International Journal of English Literature and Social Sciences
Vol-6, Issue-1; Jan-Feb, 2021

Journal Home Page Available: https://ijeab.com/
Journal DOI: 10.22161/ijels

CohLitheSP: A new technique to study quality in Specialized Translation

Nicolás Montalbán[1], Juan Manuel Dato[2]

[1]Department of Languages, Centro Universitario de la Defensa de San Javier, Spain
[2]Instituto de Estudios Superiores Carlos III, Cartagena, Spain

Received: 30 Oct 2020; Received in revised form: 18 Jan 2021; Accepted: 01 Feb 2021; Available online: 14 Feb 2021

***Abstract*— *The aim of this work is to analyse the benefits of introducing a code in Python language to analyze quality in terms of the appropriateness of the texts as regards reading. In order to do this, a didactic experience was implemented in two final dissertations on the university course of Translation and Interpreting at the University of Murcia, in the academic years 2018-2019 and 2019-2020. The steps followed were:*

1. *Evaluation metrics for both MT compared to the reference translation (student´s translation)*
2. *Definition of a tool used to calculate easibility of the text*
3. *Determination of the weights*
4. *Calculation of the amplification constant for each specific corpus*
5. *Calculation of marks of easibility of texts*
6. *External evaluation attending House´s model (2015) adapted*

In accordance with the above, the following research questions are proposed: is human or MT translation better, and, is it possible to create a rubric based on significant grounds to calculate an approximate mark for quality in Translation?

Keywords*— *Computer-based studies, English for Specific Purposes, Linguistics, Literary translation, Quality in Translation processes, Scientific-technical translation.

I. INTRODUCTION

House (1981:127) starts most of her works with questions such as "What is a good translation?". In fact, it should be "one of the most important questions to be asked in connection with translation". For Halliday (2001:14) it is notoriously difficult to say why or even whether, something is a good translation". Quality translation should be mentioned here associated to the goals of MT and new ’interactive’ and/or ’adaptive’ interfaces have been proposed for post-editing (Green, 2015; Vashee, 2017). Therefore, in this case, human and MT are inextricably linked. Some recent studies mention that MT is almost ’human-like’ or that it ’gets closer to that of average human translators’ (Wu et al., 2016) and, also that MT quality is at human parity when compared to professional human translators” (Hassan et al., 2018). Ahrenberg (2017:1) states that the aim of MT is ‘overcoming language barriers’, although human translation is aimed at producing ‘texts that satisfy the linguistic norms of a target culture and are adapted to the assumed knowledge of its readers’. In order to do that, MT is used with human post-editing (O’Brien et al., 2014).

Other authors, Popovic´ and Burchardt (2011) emphasize the fact that errors produced by MT are useful since the comparison of human and MT can be an excellent exercise, and they claim for automatic error classification. Moreover, we should include here studies on effects of mentioned tools on translations (Jimenez-Crespo, 2009;

Lapshinova- Koltunski, 2013; Besacier and Schwartz, 2015).

One of the most required standards when comparing translations as mentioned before is quality. Mateo (2014), referring to Nord (1997) defines it as “appropriateness of a translated text to fulfil a communicative purpose”. Following Mateo et al. (2017) the results of this quality should be ’Very good’, ’Satisfactory’, or ’Unacceptable’.

Nevertheless, there are authors who claim that it is almost impossible to overcome the perfection of human translation (Melby with T. Warner, 1995) and Giammarresi and Lapalme (2016). MT Translation has gone through three stages ’from early dictionary-matched machine translation to corpus-based statistical computer-aided translation, and then to neural machine translation with artificial intelligence as its core technology in recent years’ (Zhaorong, 2018). Papineni et al. (2002) focus mainly on ’developing metrics whose ratings correlate well with human ratings or rankings’ Ahrenberg (2017:2).

House (2017:2) defines translation as ’the result of a linguistic-textual operation in which a text in one language is re-contextualized in another language’. For her, there are some interaction factors which should be taken into consideration (House, 2017:2-3):

- the structural characteristics, the limitations of two languages (source and target language);
- the extra-linguistic world
- the source text with its features;
- the linguistic-stylistic-aesthetic norms of the target language;
- the target language rules;
- intertextuality in the target text;
- traditions, principles, etc., in the target language;
- the translation company´s instructions given to the translator;
- the translator’s workplace conditions;
- the translator’s knowledge and expertise;
- the translation receptors’ knowledge and expertise.

House (2017:5) also insists on the cognitive aspects of translation, and specifically, the process of translation in the translator´s mind; a matter studied over the last 30 years, but certainly recently updated (cf. Shreve and Angelone 2011; O’Brien 2011; Ehrensberger-Dow *et al.* 2013). O’Brien (2013:6) states that any translation process has a lot of connections with other disciplines such as linguistics, psychology, cognitive science, neuroscience, reading and writing research and language technology.

Equivalence is another key point in translation, and authors such as Jakobson (1966) and Nida (1964) stating on ‘different kinds of equivalence’, and Catford (1965); House (1977, 1997); Neubert (1970, 1985); Pym (1995); and see Koller (1995, 2011). On the contrary, Hatim and Mason (1990) and Reiss and Vermeer (1984) do not give equivalence much importance. Following this line Vermeer 1984; Snell-Hornby 1988 and Prunč 2007 simply ‘reject it completely’ (House (2017:6), as do Munday (2012: 77) and Baker (2011: 5) more recently. Riccardi (2010, p.86) says, “The translated text is well anchored in the target culture and, in transposing the original; the translator will often be confronted with culture-bound expressions or situations”, and for Ahikary (2020) this means that “the equivalence is one of the most important aspects or goals of translation; translator has to focus on searching for the best equivalent terms between two different languages or dialects”.

In accordance with the present experiment, which is based upon the study on the human translation and MT quality in two final dissertations in the university course of Translation and Interpreting at the University of Murcia, the following research questions are proposed: is human or MT translation better, and, is it possible to create a rubric based on significant grounds to calculate an approximate mark for quality in Translation?.

II. METHODOLOGY

2.1 Contextualization and sample

This didactic experience was implemented in two final dissertations on the university course of Translation and Interpreting at the University of Murcia, in the academic years 2018-2019 and 2019-2020.

2.2. Development of the experiment

To carry out this work, different types of materials were used. First a suitable text in English which has never been translated before. The first final dissertation was a translation of a collection of texts dealing with: Quantum Physics, Technology, Medicine, Environment and Geology, with an extension of 600 words for each one. These mentioned scientific-technical texts have been taken from scientific publications and specialized magazines. The second one is an extract from *Red Dirt* (2016), a literary text from the narrative genre, whose main feature is the use of colloquial language, and is full of phraseological units and insults, with an extension of 2,500 words. For the MT two different tools were used: Matecat for the scientific-technical texts and Wordfast Anywhere for the literary text.

Apart from that, representative texts in Spanish were selected for comparison purposes: a selection of 5 scientific-technical texts from well-known international scientific publications. As far as the literary text, an extract was chosen from the book «Escritos de un viajo indecente» by Bukowski (2006), from the same genre and full of phraseological units, including insults. The steps followed were:

1. Evaluation metrics for both MT compared to the reference translation (student´s translation)
2. Definition of a tool used to calculate easibility of the text
3. Determination of the weights
4. Calculation of the amplification constant for each specific corpus
5. Calculation of marks of easibility of texts
6. External evaluation attending House´s model (2015) adapted

2.2.1 Evaluation metrics for both MT

At this point it is important to reiterate that we are comparing a reference translation with a machine translation within the context of the underlying idea that “the closer a machine translation is to a professional human translation, the better it is” (Papineni, Roukos, Ward & Zhu 2002: 311-318).

The first evaluation metrics we are introducing here are Precision and Recall. First, we must count the number of words in both the machine and the reference translation. In order to do a calculation with Precision, the number of common words is divided by the number of words in the machine translation. The calculation of Recall is achieved by dividing the number of shared words by the number of words in the reference translation. We consider a system to be good if scores are high, so the best system is the one with the highest scores.

WER (Word Error Rate) is another metric we are implementing. In this method, differences such as substitutions, insertions and deletions are taken into account. This metric is based on Levenshtein distance calculated at word level. In this case, the lower the WER result, the better.

The most common metric used is BLEU (Bilingual Evaluation Understudy). This method discovers how many n-grams are overlapping between the machine translation and the reference translation. This metric is based upon the idea that the larger the number of n-grams overlapping between the machine translation and the reference translation, the better the machine translation is. The machine translations should be as near to 1 as possible to be considered good translations. The formula to calculate BLEU is:

$$\mathrm{BLEU} = min\left(1, \frac{\text{number of words in MT}}{\text{number of words in ref}}\right) \prod_{i=1}^{4} precision_i$$

In order to obtain the results, a programme[1], written in Python language, was used to implement the WER, BLEU, Precision and Recall functions from the information dumped in a file. The file recognized a header, followed by different text segments corresponding to the original, a reference translation and several translations to be compared. The code proceeded to calculate each function by combining the reference with each translation to generate another file in table format that could be used directly and sent to a spreadsheet. When performing translation tasks, three different machine translations were offered. The average is calculated for each suggestion offered by the machine, taking into account the above metrics. We can go a step further and consider students´ translations as a reference translation and compare them to the MT. Then, when calculating the above-mentioned evaluation metrics (WER, BLEU, Precision and Recall), the results are refined. Following this, a mark can be calculated using this formula:

(3*(1-W) +1*B+1 *P+1 *R) ·10/6

When W=0 no mistakes, maximum mark 1-W

Following the above results, a mark can be calculated using this formula:

(3*(1-W) +1*B+1 *P+1 *R) ·10/6

When W=0 no mistakes, maximum mark 1-W

2.2.1.1 Matecat

At this point, it is important to note that Matecat has been used to translate the scientific-technical texts from the first final dissertation. According to Matecat’s site: “Matecat is a free and open source online CAT tool. It is free for translation companies, translators and enterprise users.” (Matecat, 2014). The founders and main contributors of Matecat are the international research center FBK (Fondazione Bruno Kessler), the translation company Translated srl, the Université du Maine and the University of Edinburgh.

In Matecat translation, assignments are organized into projects in which the user specifies the source language

[1] This programme was developed by Juan Manuel Dato Ruiz (qualified computer technician) taking into consideration the evaluation metrics mentioned above.

and the target language. One project comprises one or several texts to be translated, and each project has a translations memory. Matecat provides, by default, a connection with Google Translate as a machine translation system, and a connection with MyMemory as a public translation memory. It is important to mention that MyMemory is an open, available translation memory including the translation memories of the European institutions, the United Nations and automatically extracted data from multilingual websites. The first operation to be carried out is the analysis of the project. By clicking Analyze, Matecat shows how many words need to be translated in the preliminary analysis report it produces. In this report, the total number of words of the source text is displayed under Total Word Count. Then the post-editing is started and it is possible to see some translation suggestions. The translator has to decide how to adjust the translation and click Translated when the work is done. Matecat also offers the concordance function to look up words and phrases in the active translation memories. Once the post-editing is finished in the last segment, we can download the translated text and the translation memory. The Editing Log allows the translator to view adjustments made to the MT suggestions in the whole process. Finally, the average Post-Editing Effort (PEE) can be observed. It is important to mention that Matecat counts words according to industry standards, so "words or phrases with a 100% Translation Memory match are given a weighting of 30% and words or phrases with a partial TM match are given a weighting of 60%" (Matecat, 2014).

2.2.1.2 Wordfast Anywhere

As far as the second final dissertation on the literary text, the CAT tool used was Wordfast Anywhere, which is a Translation memory of the company Word. The procedure to use it is as follows: the text is divided into segments that are being translated and stored, creating glossaries and translations, which will appear in future translations depending on the index of coincidence of the words. It is necessary to create an account with an e-mail to Access a protected area, which acts as a cloud, where the translation memories, the glossaries and files of the project are stored. It is possible to access from any search engine and is offering the option of MT. This is the free version of Wordfast, the second memory translation used most in the world, after SDL Trados.

2.2.2 Definition of the tool used to calculate easibility of the text

To analyze the appropriateness of the texts as regards reading, a code in Python language has been developed. The first operation carried out by this code is sequencing words of the text to recover the number of paragraphs, sentences, words and syllables in total, and later, it determines five metrics based on the studies in Coh-Metrix, but simplified. Coh-Metrix, in accordance with its web page, is «a Computer tool which produces indexes in linguistic and discourse representations of a text». It is important to say that these mentioned indexes «are used in many different ways to research cohesion of the explicit text and coherence of the mental representation». Cohesion is understood here as «the features of an explicit text which plays a role helping the reader to connect ideas in the text mentally » (Graesser, McNamara, & Louwerse, 2003). Coherence, in this context, is «the interaction among linguistic and knowledge representations». When the focus is in the text, coherence coincides with the characteristics of the text which can contribute to the coherence of the mental representation.

This new technique is called CohLitheSP since it is based upon Coh-Metrix, and does not need large dictionaries nor corpuses formed by thousands of words to offer consistent results. Furthermore, on the other hand, specific formulae have been introduced for tests written in Spanish, when just a few changes have to be made to adapt it to any language without any extra cost.

For example, to calculate the number of syllables in a text, it is imperative to know the language it belongs to. In this case, it is needed to calculate how many vowels there are and subtract the diphthongs that, according to Spanish language, are formed by open and close vowels. Furthermore, additional exceptions should be calculated, bearing in mind the rule of hiatus when there is stressed close vowel, among others.

To apply the aforementioned metrics, the following are needed:

- A reference text conforming to a valid corpus,
- A glossary of technical or specific terms which is helping to know which words are specific within a corpus. These terms will not include measurement units nor "words of stop" (prepositions, determiners, etc), and
- A set of connectors allowing to know when, in a sentence, something is being inferred from something previously said.

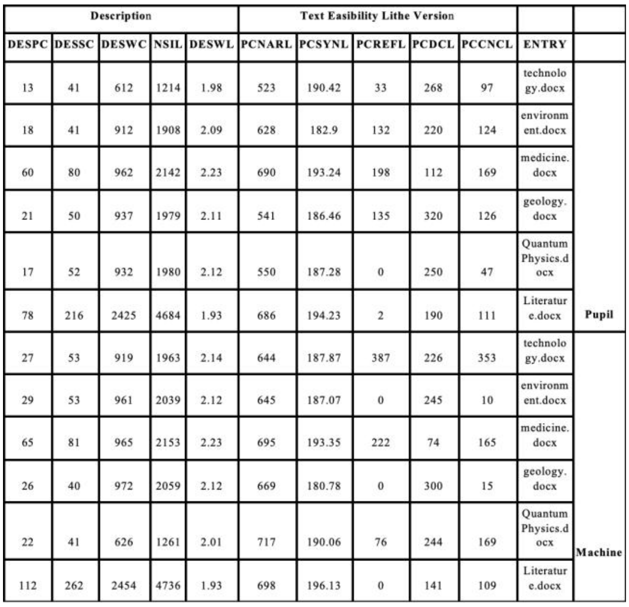
The selected metrics and their changes are:

- PCNARL. Narrativity. It is calculated determining which words of the text to be evaluated are already being recognized in the reference text.
- PCSYNL. Readability. It determines the simplicity of the text in its language. In the case

of Spanish, the readability of Fernández (1959) has been chosen (based on Flesch), which is using a number of sentences, syllables and words. If someone wants to do it for the English language, it only needs to be changed with the Flesch-Kincaid[2], whose formula is also based on a similar calculation.

- PCREFL. Referential Cohesion. In this version, the same referential cohesion as in Coh-Metrix is calculated; but instead of considering all nouns, it is only applied in technical or specific terms recognized in the glossary.
- PCDCL. Deep Cohesion. It determines the incidence of the connector over the recognized sentences.
- PCCNCL. Concreteness. In this version, instead of calculating the concreteness over the whole corpus of the language, the incidence of the terms of the glossary is determined from the recognized words in the reference text within the text to be evaluated.

This reduction in the cost of programming also requests to adopt mechanisms of compromise to be able to recognize the belonging of a word within large sets in such a way that the closest word is given back within some margins of tolerance etc.



Description					Text Easibility Lithe Version						
DESPC	DESSC	DESWC	NSIL	DESWL	PCNARL	PCSYNL	PCREFL	PCDCL	PCCNCL	ENTRY	
13	41	612	1214	1.98	523	190.42	33	268	97	technology.docx	Pupil
18	41	912	1908	2.09	628	182.9	132	220	124	environment.docx	
60	80	962	2142	2.23	690	193.24	198	112	169	medicine.docx	
21	50	937	1979	2.11	541	186.46	135	320	126	geology.docx	
17	52	932	1980	2.12	550	187.28	0	250	47	QuantumPhysics.docx	
78	216	2425	4684	1.93	686	194.23	2	190	111	Literature.docx	
27	53	919	1963	2.14	644	187.87	387	226	353	technology.docx	Machine
29	53	961	2039	2.12	645	187.07	0	245	10	environment.docx	
65	81	965	2153	2.23	695	193.35	222	74	165	medicine.docx	
26	40	972	2059	2.12	669	180.78	0	300	15	geology.docx	
22	41	626	1261	2.01	717	190.06	76	244	169	QuantumPhysics.docx	
112	262	2454	4736	1.93	698	196.13	0	141	109	Literature.docx	

Fig.1: Results of the programme

In order to do that, a structure (a decision tree) has been created to order words in such a way that we know instantly whether words are included in the structure or not: we are interested in this version not only in the lexemes of Spanish, but also in their cases. That is, considering that we have not been working with an extensive dictionary of Spanish language, nor the rules determining its lexemes, when it is masculine or feminine, in singular or plural. Furthermore, if it is a verb, it should recognize its verbal tense (present, past, future, conditional, etc.). The algorithm proceeds to repeat, as it were, a process of stressing a word, the first characters in every word several times, and more times than the last ones. In this way, when calculating the movements (Levenshtein's distance), errors will have less weight at the end of the word (morphemes) and more weight at the beginning (root).

By using this mechanism under a tolerance of 25% (the words whose ratio of Levenshtein is not below 75% are accepted) an approximation closely related to a process of lematization is obtained.

For the calculation of the narrativity, it is necessary to use these techniques, as well as for the calculation of the concreteness – to be able to generate two decision trees.

The following ideas have been considered to separate in sentences:

1. A sentence is formed by more than ONE word.
2. After a dot a sentence begins in upper case.
3. The sentences that do not comply with 1 and 2 will be separated by "**; -¿? ¡!.:**"
4. If a sentence complies with 1 or 2, it will be added to next sentence.

After applying this simplified version of Coh-Metrix over the produced texts in Spanish, it is possible to see how, after being evaluated separately with a mark from 0 to 10, they seem to describe a similar curve:

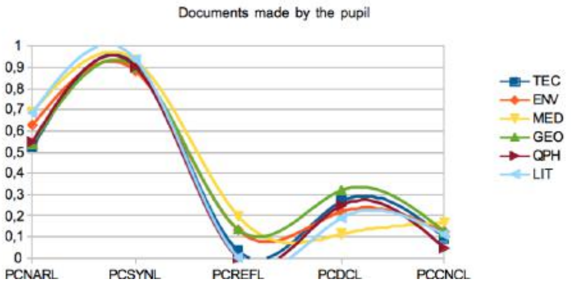


Fig.2: Text Easibility Lithe Version in Percentages

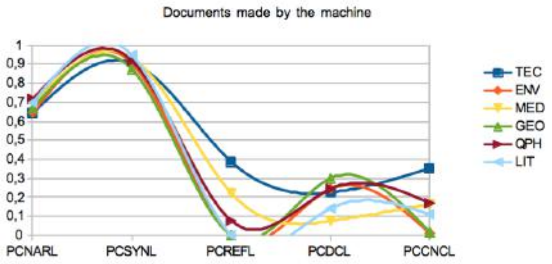
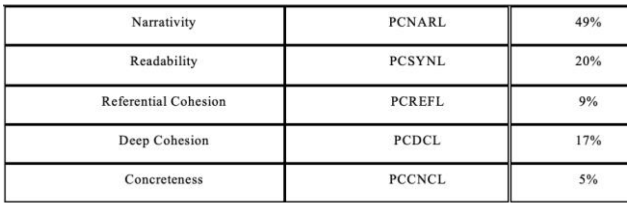


Fig.3: Text Easibility Lithe Version in Percentages (MT)

As can be seen in the above figures, different types of written texts for different technical corpuses seem to be minor differences in marks, but with a pattern that seems to say that measurements are not random. Therefore, it seems that, in addition, the texts used as references, representing a corpus without errors, have a mark below below 10 so students can never get that mark. Therefore, not only must each Coh-Lithe metric be weighted in such a way that favours the distinction among students' faculties, but, in addition, the results must be amplified so the reference texts have the same mark. For this reason, now there is an explanation on how to calculate the weighting of each metric and the constant used to amplify the mark.

2.2.3 Determination of the weights

By analysing the different students' texts, it is interesting to point out that the best marks should come from metrics where each student has the most dissenting marks and those metrics where students have better marks should weigh more. Therefore, after multiplying the media and standard deviation of each metric and normalizing the results, the following weights have been generated:

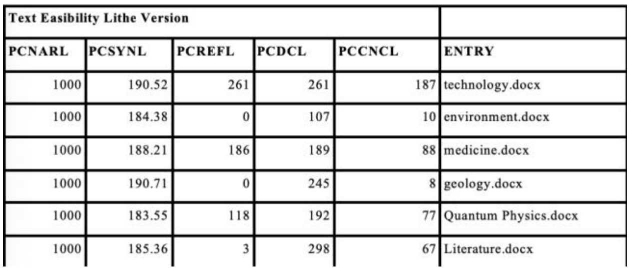


Narrativity	PCNARL	49%
Readability	PCSYNL	20%
Referential Cohesion	PCREFL	9%
Deep Cohesion	PCDCL	17%
Concreteness	PCCNCL	5%

Fig.4: Percentage of Weights

2.2.4 Calculation of the amplification constant for each specific corpus

Below, the results of evaluating the reference texts can be seen.



Text Easibility Lithe Version					
PCNARL	PCSYNL	PCREFL	PCDCL	PCCNCL	ENTRY
1000	190.52	261	261	187	technology.docx
1000	184.38	0	107	10	environment.docx
1000	188.21	186	189	88	medicine.docx
1000	190.71	0	245	8	geology.docx
1000	183.55	118	192	77	Quantum Physics.docx
1000	185.36	3	298	67	Literature.docx

Fig.5: Results of evaluating reference texts

As we can observe, with the exception of Narrativity, the maximum mark is not achieved in each parameter, so, first, the weights for each case are applied and, later, a rule of three with the maximum mark (10). The result will be the constant by which all texts using this reference document are multiplied. For example, if the amplification constant over the texts of the technological corpus as reference is needed, then this formula is being used, after calculating the coefficients from the programme:

$$K_{TEC} = \frac{10}{\frac{PCNARL_{TEC}^{REF}}{1000}\cdot 0.49 + \frac{PCSYNL_{TEC}^{REF}}{206.82}\cdot 0.2 + \frac{PCREFL_{TEC}^{REF}}{1000}\cdot 0.09 + \frac{PCDCL_{TEC}^{REF}}{1000}\cdot 0.17 + \frac{PCCNCL_{TEC}^{REF}}{1000}\cdot 0.05}$$

Under these weights, marks of the six reference texts have been studied, and it has been found an amplification of **1.39**.

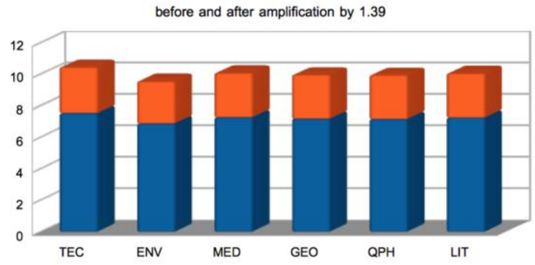


Fig.6: Marks of reference texts

For that reason, if we do not want to multiply the amplifier within its corpus, it seems that it is not inexact to multiply by **1.39**, regardless of the reference text.

2.2.5 Calculation of marks of easibility of texts

Regarding the calculation of the marks of the texts, the amplification constant must be applied by the addition of each metric divided by its maximum and multiplied by its weight. For example, the following formula can be observed over the technology texts:

$$Score_{TEC}^{PUPIL} = K_{TEC}\cdot\left(\frac{PCNARL_{TEC}^{PUPIL}}{1000}\cdot 0.49 + \frac{PCSYNL_{TEC}^{PUPIL}}{206.82}\cdot 0.20 + \frac{PCREFL_{TEC}^{PUPIL}}{1000}\cdot 0.09 + \frac{PCDCL_{TEC}^{PUPIL}}{1000}\cdot 0.17 + \frac{PCCNCL_{TEC}^{PUPIL}}{1000}\cdot 0.05\right)$$

2.2.6 External evaluation

For the external evaluation, the House´s model (2015) refined and adapted have been implemented in this new model. House distinguishes between overt and covert translations. According to House's definition 'a covert translation is a translation which enjoys the status of an original source text in the target culture´ (2015: 56). This means that the translation is not marked pragmatically by its source text, so it could have been created independently; therefore they are 'pragmatically of equal concern for source and target language addressees´ (2015: 56). Meanwhile, an overt translation has to cope with the cultural assumption of the target language to be able to translate the text appropriately. In the final dissertations that we are analysing, the first one is a covert translation, and the second one an overt translation.

House (2015:63) states clearly that translation is 'the replacement of a text in the source language by a semantically and pragmatically equivalent text in the target language´; therefore, it must be equivalent. House agrees with Halliday's assumption (1989:11) that the text and the context of the situation should be separated. In addition, the concepts of Field, Mode and Tenor from Halliday are also used (House 2015: 64). The Mode refers to both the channel (in this case, the text is written to be read) and the degree to which potential or real participation is allowed for between writer and reader. The Field refers to the content, the subject matter. The Tenor is the nature of participants, the addresser and the addressees, whether the author's personal (emotional and intellectual) stance help to transmit the message. However, in her work, House incorporates the idea of Genre, 'It connects texts with the 'macro-context' of the linguistic and cultural community in which the text is embedded´ (2015:64). The following Figure by House (2015:65) summarizes the whole model.

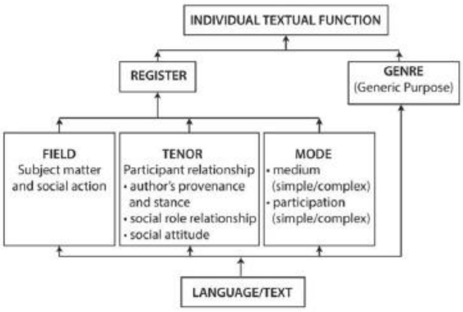
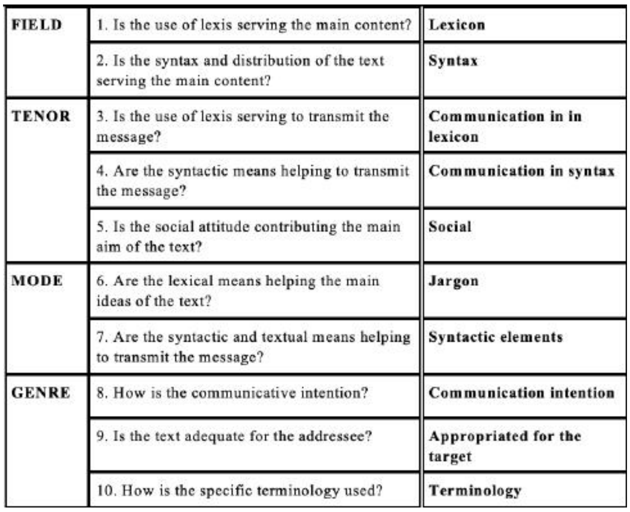

INDIVIDUAL TEXTUAL FUNCTION
REGISTER
GENRE (Generic Purpose)
FIELD Subject matter and social action
TENOR Participant relationship • author's provenance and stance • social role relationship • social attitude
MODE • medium (simple/complex) • participation (simple/complex)
LANGUAGE/TEXT

Fig.7: House´s model for analysing and comparing original and translation texts

The cultural filter is another important concept introduced by House. As defined by the author (2015:68) it 'is a means of capturing socio-cultural differences in expectation norms and stylistic conventions between the source and target linguistic-cultural communities. Therefore, to compare both texts (source and target language) it is necessary to bear the cultural concept in mind.

To apply House´s considerations, the authors of this work have created a questionnaire with 10 questions that have been posted to a class of the 4th year of a university course of Translation and Interpreting, who have already finished a subject on Specialized Translation and have the knowledge to analyse and evaluate translations of this type.



FIELD	1. Is the use of lexis serving the main content?	**Lexicon**
	2. Is the syntax and distribution of the text serving the main content?	**Syntax**
TENOR	3. Is the use of lexis serving to transmit the message?	**Communication in in lexicon**
	4. Are the syntactic means helping to transmit the message?	**Communication in syntax**
	5. Is the social attitude contributing the main aim of the text?	**Social**
MODE	6. Are the lexical means helping the main ideas of the text?	**Jargon**
	7. Are the syntactic and textual means helping to transmit the message?	**Syntactic elements**
GENRE	8. How is the communicative intention?	**Communication intention**
	9. Is the text adequate for the addressee?	**Appropriated for the target**
	10. How is the specific terminology used?	**Terminology**

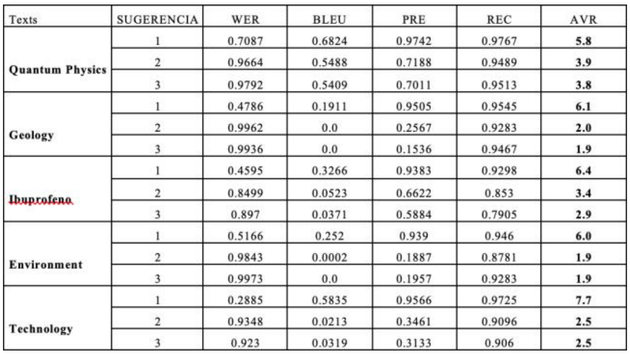
Fig.8: Questionnaire

This questionnaire was posted in Google Forms after having successfully finished the subject on Specialized Translation, as a class activity on-line. The second final dissertation with a literary text has only 9 questions since we considered two of them as one.

III. RESULTS

3.1 Evaluation metrics

The scientific-technical texts had the following results:



Texts	SUGERENCIA	WER	BLEU	PRE	REC	AVR
Quantum Physics	1	0.7087	0.6824	0.9742	0.9767	**5.8**
	2	0.9664	0.5488	0.7188	0.9489	**3.9**
	3	0.9792	0.5409	0.7011	0.9513	**3.8**
Geology	1	0.4786	0.1911	0.9505	0.9545	**6.1**
	2	0.9962	0.0	0.2567	0.9283	**2.0**
	3	0.9936	0.0	0.1536	0.9467	**1.9**
Ibuprofeno	1	0.4595	0.3266	0.9383	0.9298	**6.4**
	2	0.8499	0.0523	0.6622	0.853	**3.4**
	3	0.897	0.0371	0.5884	0.7905	**2.9**
Environment	1	0.5166	0.252	0.939	0.946	**6.0**
	2	0.9843	0.0002	0.1887	0.8781	**1.9**
	3	0.9973	0.0	0.1957	0.9283	**1.9**
Technology	1	0.2885	0.5835	0.9566	0.9725	**7.7**
	2	0.9348	0.0213	0.3461	0.9096	**2.5**
	3	0.923	0.0319	0.3133	0.906	**2.5**

Fig.9: Evaluation metrics for the scientific-technical texts

And the literary text:

Text	SUGERENCIA	WER	BLEU	PRE	REC	AVR
Literary	1	0.72288	0.0352	0.8063	0.7995	**4.1**

Fig.10: Evaluation metrics for the literary text

3.2 Evaluation of the scientific-technical and the literary texts

After applying the corresponding formulas already described in 2.2.2, 2.2.3 and 2.2.4, the following results are achieved:

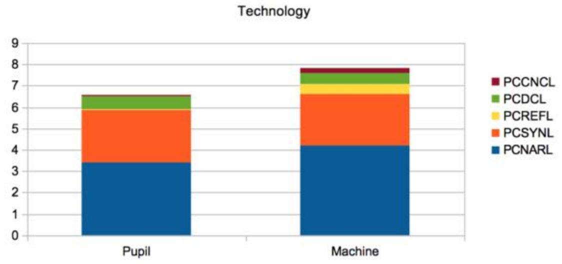


Fig.11: Evaluation amplified by its reference

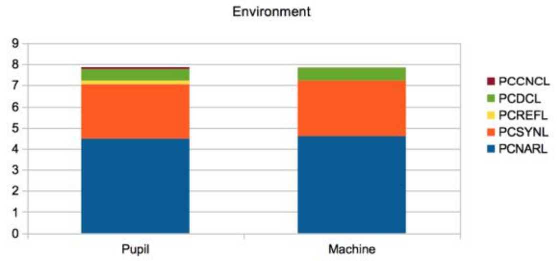


Fig.12: Evaluation amplified by its reference

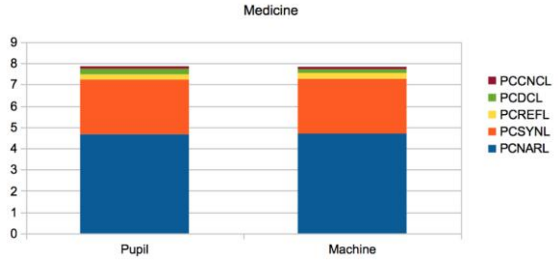


Fig.13: Evaluation amplified by its reference

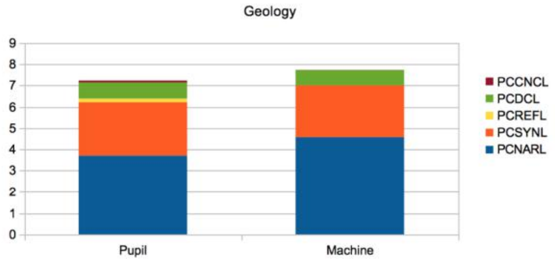


Fig.14: Evaluation amplified by its reference

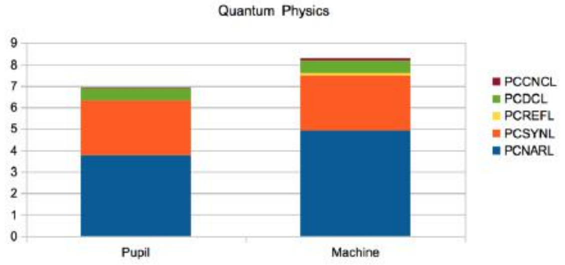


Fig.15: Evaluation amplified by its reference

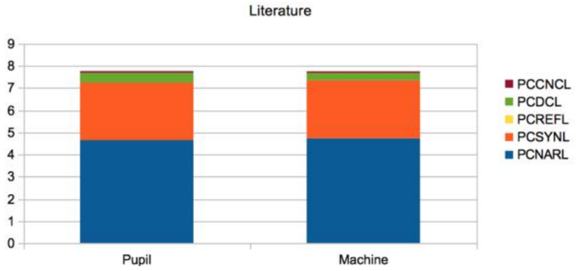


Fig.16: Evaluation amplified by its reference

IV. DISCUSSION

4.1 Evaluation metrics

An analysis of the evaluation metrics in 3.1 shows us the following results:

Regarding the final dissertation on scientific-technical texts:

- MT Suggestion 1 is the best one in the 5 texts (Matecat can offer up to 3 MT suggestions), having 5.8. 6.1, 6.4, 6 and 7.7, which is an excellent result.

Considering the final dissertation on a literary text:

- MT suggestion on the literary text had a mark of 4.1, which is not so negative if we consider that it is an overt text and the human translator had to adapt precisely to the target culture, so it means that more changes were made in the MT than in the final dissertation suggestion to post edit the text.

4.2 Evaluation of the texts

As can be seen, MT gets better results than reference translations (student's translations). In fact, considering the value these questionnaires have, these ones could be contrasted to the previous results. To be able to understand the value of questionnaires, first we observe the questions, bearing in mind that each student could evaluate their results corresponding with different degrees of relevance.

By adding the evaluations made by students in the previous questions the following results are obtained for a text within the 5 first of scientific-technical content:

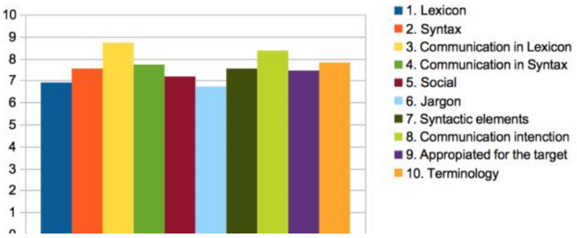


Fig.17: Survey on scientific-technical texts

It is observed that students' evaluation approximately coincides with the Coh-Lithe-SP's evaluation. In addition, a similar evaluation is achieved with the literary text compared with Coh-Lithe-SP's.

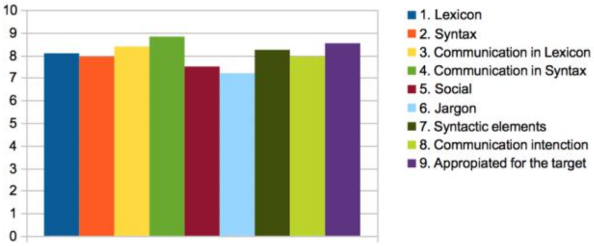


Fig.18: Survey on a literary text

V. CONCLUSION

In this work, a new and different tool has been shown which adds a supplementary challenge for students: the possibility of improving the readability of their own translations from English into Spanish.

Given the facts, the technique explained before is working properly mainly due to two results: on the one hand, it is proved that different texts coming from different typologies, including MT texts, get good or bad marks in the same metrics. On the other hand, the tests also show that, after refining the final mark, the result is approximate to a student's evaluation.

Moreover, it is important to stress the easy programming, which does not require large corpuses, despite the fact it comes from systems needing an enormous extra charge in the development of programming. This last feature is complemented by the fact that it is easily transformed to be working in any language.

The procedure used to test the new tool implemented with the external evaluation questionnaire should also be highlighted. This questionnaire updates and implements House's and Halliday's considerations by testing the new tool considering the pragmatic and cultural aspects of both source target texts.

VI. SOFTWARE

The programme written in Python used to calculate the statistics with commentaries in English can be found in the following address: https://archive.org/details/coh-lithe-sp-012

REFERENCES

[1] Adhikary, R. P. Degrees of equivalence in translation – a case study of Nepali novel 'seto bagh' into English as 'the wake of the white tiger'. *European Journal of Multilingualism and Translation Studies*, [S.l.], v. 1, n. 1, May 2020. Available at: <https://oapub.org/lit/index.php/EJMTS/article/view/173>. Date accessed: 21 july 2020

[2] Ahrenberg, L. Comparing machine translation and human translation: A case study, In Irina Temnikova, Constantin Orasan, Gloria Corpas and Stephan Vogel (eds), RANLP 2017 The First Workshop on Human-Informed Translation and Interpreting Technology (HiT-IT) Proceedings of the Workshop, September 7th, 2017; 2017, pp. 21-28

[3] Basil H. and Mason,I. (1990) *Discourse and the Translator*. London: Longman.

[4] Catford, John (1965) *A Linguistic Theory of Translation*. Oxford: Oxford University Press

[5] Fernández Huerta J.(1959). Medidas sencillas de lecturabilidad. *Consigna* 1959; (214): 29-32

[6] Giammarresi, S. and Lapalme, G. (2016). Com- puter science and translation: Natural languages and machine translation. In Yves Gambier and Luc van Doorslaer, editors, *Border Crossings: Translation Studies and other disciplines*, John Benjamins, Am- sterdam/Philadelphia, chapter 8, pages 205–224.

[7] Gregory, S. and Angelone, E. (2011) 'Uncertainty Management, Metacognitive Bundling in Problem Solving, and Translation Quality', in Sharon O'Brien (ed.) *Cognitive Explorations of Translation*. London: Continuum, 108–30.

[8] Hassan, H., Aue, A., Chen, Chowdhary, C. V., Clark, J., Federmann, C. , Huang, X., Junczys-Dowmunt, M., Lewis, W., Li, M., Liu, S., Liu, T-Y., Luo, R., Menezes, A., Qin, T., Seide, F., Tan, X., Tian, F., Wu, L., Wu, S., Xia, Y., Zhang, D., Zhang, Z., and Zhou, M. 2018. Achieving Human Parity on Automatic Chinese to English News Translation. *Computing Research Repository*, arXiv:1803.05567.

[9] Flesch, R. *(1948). "A new readability yardstick". Journal of Applied Psychology.* ***32*** *(3): 221–233.* doi:10.1037/h0057532. PMID 18867058.

[10] Green, S. (2015). Beyond post-editing: Advances in interactive translation environments. *ATA Chronicle* Www.atanet.org/chronicle-on-line/

[11] House, J., (2015) *Translation Quality Assessment: Past and present*. Routledge

[12] Jakobson, R. (1966) 'On Linguistic Aspects of Translation', in Reuben Brower (ed.) *On Translation*. New York: Oxford University Press

[13] Koller, W. (1995) 'The Concept of Equivalence and the Object of Translation Studies', *Target* 7: 191–222

[14] Koller, W. (2011) (11th ed.) *Einführung in die Übersetzungswissenschaft*. Tübingen: Francke

[15] Maureen, E-D., Göpferich, S. and O'Brien, S. (eds) (2013) Special Issue: Interdisciplinarity in Translation and Interpreting Process Research. *Target* 25:1

[16] Graham, Y.; Baldwin, T.; Moffat, A.; and Zobel, J.. 2013. Continuous Measurement Scales in Human Evaluation of Machine Transla- tion. In *Proceedings of the 7th Linguistic Anno tation Workshop & Interoperability with Discourse*, pages 33–41, Sofia, Bulgaria

[17] Jiménez-Crespo, M. A. 2009. Conventions in localisation: a corpus study of original vs. translated web texts. *JoSTrans: The Journal of Specialised Translation* 12:79–102

[18] Kirti Vashee. 2017. A closer look at sdl's adaptive mt technology. Http://kv- emptypages.blogspot.se/2017/01/a-closer-look- at-sdls-adaptive-mt.html

[19] Kishore Papineni, K., Roukos, S., Ward, T. and Zhu, W-J. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of 40th Annual Meeting of the Association for Computational Linguis- tics*. Association for Computational Linguistics, Philadelphia, Pennsylvania, USA, pages 311–318. https://doi.org/10.3115/1073083.1073135

[20] Koller, W. (1995) 'The Concept of Equivalence and the Object of Translation Studies', *Target* 7: 191–222

[21] Koller, W. (2011) (11th ed.) *Einführung in die Übersetzungswissenschaft*. Tübingen: Francke

[22] Läubli, S, Senrich, R, Volk, M. (2018). arXiv:1808.07048 **[cs.C]**

[23] Martıinez Mateo, R. (2014). A deeper look into metrics for translation quality assessment (TQA): A case study. *Miscelanea: A Journal of English and American Studies* 49:73–94

[24] Martıinez Mateo, R.; Martinez, M.and Moya Guijarro, A. J.. (2017). The modular assessment pack a new approach to translation quality assessment at the directorate gen- eral for translation. *Perspectives: Studies in Translation Theory and Practice* 25:18–48. Doi 10.1080/0907676X.2016.1167923

[25] Matecat (2014). https://site.matecat.com/benefits/?gclid=Cj0KCQjw6uT4BRD5ARIsADwJQ1-s67PUj9ENqpo1g9Yl5kP2SQyfikdDv3DU_dHUMG-rxM2J_XsFG6QaAk0yEALw_wcB

[26] Melby, A. and Warner, T. 1995. *The Possibility of Language*. John Benjamins, London and New York. https://doi.org/10.1075/btl.14

[27] Munday, J. (2008) (3rd ed. 2012) *Introducing Translation Studies: Theories and Applications*. London: Routledge

[28] O'Brien, S. (2011) *Cognitive Explorations of Translation*. London: Continuum

[29] Popovic, M.´ and Burchardt, A. 2011. From human to automatic error classification for machine translation output. In *Proceedings of the 15th Inter- national Conference of the European Association for Machine Translation*. Leuven, Belgium, pages 265– 272

[30] Mona, B. And Pérez-González, L. (2011) 'Translation and Interpreting', in James Simpson (ed.) The Routledge Handbook of Applied Linguistics. London: Routledge, 39–52

[31] Neubert, A. (1970) 'Elemente einer allgemeinen Theorie der Translation', in *Actes du Xe Congrès International des Linguistes*, Bucharest, 1967, 451–56

[32] Neubert, A. (1985) *Text and Translation*. Leipzig: Enzyklopädie

[33] Nida, E. (1964) *Toward a Science of Translation*. Leiden: Brill

[34] Nord, C. 1997. *Translation as a Purposeful Activity*. St Jerome, Manchester, UK

[35] O'Brien, S., Balling L. W., Carl, M., Simard, M. And Specia, L. 2014. *Post- Editing of Machine Translation: Processes and Applications*. Cambridge Scholars Publishing, New- castle upon Tyne

[36] O'Brien, S. (2013) 'The Borrowers: Researching the Cognitive Aspects of Translation', *Target* 25: 5–17

[37] Pym, A. (1995) 'European Translation Studies, *une science qui dérange*, and why Equivalence Needn't Be a Dirty Word', *TTR* 8: 153–76

[38] Reiss, K. and Vermeer, H. (1984) *Grundlegung einer allgemeinen Translationstheorie*. Tübingen: Niemeyer

[39] Riccardi, S. K. (2010). *Translation studies: Perspectives on an emerging discipline*. Cambridge: Cambridge University Press

[40] Wu, Y.; Schuster, M.; Chen, Z.; V. Le, O.; Norouzi, M.; Macherey, W.; Krikun, M.; Cao, Y.; Gao, Q.; Macherey, K.; Klingner, J.; Shah, A.; Johnson, M.; Liu, X.; Kaiser, L.; Gouws, S.; Kato, Y.; Kudo, T.; Kazawa, H.; Stevens, K.; Kurian, G.; Patil, N.; Wang, W.; Young, C.; Smith, J.; Riesa, J.; Rudnick, A.; Vinyals, O.; Corrado, G.; Hughes, M.; and Dean, J. (2016). Google's neural machine translation system: Bridging the gap between human and machine translation. *CoRR* abs/1609.08144. Http://arxiv.org/abs/1609.08144

[41] Zhaorong Zong 2018 *J. Phys.: Conf. Ser.* **1087** 062046