Formative Evaluation in TS  
Towards an Empirical Assessment Model in the Process and Product of Translation

Paper presented at TNR 2012, LU Leuven (Lessius University College)  
Antwerp 13-14 December 2012

Evangelia Kazantinou, Ionian University <ev_kazan@otenet.gr>  
Ioannis E. Saridakis, Ionian University, University of Athens <iesaridakis@gmail.com>

1. Introduction

One of the issues deemed to be critical in translation ethics is intervention (Pym 2009; 2012: 87-108). In translation practice, intervention suggests modifying aspects of a ST textuality, as a motivated and non-binary option in the translation practice, running counter to traditional “axioms” of our discipline, such as ST-TT equivalence and ST fidelity.

It is a challenge for Translation Studies to focus on the ethical aspects of translation practice, with the aim to elaborate on nuances of translation didactics, showing a particular interest in the evaluation models as critical means of translation intervention, both on the basis of translational competence formation and translation product assessment. Pym (ibid.) posits that intervention actually refers to things that are not textually available, i.e. intentions, and that in this sense, interventionism is not a sound basis for quantitative descriptive textual research. We regard translators as people who have to think critically in the communicative act in which they actively intervene as communicators. We argue that it is the means and the methods towards selecting the appropriate elements that can be actually moderated and taught, at least for translating certain text genres. Thus, a quantitative approach can supplement a descriptive hermeneutic model of translation analysis, at least in studying translation processes and products that are not markedly and a priori culturally-bound.

2. Intervention as Translation Assessment

Our paper focuses on formative assessment, which Hatim & Mason (1997: 199–200) define as a continuous process of interaction between translation teacher and trainee, with the aim to outline the validity of translation options (and therefore of translation errors, in whatever sense these are perceived), as a continuum between the two extremes, which are “acceptable” and “erroneous” in the translation practice. In the latter sense, this corresponds to what Pym (1992: 284) defines as non-binarism, i.e. when there are at least two right answers to which the wrong ones are opposed.

Federica Scarpa defines this type of assessment as having a pragmatic nature and posits that it is part of quality assurance, and therefore focuses not only on the translation product, but also on the translation process (Scarpa 2010: 254–255).

“Contrairement aux autres types de révision déjà présentés, où la traduction est considérée comme un produit, l’assurance de la qualité est plutôt un procédé. Il s’agit en effet d’une série d’activités systématiques et planifiées comprenant le contrôle et l’évaluation de la qualité (révision), appliquées avant, durant et après le processus de traduction [...] Dans ce cas, la qualité à évaluer n’est plus seulement celle du produit de la traduction, mais également la qualité du service de traduction offert. Elle vise tant les échéances, les interactions entre le client et le traducteur et la rétroaction sur le déroulement du travail que le choix des traducteurs pour un mandat donné (Scarpa 2010: 254–255, emphasis added).

To summarise, formative assessment and its teaching in the translation classroom is seen as a two-directional and modular activity, complementing the traditional summative assessment and is being implemented, on a preliminary basis, in our translation classes.
3. Empirical Findings

3.1. Corpus Description
To test and evaluate our conceptual approach, we compiled a bilingual, mono-source, mono-directional, parallel translation corpus (Laviosa 2002: 33-38), comprising 24,205 words of students' translations (6th semester students at DFLTI, Ionian University, final semester examination), assessing the students' performance in the framework of the Technical translation from English into Greek course. We have included two English STs of 760 words in total and 56 Greek translations.

It can also be described as a sample corpus, since the two STs were abridged to meet the time and word limit restrictions of the examination, synchronic as the two texts were produced a decade apart (1997 and 2006) and terminological, since it includes texts categorised in the same specialised subject field.

The two source texts are scientific texts. The first source text (ST_A) is an extract from a technical report for the Institute for Artificial Intelligence (IDSIA) in Switzerland. It is entitled A Formal Measure of Machine Intelligence and was published in April 2006. From this report the students were asked to translate the Abstract and part of the Introduction. In Hallidayan terms (Halliday 1978), the field metafunction of ST_A can be defined as a presentation of the fact that our understanding of Artificial Intelligence is based on our perception of human intelligence. As regards tenor and mode, the writers are expert scientists who address an expert audience. In this sense, their aim is mainly to inform on the results of this survey in a formal and personally unbiased manner. The language is standard, contemporary, British English. The medium is simple (written to be read). However, the participation required is apparently complex as the writers invite their audience to reflect on their own experiences regarding intelligence.

The second source text (ST_B) is also an extract from a 1997 article entitled Does Machine Learning Really Work. From this article the students translated the part entitled The Niche for Machine Learning. Concerning field, the text sample used in our corpus discusses the areas in which machine learning is used with the final aim to convince the audience that there is a market for machine learning applications. Regarding mode and tenor, the article addresses a specialised audience, it is written by a specialist in Artificial Intelligence (a Professor in computer science and robotics), in standard, contemporary American English with the aim to inform them in a formal style. Finally, the text, is a simple, written to be read, document, not involving any direct participation from the addressees.

3.2. Methodology of Analysis
Our analysis of the parallel corpus is based on two axes of approach:
(a) On the first axis, text chunks are chosen so as to be able to present and grade instances of lexico-semantic and morphosyntactic variation from a TT rendition that has been previously documented to be valid. Such chunks are considered as error-prone and possibly require the translator's (or the reviewer's) intervention.
(b) On the second axis, our formative approach seeks to generalise, from the typology of errors and variants discussed above, on a more descriptive and hermeneutic model. To this end, we resort to our custom classification of erroneous rendition strata, based on an SFL-derived approach (Halliday 1978: 62 ff). The model draws on the guidelines of Institute of Linguists' Diploma in Translation (2006), and can be depicted as follows:

1 Istituto Dalle Molle di Studi sull'Intelligenza Artificiale, see: <http://www.idsia.ch>.
### Table 1. Classification of Erroneous Translation Renditions (level 2) (Adapted from Saridakis 2011).

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEC</td>
<td>Serious deficit in decoding the sentential or textual meaning of the ST. It is often educed that the deficit is attributable to erroneous decoding of the morphosyntactic structure of the ST in the discourse segment (chunk) examined. When this applies, the chunk is annotated as &lt;GR+DEC&gt;. Correspondingly, when the deficit is considered or educed to be attributable to erroneous decoding of the signified of a ST lexeme, it is annotated as &lt;TERM+DEC&gt;. In the latter case, there is a borderline and often difficult distinction from instances marked as &lt;TERM+ENC&gt;. However, the didactic, and hence formative approach is quite different, given that the deficit arises at a different stage of the translation process, and requires clarification.</td>
</tr>
<tr>
<td>ENC</td>
<td>Serious deficit in the utterance of the sentential or textual meaning in the chunk examined, pinpointed on the level of reformulation in the TL. The shift is often revealed on the semantic and morphosyntactic levels, and cannot be attributed to deficient decoding of the text segment in the ST (DEC). Essentially, this category is a superset of the &lt;REG&gt; category, including also the metafunction of field, i.e. the ideational level of the texteme, in Hallidayan terms of discourse semantics. This category can be combined causally or cumulatively with the &lt;TERM&gt; and/or &lt;GR&gt; categories.</td>
</tr>
<tr>
<td>GR</td>
<td>Syntactic structures of the ST or TT, having a minor, but not negligible, impact on the translator's textual performance.</td>
</tr>
<tr>
<td>GR+DEC</td>
<td>See &lt;DEC&gt;</td>
</tr>
<tr>
<td>GR+ENC</td>
<td>See &lt;ENC&gt;</td>
</tr>
<tr>
<td>GR+REG</td>
<td>See &lt;REG&gt;</td>
</tr>
<tr>
<td>REG</td>
<td>Incompatibility of discourse register between ST and TT, particularly in terms of tenor (Halliday 1978: 62), i.e. on the level of the interpersonal and textual functions. In short, this category corresponds to an utterance of translation discourse equalling the expectancy of the assumed primary readership (cf. Pym 1992). Depending on the assumed cause (or the significance) of the incompatibility, this category can be combined with the &lt;GR&gt; and &lt;TERM&gt; categories. It is further combined with category &lt;ENC&gt;, to denote the unsuccessful balance, on the level of the TL utterance, between field, tenor and mode (Hatim &amp; Mason 1990: 64–65; cf. Saridakis 2010: 72–74).</td>
</tr>
<tr>
<td>TERM</td>
<td>Inadequate or erroneous use of a lexeme, with reference to the textual meaning of either the ST or the TT and in relation to either the signified or the signifier. This category covers mainly issues of terminology and terminological/lexical equivalence and can be related causally to &lt;GR&gt; (i.e. denoting semantico-syntactic shift); &lt;REG&gt; (i.e. when lexical choice impacts discourse register); &lt;DEC&gt; (i.e. when the deficient decoding of the lexeme examined in the ST influences the decoding of the extended unit of meaning (s. Sinclair 1996; cf. Zethsen 2009)); &lt;ENC&gt; (i.e. when the deficient codification of the lexeme alters the sentential or textual meaning in the TT).</td>
</tr>
<tr>
<td>TERM+ENC</td>
<td>See &lt;TERM&gt;</td>
</tr>
<tr>
<td>TERM+REG</td>
<td>See &lt;TERM&gt;</td>
</tr>
</tbody>
</table>

#### 3.3. Analysis of data

Our approach can be exemplified in the following text extracts. In the examples shown, emphasis is on the problematic text chunks (or lexemes), while the error classification reflects the majority of observations with regard to the chunk examined. The chunk presented as appropriate in the “trainer suggestion” part of the following examples has relied, for the most part, not on the trainer's intuition but on attested data (concordances) derived from simple corpus linguistic analysis of Greek texts from the same genre (see Saridakis 2010: 260-262). Overall, this suggestion is deemed to be representative of the textual and registerial norm of the TL.

**Ex. 1 ST_B [TERM_ENC]**

<SS1> *The Niche for Machine Learning* </SS1>

<TTB1.01> Η ανάγκη [need] εμπόδιας μηχανών [for learning machines – ambiguity]  
<TTB1.02> Η θέση [place] για την μηχανική εμπόδια  
<TTB1.03> Η θέση [place] της μηχανικής μάθησης  
<TTB1.04> Μια γωνία [corner] για τη Μηχανική Μάθηση  

3
The word *niche* is used here to refer to the “*niche in the market as a specific area which has its own particular requirements, customers and products*” (Sinclair et al. 1998: 1113-1114). In Greek the semantic equivalent can be given in a phrase, since there is not a single term for this meaning (a place in the market). Nevertheless, this is a case of (1:1) semantic equivalence, as no other translation can convey the full meaning of SS1. Two of the students (TTB1.02, TTB1.03) managed to produce a quasi correct semantic transfer of the word in Greek (όθησι – place) but they failed to translate the full sentential meaning. Almost all of them suspected that the term *niche* is used metaphorically. Some of them used quotation marks to mark the metaphor (‘όθησι’, ‘κλειδί’), whereas others used different terms. The lack of semantic equivalence led to an indirect interference on the part of the translator (cf. Batsalia & Sella 1997: 169).

Student renditions in Example 1 can be broadly categorised under [TERM+ENC]. The students realised that the SS1 key sememe (*niche*) is not used in its prime meaning in the ST, but failed to produce a semantically and stylistically adequate translation into Greek, therefore altering the sentential meaning in the TT.

### Ex. 2 ST_A [TERM+DEC]

*Most of us think that we recognise intelligence when we see it, but we are not really sure how to precisely define or measure it. We informally judge the intelligence of others by relying on our past experiences in dealing with people. Naturally, this naive approach is highly subjective and imprecise.* <SS2> *A more principled approach would be to use one of the many standard intelligence tests that are available.*</SS2> *Contrary to popular wisdom, these tests, when correctly applied by a professional, deliver statistically consistent results and have considerable power to predict the future performance of individuals in many mentally demanding tasks.*

In this context the qualifier *principled* can be considered as denoting an entity that is “*based on a given set of rules*” (Soanes et al. 2003: 1399). The co-text of the utterance positions the meaning of intelligence as “*relying on our past experiences in dealing with people*” and the author maintains in his paper that a more scientific and formal definition is needed. Many students (11 out of a total 47) translated this term as *principles in general*. Six (6) mistook it as meaning *theory*: Four (4) understood it as meaning *moral principles*. Only two (2) students translated it as *scientific approach*, being closer to the ST intended meaning. Semantically, the <TTA2.01> rendition (based on principles in general) is rather ambiguous, and this over-translation can be explicated as reflecting uncertainty primarily on the level of SS decoding. This is also a case of a (1:1) lexico-semantic equivalence. The students failed to
contextualise the meaning of the utterance and opt for the more appropriate qualifier 'επιστημονική' [scientific]. The remaining renditions of our example can also be considered cases of direct lexical transfer, since the students were unable to contextualise the semantic field of the signifier. Overall, Example 2 (SS2: the principled approach) is typical of the [TERM+DEC] category. The students were unable to decode the signified at the level of the source text, i.e. in its immediate pragmatico-semantic context. The phenomenon of erroneous TT encodings is presumably rooted at the stage of decoding the meaning and not at the stage of encoding it in the TL.

Ex. 3 ST_B [GR+ENC]

However, for other types of problems, machine-learning methods are already emerging as the software development method of choice. In particular, machine learning is beginning to play an essential role within the following three niches in the software world:
(1) data mining,
(2) difficult-to-program applications, and
(3) customized software applications.

<SS3> Difficult-to-program applications: </SS3>

Machine-learning algorithms can play an essential role in applications that have proven too difficult for traditional manual programming – such applications include face recognition and speech understanding.

The ST structure does not exist in modern written Greek. In English it is an alternative to a relative clause\(^5\) and can be translated as such. In any case it requires a syntactic shift. Moreover, there is an additional difficulty for the translator because it is actually the title of a paragraph and it must be translated in a laconic way. Two of the students decided to preserve the dashes, therefore resorting to a syntactic calque (TTB3.01 and TTB3.08). Two of them provided an explanation (TTB3.02 and TTB3.06). Two also translated the infinitive to program as programmes instead of programming (TTB3.06 and TTB3.09). One of them used an adverbial neologism (TTB3.03) and one shifted semantically, translating difficult as complex.

Example 3 is typical of the [GR+ENC] category. The TT addressees are not familiar with this kind of syntax. Even though a syntactic shift would be reasonable in this case, the (non binary) error lies in the students' calquing of a TL-irrelevant syntactic structure, thus impeding encoding correctness and aptitude.

Ex. 4 ST_A [TERM+REG]

Most of us think that we recognise intelligence when we see it. But we are not really sure how to precisely define or measure it. We informally judge the intelligence of others by relying on our past experiences in dealing with people. Naturally, this naive approach is highly subjective and imprecise.

Example 4 requires commentary on two points: the first relates to the translation of the non-standard usage of the phrase recognise something when we see it. The ST author apparently decreases the level of formality, something that is stylistically inapt in the Greek textual norm. Translating this phrase literally in Greek might create a meaningful phrase, but it is a stylistically marked choice and a phrase that would not be normally used in this context. The semantic equivalent would be the quasi-metaphorical rendition με την πρώτη ματιά (at first glance). Nevertheless, 34 out of 47 students (72%) translated this phrase with a syntactic loan, a word-for-word translation of the original. The remaining students decided to use a varied form of see (συναντώ) or meet (συναντώ). One student chose to skip this part (TTA4.40), another to translate it as around us (TTA4.21) or come into contact with it (TTA4.42). Secondly, in many cases the person of the subject of the relative clause does not correspond to the person of the subject of the main clause. In nine (9) cases the students failed to properly show in the TT that the subject of both sentences is the same. Instead of translating most of us [we] believe that [we] recognise, in five (5) cases they translated most of us, [they] think that [they] recognise and in 4 cases they translated most of us, [they] think that [we] recognise. This is probably an error the students made because they failed to see that the subject of the verb think is deduced from the pronoun us.

Overall, Example 4 can be classified as typical of the [TERM+REG] category. By translating the sentence word-for-word and, thus, failing to produce a semantic equivalent in Greek, many students altered the anticipated register of the ST.
Machine-learning algorithms can play an essential role in applications that have proven too difficult for traditional manual programming—applications such as face recognition and speech understanding. The most accurate current programs for face recognition, for example, were developed using training examples of face images together with machine learning algorithms. In a variety of applications where complex sensor data must be interpreted, machine-learning algorithms are already the method of choice for developing software.

Example 5 typically falls into the [GR+DEC] category. Here, a common English structure of [gerund + noun] has been erroneously decoded and transferred in an [adjective + noun] structure. More specifically, the error lies in decoding the sentential meaning of the ST, which results in an erroneous TL translation.

### 3.4 Generalisation

Our two-level analysis showed that the categories of Decoding (DEC) and Encoding (ENC) may be placed one level above the others, as they can be combined with them. One way to explain this is that both decoding and encoding are stages at the translation process (Batsalia/Sella 1997: 126-131) during which decisions are made at the level of syntax, terminology and register. Moreover, they are almost always mutually exclusive. Therefore, an error either at the stage of decoding or at the stage of encoding needs a further description.

Most probably, students make errors at the stage of decoding (GR or TERM) because they are not very proficient in the SL (in the case of syntax) or because they are unfamiliar with the subject field of the text (the field of the utterance). On the other hand, students make errors at the level of encoding (GR, TERM or REG) because although they are able to decode the meaning of the ST, they fail to produce an adequate result in the TL, i.e. their mother-tongue. Decisions at the level of register can be further specified as regarding syntax or terminology and the stage of encoding. Errors here show insufficient knowledge in terms of text types of the TL and the norms governing them.

### 4. Formative and Summative Assessment. A Converging Approach

The above generalisation has been shown to facilitate the students' formal classification of unsuccessful renditions and thus to enable the much sought after interface between teacher and trainee.
In other words, our model seeks to describe, delineate and exemplify the (formative) aims suggested by Juan Sager (1994: 239) as follows:

- control of the accuracy of translation equivalents;
- quality control of the style, eliminating sources of language interference;
- adjust the translation product to a particular, and defined in detail, level or style of expression (register); and
- achieve stylistic/textual harmonisation, in the case of texts translated by various translators.

In this sense, our model aims to foster and exemplify the sound ethical motivation of students, in terms of requisite textual intervention.

Our empirical, error-driven analysis of translation trainee renditions, obviously goes hand-in-hand with the summative assessment models that are typically used by the translation industry during the revision stage (Kostopoulou & Saridakis 2011: 207). Kiraly's model (1995: 83-84), moves in our opinion more towards a summative assessment model, rather than towards a formative one, as he suggests a five-point scale, with one at the lowest rank, i.e. an unacceptable translation and 5 at the top, i.e. a very good translation.

<table>
<thead>
<tr>
<th>Scale rank</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>This is a totally unacceptable translation</td>
</tr>
<tr>
<td>2</td>
<td>This is a poor translation. It would require major improvements before it could be submitted to an employer</td>
</tr>
<tr>
<td>3</td>
<td>This translation is marginally adequate. It has several errors and would require a moderate amount of work to prepare it to be submitted to any employer</td>
</tr>
<tr>
<td>4</td>
<td>This is basically a good translation. It does have some minor errors, but they could be eliminated quite easily</td>
</tr>
<tr>
<td>5</td>
<td>This is a very good translation. It contains no errors with respect to the norms of the TL and it is a functionally acceptable translation of the source text</td>
</tr>
</tbody>
</table>

Table 2. Five-point scale for rating translations (Kiraly 1995: 83).

In other words, we believe that a register-based error analysis didactic approach, as outlined in this paper, is readily adaptable to Kiraly's macro-textual (and to some extent Skopos-theory imbued) model of translation product evaluation. In this sense, the two approaches are mutually complementary and can be easily combined in the translation teaching classroom, exemplifying also appropriateness of TL renditions through the analysis of TL genre-comparable disposable corpora.

5. Conclusions

Utilising our combined formative evaluation approach in the translation classroom, it can be said that students soon become aware of their deficiencies, but they consider them as translation difficulties. In most cases they recognise them as errors at the level of syntax, vocabulary, at the semantic or at the pragmatic level. When such errors are explained in the framework of a comprehensive model that focuses on the stage of translation process where the error is actually rooted, then they become more conscious of these translation difficulties and can be helped to overcome them. They take more deliberate and descriptively well-documented decisions on the TT lexicogrammar. For the teacher seeking a formative approach, this analysis may serve as a tool in the translation teaching classroom to be able to focus on the problematic areas and provide them with source texts that emphasise these areas.

Hence, the teacher-trainee interaction proposed with the exemplification of our model for the analysis of erroneous renditions, aims to make the trainees’ options more considered, more conscious and deliberate, less error-prone (and hence less norm-divergent), and meta-textually more transparent. As
Pym suggests (2012: 100) our aim is thus to link the trainee translator to the contextual elements, i.e. the textual and the extra-textual background of his effort. Such elements are the “things, the client’s instructions, current translation norms and their own work conditions” (ibid.).

On the level of causality, no theory would ever grasp the full dynamics of the translation process. However, as Pym (ibid.) argues, recognising the “multiple causation is not enough to [...] orient the translator’s decisions. It does, however, open up the terrain we need”. In conclusion, the more extensive the analysis of the systemic norms that govern the particulars of the ST (and hence of the TT), the more elaborate and efficient this needed terrain.

6. References