In Times of Neural Machine Translation:
Linguistic Competence Back on the Translation Stage

Bulgaru Gabriela

https://orcid.org/0000-0002-5089-3252

Babeș-Bolyai University, Romania

Abstract

Long time ago, translation competence was seen as a mere linguistic asset, eventually a bilingual one. As society possibilities of communication developed and technology progressed, translation competence has been defined as a more and more complex set of knowledge and skills. It includes now an increasing number of sub-competences, depending on the perspective of approaching translation: either as a process, as a product, or as a profession. Nowadays machine translation and post-editing of automatic output, which have steadily entered the language service providers market, has brought a reorganization within the translation competence framework. In this paper we explore, thus, how a paradigmatic change has taken place: the linguistic sub-competence, which, as fundamental as it is, became rather shadowed by the other newer, modern sub-competences, is again at the core of translation competence. We bring in findings of our ongoing research on a corpus of machine-translated news texts, that we post-edited and quality assessed and evaluated, and discuss how machine translation fluency errors call for a refocus on linguistic competence in students' translation training.

Keywords: *translation competence, linguistic sub-competence, machine translation post-editing, machine translation fluency errors*

1. Translation competence, past and present

Translation had had a long-time status of a solely linguistic activity, until information and communications technology started to develop and slowly change society. Academic and professional research into translation flourished and, several decades ago, Translation Studies set itself up as a scientific discipline. Since then, special attention was given to translation competence, i.e. the knowledge and skills translators need to be trained in, so that they deliver a qualitative translation and meet clients' expectations. In the following lines, we only sketch some of the most prominent approaches to translation competence, in order to show how models of conceptualization progressed from a unique component to a multicomponent view.

In the post-industrial or postmodern era, researchers such as Vinay and Darbelnet 1958/1995, Catford 1965, Mounin 1976, Pergnier 1976, and others considered translation a matter of thorough or main linguistic competence, but in any way, a linguistic asset. It is important to note that, starting with the same period, first steps were made into machine translation (MT) research, with Warren Weaver's "Translation" memorandum in 1949 [1]. The world had already experienced the terrible World War II and the nuclear bombing; thus, the idea of making the process of translation faster and cheaper, in order to facilitate communication between people, for a peaceful future of the planet, rapidly grew among experts. However, research in MT made rather little progress in the beginning and experiments were even disappointing. Translation Studies, on the contrary, developed a lot and, in 1972, Holmes, the first one to legitimize the discipline, made up a taxonomy [2] that showed its large borders: descriptive translation studies, translation theory, and applied translation studies.

Holmes' 1972 paper officially marked the beginning of a fruitful period of research in all Translation Studies branches. Numerous books, papers, analyses, nowadays considered benchmarks in the Translation Studies history (Reiss 1971, Steiner 1975, Toury 1978/1995, Vermeer 1978/1989, Wills 1982, Berman 1984, Baker 1992, Nord 1997, Hatim and Mason 1997, Venuti 2000, for instance), pushed forward the investigations. Translation competence was one of the topics greatly debated.

As mentioned above, scholars considered at the beginning that translation competence is an innate skill or a bilinguals' asset [3]. However, the information era started with three different

definitions of translation competence: ideal bilingual competence, expert competence, and communicative competence [4]. Translation is tackled as a process and the interest is in how this process takes place, but language is eventually the unique substance the process is focused on.

Another representative approach was the functionalist theory [5], which presents translation as a skopos-driven activity. This perspective triggered the conceptualization of translation competence as a set of interrelated competences in terms of knowledge, skills, awareness, and expertise [6, 7]. Some notable translation competence frameworks or models have been elaborated since then. PACTE's holistic, dynamic model [8] comprised the strategic, bilingual, extra-linguistic, psycho-physiological, and instrumental sub-competences, and knowledge about translation; the central sub-competence was the strategic one.

Göpferich's model [9] encompassed, around the central strategic sub-competence and motivation, the following sub-competences: domain, psychomotor, translation routine activation, tools and research, and communicative sub-competence in at least two languages. It also presented other factors, such as translator's self-concept, psycho-physical disposition, translation norms etc. EMT network's latest translation competence framework includes five sub-competences: language and culture, translation, technology, personal and interpersonal, and service provision [10].

It results that the linguistic element is not at the center anymore, but on the same line with other components, which are more and more numerous, as the conceptual framework of translation competence continued to expand. Since translation is not only an academic discipline, but also a business, whose landscape has tremendously changed in the past few years, new sub- or interrelated competences required on the market were therefore englobed in the competence framework (translation business, business management, digital intelligence, for instance).

2. Neural machine translation: linguistic competence back on stage

The rise of neural machine translation (NMT) in 2015 [11] was above expectations and surpassed the level of performance of machine translation systems already in use (rule-based, example-based, statistical and hybrid MT systems), due to combined recurrent neural networks and deep learning algorithms, plus other algorithms that improve the process of automatic translation. The output can have different degrees of quality, depending on the source and target language pair, the linguistic data sets the MT engines can train with, and the type of text to translate. Nevertheless, post-editing is still necessary in order to attain the aimed quality: full post-editing for quality similar to human translation, or light post-editing for the gist of the information only.

Many language companies adopted MT in view of its high-speed processing that can handle large volumes of texts. ELIS 2022 survey results [12] show that 58% of language company respondents have implemented MT and other 20% plan to do so; also, more than 70% of independent language professionals already use MT to some extent, while more than 60% of the academia or training institute respondents say that MT is implemented, with other 10% planning to do so. Consequently, post-editing has been lately the fastest growing service line on the language market. Light post-editing is regarded as an opportunity by more than 40% of language companies, language departments (in public agencies or private enterprises), training institutes, and by independent language professionals as well, although in a smaller percentage [12].

We therefore set to analyze a corpus of English news translated in Romanian with neural machine translation. Our purpose is to quality assess and evaluate the output of neural machine translation on this language combination, to full post-edit it and see what efforts this process requires from the part of translators/post-editors. This approach will bring a practical insight into the competences translators need to rely upon and employ more than they used to in the habitual context. It will also shed light on MT phenomena and errors and on what knowledge and skills future translators need to acquire and develop.

Google Translate, the application we used, is said to have access to the largest linguistic database to train with and offer features for over 100 languages. There are also experiments that confirm our research-built opinion that, due to its database, this engine may work better than other MT software like those developed in-house [13, for instance], with low-resourced languages, such as Romanian. We assessed and evaluated the quality of MT output using the harmonized MQM-DQF model elaborated by QTLaunchPad and TAUS [14], which proposes eight dimensions for assessing errors in translation: Accuracy, Design, Fluency, Locale Conventions, Style, Terminology, Verity, and any Other.

In our ongoing research, we covered so far a corpus of 6960 words, distributed in 558 segments. Initial segmentation of the source and target corresponding texts was made by sentence, but the assessment and evaluation process required a more fragmentary approach, hence a sub-

segmentation dictated by error. At this point, we need to mention that our research also includes complementary analyses carried out in the same framework, serving other specific research objectives. In this paper, we present only the findings concerning the NMT performance in terms of Fluency. In the NMT output of 6960 words, we identified and analysed 501 errors thus far. Each error was annotated with different degrees of severity, depending on the impact it had on the translation functionality and consequently on its potential readers. Table 1 shows how the highest error percentages are related to Accuracy – 170 and Fluency – 164. While Accuracy refers to content equivalence made in translation, Fluency deals with language and its good functioning. As one can see, the great majority of Fluency errors are grammar linked (part of speech, word form, verbal tense, word order etc.). Other important errors are those related to cohesion and punctuation.

Count of Error Category	Column Labels				
Row Labels	Critical	Major	Minor	Neutral	Grand Total
⊞ Accuracy	64	42	45	19	170
⊟ Fluency	14	56	82	12	164
⊞ Grammar	10	49	55	10	124
⊞ Inconsistency			2		2
⊞ Punctuation	1	1	10		12
⊞ Spelling			8		8
⊟ (blank)	3	6	7	2	18
ambiguity	1	1	2		4
coherence				1	1
cohesion	2	5	5	1	13
Grand Total	**78**	**98**	**127**	**31**	**334**

Table 1. Number and types of Fluency errors

Although there are fewer critical Fluency errors as compared to Accuracy, the number of major ones was bigger, while the number of minor ones was almost double (Fig. 1).
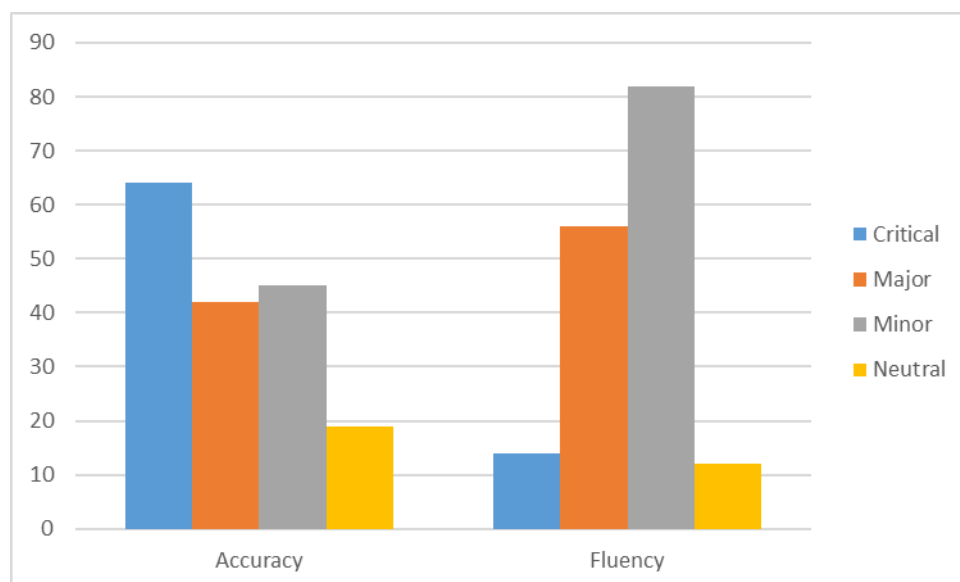
Fig. 1. Severity of Accuracy and Fluency errors

This means that, despite the small number of critical Fluency errors, the great number of the other major and minor ones and their diversity make the post-editing process cumbersome and the workload considerable. Moreover, translators need to demonstrate strong linguistic competence in

order to deal not only with grammar, but also with ambiguity, cohesion and coherence issues and reformulation. If MT is chosen as a facility in their translation project, they have to concentrate constantly on language as much as on the transferred content. This triggers a refocus on linguistic competence in students' translation training.

References

[1] Weaver, W., "Translation", 1949, https://repositorio.ul.pt/bitstream/10451/10945/2/ulfl155512_tm_2.pdf, last accessed on 7th May 2022.

[2] Holmes, J. S., "The Name and Nature of Translation Studies" 1972, in *The Translation Studies Reader*, Venuti L. (ed.), London/New York, Routledge, 2000, pp. 172-185.

[3] Harris, B., and B. Sherwood, "Translating as an innate skill", in *Language, Interpretation and Communication*, Gerver D. and H. W. Sinaiko (eds), New York, Plenum, 1978, pp. 155-170.

[4] Bell, R. T., *Translation and Translating: Theory and Practice*, London/New York, Longman, 1991, pp. 38-42.

[5] Nord, C., *Translating as a Purposeful Activity: Functionalist Approaches Explained*, Manchester, St. Jerome Pub., 1997.

[6] Schäffner, C., and B. Adab, "Developing Translation Competence: Introduction", in *Developing Translation Competence*, C. Schäffner, and B. Adab (eds), Amsterdam/Philadelphia, John Benjamins Pub., 2000, pp. ix-x.

[7] Kiraly, D. C., *A Social Constructivist Approach to Translator Education: Empowerment from Theory to Practice / Don Kiraly*, Manchester, UK/Northampton MA, St. Jerome Pub., 2000, p. 3.

[8] PACTE group, "Building a Translation Competence Model", in *Triangulating Translation: Perspectives in Process Oriented Research*, F. Alves (ed.), Amsterdam/Philadelphia, John Benjamins Pub., 2003, pp. 43-66.

[9] Göpferich, S., "Towards a Model of Translation Competence and Its Acquisition: the Longitudinal Study *TransComp*", in *Behind the Mind: Methods, Models and Results in Translation Process Research*, S. Göpferich, A. L. Jakobsen, and I. M. Mees (eds), Copenhagen, Samfundslitteratur Press, 2009, pp. 11-37.

[10] "European Master's in Translation Competence Framework 2017", European Commission website, European Master's in Translation (EMT), 22 May 2017, https://ec.europa.eu/info/sites/info/files/emt_competence_fwk_2017_en_web.pdf, last accessed on 10th May 2022.

[11] Bahdanau, D., K. Cho, and Y. Bengio, "Neural Machine Translation by Jointly Learning to Align and Translate", Conference Paper at ICLR, 2015, in *arXiv* – Cornell University, https://arxiv.org/pdf/1409.0473.pdf, last accessed on 11th May 2022.

[12] ELIS 2022, "European Language Industry Survey 2022. Trends, expectations and concerns of the European language industry", pp. 11, 25, https://elis-survey.org/#results, last accessed on 11th May 2022.

[13] Gröhn, A., "Suitability of Neural Machine Translation for Different Types of Texts: A Study on Potential Predictors", 2019, p. 26, https://core.ac.uk/download/pdf/224641759.pdf, last accessed on 11th May 2022.

[14] Lommel, A. et al. (ed.), "Multidimensional Quality Metrics (MQM) Issue Types", German Research Center for Artificial Intelligence (DFKI) and QTLaunchPad, 2015, http://www.qt21.eu/mqm-definition/issues-list-2015-12-30.html, last accessed on 11th May 2022.