Description of RusLTC translation error typology

Maria Kunilovskaya
mkunilovskaya@gmail.com
Tyumen State University, Russian Federation

RusLTC has an associated on-line error-tagging facility, which is in everyday use as a teaching tool.

It was designed

1) to streamline and facilitate the assessment process,
2) probably reduce its subjectivity and, more importantly,
3) enable us to store the results of error correction in a machine-readable format. Last, but not least we aspired
4) to introduce an online blended learning system that would promote learner-centered approach and increase students' autonomy.

To this end we have developed a typology of translation errors based on the analysis of the students' translations and drawing on the experience accumulated in error-based translation quality assessment (TQA) both at home and abroad. On the one hand it is intended to reflect the practices used at our universities, but on the other hand it departs from the traditional approach in **emphasizing textual and pragmatic issues in translation**. And this transgression is in step with the corpus linguistic influence over translation studies (TS). It seems that introduction of corpus methods, that are essentially empirical, to TS has brought the researchers back to text-oriented approach in TQA advocated by Juliane House (1997), and characterized by more attention to beyond-sentence-level macrotexual features of translations. Many of the translation error classifications including the one developed by Dr. Hurtado (Hurtado 1995 via Waddington 2001) feature pragmatic errors as a separate type (Hansen 2009; Lee-Jahnke: 2001 (via Secara 2005); Williams 2001/2009).

Though we are opposed to the term 'translation error' as too negative and categorical, and would prefer to speak of more or less appropriate solutions/renditions, we use this term for the sake of brevity and tradition.

Skipping the overview of the many approaches to error classification, described at length by better scholars (House 1997; Secara 2005; Williams 2001/2009), but trying to describe our methodological stand, we would define it as target text-centered (i.e. it is the TT which is the object of evaluation) and **the major criterion of evaluation is ST-TT relations**. Though we offer a way to rank students' results we are not so much into measuring results as into helping students develop individual aptitudes (we pursue the goals of **formative rather than summative** assessment). It means that we mostly seek to describe a mistake in terms of **the harm it does to the faithfulness of translation and its textual and linguistic quality**.

So, the basic principle behind this classification is the type of negative effect to the TT quality caused by translators' renderings. They can either be inappropriate in terms of content transfer or in terms of target language expression.

Though microtextual imperfections in translations which deal with target text production (or expression stage of the translation process and have to do with the TT-knowledge mostly) are easier to spot and can be annotated without ST-TT comparison, by a non-translator reviser, we agree that **more attention should be given to translator solutions that prevent the reader from extracting the message intended by the ST author** (i.e. content transfer errors).

In devising our content error classification (i.e. trying to find a reasonable and workable typology of text meanings) we relied on **the three semiotic dimensions of a linguistic sign** proposed by Charles Peirce and Charles Morris: semantics, syntactics (syntax) and pragmatics, which we have adopted for the purposes of translator-teaching-oriented error analysis and interpreted as the following error tags:

    content_reference
        omission
        distortion
        nonsense
        inexact
        unclear

```
content_cohesion
        theme-rheme
        logic
content_pragmatics
        tenor
        field
```

These error types also correlate with the known text linguistics notions of the situational context and texture (Halliday, Hasan 1985) and are assigned in all cases when the TT deviates from the original in denotative/experiential meaning (reference), in textual and interpersonal aspects. The three categories are interconnected - we have noticed that such deviations are rarely covert – in most cases inability to render, say, experiential meaning creates "ripples" in the text structure and affects cohesion or pragmatics of the TT, preventing its effective understanding (to be tested, though). But in annotating errors we tried to define the type of ST content that is immediately affected.

Inadequacies that have to do with the text semantics (who-does-what-to-whom) are varied as to the degree of disinformation on the part of the reader. The latter can be deprived of the relevant information about ST content because it

is simply not in the TT (omission),

is substituted for something completely or partially different (distortion and inexact),

is expressed so badly that can hardly be derived from TT (unclear) or cannot be understood at all even given the context (nonsense).

*For example, while translating the text by Adam Hartung on the decreasing role of printed media and TV in creating and promoting brands, 7 out of 10 students made a semantic mistake (**wrong reference, inexact**) in the sentence below, because they failed to interpret the verb "to see" in this context as "to experience something" (Macmillan Dictionary). Instead they chose more frequent lexico-semantic variants denoting mental or sense perception, which led to lack of TT coherence and left the target reader wondering as to the message of this phrase.*

(1)	Source	Target	Back translation
	And investors in these companies either **saw** their values soar, or practically disintegrate.	Инвесторы тоже понимали, в каком случае их прибыль возрастет, а в каком значительно сократится.	Investors also **understood** in which case their profits will rise and in which they will drop.
		Инвесторы видели, что ценность компаний взлетала или же, наоборот, падала.	Investors **have seen that** the companies' value went up or, vice versa, plummeted.

*The next example is from the text "Diamonds are forever" which tells about the marketing genius of the De Beers company which manages to keep prices for diamonds artificially high due to its exclusive dominance in the world market, particularly through heavy advertising. It is the case of obvious **distortion** where the semantics of the emphasized part runs counter the original meaning and contradicts the message of the whole text, which should have alarmed the translator.*

(2)	Source	Target	Back translation
	Invented by one of the richest companies in the world, Diamonds are Forever is a slogan **which does not bear close examination.**	Слоган "Бриллианты бессмертны", использованный одной из самых успешных компаний в мире, говорит сам за себя.	The slogan "Diamonds are immortal" used by one of the successful companies in the world, **is self-explanatory**.

Content mistakes of the second type are those which deal with **TT cohesion and coherence** and transfer or modification of cohesive ties in translation. Disregard for the necessary shifts in cohesion

dictated by TT genre conventions and interlingual differences in grammatical structures (Blum-Kulka 1986; Hatim 1997; Kachroo 1984) lead to lack of texture in TTs. To be readable and acceptable (in terms of TQA) TT should "hang together", have reasonable thematic and information structure, be based on the semantic and pragmatic interaction between sentences. Yet, students find it difficult to provide acceptable linkage between sentences in the TT. In case of English-Russian translation this problem arises due to the interlingual differences – Russian unlike English has less formal, information-structure-driven syntax, there are differences in the use of connectives, saying nothing about more subtle differences in text structure norms. Typical examples here would be sentence-by-sentence translation process when the translator does not account for changes made in the previous sentence while translating the next one.

For example, in the Russian translation of the first sentence the indefinite adverb "somehow" which means "in some way, nobody knows how" disappeared, but the connective "in fact" with its contrastive meaning of "introducing something opposite to what has been asserted" was retained, though there is nothing in the previous text that could support this opposition.

(3)	Source	Target
	When General Electric discovered a way to produce high grade synthetic diamonds, De Beers still managed **somehow** to prevent GE undercutting their prices. **In fact,** the Oppenheimers have spent more than $160 million a year repeating their message that "diamonds are forever", probably the most successful advertising slogan of all time.	Когда компания Дженерал Электрик изобрела способ производства синтетических бриллиантов высокого качества, De Beers сумели предотвратить снижение стоимости своих бриллиантов. В действительности, Оппенгеймеры потратили более 160 миллионов долларов на раскрутку своего слогана, который возможно является самым успешным за всю историю маркетинга.

Or else students ignore the existing extralinguistic presuppositions and contradict the logic of the world as in the example 4 below. *The student failed to solve the problem with "outlawed", substituted it for some other meaning, after which resumed calquing the syntax of the ST. In the sentence immediately after, the translator erroneously interprets "and" as simple additive conjunction while in fact there are causative-consecutive relations between the two clauses.*

(4)	Source	Target	Back translation
	Appalled by De Beers' business ethics, **America outlawed the company, effectively preventing it from opening its own outlets in the United States**.	Американское правительство было потрясено деловой этикой компании De Beers и **обвинило ее в нарушении закона, тем самым не позволив ей** открывать филиалы в Соединенных Штатах.	The American government was shocked by the business ethics of De Beers and **accused it of breaking the law, and by doing it forbade opening branches in the USA**.
(5)	Ironically, America remains by far De Beers' single biggest market **and** the company operates through American dealers unhindered.	Как ни странно, именно в Америке находится самый крупный рынок "Дэ Бирс", **кроме того** компания беспрепятственно осуществляет свою деятельность через американских поставщиков.	However strange, it is America which is the largest De Beers market, **besides** the company freely operates through its American dealers.

Pragmatic mistakes (Nord 1997) have to do with the functional aspect of the translation project which to some extent is described in the translation brief. It is an umbrella category for **the inadequacies as far as the text (intended) function, lack of message (unclear what it boils down to), inappropriate rendering of cultural reference, disregard to the intended audience, lack of ST rhetoric effect, missing connotations and the ST author's attitudes, inadequate rendering of situational aspects of texts (time, place, medium).**

3

In our experience most of these mistakes occur because translators fail to understand the ST message in the first place; they forget that any text is set in a particular time-space frame, and are unable to produce texts in the TT (though L1) which would respect the intended register (field, tenor, mode). As a result the TT fails to "strategically direct its intended audience to an interpretation of it - generally one which in certain respects matches the interpretation given to the source text" (Aston, 1999).

The clause in bold in the context below (6) has an implication which appears to be totally lost in translations, though the semantics is intact. Similarly, a semantically equivalent translation of the popular wisdom "It's a good horse that never stumbles" in Russian totally destroys the implication – the need to accept the imperfection of the world, everyone makes mistakes, a consolation of kinds.

(6)	Source	Target	Back translation
	And even when profits are down and their share price takes a tumble, **De Beers makes sure that the legend of the diamond lives on**.	И даже когда прибыль Де Бирс падает и цена их акций стремительно снижается, компания убеждена: **легенда о бриллиантах живет.**	… De Beers is sure: the legend about diamonds lives on.
		И даже во времена меньших доходов и боле низких цен на акции De Beers **продолжают свою легендарную "алмазную" историю.**	De Beers continues its legendary "diamond" story.

The pragmatic mistake which impaired the search for background information and its use in the TT in the next example (7) consists in disregard for the publication date of the original – December 2012. In the text mentioned above, Adam Hartung refers to the bankruptcy of an American legendary snacks producer Hostess Brands, which featured heavily on the news back then. The students googled the company and found out that it rose from ashes and was modernized in late 2013, but the author could not have referred to it in 2012. And the Russian reader does require some hint as to how to interpret this reference. 6 out of 10 translations leave the reader wondering what this sentence means in the context.

(7)	Their (*about huge companies like P-and-G, Kraft, GM and Target – added by the authors*) size, hierarchy and arcane business practices will lead to huge problems. **Imagine a raft of new Hostess Brands experiences.**	И таким компаниям, как Hostess Brands снова придется принимать новые бизнес-решения.	And such companies as Hostess Brands will have to find new business solutions again.
		Представьте себе множество новых продуктов компании Хостесс (Hostess).	Imagine many new products from Hostess Brands.
		Только представьте сеть новых Hostess Brands.	Just imagine a network of new Hostess Brands outlets.
		Представьте появление ряда новых компаний напоминающих "Хостесс Брэндс".	Imagine the rise of a number of new companies resembling Hostess Brands.

Like it or not, but we refer **register mistakes** to the same category of pragmatic mistakes. When students use vocabulary or structures that are inconsistent with the field/subject of the text, or lead to unmotivated change in the tenor, or are inappropriate in the mode specific for the target we annotate them as pragmatic errors in the content error category. *For example, in the text below a translator uses a phrase that can be back-translated as 'maintenance costs' which in Russian is used to talk about machinery, premises, production lines, but not about administrative costs/expenditure.*

(8)	Source	Target

The improvements in efficiency that these projects will bring will also mean real savings in **operating costs**.	Улучшения в продуктивности, которые принесут эти проекты, также будут означать существенную экономию в **текущих эксплуатационных расходах**.

Language errors (or errors of expression) are classified according to the traditional description of levels in the language hierarchy and inherit the tradition in foreign language teaching practices.

language_lexical
    choice-of-word
    combinability
language_morphology
    wrong_wordform
language_syntax
    incomplete_structure
    ungrammatical
    word_order
    preposition
language_spelling
    capitals
    typo
language_punctuation

The classification is a **tree-like hierarchy** which allows to use of the upper-level catch-all categories in case of doubt and to avoid introducing user-defined classes. Well, of course, in applying any classification like this there will be borderline cases and overlaps (this issue is described at length in the literature on error-tagging in learner corpora – see Meurers 2011; Dagneaux, Denness, Granger 1998), but we tried to stick to **a purely descriptive approach** to avoid hybrid and overlapping categories (the same approach is adopted in Mellange – see (Kübler 2008).

As it can be seen, the classification shows neither the reasons behind mistakes (flaws in the translation competence), nor the gravity of them, because these seem to be rather interpretative and most subjective elements in TQA. But at the same time we felt that these characteristics of translation mistakes are important for a well-rounded description and can be useful for didactic purposes. Therefore, we created **extra tagging facilities** which enable the annotator to add two attributes to each mistake tagged within the major classification: **1) those describing the gravity of mistakes and 2) those which reflect the annotator's speculations as to the causes of the mistakes**.

The latter describes the low level of a specific component of the translation competence or the breach of an established translation norm (this tag set is arbitrary and optional and can be compensated in the annotator's commentaries to each mistake):

• background_info (low extralinguistic competence)
• SL (flaws in linguistic competence, associated with comprehension)
• TL (poor command of the TL as regards productive skills)
• too_literal (insufficient level of transfer competence/strategic competence/knowledge of translation – inability to detect a problem)
• too_free (inability to find appropriate ways to solve problems)
• proper_name (lack of understanding of respective translation norms; it is included due to the number of such mistakes)
• inconsistency (self-explanatory – inability to stick to the same strategy throughout the text).
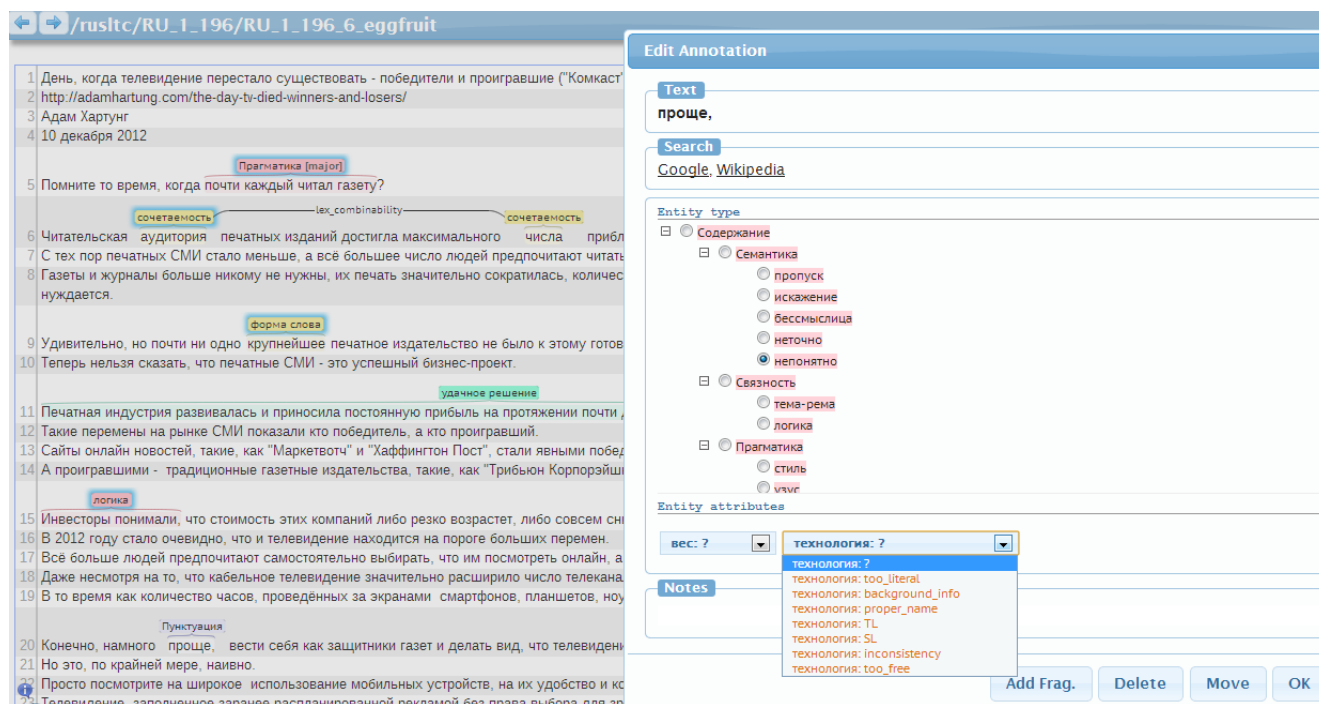
The Weight attribute, which describes the gravity of mistakes, is a three-member scale which differentiates **critical, major and minor** errors. Both content and language errors can be marked as critical; the former in case of significant harm to the ST-TT relations, when the key information of any type appears lost or misunderstood; the latter are critical if they are obviously binary (Pym 1992), violate a basic rule of the TL and immediately spring to the reader's eye without being typos, which are easily corrected by a spell-checker. Mistakes marked up as minor (or not marked up at all) are rather recommendations or non-binary mistakes.

Beside mistakes proper we have three more tags that can be assigned to a highlighted text span: **Delete, Good Solution and Note**. It is important that we highlight adequate solutions that are exceptionally creative, or just good in cases when most students failed to overcome the problem. It is a part of positive evaluation that we practice.

All the error tags and attributes are documented and exemplified in **the error tagging manual**, which was revised after a series of interrater reliability experiments (Kunilovskaya 2015) and discussions.

Technically we use the adopted installation of the open-source text annotation programme called *brat* (Stenetorp et all 2012), which runs at our server. This is the teacher's interface of this tag-editor (Fig.3).

Figure 3. BRAT-based online error tag editor

To reduce the known 'source effect' which contributes to the subjectivity of assessment before correction **all translations are anonymized**, but we keep a support file which helps to mail students links to their individual translations.

For each marked up text-file the programme creates a plain txt which contains machine-readable error tags. These files are used to generate individual and group statistics, including those that are used to rank the translations on the basis of their quality.

References

1. Aston G. Corpus use and learning to translate. In Textus 12, 1999, 289-314.
2. Blum-Kulka, Sh. Shifts of cohesion and coherence in translation. In Interlingual and Intercultural Communication: Discourse and Cognition in Translation and Second Language Acquisition Studies, J. House and S. Blum-Kulka (eds), 1986/2000, 17–35.
3. Dagneaux E., Denness Sh., Granger S. Computer-aided error analysis. In System 26, 1998, 163-174.
4. Halliday M.A.K., Hasan R. Language, context, and text: aspects of language in a social - semiotic perspective. Oxford: Oxford University Press, 1985.
5. Hansen G. A Classification of errors and revision. In CIUTI-Forum 2008: Enhancing Translation Quality : Ways, Means, Methods edited by Martin Forstner, Hannelore Lee-Jahnke, Peter A. Schmitt, 2009.
6. Hatim B. Communication across cultures. Translation Theory and Contrustive Text Linguistics. Exeter University Press, 1997.

7. House J. Translation quality assessment: A model revisited. Gunter Narr Verlag, 1997.
8. Kachroo, B. Textual Cohesion and Translation. In Meta: Translators' Journal, vol. 29, No. 2, 1984, 128-134.
9. Kübler N. A comparable Learner Translator Corpus: creation and use. In Proceedings of LREC Workshop on Comparable Corpora, 2008.
10. Kunilovskaya, M. (2015). How far do we agree on the quality of translation? In English Studies, 1(1), 18-31.
11. Meurers D. On Annotating Learner Corpora: Why? Which annotations? How? Symposium "What's Hard in German? Structural Difficulties, Research Approaches and Pedagogic Solutions", Bangor University, July 18/19 2011.
12. Nord Ch. Translating as a purposeful activity. Functionalist approaches explained. Manchester: St. Jerome, 1997.
13. Nord Ch. Translating as a purposeful activity. Functionalist approaches explained. Manchester: St. Jerome, 1997.
14. Secară A. Translation evaluation: A State of the Art Survey. In Proceedings of the eCoLoRe/MeLLANGE Workshop. 21-23 March 2005, Leeds, UK, 39-44.
15. Stenetorp P., Pyysalo S., Topić G., Ohta T., Ananiadou S. and J. Tsujii (2012). BRAT: a Web-based Tool for NLP-Assisted Text Annotation. In Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics. P.102–107.
16. Waddington C. Different Methods of Evaluating Student Translations: the Question of Validity. In Meta, 46 (2), 2001, 312-325.
17. Williams M. The application of argumentation theory to translation quality assessment. Meta 46 (2), 2001, 326-44.
18. Williams, M. Translation Quality Assessment. - Mutatis Mutandis. Vol 2, No 1, 2009, 3 – 23.