard shortcut that can be used when annotating).

The screenshot shows a web interface for managing error typologies. At the top, there is a navigation bar with the logo 'Accolé' and several menu items: 'Tableau de bord', 'Supervision', 'Administration', and 'Emmanuelle Esperanca-Rodier (emmanuelle)'. The main heading is 'Modèle d'erreurs : TC41' with two small icons. Below the heading is a blue button labeled 'Retour à la liste de modèles d'erreurs'. The 'Description' field contains the text 'Demonstration at the conference'. Below this is a section titled 'Les types d'erreurs' which contains four input fields: 'Catégorie', 'Sous-catégorie', 'Nom', and 'Code'. To the right of these fields is a blue button labeled 'Ajouter un type d'erreur'. At the bottom of the page, there is a footer with 'Développé par FBM' on the left and 'Copyright 2014 - GETALP LIG' on the right.

Figure 2. Error Typology Creation.

ACCOLÉ offers Vilar and DQF-MQM typologies. We are planning to add MeLLange.

### 2.4 Error Annotation

Once in a project, one can visualize the source and target sentences and identify the error manually by clicking on the source occurrence that has not been correctly translated, and on the corresponding target error occurrence. Figure 3 shows the interface proposed to the annotator.

## Annoter les erreurs du segment 1 - validé

Tableau des couples Valider le couple courant Aller au couple suivant

Etape 1 : sélectionner les mot(s)

Phrase source

Mais ceci n'est possible que si le rôle de la subsidiarité horizontale est clairement énoncé, ce qui n'a pas été le cas dans les traités européens, la Charte des Droits fondamentaux ou le travail de la Convention européenne.

Phrase cible

But this is possible if the role of the horizontal subsidiarity is clear, this was not the case in the case of the European Union, human rights or the work of the European Court of Justice.

Etape 2 : créer l'erreur

Source	Cible	Actions
clairement	clear	Ajouter l'erreur

Récapitulatif Supprimer des erreurs

Source	Cible	Erreur	Actions
ø		Inconnu	
européens	EU	Mauvais choix lexical	
énoncé	.	Mots signifiants	
que		Mots signifiants	

Figure 3. Error Annotation Process.

The annotation is done in two steps.

The first step is to select, with the mouse, words in the source sentence, and their equivalent in the target sentence, presenting a translation error. Single or Multiple selection can be done. The selection of non-adjacent words is possible. One can also select a word and its position in the translation hypothesis.

The second step is to choose the type of errors either with the help of the mouse, or with the aid of keyboard shortcuts, to be associated with the pair of previously selected source/target words. It is important to associate the type of errors with both the source and the target in order to identify error models that will ultimately propagate annotations to other corpora.

### 2.5 A collaborative platform

To answer the inter-annotator agreement issue (Popović, 2018), ACCOLÉ offers two mechanisms to help the annotator in her/his task: a supervisory mechanism allowing a manager to control the progress of the task. This mechanism encourages especially the communication between supervisor and annotator by the possibility of creating threads for a specific source/target phrase pair (asking for clarification on an error type, pointing out an annotation error, etc.); and a collaborative mechanism allowing annotators to communicate about a precise source/target phrase pair among themselves. This latest mechanism is an option to be activated in the project while creating it.

### 2.6 SSTC representation of errors

The platform uses a data representation based on the Structured String-Tree Correspondences (SSTC) (Boitet and Zaharin, 1988). An error consists of a label and a set of SNODEs (range representing the substring in the corresponding source or target phrase). For example, in Figure 4, the error concerning "all" and "any" is described by its label Poor lexical choice (cat. Object word, sub-cat. Sense), by its position in the source sentence (SNODE [49-

56] - substring between the 49th and 56th characters) and the target phrase (SNODE [46-48]). The advantage of using SNODEs in this way is to avoid the use of a syntactic structure to describe the error.

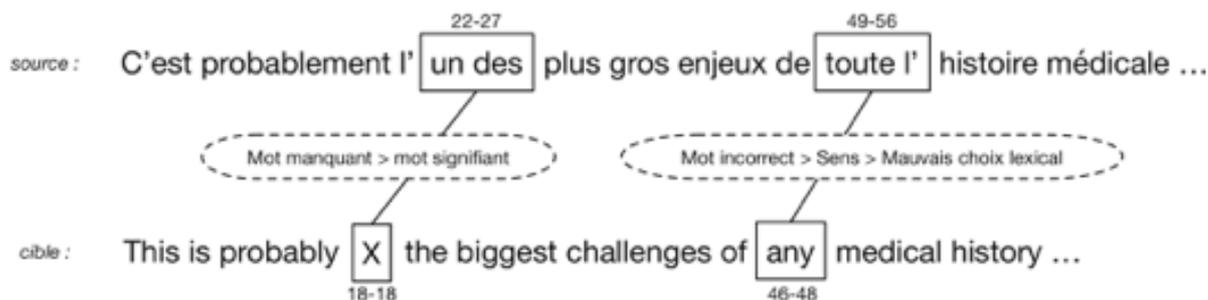


Figure 4. Annotation Representation Example.

SNODEs also permit to spot the position of a missing word in the target, as we can see in Figure 1 with “un des” not translated, thus represented by SNODE [16-18] in the target. They reversely permit to associate a position in the source to extra words appearing in the target.

The other advantage is that it permits to add a posteriori morphosyntactic analysis on the source and target sentences. One or more analyses (Stanford Tagger, Xerox Incremental Parser, etc.) can be attached using SNODE to phrases. At the end of the annotation task, its analyses will be used to search for models of errors (morphosyntactic patterns for example). The idea is to use errors as a pivotal representation in the search mechanism.

## 2.7 Looking for Errors

To study the behaviour of MT systems, morphosyntactic analyses or error types (as illustrated in figure 5) are used to search for error models (morphosyntactic patterns for example) within a project. The errors are therefore used as a pivotal representation in the search mechanism. It will help looking for any consistency in the errors of MT systems.

Numéro	Source	Cible	Annotateurs
1	industrie canada centre d' innovation en technologies de l' information travail et technologie	industry canada centre for innovation in the information technology , and technology .	tilly 3, emmanuelle 0, sophie 4, maitreyi 0, capucinea 0
12	introduction : positionnement et enjeux	the introduction and social issues .	tilly 3, emmanuelle 0, sophie 5, maitreyi 0, capucinea 0
15	le changement technologique comme processus global : un exemple	the technological change as a global process .	tilly 2, emmanuelle 0, sophie 3, maitreyi 0, capucinea 0
18	le contexte de l' organisation	the organization of the	tilly 2, emmanuelle 0, sophie 2, maitreyi 0

Figure 5. Example of an error type search.

## 2.8 Corpora

At the time of writing this article, ACCOLÉ proposes 2 typologies of errors (those of Vilar et al (2006) and DQF-MQM (Lommel and Melby, 2018) as well as 15 French-English corpora (ranging from journalistic news, technical documents, patents, French monolingual texts, excerpts from the Basic Travel Expression Corpus (BTEC), up to climate documents), which led to the creation of 19 projects. These correspond to 6,817 sentences, 134,273 source words, 114,511 target words, for 23,525 annotations made by native English annotators, as shown in table 3.

Projects	Corpora	Error Typologies	Phrase #	Word # (source/target)	Annotation #
citi1 - Vilar	citi1	Vilar	516	13218 / 11857	3265
citi2 - Vilar	citi2	Vilar	1382	24402 / 21320	5569
News 156 LIG-Moses - Vilar	News 156 Moses	Vilar	156	4648 / 3831	982
News 156 Google - Vilar	News 156 Google	Vilar	156	4648 / 3715	616
News 2000 - Vilar	News 2000	Vilar	2000	55368 / 45525	11224
BTEC1 LIG-Moses - DQF-MQM	BTEC1-LIGMoses	DQF-MQM	469	3445 / 3048	0
BTEC1 NMT LIG - DQF-MQM	BTEC1-NMTLIG	DQF-MQM	469	3445 / 3004	0
BTEC1 LIG-Moses - Vilar	BTEC1-LIGMoses	Vilar	469	3445 / 3048	807
BTEC1 NMT LIG - Vilar	BTEC1-NMTLIG	Vilar	469	3445 / 3004	801
Démonstration	citi1	Vilar	516	13218 / 11857	0
Altica test 1	doc1_Fujitsu_180118 (monolingue)	Vilar	41	1104 / 0	2
Altica test 2	doc2_Fujitsu_20180109 (monolingue)	Vilar	26	723 / 0	0
Brad Pitt	WIPOTranslateN	DQF-MQM	9	272 / 305	11
George Clooney	WIPOTranslateS	DQF-MQM	9	272 / 302	25
Jude Law	CANADA_eTranslation	DQF-MQM	28	519 / 768	35
Robert	CANADA_LegacyMT@EC	DQF-MQM	28	519 / 776	51
Hawaii	WIPOTranslateN	Vilar	9	272 / 305	13
Madagascar	WIPOTranslateS	Vilar	9	272 / 302	29
Nouvelle Calédonie	CANADA_eTranslation	Vilar	28	519 / 768	36
New Zealand	CANADA_LegacyMT@EC	Vilar	28	519 / 776	59
<b>TOTAL</b>	<b>13</b>	<b>2</b>	<b>6817</b>	<b>134273/114511</b>	<b>23525</b>

Table 2: Corpora List

These corpora are structured according to SNODEs and are available on request in XML or JSON format.

### 3 Conclusion and further work

ACCOLÉ allows the annotation of translation errors according to built-in error typologies or uploaded ones, on several annotated corpora of different texts, translated by different Statistical or Neural MT systems. ACCOLÉ offers 15 corpora, 19 projects consisting in 248,784 words and 23,525 annotations.

Some annotated corpora available have already been used for a linguistic comparison of the translation quality of different MT systems (Esperança-Rodier and Becker, 2018).

A study is planned to use ACCOLÉ for the bilingual annotation of MWEs.

The error model (morphosyntactic patterns) search in the annotations, essential asset of ACCOLÉ, will allow us in a future version, to use the models of errors in order to propagate, thanks to neural network, the annotations of errors in a semi-automatic way on new corpora. This will allow to build more rapidly annotated corpora and consequently assessing almost in real-time new MT systems. We will then be able to compare the behaviour of MT systems on different domains, thus providing to the community error-annotated bilingual parallel corpora.

We also want to improve it by allowing the association of trees of dependencies.

Further work evaluating inter-annotator agreement is still being conducted at the time of writing this paper.

### Acknowledgements

The work reported above has been granted by LIG/Emergence funding in 2017 and 2019.

### References

- Aziz, Wilker, Sheila Castillo Maria de Sousa, and Lucia Specia. 2012. PET: a Tool for Post-editing and Assessing Machine Translation. In *Proceedings of the 16th Annual Conference of the European Association for Machine Translation*, pages 3982-3987.
- Boitet, Christian and Yusoff Zaharin. 1988. Representation trees and string- tree correspondences. In *Proceedings of international Conference on Computational Linguistics COLING-88*, pages 59-64.
- Castagnoli, Sara, Dragos Ciobanu, Natalie Kübler, Kerstin Kunz, Alexandra Volanschi. 2011. Designing a Learner Translator Corpus for Training Purposes. In *Corpora, Language, Teaching and Resources: From Theory to Practice*. Edited by Kübler N. Bern: Peter Lang.
- Esperança-Rodier, Emmanuelle and Nicolas Becker. 2018. Comparaison de systèmes de traduction automatique, probabiliste et neuronal, par analyse d'erreurs. Proceedings of the 4<sup>th</sup> day on « *Traitement Automatique des Langues et Intelligence Artificielle* » - *TALIA 2018 Day of the plate-forme Intelligence Artificielle (PFIA 2018)*. Nancy, France, 6 juillet 2018. Edited by Didier Schwab et Pierre Zweigenbaum.
- Esperança-Rodier, Emmanuelle, Caroline Rossi, Alexandre Bérard, Laurent Besacier. 2017. Evaluation of NMT and SMT systems: A study on uses and perceptions. In *Proceedings of the 39th Conference Translating and the Computer*, Nov 2017, Londons, United Kingdom. Translating and the Computer 39.
- Felice, Mariano and Lucia Specia. 2012. Linguistic Features for Quality Estimation. In *Proceedings of the 7th Workshop on Statistical Machine Translation*, Montréal, Canada, June 7-8, 2012 Association for Computational Linguistics, pages 96–103
- Germann, Ulrich. 2008. Yawat: Yet Another Word Alignment Tool. In *Proceedings of 46th Annual Meeting of the Association for Computational Linguistics*. June 15-20, 2008. Columbus, Ohio, USA. Demo Papers.
- Kübler, Natalie, Maria Zimina, Serge Fleury. 2016. Origines des erreurs en Traduction Spécialisée : différenciation textométrique grâce aux corpus de textes cibles annotés. In *Actes de la conférence conjointe JEP-TALN-RECITAL 2016*, volume 09: ELTAL, Paris.

- Lommel, Arle, and Alan, K. Melby. 2018. Tutorial: MQM-DQF: A Good Marriage (Translation Quality for the 21st Century). In *Proceedings of 13th Conference of the Association for Machine Translation in the Americas (Volume 2: User Papers)*. Vol. 2.
- Popović, Maja, Hermann Ney, Adrià de Gispert, José B., Mariño, Deepa Gupta, Marcell Federico, Patrik Lambert and Rafael Banchs. 2006 Morpho-syntactic Information for Automatic Error Analysis of Statistical Machine Translation Output. In *Proceedings of the Workshop on Statistical Machine Translation*. 1-6. New York City. June 2006 ©2006 Association for Computational Linguistics
- Popović, Maja. 2018. Error Classification and Analysis for Machine Translation Quality Assessment. *Moorkens J., Castilho S., Gaspari F., Doherty S. (eds) Translation Quality Assessment. Machine Translation: Technologies and Applications, vol 1. Springer, Cham*
- Potet, Marion, Emmanuelle Esperança-Rodier, Laurent Besacier, Hervé Blanchon. 2012. Collection of a large database of French-English SMT output corrections. In *CHAIR, N. C. C., CHOUKRI, K., DECLERCK, T., DOĞAN, M. U., MAEGAARD, B., MARIANI, J., ODIJK, J. et PIPERIDIS, S., éditeurs : Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey. European Language Resources Association (ELRA).
- Specia, Lucia, Gustavo H. Paetzold, and Carolina Scarton. 2015. Multi-level Translation Quality Prediction with QuEst++. In *Proceedings of 53rd Annual Meeting of the Association for computational linguistics and the 7th Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing. System Demonstrations*. Beijing, China. Pages 115-120.
- Steurs, Frieda, Winibert Segers and Hendrik Kockaert. 2015. Translation Expert (TranslationQ & RevisionQ): Automated translation process with real-time feedback & evaluation/ revision with PIE. (Abstract No. Keynote speech 4). *Talking to the World, University of Newcastle, UK*. 09 Sep 2015-11 Sep 2015.
- Stymne, Sara. 2011. Blast: A tool for error analysis of machine translation output. In *the 49th Annual Meeting of the Association for Computational Linguistics. System Demonstrations*. Pages 56–61. Portland, OR, USA,
- Tsoumari, Mara, and Georgios Petasis. 2011. Coreference Annotator A new annotation tool for aligned bilingual corpora. In *Proceedings of Recent Advances in Natural Language Processing*. Pages 43–52. Hissar, Bulgaria. 12-14 September 2011.
- Vilar, David, Jia Xu, Luis Fernando D'Haro *et al.*. 2006. Error analysis of statistical machine translation output. In *the Proceedings of 5th International Conference on Language Resources and Evaluation*. Pages 97-702.
- Wisniewski, Guillaume, Anil Kumar Singh, Natalia Segal and François Yvon. 2013a. Un corpus d'erreurs de traduction. *TALN*
- Wisniewski, Guillaume, Anil Kumar Singh, and François Yvon. 2013b. Quality Estimation for Machine Translation: Some Lessons Learned, *MT Journal*