



HAL
open science

ACCOLÉ: Annotation Collaborative d'erreurs de traduction pour CORpus aLignÉs

Emmanuelle Esperança-Rodier, Francis Brunet-Manquat

► **To cite this version:**

Emmanuelle Esperança-Rodier, Francis Brunet-Manquat. ACCOLÉ: Annotation Collaborative d'erreurs de traduction pour CORpus aLignÉs. GDR LIFT LIFT2019, Nov 2019, Orléans, France. hal-02386413

HAL Id: hal-02386413

<https://hal.science/hal-02386413>

Submitted on 29 Nov 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ACCOLÉ : Annotation Collaborative d'erreurs de traduction pour CORPUS aLignés

Emmanuelle Esperança-Rodier - Francis Brunet-Manquat

Prénom.Nom@univ-grenoble-alpes.fr

Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, 38000 Grenoble, France

Objectifs

- ◆ Aider l'utilisateur dans le choix d'un système de TA
- ◆ Détecter les phénomènes linguistiques incorrectement traités par le système de TA étudié
- ◆ ACCOLÉ : annotation manuelle des erreurs de traduction selon des critères linguistiques

Annotation Collaborative

Annotation ...

Typologies → Vilar^[1]

→ DQF-MQM^[2]

→ autres

... Collaborative

→ Supervision Annot./Admin.

→ Collaboration Annot./Annot.

The screenshot shows the ACCOLÉ annotation interface. It displays a source phrase in French and its target in English. A dropdown menu lists various error types such as 'Ponctuation', 'Mots incorrects', 'Forme non vue', 'Radical', 'Inconnu', 'Forme incorrecte', 'Idiome', 'Mots supplémentaires', 'Style', 'Sens', 'Mauvaise désambiguïsation', 'Mots manquants', 'Mots outils', and 'Mots signifiants'. A discussion panel on the right shows a user's comment and a response.

Recherche

Projet : citi1 - Vilar

Numéro	Source	Cible	Annotateurs
33	la problématique du changement technologique dans les organisations : quelques repères théoriques	the problem of technological change in the organisations : a few benchmarks theoretical	tilly 2, emmanuelle 0, sophia 3, smaltrey 0, capucine 0
54	l'évaluation d'implantation	the assessment of establishment	tilly 1, emmanuelle 0, sophia 0, smaltrey 0, capucine 0
65	l'auteur tient à remercier ses collègues du programme travail et technologies, david tippin, richard lavoie, duncan sanderson, anne - marie - tessier et, plus particulièrement mme judith légaré, pour leur contribution au présent document.	the rapporteur would like to thank our colleagues in the work programme, and technology, david tippin, richard lavoie, duncan sanderson, anne - marie - tessier, and especially to mrs judith légaré, for their contribution to this document.	tilly 3, emmanuelle 0, sophia 9, smaltrey 0, capucine 0
69	la question du changement technologique dans le travail n'est pas nouvelle.	the issue of technological change in the work is not new.	tilly 2, emmanuelle 0, sophia 1, smaltrey 0, capucine 0
70	depuis les débuts de ère indust rielle, le développement de	since the beginning of time indust rielle, the development of	tilly 3, emmanuelle 0

- Recherche par :
- Types d'erreurs
 - Modèles
 - Forme de surface

Structure de données SSTC^[3]

Snode

Erreur comme pivot de recherche

Corpus

PROJETS	Modèles d'erreurs								
	Vilar	Nb phrases	Nb mots (source/cible)	Nb annotations	DQF-MQM	Nb phrases	Nb mots (source/cible)	Nb annotations	
citi1	citi1 - Vilar/Demonstration	516	13218 / 11857	3265					
citi2	citi2 - Vilar	1382	24402 / 21320	5569					
News 156 Moses	News 156 LIG-Moses - Vilar	156	4648 / 3831	982					
News 156 Google	News 156 Google - Vilar	156	4648 / 3715	616					
News 2000	News 2000 - Vilar	2000	55368 / 45525	11224					
BTEC1-LIGMoses	BTEC1 LIG-Moses - Vilar	469	3445 / 3048	807	BTEC1 LIG-Moses - DQF-MQM	469	3445 / 3048	0	
BTEC1-NMTLIG	BTEC1 NMT LIG - Vilar	469	3445 / 3004	801	BTEC1 NMT LIG - DQF-MQM	469	3445 / 3004	0	
WIPOTranslateN	Hawaii	9	272 / 305	13	Brad Pitt	9	272 / 305	11	
WIPOTranslateS	Madagascar	9	272 / 302	29	George Clooney	9	272 / 302	25	
CANADA_eTranslation	Nouvelle Calédonie	28	519 / 768	36	Jude Law	28	519 / 768	35	
CANADA_LegacyMT@EC	New Zealand	28	519 / 776	59	Robert	28	519 / 776	51	
doc1_Fujitsu_180118 (monolingue)	Altica test 1	41	1104 / 0	2					
doc2_Fujitsu_20180109 (monolingue)	Altica test 2	26	723 / 0	0					
Total		13	5289	112583 / 94451	23403	6	1012	8472/8203	122
						19	6301	121055/102654	23525

Références

- [1]Vilar, D., Xu, J., D'Haro, L.F. and al. (2006) Error analysis of statistical machine translation output. *5th International Conference on Language Resources and Evaluation*. 97-702.
 [2]Lommel, A., and Alan, K. M. (2018). Tutorial: MQM-DQF: A Good Marriage (Translation Quality for the 21st Century). *13th Conference of the Association for Machine Translation in the Americas (Volume 2: User Papers)*. Vol. 2.
 [3]Boitet, C. and Zaharin, Y. (1988). Representation trees and string- tree correspondences. In *Proceedings of international Conference on Computational Linguistics COLING-88*, 59-64.

Remerciements

ACCOLÉ et les corpus ont été réalisés grâce aux projets LIG/Émergence 2017 et 2019.

Liens

<http://lig-accole.imag.fr>



GETALP

Traitement de Données et de Connaissances à Grande Echelle

