

**Learner Translator Corpora  
Related Research (March 14, 2013)**

№	Corpus Name and Time of Compilation	Author (coordinator) and Affiliation	Online Availability / Size	Genres	Languages Directionality	Linguistic Mark-up / Query Interface	Error Typology and Annotation	Metadata
1.	<b>Corpus of Student L1–L2 Translations</b> (1988-1992)	Robert Spence, University of Leipzig, <b>Germany</b>	unavailable 280 thousand tokens, 49 sources, 1232 targets	popular science, economics, politics, tourism, literature	German (L1) to English (L2)	No /	no	
2.	<b>Student Translation Archive</b> (2002-2003)	Lynne Bowker Peter Bennison University of Ottawa, <b>Canada</b>	unavailable	Legal, medical, etc	French/Spanish (L2) to English (L1)	POS, semantic / конкорданс, частотность by Translation Tracking System	Not reported	Extensive data on students, conditions of translation
3.	<b>A learner corpus</b> in the framework of PELCRA project, (2001)	Raf Uzar Jacek Waliński University of Łódź, <b>Poland</b>	Downloadable <a href="http://pelcra.pl/resources/parallel/pelcra_par_2.tgz">http://pelcra.pl/resources/parallel/pelcra_par_2.tgz</a> . 257 sources Polish 182 530 English 177 890	academic	Polish (L1) into English (L2)	POS, semantic / конкорданс, частотность manual alignment	yes	No?
4.	<b>ENTRAD</b> University of Zaragoza, Aragon, Spain (2005-2008)	Celia Serrano Florén and Rosa Lorés	<a href="http://ice.unizar.es/entrad/">http://ice.unizar.es/entrad/</a>		Spanish (L2) to English (L1) mixed	No / text-level alignment	no Errors are highlighted (using a colour code) and marked graphically	
5.	Russian Translation Learner Corpus ( <b>RuTLC</b> ) (2004-2008)	Ekaterina Sosnina Ulyanovsk Technical University, <b>Russia</b>	unavailable	technical	English (L2) to Russian (L1)	no/ self made	a typology of 40 items	no Not translator trainees!
6.	<b>MeLLANGE LTC</b> (2006)	Sara Castagnoli et al Partnership of <b>European</b> Universities	<a href="http://corpus.leeds.ac.uk/mellange/ltc.html">http://corpus.leeds.ac.uk/mellange/ltc.html</a> (ca. 350 tokens each) 429 targets	legal, technical, administrative and journalistic	Multilingual Catalan, English, French, German, Italian and Spanish	tokenisation, POS tagging and lemmatisation/ self-made Mini-context level alignment	Hierarchical typology of 30 types in Content and Language categories; a customised version of the manual annotation tool MMAX2, originally developed by Mark-Christoph Müller at the European Media Laboratory, Heidelberg (2006) 232 targets	Extensive searchable Reference translation translation kits

							annotated	
7.	Multiple Italian Student Translation Corpus ( <b>MISTIC</b> ) (2009)	Sara Castagnoli, University of Bologna, Forli, <b>Italy</b>	unavailable 59 sources, 480 targets 176,000 target language words	specialised	English/French to Italian	POS-tagging / aligned – to both the relevant ST and all concurrent TTs – using the IMS Corpus Workbench (Christ 1994)	no	107 Italian-native post-graduate students of specialised translation
8.	<b>UPF - LTC</b> (2006 – present)	Anna Espunya Pompeu Fabra University in Barcelona, <b>Spain</b>	unavailable 10 sources, 194 targets English - 8,448 tokens Catalan - 200,187	non-fiction (informative/instructive and essay), fiction (narrative, drama and screenplay)	English to Catalan	the English component (lemma, word and POS), on the Catalan component (lemma, word, POS, morphosyntactic information, syntactic function and error type) / IMS Open Corpus Workbench tools	25 simple categories with no subdivisions	first- and third-year students doing the Translation and Interpreting degree between 2006 and 2011
9.	Korpus studentischer Übersetzungen ( <b>KOPTE</b> ) since 2009	Andrea Wurm Saarland University <b>Germany</b>	unavailable 77+ sources 971+ targets corpus site <a href="http://fr46.uni-saarland.de/index.php?id=3702">http://fr46.uni-saarland.de/index.php?id=3702</a>	mainly newspaper articles	French(L2) to German (L1)	Automatic tokenization, lemmatization and POS-tagging TreeTagger (Schmid 1994) the STTS tagset for German, Achim Steins tagset for French (Stein 2003) / <b>No alignment</b>	Manual annotation of the <b>teacher's evaluation</b> with UAM Corpus Tool (O'Donnell 2008)	
10.	Norwegian-English Student Translation corpus ( <b>NEST</b> ) since 2008	Anne-Line Graedler, Hedmark University College, <b>Norway</b>	unavailable 18 sources (200 to 900 words each), around 214 translations; approximately 121,000 tokens in total	factual prose for a general readership; instructional texts; information pamphlets; fictional prose; a few have been <b>slightly altered</b> to incorporate specific translation challenges, or to avoid particular problems	Norwegian (L1) to English (L2)	not yet in place	In progress	Extensive meta-data Not translator trainees! more than 100 students
11.	<b>RusLTC</b>	Andrey Kutuzov National Research University Higher School of Economics	<a href="http://www.rus-ltc.org/">http://www.rus-ltc.org/</a> 260 sources+	1. Academic 2. Informational 3. Tech	English to Russian Russian to English	Tokenization, POS/self-made JAVA-based (planning to migrate to a Python-	Hierarchy of 26 types in Content/Language colour-coded	<ul style="list-style-type: none"> <li>• Trainee's gender</li> <li>• Trainee's experience (year of study or contest)</li> </ul>

	(Moscow, <b>Russia</b> ) Maria Kunilovskaya Tyumen State University (Tyumen, Russia)	1795 targets + ca. 1 200 000 tokens in total (almost equally split btw EN and RU)	4. Fiction 5. Educational 6. Encyclopaedia 7. Speech 8. Letters 9. Advertisement 10. Review		based interface that would incorporate all types of queries)	categories, inc. good decisions Implemented with customized brat text annotation program <a href="http://brat.nlplab.org/index.html">http://brat.nlplab.org/index.html</a>	translation) • Grade for the translation • Conditions of production (routine/exam; home/classroom) • Year of production • Trainee's affiliation
--	--------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------	--	--------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

## References

1. Spence, R. 1998. "A Corpus of Student L1-L2 Translations". In Proceedings of the International Symposium on Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching, S. Granger and J. Hung (eds), 110-112. Hong-Kong: The Chinese University of Hong Kong.
2. Bowker L. & P. Bennison (2003) Student translation archive: design, development and application. In Zanettin F., S. Bernardini & D. Stewart (eds.) Corpora in Translator Education. Manchester: St Jerome, 103-117. Also referred to in Bowker, L., & Bennison, P. (2002). Translation Tracking System: A tool for managing translation archives. In Proceedings of LREC 2002 (pp. 503-507). <http://gandalf.aksis.uib.no/lrec2002/pdf/115.pdf>. Accessed March 12, 2014.
3. Uzar R. & J. Waliński (2001) Analysing the fluency of translators. International journal of corpus linguistics, 6, 155-166. <http://pelcra.pl/>
4. Florén C. (2006) ENTRAD, an English Spanish parallel corpus created for the teaching of translation. Paper presented at the 7th Teaching and Language Corpora conference (TALC 2006), Paris, 1-4 July 2006.
5. Sosnina E. Russian Translation Learner Corpus: The First Insights. The proceedings of the 6 international scientific conference «Interactive systems: problems of human-computer interaction», Ulyanovsk: UISTU, 2005. Sosnina E.P. (2006) Development and application of Russian Translation Learner Corpus. Presented at Corpus Linguistics – 2006, St. Petersburg, 10-14 October 2006.
6. Castagnoli, Sara, Dragos Ciobanu, Kerstin Kunz, Alexandra Volanschi, Natalie Kübler (2006). Designing a Learner Translator Corpus for Training Purposes
7. Castagnoli, S. (2009). Regularities and variations in learner translations: a corpus-based study of conjunctive explicitation. PhD Dissertation, University of Pisa. Also described in in Castagnoli, Sara (2011) Exploring Variation And Regularities In Translation With Multiple Translation Corpora // Rassegna Italiana di Linguistica Applicata 1-2/2011.
8. Espunya, A. Investigating lexical difficulties of learners in the error-annotated UPF learner translation corpus // Twenty Years of Learner Corpus Research. Looking Back, Moving Ahead / edited by Sylviane Granger, Gaëtanelle Gilquin, Fanny Meunier. P.129-136. Espunya, A (2013) The UPF learner translation corpus as a resource for translator training // Language Resources and Evaluation <http://link.springer.com/article/10.1007/s10579-013-9260-1> Accessed March 12, 2014.
9. Wurm, Andrea. Proper names and specific cultural items in a corpus of student translations (KOPTE) // trans-kom 6 [2] (2013): 381-419 <http://fr46.uni-saarland.de/index.php?id=3702> Accessed March 10, 2014.
10. Graedler, Anne-Line (May 2010). NEST – a corpus in the brooding box <http://www.helsinki.fi/varieng/series/volumes/13/graedler/> Accessed March 10, 2014.
11. Kutuzov, A. Kunilovskaya M. Russian Learner Translator Corpus: Design, Functionality and Perspectives (Introducing a Resource for Studying and Teaching Translation) (in submission)