DOI: 10.23817/strans.16-8

Received: 27.02.2025
Accepted: 20.06.2025

STUDIA TRANSLATORICA
2025 • vol. 16 143–159
ISSN 2084–3321 e-ISSN 2657-4802

Magdalena Krawiec
University of Rzeszow/ Poland

Translational reality and technical documentation: A case of machine-translated online content in Microsoft Azure

Abstract

Translational reality and technical documentation:
A case of machine-translated online content in Microsoft Azure

Neural Machine Translation (NMT) represents a paradigm that is being integrated within the IT sector, where scalability, consistency, and accessibility are crucial for disseminating vast amounts of technical content to global audiences. While NMT offers rapid production of technical documentation in multiple languages, it raises questions about the quality of machine-translated specialist texts and, in consequence, the quality of specialist knowledge transferred across language communities based on those texts. This paper explores the emergent paradigmatic shift in technical content translation through a comparative analysis of Microsoft Azure's English-language documentation and its machine-translated counterpart in Polish. The paradigmatic shift from a human-mediated to a machine-enhanced approach prompts an exploration into the evolving dynamics of translation as a practice and a reconsideration of the conceptual underpinnings of translation as a multifaceted phenomenon. The study aims to account for the capacities and limitations of NMT in preserving fidelity and source intelligibility in machine-translated outputs and consequently, to provide insights into the role of NMT in technical contexts.

Keywords: Neural Machine Translation, NMT, technical content, Microsoft Azure documentation, specialist language

1. Introduction

In an era marked by technological leaps, Neural Machine Translation (NMT) stands as a transformative force reshaping the very fabric of the concept of

translation. The development of neural technologies has propelled translation engines to new heights and opened up new avenues for research, fundamentally altering the landscape of natural language processing. In view of the emergent need for a rapid production of content in multiple languages with increased efficiency and less exertion (Briva-Iglesias et al. 2023: 61; Sánchez-Gijón et al. 2019; O'Brien 2006), NMT has exhibited promising performance across diverse multilingual translation tasks, in that it fosters more nuanced and interconnected global communication.

Nevertheless, progress in the NMT field may be hindered by the paucity of large-scale specialised parallel data (Faheem et al. 2024; Ranathunga et al. 2021). It remains a fact that the body of research available has been predominantly conducted on high-resource languages and contemporary accounts of the performance of NMT are largely based on the findings formulated in relation to these (Rosa-Sorlozano & Candel-Mora 2025; Terribile 2024; Kübler et al. 2024). This holds particularly true for the bilingual Polish-English specialised corpora which, due to the rich inflectional morphology, extensive use of grammatical cases and relatively free word order in Polish, do not cease to present challenges to translation engines (Jassem & Dwojak 2019). The lack of extensive, high-quality and, above all, specialised parallel corpora for this language pair hinders effective neural training and frequently renders machine-translated outputs inoperative.

The ensuing text presents an account of machine-translated online content in Microsoft Azure documentation and is intended firstly to fill in the obvious lacuna in the uncharted field of linguistic analysis of machine-translated outputs in the Polish-English language pair; secondly, to contribute to a better understanding of the still scarcely researched interface of the specialist language of IT and translation; and thirdly, to outline the gradual yet inexorable advent of a new translational reality in specialist settings. Therefore, the paper seeks to explore the bearings of the postulated paradigmatic shift in technical translation on the fidelity and intelligibility of the target texts and, as a consequence, the quality of specialist knowledge transferred across the Polish-English language pair based on those texts. This shift, marked by a progression from human-mediated processes to machine-enhanced translation outputs, merits further attention through a comparative analysis of Microsoft Azure's English-language documentation and its machine-translated counterparts. Owing to the fact that specialist knowledge is "reconstructed" (Grucza 2008: 82) based on a reading of specialist texts, knowledge transference seems to be the primary purpose of technical documentation, especially in multilingual contexts[1].

1| A distinction needs to be drawn between the terms *specialised* language and *specialist* language (among the plethora of other terms), with the former representing a broader

2. Neural Machine Translation

The concept of using computers for the translation of natural languages dates back to the inception of computing itself (Hirschberg/ Manning 2015; Hutchins 2004). Although it was not until the 20th century that the first concrete proposals were made, a good point of departure for outlining historical timeframes with regard to the evolution of the concept of machine translation might be anchored as far back as the 16th century, and traced through the lens of Descartes' proposal of a universal language in the form of a cipher aimed at establishing inter-lingual equivalencies sharing one code number (Hutchins 1986: 21). This early vision of representing meaning through a universal code anticipates later ideas of mechanical dictionaries, which, as Hutchins (1986: 21) notes, arose from aneed to overcome the barriers of languages, so as to create "rational" and "logical" means of scientific communication.

Writers such as Beck (1657), Becher (1661), Kircher (1663) and Wilkins (1668) concurred in the recognition of genuine differences between languages and continued to make suggestions for mechanical dictionaries on numerical bases. Most certainly, none of these suggestions incorporated the actual construction of translation machines; rather, human translators were construed as simulating machines through the "use of the tools provided in a mechanical fashion" (Hutchins 1986: 22). The actual concept of translating machines was not explicitly proposed until 1933, when George Artsrouni and Petr Smirnov-Troyanskii independently issued patents for such devices (Hutchins 2005). Nevertheless, these attempts did not obtain sufficient recognition, and the idea of translation machines was only brought to general notice following the memorandum of 1949 by Warren Weaver, an American scientist, mathematician, and science administrator, on the use of the then newly-invented digital computers for the purpose of translating documents. Since the small-scale Georgetown experiment in Russian-English translation in the 1950s (Gordin 2015: 213–217) – the first real demonstration of MT on a computer –

concept that implies "specialisation through topic" and "specialisation through special characteristics where the exchange of information takes place" (Cabré 1995: 135), and the latter referring more narrowly to the language used by domain experts within specific professional or disciplinary contexts (Grygiel 2015: 8). Moreover, the adjective *specialist* underscores the human-centred dimension of languages used in professional socio-cultural surroundings; it conceptually links to expertness, expertise and domain-specific knowledge related to work-oriented settings. For Grucza (2013: 108), from an anthropomorphic perspective, specialists represent the primary subject of research in the linguistics of specialist languages, which investigates their idiolects, that is, the language properties and peculiarities that characterise and distinguish these specialists as specialists.

machine translation (MT) has periodically[2] drawn researchers' attention over the past seventy years.

Human translation was naturally believed to be the upper bound of achievable performance, unattainable for computer translation systems (Popel et al. 2020: 1). Nevertheless, throughout the years, machine translation has evolved towards the application of deep-learning neural-based methods (Junczys-Dowmunt et al. 2016). Owing to their ability to capture contextual nuances, neural methods have superseded previous approaches, such as rule-based and statistical methods (Koponen et al. 2019: 63–64). NMT leverages extensive multilingual capacities – rather than predefined linguistic rules or statistical models – to model complex translation patterns. Unlike its predecessors, NMT processes entire source inputs and learns which elements – lexical, syntactic, or semantic – are most relevant at each decoding step, thereby optimising the flow of information into the target-language output (Koehn 2020).

3. Rationale and research questions

The present inquiry into the translational reality within the highly specialist socio-pragmatic setting of IT and the evolving dynamics of end users' experience with machine-translated content in multilingual contexts is anchored in the author's ongoing professional interaction with software developers and her extensive background in IT translation. This practical engagement has revealed a gap in the literature, where numerous studies to date have predominantly focused on translators' perspectives (Chatzikoumi 2020; Läubli et al. 2018), rather than on actual user experience as regards interacting with machine-translated content. The baseline for including translators in research design is frequently governed by their comprehensive understanding of both source and target languages and their ability to identify nuanced linguistic and pragmatic errors that may elude non-linguists. Nevertheless, a recent study conducted by Krawiec (2024) highlighted discrepancies in the interpretation of the text's intended meaning between translators and domain experts. Although translators are well-equipped linguistically, their expertise may not fully encompass the domain-specific knowledge required to interpret and act upon specialised (and specialist) content in line with subject-matter expectations. Accordingly, their ability to assess the comprehensibility of translated content from the standpoint of the target user group may be limited. In light of the foregoing considerations and based on the observations outlined, the core research questions guiding this paper may be formulated as follows:

2| While early enthusiasm was tempered by setbacks such as the critical 1966 ALPAC report, which led to a decline in funding and interest, with the advent of neural models MT has since re-emerged as a promising field (Hutchins 1997).

RQ1: How do domain professionals assess the quality of machine-translated technical texts within a professional IT context?

RQ2: To what extent does machine translation effectively convey domain-specific knowledge, as perceived by domain professionals?

Moreover, the following hypothesis was tested:

When assessing machine-translated texts, domain professionals may demonstrate a tendency to prefer the original source text, especially in cases where the translation does not meet their expectations for intelligibility or fidelity.

4. Corpus

The present investigation relies on close hands-on language analysis of corpora-extracted attestations. The baseline for compiling the data is anchored in Charteris-Black's (2004) conception of corpus analysis, in line with which any corpus-based investigation should strive for a text's *naturality* (Charteris-Black 2004: 31). To that end, the exploration was conducted with the aid of *authentic* and *authorised* specialist texts[3]. For the sake of clarification, within the canvas of this paper, the criterion of *authenticity* is met when a text is produced by specialists in professional socio-pragmatic settings, whereas *authorisation* pertains to specialist texts issued by accredited providers, signifying their official status, credibility and reliability. The rationale behind elaborating on technical concepts with the aid of specialist *go-to* online content, accessed worldwide by domain professionals as their primary source of reference, lies in its technical insightfulness which refers to the depth, accuracy, and specificity of information that such content offers[4].

In the foregoing, Microsoft Azure documentation available online at https://learn.microsoft.com/en-us/ was utilised as a reference base to analyse the dynamics governing the translational reality within the field. A sample of 150 segments originally published in English and machine-translated into Polish was selected and subjected to quantitative and qualitative analysis. The decision

3| Inasmuch as "activity and discourse are always tightly linked" (Gotti 2017: 2), the proposed integration of texts' authenticity and authorisation allows us to frame the discussion around the optics of actual linguistic practices performed within particular professional settings (Krawiec 2022; 2024).

4| It is postulated that a text that is technically insightful reflects up-to-date domain knowledge and industry standards and conveys fundamental specialised concepts as well as addresses nuanced, context-dependent details essential for expert understanding and practical application within the professional setting. Consequently, this type of content serves as a reliable foundation for research and evaluation in specialist communication contexts.

to analyse exactly 150 machine-translated segments balances methodological feasibility with sufficient coverage of diverse error manifestations. A sample of this size allows for the observation of recurring patterns in translation quality while remaining feasible for detailed annotation and commentary. The data were selected through a random sampling procedure to ensure an unbiased and representative subset of source segments and their machine-translated counterparts. It was also acknowledged that domain professionals may experience fatigue or reduced motivation during extensive evaluations, potentially affecting rating consistency and reliability. To mitigate this, the assessment was structured to minimise cognitive load by deliberately limiting the number of segments to 150, including clear instructions and opportunities for breaks. Given the still preliminary nature of this study, these measures aim to ensure reliable and insightful findings.

5. Participants and procedure

In the present contribution, an experimental research design (cf. Krawiec 2024) is pursued with minor adjustments tailored to the needs of the ensuing paper, in that actual texts' end users – interchangeably referred to as domain professionals – rather than linguists or translators, were asked to participate in the assessment and evaluation of machine translations. The research is structured around two strands, quantitative and qualitative. To begin with, Qualtrics XM was selected as an online tool to assess and elaborate on the source and target segments. Below is a presentation of the steps followed:

1. Two[5] domain professionals in the field of IT, hereinafter referred to as "Raters"– each with a degree in computer science and at least ten years of documented hands-on experience in the IT sector, but without formal training in translation or linguistics – were recruited to assess 150 randomly selected source segments in English and their machine-translated counterparts in Polish.
2. For assessment purposes, Lommel's (2018: 12) typology of *error severity* drawn from a broader methodology referred to as the Multidimensional

5| While the number of Raters is clearly limited, it aligns both with the study's preliminary nature and mixed-methods design, in which the qualitative component prioritises in-depth, expert-informed insights, and the quantitative analysis focuses on pattern identification rather than statistical generalisability. The goal was not to produce a broad consensus but to elicit detailed evaluations rooted in authentic professional experience, which is particularly valuable in assessing the functional adequacy and intelligibility of machine-translated specialist texts. The paper seeks to balance empirical observations with contextual interpretation, drawing on domain-specific expertise and direct familiarity with the type of technical content under examination.

Quality Metrics (MQM) framework was applied. By default, Lommel's (2018: 12) concept of severity encompasses four levels, each originally defined in relatively succinct terms, namely *critical* – an error type that by itself renders a translation unfit for purpose; *major* – an error type that makes the intended meaning unclear, in that a recipient is incapable of recovering it from the text, but the error itself is unlikely to cause harm; *minor* – an error type that does not impact usability; *null* – a change that is not an error[6].

3. The Raters were offered a comment box and asked to elaborate on their assessments, with particular emphasis placed on their experience as regards the interaction with both texts.
4. Raters' scores were calculated and presented in the form of a contingency table.
5. Inter-rater reliability was calculated to quantify the level of agreement between the Raters. To that end, the weighted Cohen's Kappa κ was used (Cohen 1960; Cohen 1968).
6. Raters' assessments and reflections were juxtaposed and discussed.

6. Results and discussion

To assess inter-rater agreement in error severity classification, the following contingency table (Table 1, s. 150) was constructed based on the judgments of two independent Raters[7].

Weighted Cohen's Kappa was selected to quantify the agreement between two ordinally scaled samples, yielding a score of approximately 0.975, which indicated an almost perfect level of agreement[8]. So as to arrive at an even fuller

6| The preference for Lommel's (2018) typology of error severity was intentional. The decision was guided by the fact that Lommel's framework is specifically tailored to machine translation quality assessment. Unlike typologies such as those proposed by Piotrowska (2007) or Hejwowski (2004) – designed primarily for translator training or the analysis of human translation processes – Lommel's model offers a systematic and scalable approach suited to the evaluation of MT output. Its focus on error severity rather than error type makes it particularly appropriate for cross-system comparisons and for capturing the functional impact of errors in automated translation.

7| The table organizes 150 machine-translated segments into four ordinal error categories, i.e. critical, major, minor, and null, mapping out the severity of translation errors. Each cell in the table represents the frequency of segments assigned a particular error severity level by Rater 1 and Rater 2, offering insight into how consistent they were in evaluating translation quality.

8| It needs to be stated that weighted Cohen's Kappa may follow two directions, i.e. linear and quadratic. In this study, I settled upon the latter, where the penalty increases quadratically with the distance between the categories, meaning that it more heavily penalizes larger disagreements, assuming that bigger differences in categories should matter more.

Table 1: The comparison of Rater assessments for 150 machine-translated segments across four error severity levels

	Critical (Rater 2)	Major (Rater 2)	Minor (Rater 2)	Null (Rater 2)
Critical (Rater 1)	38	1	0	0
Major (Rater 1)	2	19	0	0
Minor (Rater 1)	0	0	29	5
Null (Rater 1)	0	0	3	53
Column totals	40	20	32	58

picture of rater consistency, the percentage agreement for each error category was calculated, with 95.0% for *critical errors*, 95.0% for *major errors*, 90.6% for *minor errors* and 91.04% for *null errors*. These findings show high agreement across all error categories, suggesting that the Raters were generally consistent in their evaluations. Slightly lower agreement for null errors (91.4%) and minor errors (90.6%) may indicate a degree of subjectivity in distinguishing these error types from each other. Consider the following segment that was marked as *critical* by both Raters (26.0% of all cases – Rater 1; 26.6% of all cases – Rater 2):

> [ENG] Each dependency will adhere to Intune Win32 app retry logic (try to install three times after waiting for five minutes) and the global reevaluation schedule.
>
> [PL] Każda zależność będzie zgodna Intune logiką ponawiania prób aplikacji Win32 (spróbuj zainstalować trzy razy po odczekaniu pięciu minut) i globalnym harmonogramem ponownej ewaluacji.
>
> [Back translation[9]: Each dependency will be compliant Intune logic of retrying attempts of application Win32 (try to install three times after waiting five minutes) and global schedule of re-evaluation].

The Raters were unanimous in disapproving of the intelligibility and fidelity of the output, as in their comment boxes, they elaborated on the passage as convoluted (both Raters) and *counterproductive* (Rater 1). What rendered the segment inoperative for the recipient are several – seemingly minor – linguistic inaccuracies and stylistic inconsistencies which, while combined, jointly affected the final reading of the target segment and the end users' learning experience. Linguistically, the awkward phrasing of *będzie zgodna Intune logiką ponawiania prób aplikacji Win32* [*will be compliant Intune logic of retrying*

9| Back translations are not intended as corrected or natural renderings but as literal reflections of the flawed translations, aiming to illustrate errors as faithfully as possible.

attempts of application Win32] lies, first, in the incomplete translation of the verb *to adhere to* as *zgodna* instead of *zgodna z* and second, in the impersonal construction of agency in the source text, i.e. attributing an inanimate agent in the subject position to the verb *to adhere to* which in non-specialist settings is more likely to be preceded by a human agent. Arguably, also translating *to adhere to* as *zgodny* does not capture the core meaning of the source verb, inasmuch as the adjective *zgodny* implies alignment with the specified process or mechanism rather than adherence to a set process. Interestingly, this aspect was also highlighted in the comments section, where the following translation solution – deemed workable by the Raters – was set forth: "każda zależność działa w oparciu o…" [*Each dependency adheres to…*].

For the sake of clarification, although *zgodny z* may in fact be suitable in certain contexts and deemed a workable translation solution for the verb *to adhere to*, a statement may be ventured that owing to the lack of the nuanced understanding required to discern when to apply each expression appropriately, NMT engines happen to produce translations that may fail to capture the intended meaning or contextual appropriateness. What adds to the confusion in the target segment, and decreases its technical accuracy, is the imprecise rendition of *retry logic* as *ponawiania prób aplikacji Win32 app* [*of retrying attempts of application Win32*] rather than *ponawiania prób instalacji* [*of retrying installation attempts*] as well as the use of imperative rather than declarative mood in the translation of *try to install three times after waiting for five minutes.* The mismatch of moods in the machine-translated output wrongly prompts the recipient to pursue a specific course of action, whereas an accurate rendering would employ a descriptive and explanatory phrase such as *podejmuje trzy próby instalacji* [*tries/will try to instal three times*] to align with the original text intent. The segment was commented upon as *low-quality* (Rater 1), *unreadable* (Rater 2), and cognitively taxing to process (both Raters). The Raters reported the need to consult the source text in order to reconstruct the segment's original purpose with minimal cognitive effort, mitigate the risk of misinterpretation, and reduce the likelihood of software misconfiguration.

With regard to the next category of errors, the following segment was classified as exhibiting *major severity* (14.0% of all cases – Rater 1; 13.3% of all cases – Rater 2):

[ENG] Install behavior

[PL] Zachowanie instalacji

[Back translation: behaviour of installation; preservation (keeping) of installation]

While *behavior* does translate to *zachowanie*, in technical settings the noun fails to convey the source idea. Settling upon a more natural rendering is

contingent upon the context, in that the noun *behavior* may refer to the mode or method of installation, the way the installation is performed, system actions during the installation process or user-selectable installation settings. Inasmuch as the source segment revolves around the way installation is carried out, the most probable translations involve *tryb instalacji* [*mode of installation*] and *sposób instalacji* [*installation method*]. Interestingly enough, in the comment boxes, the Raters presented a substantial degree of uncertainty as far as what the exact wording should be, nevertheless they demonstrated comprehensive and, above all, actionable understanding of the source-text phrase. By building upon such observations, the source-text noun *behaviour* does not cause inferential problems owing to its contextual fit. The noun in English is standardised in technical contexts and forms part of strong collocations, e.g. "system behavior" or "application behavior", that constitute an essential component of the repertoire of technical terms, therefore it immediately aligns with the expected understanding of how a system functions. *Zachowanie*, on the other hand, introduces a conceptual misalignment in that it brings a more human-centred understanding that does not fit neatly into a technical framework. For the Raters, a contextless phrase *zachowanie installacji* lacks immediate conceptual clarity and fails to unambiguously signal a technical concept like installation settings or expected actions during software deployment. The very first association the Raters made for *zachowanie* was that of *preservation*, which prompted them to pause to infer what *zachowanie* – while combined with *instalacji* – is meant to denote. Therefore, even while placed in the vicinity of a deeply ingrained concept of installation, it failed to operate as a readily available conceptual shortcut to access the way the process is carried out.

The next category of errors encompasses those referred to as *minor* (22.6% of all cases – Rater 1; 21.3% of all cases – Rater 2):

> [ENG] Within that folder, create a PowerShell script file called Install.ps1 and add the following content, replacing <RemoteDesktop> with the filename of the .msi file you downloaded.
>
> [PL] W tym folderze utwórz plik skryptu programu PowerShell o nazwie Install.ps1 i dodaj następującą zawartość, zastępując <RemoteDesktop> ciąg nazwą pobranego .msi pliku.
>
> [Back translation: In this folder create a script file of the PowerShell program named Install.ps1 and add the following content, replacing <RemoteDesktop> string with the name of the downloaded file .msi.]

The Raters concurred in the recognition that the machine-translated text is functional for its purpose, and rather unaffected by stylistic, grammatical or spelling errors. Despite the overall accuracy, the addition of *ciąg* [*string*, *sequence*] was reported as unnecessary, as it is not commonly used in this

context for a filename in Polish technical language. Owing to the minor potential for confusion, the addition of *ciąg* does not detract from the clarity to a great degree but is considered unnatural. Linguistically, the final part of the sentence *nazwą pobranego .msi pliku* [*with the filename of the file .msi*] would benefit from a minor stylistic adjustment to *nazwą pobranego pliku .msi* [*with the filename of the .msi file*].

Relying on the insights from the qualitative data provided in the comment boxes, it seems that the two Raters paid little – if any – heed to stylistic, grammatical or spelling errors in the segments as long as they affected neither its readability nor learnability – and classified such instances as *null errors* (37.3% of all cases – Rater 1; 38.6% of all cases – Rater 2) – which seems like a natural course of action:

> [ENG] If the script exits with a nonzero value, the script fails (…).
>
> [PL] Jeśli skrypt zakończy działanie z wartością inną niżzerowa, skrypt zakończy się niepowodzeniem (…).
>
> [Back translation: If the script ends operation with a value other thanzero, the script will end with failure (…).]

> [ENG] (…) command in a PowerShell script (…).
>
> [PL] (…) polecenia w skry skryptzie programu PowerShell (…).
>
> [Back translation: (…) command in a scrip script of the PowerShell program (…).]

Upon closer scrutiny of the corpus material, it may be observed that some errors arise from inaccurate direct translations – which seems like a recurrent issue – and thereby manifest as syntactic calques. Consider the following target segments where the Polish translation excessively adheres to the English word order, resulting in an unnatural structure in which the machine-translated output replicates the English syntax rather than adapting it to Polish linguistic norms, in particular with regard to the placement of modifiers:

> [ENG] The Intune management extension supports devices that are Microsoft Entra joined, Microsoft Entra registered (…).
>
> [PL] Rozszerzenie do zarządzania Intune obsługuje urządzenia, które są Microsoft Entra przyłączone, Microsoft Entra zarejestrowane (…).
>
> [Back translation: The extension for managing Intune supports devices that are joined Microsoft Entra, registered Microsoft Entra (…).]

The phrases *Microsoft Entra przyłączone* and *Microsoft Entra zarejestrowane* are syntactically odd (both Raters) and *sound nonsensical* (Rater 1). As regards technical Polish, more concise non-finite or participial constructions such as *urządzenia przyłączone lub zarejestrowane w Microsoft Entra* are preferred.

Probing into the qualitative data revealed the Raters' inclination towards settling upon Anglicisms, either phonetically-adapted or not, rather than correcting the outputs and *translating them at a push* (both Raters) – a tendency rooted in both cognitive efficiency and practical necessity. This is owing to Anglicisms becoming an in-group code or a conceptual shorthand that prioritises speed, mutual understanding, interoperability and task execution. In IT, the use of Anglicisms serves to reduce cognitive load, as domain professionals are spared the mental effort of mapping a translated term back onto its English counterpart – a pragmatically suboptimal step. Arguably, such forced translations might be overly broad, therefore add to the need for further clarifications. An example of this tendency is illustrated by the noun *instance*, unanimously rendered by the Raters as *instancja* in the comment boxes:

> [ENG] At least one Windows OS image available on the instance.
>
> [PL] Co najmniej jeden obraz systemu operacyjnego Windows dostępny w wystąpieniu.
>
> [Back translation: At least one image of the Windows operating system available in the instance.]

Another case in point that lends support to this observation is as follows:

> [ENG] If you want to use existing tools and processes, such as automated pipelines, custom scripts, or external partner solutions, you need to use the standard host pool management type.
>
> [PL] Jeśli chcesz użyć istniejących narzędzi i procesów, takich jak zautomatyzowane potoki, niestandardowe skrypty lub rozwiązania partnerów zewnętrznych, musisz użyć standardowego typu zarządzania pulą hostów.
>
> [Back translation: If you want to use existing tools and processes, such as automated streams, nonstandard scripts or solutions of external partners, you must use the standard type of management of the pool of hosts.]

In the machine-translated segment above, the Raters continued to share unanimity and opted for a phonetically-adapted version of the noun *pipelines*, i.e. *pajplajny*. While it does not take much in terms of phonetic adaptation for the term to fit the phonetic system of Polish, in the comment boxes the noun is noted down following a phonetic spelling based on its pronunciation in the Polish language and conjugated where necessary. Such an inclination seems to offer a shortcut to ensure that terms are quickly understood by their fellow colleagues and is perhaps anchored in the need to maintain the concept's clarity and transparency without losing its essence.

This holds particularly true for cases when a concept is already so deeply ingrained in the target audience's understanding through the term that is

attached to it, that translation might either distort it or cause a lack of recognition. Interestingly, this long-established recognition is, in turn, grounded in the paucity of translation ideas right at the very emergence of the concept, so the term did need to eventually become localised. Since the name in Polish is tied closely to the original pronunciation, mapping the term to the associated concept is effortless and ensures that the conceptual link between the two remains intact.

7. Conclusions

In response to RQ1, the data showed an almost perfect (κ=0.975) inter-rater agreement in error severity classification, which proves that the actual consumers of technical content largely concurred in their assessments of the quality of machine-translated outputs. The Raters achieved a high percentage agreement across all error categories (95.0% – *critical* and *major* errors, 90.6% – *minor* errors, 91.4% – *null* errors), supporting the reliability and representativeness of the findings. The answer to RQ2 follows partially from the numerical insights gained in RQ1 as well as from the qualitative data provided by the Raters. Probing into the ratings revealed that machine-translated content lacked acceptance in 40% of the cases subjected to analysis (critical and major errors), as it failed to enable its end users to reconstruct specialist knowledge. The Raters concurred in the recognition that what primarily contributed to flagging a segment as *critical* or *major* was a high density of mistranslations, frequently combined with stylistic or grammatical inconsistencies of varied degrees. Since technical documentation is purpose-driven by nature, in that it allows its end users to act upon the content provided, browsing densely mistranslated content renders the whole process *counterproductive* – to adopt one of the Raters' exact phrasings – to its original purpose, i.e. reconstructing specialist knowledge and acting upon it.

The Raters noted that in the case of critical errors (26.0% – Rater 1, 26.6% – Rater 2) and also frequently in the case of major errors (14.0% – Rater 1; 13.3% – Rater 2), even referring back to their specialist knowledge proved insufficient to deduce the contextual meaning – an observation especially relevant given that acting on mistranslated instructions can lead to failed troubleshooting attempts, ultimately hindering successful task completion in professional settings. Although in slightly less than 60% of the cases the Raters marked the segments as instances of minor (22.6% – Rater 1; 21.3% – Rater 2) or null (37.3% – Rater 1; 38.6% – Rater 2) severity, there remained a reluctance towards the use of machine-translated documentation. The hypothesis that actual texts' end users may demonstrate a tendency to prefer the original source text over its machine-translated output was therefore supported. Apart from

the said cases of critical and major severity, a *precautionary* need to reference the source text was also simultaneously mentioned by the Raters in 25% of the segments with lower error severity, so as to successfully reconstruct the segment's original context which was inadequately rendered or obscured in the machine-translated version. Low-quality content lacking in intelligibility contributes to increased cognitive load anchored in the need to reconcile inconsistencies, infer missing information, or reinterpret misleading content. In technical settings, where clarity and precision are paramount, such disruptions affect the learnability of new concepts and may lead to misconfigurations of software. In this regard, a pivotal finding of the qualitative part of the present scrutiny is that domain professionals lean towards consulting the source text, rather than its machine-translated counterpart, as it helps them proceed with minimal cognitive strain by mitigating the risk of misinterpretation. The preference for the source text over machine-translated outputs may also stem from domain professionals' strong familiarity with English. This might extend to the point where English feels more natural than the native language to account for highly specialised concepts, be it either through the observed incorporation of single loanwords or, more broadly, conducting whole conversations in specialist socio-pragmatic settings.

This study presents several limitations. Firstly, the limited number of raters might affect the generalisability of the findings. Including a larger group of judges could yield more representative quantitative results. Secondly, future research might benefit from incorporating document-level assessments to reflect broader contextual and discourse-related issues. Furthermore, increasing the number of segments assessed could improve the statistical power and facilitate a more nuanced interpretation of translation outcomes.

References

Becher, Johann Joachim (1661). *Character, Pro Notitia Linguarum Universali*. Frankfurt.

Beck, Cave (1657). *The universal character, by which all the nations in the world may understand one another's conceptions*. London.

Briva-Iglesias, Vincent/ O'Brien, Sharon/ Benjamin, R. Cowan (2023). "The impact of traditional and interactive post-editing on Machine Translation User Experience, quality, and productivity". In: *Translation, Cognition & Behavior* 6(1). Pp. 60–86.

Cabré, Teresa Maria (1995). "On diversity and terminology" In: *Terminology* 2(1). Pp. 1–6.

Charteris-Black, Jonathan (2004). *Corpus Approaches to Critical Metaphor Analysis*. London.

Chatzikoumi, Eirini (2020). "How to evaluate machine translation: A review of automated and human metrics". In: *Natural Language Engineering* 26(2). Pp. 1–25.

Cohen, Jacob (1960). "A coefficient of agreement for nominal scales". In: *Educational and Psychological Measurement* 20(1). Pp. 37–46.

Cohen, Jacob (1968). "Weighed kappa: Nominal scale agreement with provision for scaled disagreement or partial credit". In: *Psychological Bulletin* 70(4). Pp. 213–220.

Faheem, Mohamed Atta/ Wassif, Khaled Tawfik/ Bayomi, Hanaa/ Abdou, Sherif Mahdy (2024). "Improving neural machine translation for low resource languages through non-parallel corpora: a case study of Egyptian dialect to modern standard Arabic translation". In: *Scientific Reports* 14(1). Pp. 1–10.

Gordin, Michael (2015). *Scientific Babel: How Science Was Done Before and After Global English*. Chicago.

Gotti, Maurizio (2017). "Interdisciplinary Cooperation in the Analysis of Specialized Discourse: Challenges and Prospects". In: Vargas-Sierra C. (ed.) *Professional and Academic Discourse: an Interdisciplinary Perspective* 2. Pp. 1–13.

Grucza, Sambor (2008). *Lingwistyka języków specjalistycznych. Języki. Kultury. Teksty. Wiedza.* Warszawa.

Grucza, Sambor (2013). *Lingwistyka języków specjalistycznych*. Warszawa.

Grygiel, Marcin (2015). "Business English from a linguistics perspective". In: *English for Specific Purposes – World*. Special Issue 1. Pp.1–12.

Hejwowski, Krzysztof (2004). *Translation: A cognitive-communicative approach.* Olecko.

Hirschberg, Julia/ Manning, Christopher (2015). "Advances in natural language processing". In: *Science* 349. Pp. 261–266.

Hutchins, John (1986). *Machine translation: past, present, future*. Chichester/ New York.

Hutchins, John (1997). "From first conception to first demonstration: the nascent years of machine translation, 1947–1954". In: *Machine Translation* 12. Pp. 195–252.

Hutchins, John. (2004). "Two Precursors of Machine Translation: Artsrouni and Trojanskij". In: *International Journal of Translation* 16(1). Pp. 11–31.

Hutchins, John (2005). *Early years in machine translation: Memoirs and biographies of pioneers*. Amsterdam.

Jassem, Krzysztof/ Dwojak, Tomasz (2019). "Statistical versus neural machine translation a case study for a medium size domain specific bilingual corpus". In: *Poznan Studies in Contemporary Linguistics* 55(2). Pp. 491–515.

Junczys-Dowmunt, Marcin/ Dwojak, Tomasz/ Hoang, Hieu (2016). "Is neural machine translation ready for deployment? A case study on 30 translation

directions". In: *Proceedings of the Ninth International Workshop on Spoken Language Translation (IWSLT)*.

Kircher, Athanasius (1663). *Polygraphia nova.* Rome.

Koehn, Philip (2020). *Neural Machine Translation*. Cambridge.

Koponen, Maarit/ Salmi, Leena/ Nikulin, Markku (2019). „A product and process analysis of post-editor corrections on neural, statistical and rule-based machine translation output". In: *Machine Translation* 33. Pp. 61–90.

Krawiec, Magdalena (2022). *Conceptual Metaphors as an Organisational Framework of the Specialist Language of IT. An Analysis of Cloud Computing Terminology.* Göttingen.

Krawiec, Magdalena (2024). "Zooming in on Neural Machine Translation in the specialist language of IT: A case of Microsoft Azure documentation". In: Migodzińska, M./ Pietrzak, A. (eds.). *Fachsprachen – Fachkommunikation – fachdidaktische Diskurse.* Göttingen. Pp. 77–87.

Kübler, Natalie/ Martikainen, Hanna/ Mestivier, Alexandra/ Pecman, Mojca (2024). "The role of corpora in the translation of phraseological structures". In: *Recent Advances in Multiword Units in Machine Translation and Translation Technology.* Pp. 57–78.

Läubli, Samuel/ Rico, Sennrich/ Martin, Volk (2018). "Has Machine Translation Achieved Human Parity? A Case for Document-Level Evaluation" In: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. Pp. 4791–4796.

Lommel, Arle (2018). "Metrics for Translation Quality Assessment: A Case for Standardising Error Typologies". In: Moorkens, J./ Castilho, S./ Gaspari, F./ Doherty, S. (eds.) *Translation Quality Assessment From Principles to Practice*. Cham. Pp. 109–127.

O'Brien, Sharon (2006). "Pauses as Indicators of Cognitive Effort in Post-Editing Machine Translation Output". In: *Across Languages and Cultures* 7(1). Pp. 1–21.

Piotrowska, Maria (2007). *Proces decyzyjny tłumacza. Podstawy metodologii nauczania przekładu pisemnego*. Kraków.

Popel, Martin/ Tomkova, Marketa/ Tomek, Jakub/ Kaiser, Łukasz/ Uszkoreit, Jakob/ Bojar, Ondřej/ Žabokrtský, Zdeněk (2020). "Transforming machine translation: a deep learning system reaches news translation quality comparable to human professionals". In: *Nature Communications* 11. Pp. 1–15.

Ranathunga, Surangika/ Lee, Annie En-Shiun/ Skenduli, Marjana Prifti/ Shekhar, Ravi (2021). "Neural Machine Translation for Low-resource Languages: A Survey". In: *ACM Computing Surveys* 55. Pp. 1–37.

Rosa-Sorlozano, Carmen/ Candel-Mora, Miguel Ángel (2025). "Machine translation of tourism reviews". In: *Translation and Translanguaging in Multilingual Contexts* 11(1). Pp. 48–64.

Sánchez-Gijón, Pilar/ Moorkens, Joss/ Way, Andy (2019). "Post-Editing Neural Machine Translation versus Translation Memory Segments". In: *Machine Translation* 33(1–2). Pp. 31–59.
Terribile, Silvia (2024). "Is post-editing really faster than human translation?" In: *Translation Spaces* 13(2). Pp. 171–199.
Wilkins, John (1668). *An Essay Towards a Real Character, and a Philosophical Language*. London.

Magdalena Krawiec
Uniwersytet Rzeszowski
Instytut Lingwistyki Stosowanej
al. Tadeusza Rejtana 16C
35-310 Rzeszów
Poland
mkrawiec@ur.edu.pl
ORCID: 0000-0001-9541-2515