УДК 811.111’322.4
DOI https://doi.org/10.32782/tps2663-4880/2025.43.1.32

SIMPLIFIED TECHNICAL ENGLISH: AN OVERVIEW OF THE CURRENT STANDARD AND ITS RELEVANCE FOR MACHINE AND AUTOMATED TRANSLATION

СПРОЩЕНА ТЕХНІЧНА АНГЛІЙСЬКА: ОГЛЯД СУЧАСНИХ СТАНДАРТІВ ТА ЇХНЄ ЗНАЧЕННЯ ДЛЯ МАШИННОГО ТА АВТОМАТИЗОВАНОГО ПЕРЕКЛАДУ

Karpina O.O.,
orcid.org/0000-0001-9520-074X
PhD in Philology,
Associate Professor at the Applied Linguistics Department
Lesya Ukrainka Volyn National University

The 21st century has witnessed a remarkable transformation of machine translation (MT), evolving from rudimentary devices capable of word-for-word substitution to cutting-age neural architectures adopting deep learning approaches to deliver fluent and contextually appropriate performance. Recent studies on neural machine translation (NMT) demonstrate promising results, with machine-generated output not merely approaching, but in some cases even surpassing human translation in terms of accuracy and idiomaticity. Yet, persistent challenges, such as lexical ambiguity, syntactic complexity, terminology inconsistency continue to affect system output, prompting human experts – researchers and professional translators – to explore alternative solutions. One of the solutions often referred to is the application of controlled languages in highly-regulated industries, where accuracy and clarity is directly linked to safety and operational efficiency. Simplified Technical English, formally known as ASD-STE100, has become an influential international standard designed to make technical writing clear, efficient and reliable. Initially developed upon the request of the aerospace industries, STE has gradually expanded the sphere of its application from maintenance documentation to fields such as language services, translating, and interpreting.

In this paper, we provide a comprehensive overview of the 9th issue of ASD-STE100, which reflects ongoing technological and terminological advancements, and examine how its implementation can enhance the performance of MT systems. We focus on STE as a pre-editing strategy, which significantly reduces interpretative variability by establishing one-to-one correspondences between words and meanings, simplifying syntactic structures, limiting stylistic variations, making input texts more MT-friendly.

The paper begins with the historical overview of the concept of controlled languages, their application in MT and the evolution of STE specification into an international standard for technical writing. Then, we examine the structure of the current issue analysing the content of each section and illustrating the key points with the examples from the standard. Next, we provide a brief outline of the updates introduced in the current issue. Finally, we identify the primary challenges faced by current MT systems and put forward the proposals for minimising errors by drawing on the standard guidelines.

The perspectives for future research may focus on empirical evaluation of English-Ukrainian MT output from STE compliant and non-compliant input texts, with potential application in the Ukrainian localisation industry.

Key words: STE, controlled language, technical writing, machine translation quality, approved vocabulary, pre-editing, ambiguity.

21 століття стало свідком суттєвих трансформацій у галузі машинного перекладу (МП), який поступово еволюціонував від примітивних пристроїв, здатних здійснювати послівну заміну, до найсучасніших нейронних архітектур, що використовують підходи глибинного навчання для забезпечення якісного, контекстуально відповідного перекладу. Останні дослідження в галузі нейронного машинного перекладу (НМП) свідчать про обнадійливі результати: МП не лише не поступається професійному людському перекладу, а й в окремих випадках навіть перевершує його за точністю та ідіоматичністю. Проте, непереборні виклики, як-от лексична неоднозначність, синтаксична складність, термінологічна непослідовність, продовжують впливати на результати роботи систем, спонукаючи експертів – дослідників і професійних перекладачів – до пошуку альтернативних рішень. Одним із рішень, до якого почасти звертаються, є застосування контрольованих мов у галузях із високим рівнем регулювання, де точність і чіткість безпосередньо пов’язані з безпекою та операційною ефективністю. Спрощена технічна англійська (STE), офіційно відома як ASD-STE100, стала впливовим міжнародним стандартом, покликаним зробити технічні тексти чіткими, ефективними та надійними. Первинно розроблена на замовлення аерокосмічної галузі, спрощена технічна англійська поступово розширила сферу свого застосування від документації з технічного обслуговування до галузей, як-от мовні послуги, усний та письмовий переклад.

У поточній статті запропоновано вичерпний огляд 9-го випуску ASD-STE100, у якому відображено поточні технологічні та термінологічні нововведення, а також проаналізовано, у який спосіб його впровадження здатне покращити ефективність систем МП. Ми розглядаємо STE як стратегію попереднього редагування, яка суттєво знижує варіативність тлумачень через встановлення прямих відповідників між словами та значеннями, спрощення синтаксичних структур, обмеження стилістичних варіацій, що робить вхідні тексти більш придатними для МП.

Стаття розпочинається з історичного огляду поняття контрольованих мов, їх застосування в МП та еволюції STE від специфікації до міжнародного стандарту з технічного письма. У подальшому викладі розглянуто структуру

поточного випуску, проаналізовано зміст розділів й наведено приклади ключових моментів зі стандарту. Далі представлено короткий огляд оновлень, наявних у поточному випуску. Насамкінець, окреслено першочергові виклики, з якими стикаються сучасні системи МТ, і запропоновано способи мінімізації помилок, відповідно до рекомендацій зі стандарту.

Перспективу майбутніх досліджень може бути спрямовано на емпіричне оцінювання результатів англійсько-українського МП із вхідних текстів, що відповідають та не відповідають стандартам STE, з подальшим застосуванням у галузі української локалізації.

Ключові слова: STE, контрольована мова, технічне письмо, якість машинного перекладу, затверджена лексика, попереднє редагування, неоднозначність.

"Simplicity is the ultimate sophistication"
Leonardo da Vinci[1]

Machine translation (MT) has become an indispensable instrument for global communication and sharing knowledge in various spheres of human life. It has remarkably evolved from rule-based and statistical engine designs to sophisticated neural machine translation (NMT) systems powering general-purpose translation services, such as Google Translate and DeepL, as well as large language models, which show a strong potential in translation practices due to its context awareness and idiomaticity. With the advent of NMT, the domain turned a new spiral, implementing deep learning architecture that allows to model linguistic patterns with unprecedented fluency [7]. Recent research showcases that current MT technologies are utilized even in literary translation, once seen as an exclusive domain of human professionals [1]. Their excellent learning capacity, background knowledge, cultural and context sensitivity make the systems compatible with human expertise and creativity in completing the most challenging translation tasks. Despite these undeniable achievements, the effectiveness of modern translation technologies is still challenged by various types of ambiguity, consistency issues, syntactic complexity, which continue to compromise translation quality in domains requiring special clarity and precision such as technical documentation [5]. To address these issues, the concept of controlled languages has been introduced and actively exploited. It is regarded as an alternative solution, offering a standardized vocabulary and simplified grammar to mitigate the abovementioned issues and avoid misinterpretation. One of the most widely adopted forms of controlled languages in technical writing is Simplified Technical English (STE), an international standard developed to ensure clarity in technical writing.

The **goal** of this paper is to provide a comprehensive overview of the latest edition of Simplified Technical English, a standard for technical documentation, issued in January, 2025 and to examine its practical value for MT.

The effectiveness of a controlled language as a pre-editing strategy has been shown in a number of recent studies [3; 4; 8; 9; 10]. In particular, empirical research demonstrates a positive impact of the application of controlled languages across various types of MT architecture [8]. Computational analysis of STE examines its relevance in automatic parsing, with potential benefits for MT network design [4]. However, despite the existing research on the usefulness of STE and its impact on MT output, few studies address the content of the latest issue of STE. Offering a clear and comprehensive overview of the manual is beneficial not only with the reference to MT sphere in general, it can also be used in educational purposes. Specifically, a systematic explanation of the mechanisms of STE, supported by the examples from the manual, can introduce students of translation studies to the importance of human-machine collaboration operating not only at the post-editing level, but at the pre-editing level too.

The importance of pre-editing, was recognized prior to the practical implementation of MT. Erwin Reifler, an Austrian linguist and one of the pioneers of MT research, highlighted the importance of the pre-editing stage of translation, arguing that raising the explicitness of the input text by reducing its ambiguity and making it prepared for mechanical processing could significantly improve MT output [6]. The concept of pre-editing was not only associated with the elimination of morphological and syntactical structures viewed as potentially challenging for MT, but also with the idea of adapting the input text in accordance with the specifics of the target language. In this sense, the pre-editor was expected to predict possible translation by considering lexical and structural peculiarities of the target language, and adapt the source text with the available linguistic means. Although these eccentric proposals were met with scepticism, some of Reifler's ideas paved way to improving MT quality. Among them is the concept of "regularised language", designed for MT processing. Stuart Dodd suggested employing a simplified form of English, defined as the "standardization of English syntax as a means of simplifying the use of English either as a source language or as a target language" [6].

[1] This quote serves as a caption to the image of the Vitruvian Man in the Introduction section of STE.

Similar attempts to improve MT quality by means of a controlled language design were undertaken by the developers of KANT system, their research targeting multilingual translations of heavy equipment documentation. Among their proposals were assigning a single meaning to each word/part of speech pair; restricting the use of pronouns and participial forms, introducing strict rules for acronym use and spelling etc [9].

These early proposals, sometimes extreme and radical, were the precursors to the concept of controlled languages, one of the most prominent of which is officially known as ASD Simplified Technical English (STE), developed to ensure clarity of technical documentation. Although English is a global language, functioning as *lingua franca* across various domains, its complex grammar and semantic variability can challenge comprehension for non-native speakers. Misinterpretation is especially dangerous in technical translation, where precision is directly linked to human safety. Regulatory documents functioning across technical domains emphasize the importance of clarity, removing potential risks and hazards [3]. As a prerequisite of global international trade, technical translation supports technological progress via dissemination of technical knowledge. For instance, according to EU regulations, technical documentation must be localised, with all products including instruction manuals in the language of the target market. This requirement creates a steady demand for professional technical translation services. But in practice, technical texts are often pre-processed by machine translation technologies, with various types of ambiguity remaining the primary obstacle to adequate translation.

STE was initially developed to disambiguate the language of aerospace documentation [2]. Following the request of the Association of European Airlines (AEA) in the late 1970s, the European Association of Aerospace Industries (AECMA) established Simplified English Working Group, which released the first Simplified English guide. Rather than focusing on stylistic aspects, the researchers developed detailed guidelines for writing technical texts in a clear, simple and unambiguous manner to ensure global comprehensibility. The success of this project soon led to its application in other spheres, including professional translation and interpreting. One of the primary objectives of STE creation was to facilitate the process of translation of technical documentation through restricted vocabulary and controlled syntactic structures, which also makes texts more manageable for MT tools, reducing potential errors and saving post-editing effort.

STE is structured in two principal parts, covering a set writing rules and controlled vocabulary. The section with writing rules includes over 60 rules that regulate word choice, punctuation, style and syntax. Specifically, these rules govern such aspects as lexical control, according to which words must be used exclusively in an approved meaning as an approved part of speech, with a clear distinction between technical nouns and technical verbs. For instance, CLOSE (v) is allowed only in the meaning of shutting a physical object ("Close the instrument panel"), not in other contexts, such as "to stop operating" etc. Similarly, RIGHT (adj) is allowed in the meaning of the "east side when you look north", excluding the evaluative component of "being correct" from its semantics. By reducing polysemy, it helps eliminate one of the most persistent challenges of MT – lexical ambiguity. Technical nouns cannot be used as technical verbs and vice versa. For instance, the sentence "Oil the steel surfaces" illustrates an incorrect use of the term "oil", belonging to the category of technical nouns. The STE-compliant variant should instead be "apply oil", which rephrases the sentence attributing the term to the approved morphological category. Additionally, authors are not allowed to use different technical terms to refer to the same concept. These rules greatly enhance consistency of terminology in translation, – an aspect, in which MT usually struggles due to the variety of terms found in the original.

Restrictions are reinforced on noun clusters consisting of more than three words. Thus, when a long technical term comes from an official document or technical drawing, it must be written in full provided with an explanatory sentence, and then in the remaining text used in a shortened form. For example, the term "ramp service door safety connector pin" must include the explanation – "the pin that holds the ramp service door, referred to in this procedure as "safety connector pin."

Apart from noun usage, the standard also provides clear guidance on the verb application. Specifically, it recommends the use of simple tense forms, discouraging complex syntactic constructions with auxiliary verbs. The active voice is generally preferred to the passive one, however, the passive voice may be used in descriptive writing, when the agent is unknown. Past participles and *-ing* forms are permissible under specific conditions: the past participle may function as an adjective indicating the condition of something (e. g. "disassembled unit", "damaged surfaces"), while *-ing* forms are acceptable only if they appear as approved technical nouns or parts of such nouns (e. g. "testing", "cleaning", "grinding wheel" etc.). These grammatical restrictions directly benefit MT

output, as the standardized constructions make the input text easier to parse and process. The consistent use of certain terms as fixed morphological categories helps MT systems differentiate between adjectives, verbs, and nouns. The use of the active voice eliminates ambiguity related to identifying the agent, allowing translation engines to generate more accurate output.

Another essential aspect of STE standard concerns sentence structure. STE requires sentences to be short and simple, giving accurate instructions or covering one topic per sentence. The standard encourages the use of connecting words and phrases to connect the idea in one sentence with the information in the sentence that follows. To achieve consistency and coherence, the standard recommends only those connecting words that are listed in the vocabulary of approved words, such as "and", "but", "thus", "as a rule" etc. Demonstrative pronouns, such "this" or "these", are also attributed to the category of connecting words. According to the STE standard, only one idea can be expressed within one sentence, therefore, connecting words and phrases are used at sentence boundary, linking related ideas and contributing to overall textual coherence.

Although the standard promotes simplicity and compression, it warns against the use of contractions or omission of essential words. In this regard, the sentence "Can be a maximum five inches long" is considered ambiguous and should be rephrased in accordance with the standard "Cracks can have a maximum length of five inches." Such structures are not only comprehensible for human readers, but also have direct implications for MT. It is widely recognised that the absence of explicit subjects often leads to erroneous translations due to difficulties in syntactic parsing and semantic alignment. By reinforcing complete sentence structures, STE minimizes the likelihood of parsing and semantic errors, contributing to higher quality translation.

In ASD-STE100, a clear distinction is made between procedural and descriptive writing, each serving a distinct communicative purpose and provided with different writing guidance. Procedural writing deals with giving instructions, which should be devised in imperative mood, using action verbs, with the sentence length not exceeding 20 words, e. g "Put preservation oil into the unit through the vent hole" (10 words). Descriptive writing, on the other hand, contains information about an item, a product, or a system, how it is made or how it operates. Such type of writing typically employs declarative mood, with a single subject per sentence and the information structured into paragraphs, each focused on a single topic. For MT, such differentiation helps MT engines select appropriate translation strategies, i. e. translating "Press the button" as a command, not a description.

The overall comprehension, as well as translation accuracy is enhanced through punctuation guidelines, provided in ASD-STE100. It is stated that correct punctuation is essential for establishing connections between parts of a sentence and restricting ambiguity issues. To guide writers through punctuation features, the standard refers to authoritative reference books, such as "The Chicago Manual of Style" or "US Government Publishing Office Style Manual" etc. While all standard English punctuation is generally permitted, the use of semicolon is not prohibited as it is typically associated with long and complex sentences, which causes potential parsing errors for MT.

The second part of the STE standard focuses on the dictionary of controlled words and their specific meanings, which comprises 875 words most frequently used in technical writing, along with the examples of their usage in technical contexts. The dictionary adopts American English spelling as defined in Meriam Webster's dictionary. It also contains the examples of words that are not approved in the standard, with their possible STE-compliant substitutes. For example, the word "main" is not approved in STE and should be replaced by the approved word "primary". Each approved word is provided with an appropriate meaning, which does not allow any alternatives unless they are provided in the dictionary. Thus, the noun "oil" has an approved meaning of the substance used for lubrication; using it with an alternative meaning or as a verb requires a different term or construction, e. g. "put oil" or "lubricate" instead of "oil" used as a verb. Similarly, the preposition "about" is approved only in the sense of "concerned with". When the intended meaning is "approximately", approved alternative terms "around" or "approximately" should be adopted. The dictionary of approved words holds promising prospects for the users of computer-assisted translation (CAT) tools. Translators can build domain-specific glossaries to be integrated into their translation projects, which will ensure terminological consistency across projects and domains. Such integration is particularly beneficial for large-scale team projects, where maintaining consistency is challenging due to a large number of participants. By aligning terminology databases with STE guidelines, translators and project managers can develop clear standardized glossaries created in accordance with the principles of lexical control and semantic precision.

As a "living" controlled language, STE has undergone updates and amendments, to match evolving needs. The ninth issue of the STE standard reflects the evolution of the sphere of technology, leading to the expansion of the scope of vocabulary beyond the language of aerospace industry. The manual changed its status from the international specification to the international standard: STE is now framed as "standard for technical documentation", instead of the defining noun "specification" in its subtitle. This issue explicitly acknowledges its applicability in IT, transport, energy, environmental sciences, and legal writing. The addition of such categories as "Law", "Animals, plants and life forms" as well as category adaptations including a wider range of vocabulary (for example the category "Information technology and telephony terms" was reshaped into "Computer science, information and communication technology") confirms its widening status. The expansion is consistent with the growing use of the standard in academic world, language services, including interpreting and MT, which is explicitly stated in the introduction to the current issue.

The updates introduced in the revised version of STE concern both sections, i. e. the set of writing rules and the dictionary of approved words. As for the first section, specifying the way of writing technical documentation, the updates refer to language and formatting: several rules were rewritten with a clearer, more explicit wording to enhance readability, and more examples covering new technical contexts were provided to illustrate STE and non-STE compliant structures.

Another important amendment refers to restricted use of phrasal verbs and pronouns, especially pronouns *it* and *they,* in cases when their reference is ambiguous. Technical writers are also instructed to replace phrasal verbs with regular ones, e. g. to replace "give off" with "release" or "put out" with "extinguish" to avoid non-approved meanings and contribute to lexical consistency. It should be noted, that the use of articles and demonstrative adjectives "this" and "these" is encouraged in order to enhance correct interpretation and highlight the syntactic role of nouns in the sentence. For example, in the sentence "Turn ***the*** shaft assembly" adding "the" indicates that "shaft assembly" is a single unit.

The changes in vocabulary section are connected with the dynamics of technological and industrial sphere: new entries were included, covering such industries as electronics, IT, and defense, laws and regulations, biology, while obsolete terms were removed from the list. In some cases, the words changed their status from approved to non-approved, where ambiguity was identified. In other cases, the terms expanded their meaning, provided with additional approved alternatives. For example, the vocabulary entry "consecutively" was clarified through an alternative meaning "sequence" and illustrated with the following examples: "do these steps one after another" and "do these steps in sequence." The addition of such approved alternatives as "task", "annotation", "communication", "contact" and the like marks a deliberate broadening of the lexical and semantic scope of STE aligning it with broader domain contexts, where these concepts are frequently employed.

Conclusions and prospects for further research. Simplified Technical English is the international standard used to ensure clarity and conciseness in technical documentation. Initially created to address misinterpretation in the language of aerospace documentation, it has later evolved into a cross-domain standard expanding its applicability onto such fields as information technology and communication, legal writing, environmental sciences, engineering, as well as professional translation and interpreting. Simplified Technical English standard is structured into two parts, consisting of a set of 65 writing rules, which regulate lexical, morphological, syntactic, and stylistic aspects, and a dictionary of 865 approved words, each assigned a particular meaning and part of speech.

The use of Simplified Technical English offers clear benefits for machine and automated translation. Functioning as a pre-editing stage, it can significantly reduce ambiguity at all linguistic levels, which is considered to be one of the main causes of MT errors. By adopting the principles of lexical control, simplification of grammar and stylistic regularity, the STE standard can improve the input text, making it easier for MT systems to process and parse. The adoption of the STE standard can increase the productivity of computer-assisted translation (CAT) tool users, providing translators and project managers with ready-made terminology databases that can be developed into project glossaries, maintaining terminological consistency across projects.

The perspective for the future research can focus on the empirical evaluation of English-Ukrainian MT output produced from STE-compliant and non-compliant input texts, where systemic examination is still lacking. Furthermore, pre-editing translation exercises can be integrated into regular practices of translation students, raising awareness of the importance of linguistic clarity and terminology consistency. Comparing MT performance before and after pre-editing according to the STE guidelines can demonstrate how language control has a direct impact on translation quality. Such educational approach aligns

with the evolving demands of the translation industry, helping to train specialists who can easily adapt to rapid changes, critically evaluate new tools and make informed translation decisions.

REFERENCES:

1. A 2-step Framework for Automated Literary Translation Evaluation: Its Promises and Pitfalls / Sheikh S. et al. 2024. ArXiv:2412.01340v1 : website. DOI: https://doi.org/10.48550/arXiv.2412.01340

2. ASD-STE 100 Simplified Technical English. Standard for technical documentation, Issue 9, January 2025. Brussels : Aerospace,Security and Defence Industries Association of Europe.

3. Byrne J. Scientific and Technical Translation Explained: a nuts and bolts guide for beginners. London ; New York : Routledge, Taylor and Francis Group, 2014.

4. Cortés-Rodríguez F.J., Rodríguez-Juárez C. Computational analysis of adjuncts in ASD-STE100 for the NLP parser ARTEMIS. *VIAL. Vigo International Journal of Applied Linguistics*. 2023;20. DOI: https://doi.org/10.35869/vial.v0i20.4356

5. Hudecová E. Challenges of Machine Translation. *Yučovanie prvého (L2) a druhého (L3) cudzieho jazyka z pohľadu aplikovanej lingvistiky : kooperácie – kontrasty – analógie – interdisciplinarita*. 2020. URL: https://www.researchgate.net/publication/346975193_Challenges_of_Machine_Translation (access date: 06.10.2025).

6. Hutchins W. Looking back to 1952: the first MT conference // 1997 TMI. URL: https://aclanthology.org/1997.tmi-1.3.pdf (access date: 06.10.2025).

7. Koehn P., Knowles R. Six challenges for neural machine translation. *Proceedings of the First Workshop on Neural Machine Translation*. 2017. P. 28–39. DOI: https://doi.org/10.48550/arXiv.1706.03872

8. Marzouk S., Hansen-Schirra S. Evaluation of the impact of controlled language on neural machine translation compared to other MT architectures. *Machine Translation*. 2019. Vol. 33, P. 179–203. DOI: https://doi.org/10.1007/s10590-019-09233-w

9. Mitamura T. Controlled Language for Multilingual Machine Translation. *Proceedings of Machine Translation Summit VII, Singapore*, September 13–17, 1999. URL: https://aclanthology.org/www.mt-archive.info/90/MTS-1999-Mitamura.pdf (access date: 06.10.2025).

10. Shufrans TechDocs. *Simplified Technical English case study*. 2021. URL: https://www.shufrans-techdocs.com/simplified-technical-english-case-study/ (access date: 06.10.2025).

Дата першого надходження рукопису до видання: 09.10.2025
Дата прийнятого до друку рукопису після рецензування: 21.11.2025
Дата публікації: 30.12.2025