UDC 81'25:81'33:54:378
DOI https://doi.org/10.32782/2409-1154.2025.75.42

Tarasova V. V.,
Candidate of Philological Sciences,
Associate Professor of the Department of Foreign Languages
of the Chemistry and Physics Faculties,
Educational and Scientific Institute of Philology
Taras Shevchenko National University of Kyiv
https://orcid.org/0000-0002-7161-323X

TRANSLATION CHALLENGES AND EQUIVALENCE ISSUES OF CHEMICAL TERMINOLOGY IN STUDENT SCIENTIFIC DISCOURSE

Summary. This pilot study investigates how student translators handle multiword chemical terminology (MWTs) in English for Specific/Academic Purposes (ESP/EAP) and whether a brief corpus-assisted post-editing (PE) activity improves terminological accuracy relative to human-only translation. Eighteen second-year chemistry majors at Taras Shevchenko National University of Kyiv completed two stages. Stage 1 (descriptive/corpus-based) compiled a small domain corpus consisting of a student subcorpus (12,054 tokens; 83 texts) and an expert/reference subcorpus (16,265 tokens; 54 excerpts from journal articles, lab protocols, and textbooks). Ninety-six target MWTs were identified and 518 term instances were annotated using a concise two-dimensional rubric (conceptual exactness; collocational/phraseological fit). Interrater agreement met a priori thresholds: weighted $\kappa = .79$, 95% CI [.70, .88] for conceptual exactness; weighted $\kappa = .73$, 95% CI [.61, .84] for collocational fit; ICC(2,2) = .82 , 95% CI [.70, .90], p < .001 for total scores, indicating good absolute-agreement reliability. Nearly two in five term instances contained at least one equivalence failure, with non-target collocation, modifier-order issues, and register drift as dominant patterns. Stage 2 (within-subjects) compared human-only translation with corpus-assisted PE of machine-translation output using two counter-balanced source texts (≈250–300 words) containing matched MWT sets. Terminological accuracy (0–4 per term) was higher for corpus-assisted PE than for human-only translation, $M = 2.90, SD = 0.38$ vs. $M = 2.62, SD = 0.42$; $t(17) = 4.43, p < .001$, mean difference = 0.28, 95% CI [0.15, 0.41], $dav = 0.73$. Gains concentrated in collocational/phraseological fit ($\Delta = 0.22; t(17) = 3.12, p = .006$), whereas conceptual exactness improved slightly but non-significantly ($\Delta = 0.06; t(17) = 1.84, p = .083$). Time-on-task decreased (median 19.5 vs. 24.0 minutes; Wilcoxon $Z = -2.17, p = .030$), and students rated the workflow as useful (median = 4/5). No order effects emerged; a mixed-effects model confirmed the condition effect ($\beta = 0.27, SE = 0.07, t = 3.98, p < .001$). Findings indicate that discipline-preferred phraseology around MWTs – rather than concept selection per se – is the binding constraint for novice translators in chemistry. The study contributes (a) a reliable, classroom-feasible rubric for term-level assessment, (b) an empirically grounded error landscape for MWTs, and (c) evidence that brief corpus-assisted PE yields measurable quality gains with reduced effort. Implications include integrating KWIC-guided PE checkpoints, curated term bases (aligned with IUPAC and conventional usage), and targeted phraseology drills into chemistry-focused ESP/EAP curricula.

Key words: corpus-assisted post-editing; corpus-based translation studies; English for Specific/Academic Purposes; learner translation; machine translation; multiword chemical terminology; tertiary student scientific discourse.

Statement of the problem and its scientific/practical relevance. Across chemistry programmes, students are expected to read, translate, and produce discipline-specific texts in English; yet their performance often falters at the point of terminological equivalence – selecting target-language terms that preserve the source concept with appropriate genre and discourse constraints. Recent work in learner translation of (semi-)specialised texts shows persistent error patterns involving semantic drift, misuse of near-synonyms, and genre-incongruent phrasing, for instance, under-/over-specification, register shifts [1]. At the same time, ESP/CLIL research for chemistry highlights that command of chemical lexis and word-formation (Greek/Latin roots, affixes) is foundational but unevenly developed in typical curricula, which undermines both reading accuracy and L2 scientific writing [2; 3]. Theoretical debates in translation studies also show that "equivalence" remains a contested construct; nonetheless, it is central to quality in specialised translation and must be operationalised at semantic, pragmatic, and genre levels [4; 5; 6]. Practically, this gap impedes students' ability to state methods and mechanisms precisely, describe procedures without ambiguity, and align with disciplinary genre markers, specifically IMRaD sections, hedging, nominalisation. It also complicates instructors' assessment of translation quality and the design of pedagogical interventions that go beyond glossary memorisation – corpus-based term validation, morphology-driven term formation, and guided post-editing of MT outputs for LSP texts [7; 3]. Therefore, the problem this study addresses is the lack of an empirically grounded model that (a) maps typical equivalence failures in students' translation of chemical terminology to underlying linguistic and discourse causes and (b) links these failures to instructional tasks that measurably improve terminological accuracy and genre-appropriate expression in student scientific discourse.

Analysis of recent studies and publications, and the unresolved parts of the problem. Recent scholarship confirms that "equivalence" in specialised translation remains a contested, multi-level construct (semantic, pragmatic, and genre/register alignment) rather than a single measurable target. Surveys of professional translators underscore divergent operationalizations of equivalence in practice, with implications for how quality is defined and assessed in training contexts. This debate motivates empirical work that triangulates product features (accuracy, terminological fit) with process and context (task, genre, audience) [4; 8; 9].

Within chemistry-oriented ESP/EAP, studies document persistent language-related barriers for Eng+ (English-additional-language) trainees, affecting the precision of method descriptions, argumentation, and the use of disciplinary lexis. Parallel research in chemistry education shows that students struggle to "translate" among verbal, symbolic, and particulate representations – a competence closely tied to the correct construal and naming of entities and processes. Together, these strands suggest that terminological equivalence is inseparable from representational fluency and genre conventions in scientific discourse [10; 11]. At the level of terminology, two trends are salient. First, domain initiatives aim to stabilise cross-language naming, e.g., efforts to develop multilingual chemistry dictionaries to reduce ambiguity around systematic vs. trivial names and multiword nomenclature. Second, research in Terminology and related venues highlights how "popularisation" vs. "scientisation" pressures can shift term meanings across audiences, complicating one-to-one mappings for students who are learning to write "as chemists" [14; 13].

For student translators, the hardest cases typically involve multiword terms and secondary term formation (derivatives, affixation, compounds). Comparative work shows that while neural machine translation (MT) now handles some multiword units better than earlier systems, corpus-based verification remains crucial for selecting field-appropriate equivalents and collocations – an area where novice translators underuse morphology and corpus tools [14; 15].

Pedagogically, three intervention lines are prominent: (1) corpus-based translation training: recent frameworks advocate integrating learner corpora and domain corpora to surface recurrent decision points and error patterns linked to task/genre variables; this supports data-driven feedback on equivalence beyond glossary memorisation [8; 9; 17]; (2) MT post-editing (PE) for LSP: PE training demonstrably changes the kinds of revisions students make; in specialised domains, coupling PE with corpus checks is recommended to handle phraseology and term variants [13]; (3) ESP/disciplinary supports: allowing multilingual scaffolds in assessment and targeted lexical development has measurable benefits for chemistry undergraduates, pointing to the value of bridging pedagogies that connect domain knowledge, representation work, and language [18].

From a technology perspective, broad surveys of MT for specialised domains stress terminology handling as a persistent weakness – coverage, domain adaptation, and context-sensitive disambiguation remain open issues – while emerging studies test GenAI as a PE assistant to influence learners' lexical/syntactic choices. These findings reinforce the need for human-in-the-loop workflows that institutionalise corpus validation and style/genre control in student work [8; 11].

Building on the foregoing review, several unresolved facets emerge at the intersection of MT-supported translation, corpus-informed terminology work, and chemistry-specific genre conventions. First, the field still lacks a robust way of operationalising equivalence for chemistry terminology used in student genres: validated rubrics that jointly capture (a) conceptual exactness (IUPAC vs. conventional names), (b) collocational/phraseological fit, and (c) genre-specific realisations across methods, results, and procedures are scarce, and current evaluation practices tend to split "accuracy" from "style" instead of treating them as an integrated construct [4; 12]. Second, multiword terminology and secondary term formation remain pedagogically under-served; few approaches systematically connect morphology (roots/affixes), corpus evidence, and post-editing (PE) guidelines to help novices resolve equivalence for compounds, nominal groups, and derived forms common in chemistry (e.g., enzyme/substrate complexes, rate-determining step) [1; 2; 11]. Third, although research links difficulties in moving among symbolic, particulate, and macroscopic representations to language-of-science challenges, integrated interventions that simultaneously train representational fluency and terminological equivalence are still rare and insufficiently evaluated [10; 11]. A fourth unresolved area concerns assessment at scale with authentic corpora: learner-corpus and DIY-corpus approaches are promising, yet widely usable protocols for tagging equivalence failures (e.g., under-/over-specification, register drift, semantic calques) and converting those tags into teachable tasks and grading criteria remain under-specified [8; 17]. Finally, the integration of MT/GenAI in student translation calls for controlled, chemistry-specific studies that compare human-only, PE-only, and hybrid corpus-validated workflows, with outcomes measured on terminological precision and genre alignment; such evidence is presently limited [2; 16; 19].

Accordingly, this study adopts *a pragmatic, two-stage aim*: first, to compile a small domain corpus that maps recurrent equivalence failures in student translations of multiword chemical terms – focusing on conceptual exactness and collocational fit – and second, to pilot-test a brief corpus-assisted post-editing activity to determine whether it improves terminological accuracy relative to human-only translation within a single cohort.

Presentation of the main research material. This study employed a two-stage, mixed-methods pilot within a single cohort. In *Stage 1* (descriptive/corpus-based), we compiled a small domain corpus and mapped recurrent equivalence failures in student translations of multiword chemical terms, focusing on two dimensions: (a) conceptual exactness (including IUPAC vs. conventional names) and (b) collocational/phraseological fit. In *Stage 2* (mini-experimental), we ran a within-subjects, counter-balanced comparison of human-only translation versus corpus-assisted post-editing of machine-translation output on matched chemistry texts. The primary outcome was terminological accuracy; secondary outcomes included collocational fit, time-on-task, and learner acceptability.

Convenience sampling approach was used to involve 18 participants who were second-year chemistry majors of Taras Shevchenko National University of Kyiv enrolled in an ESP/EAP course. Inclusion criteria were B2–C1 English proficiency and completion of at least one laboratory methods course with no prior professional translation experience. All participants provided informed consent, and all data were pseudonymised.

Materials comprised a student sub-corpus (anonymised course outputs and the translations produced in Stage 2) and an expert/reference sub-corpus (short methods/results sections from peer-reviewed articles, lab protocols, and textbooks). A curated termbase of 80–100 multiword terms was created on the basis of frequency, pedagogical relevance, and coverage of nominal groups and derived forms. Each entry included preferred IUPAC/conventional mappings, morphological families, typical collocates, and example concordances. Concordancing was supported by a lightweight KWIC tool with searches for collocation windows (±5) and noun-phrase patterns.

To assess outcomes, we used a concise two-dimension rubric. Conceptual exactness was scored 0–2 (0 = wrong/misleading, 1 = partial/underspecified, 2 = exact), and collocational/phraseological fit was scored 0–2 (0 = unlikely/ungrammatical, 1 = acceptable but non-preferred, 2 = field-typical). Scores were summed per term (0–4) and averaged per text. For diagnostic feedback and later analysis, raters also applied error tags (under-/over-specification, register drift, semantic calque, malformed derivation, non-target collocation). Process measures included time-on-task and a short acceptability survey (five Likert items plus two open questions).

Stage 1 began with pilot annotation and rater calibration on approximately 15% of student texts. Discrepancies were used to refine the rubric and codebook; interrater agreement was then estimated using weighted κ for sub-scores and ICC(2,2) for total scores (a priori target ≥ .70). The observed reliability was ICC(2,2) = 0.82, 95% CI [0.70, 0.90], p < .001, indicating good absolute-agreement for the average of two raters. In Stage 2, two comparable source texts (A/B, ~250–300 words) were constructed to contain matched sets of target multiword terms. Participants were randomly assigned to task order (AB vs. BA). In the human-only condition, students translated without tools; in the corpus-assisted PE condition, they post-edited MT output with access to the curated concordancer and termbase, following brief guidance that prioritized validation of multiword terms and local phraseology. Time-on-task was recorded for each condition, and outputs were exported for term-level scoring.

Data processing involved automatic detection of target term spans (pattern rules for multiword noun phrases) with rater verification before scoring. The primary analysis was a within-subjects comparison of mean terminological accuracy between conditions. Using *Jamovi* statistical software (version 2.5.5), normality was checked (Shapiro–Wilk); paired t-tests with Cohen's d_av were used when assumptions held, otherwise Wilcoxon signed-rank tests with effect size r. Secondary analyses examined differences in the collocational sub-score, time-on-task, and error-tag frequencies, and fitted mixed-effects models with random intercepts for participant and term to control variability. Reliability was re-estimated on a fresh 10–15% subset from Stage 2, and feasibility was summarised via completion rates, median time, acceptability ratings (median, IQR), and brief thematic coding of open responses.

Ethical safeguards followed institutional guidelines for minimal-risk educational research: participation was voluntary, grades were unaffected, and only aggregate results were reported. To support rigor, we pre-checked text difficulty and term density with subject-matter experts, maintained a detailed rater codebook with adjudication procedures, and confined outcomes to the declared scope – multiword chemical terminology – so that effect sizes from this pilot can inform future scaling.

Further we report the results in two parts: first, the Stage 1 corpus-based mapping of equivalence failures and rubric reliability, and second, the Stage 2 within-subjects comparison of human-only translation versus corpus-assisted post-editing.

In developing the corpus profile, the pilot sample comprised 18 second-year chemistry students enrolled in an ESP/EAP course, yielding a student sub-corpus of 12,054 tokens (83 texts) and an expert/reference sub-corpus of 16,265 tokens (54 excerpts from journal articles, lab protocols, and textbooks). We curated 96 target multiword terms (MWTs) and annotated 518 term instances in student outputs ($M = 6.2$ per text). Following rubric and codebook refinement, interrater agreement met a priori thresholds. Weighted κ for conceptual exactness was $.79$, 95% CI [.70, .88], and weighted κ for collocational/phraseological fit was $.73$, 95% CI [.61, .84]. For total scores, the intraclass correlation (two-way random effects, absolute agreement, average of two raters) was ICC(2, 2) = $.82$, 95% CI [.70, .90], $p < .001$, indicating good absolute-agreement reliability. Below, Table 1 summarises Stage 1 rubric outcomes for conceptual exactness, collocational/phraseological fit, and the total per-term score.

Table 1

Rubric Scores for Conceptual Exactness, Collocational/Phraseological Fit, and Total (Stage 1)

Dimension	Scale range	M	SD
Conceptual exactness	0–2	1.41	0.54
Collocational/phraseological fit	0–2	1.34	0.57
Total per term	0–4	2.75	0.85

Note. Scores are per multiword term (MWT). Total = conceptual exactness + collocational/phraseological fit. N = 518 annotated MWT instances.

As shown in Table 1, performance was higher on conceptual exactness than on collocational/phraseological fit. Multiword terms containing three or more tokens scored lower than two-token units on the total score ($\Delta = -0.28$, 95% CI [−0.41, −0.15]). Derived forms (e.g., -ation, -ivity, -ic/-ous*) tended to reach the correct concept (higher conceptual exactness) while underperforming on field-typical phraseology (lower collocational fit), suggesting that morphology supports conceptual accuracy but not necessarily discipline-preferred phrasing.

The study also found that within the 518 annotated instances, 39.8% (≈206/518) contained at least one equivalence failure. Tag frequencies normalized per 100 term instances – and corresponding counts for this sample – were: *non-target collocation*, 14.8 (≈77); *under-/over-specification*, 11.2 (≈58); *register drift*, 8.9 (≈46); *semantic calque*, 6.7 (≈35); and *malformed derivation*, 5.4 (≈28). Structurally, pre-modifier stacks, for example, *high-temperature catalytic oxidation*, generated more collocational and register problems than of- phrases, for instance, *oxidation of X at high temperature*, and noun–noun compounds with measurement/process heads, for example, *dose–response curve, rate-determining step*, showed the highest density of collocation and under-specification tags.

Analyses by term family yielded distinct error profiles. For process/kinetics items, for example, *rate-determining step, reaction rate constant*, we most often observed under-specification, for instance, *missing hyphenation or omission of the head noun such as "step" or "constant"*, alongside collocational drift (non-target prepositional choices). For solutions/reagents terminology, for example, *aqueous working solution, stock solution, 0.1 M sodium chloride solution*, errors primarily concerned modifier order and placement; by contrast, the expert corpus consistently exhibited frames such as "prepare a 0.1 M aqueous NaCl solution." For redox/measurement terms (e.g., oxidation–reduction potential / redox potential), we documented occasional near-synonym confusion at the conceptual level and register inconsistencies (e.g., ORP value is big vs. the field-typical ORP is high).

Concordance (KWIC) analyses of the expert sub-corpus further indicated stable phraseological patterns that were underused by students. Procedure verbs – *prepare, adjust, dilute, maintain, and monitor* – recurrently co-occurred within ±5 tokens of solution/reagent multiword terms, while method/result reporting was characteristically realized through passive constructions with hedging, for example, *was determined, was monitored, was maintained at*. Together, these patterns account for a sizable portion of the collocational and register deviations identified in student outputs and delineate clear targets for corpus-informed instruction.

Because all participants were second-year students, between-year comparisons were not applicable. Instead, we ran within-cohort, exploratory associations between self-rated English proficiency (B2–C1; ordinal) and the two rubric dimensions. Using Spearman rank correlations, the association with collocational/phraseological fit was small and non-significant, $r_s = .21, p = .40, n = 18$; the association with conceptual exactness was negligible, $r_s = .12, p = .63, n = 18$. These results suggest that domain-specific phraseology, rather than general proficiency, is the more salient constraint in this cohort; however, estimates should be interpreted cautiously given the modest sample size.

These findings directly motivate Stage 2's within-subjects comparison of human-only vs corpus-assisted PE workflows, with terminological accuracy as the primary outcome and collocational fit as the most sensitive secondary indicator.

In Stage 2, terminological accuracy (0–4 per term) was higher for the corpus-assisted post-editing condition than for human-only translation, $M = 2.90, SD = 0.38$ vs. $M = 2.62, SD = 0.42$. This difference was statistically significant, $t(17) = 4.43, p < .001$, mean difference = 0.28, 95% CI [0.15, 0.41], Cohen's $dav = 0.73$, indicating a medium–large within-subject improvement in a single cohort ($n = 18$). Subscore analyses showed that collocational/phraseological fit improved from $M = 1.27, SD = 0.33$ (human-only) to $M = 1.49, SD = 0.35$ (corpus-assisted PE), $t(17) = 3.12, p = .006, \Delta = 0.22$, 95%CI[0.07,0.38], $dav = 0.52$, whereas the gain in conceptual exactness was small and not significant, rising from $M = 1.35, SD = 0.29$ to $M = 1.41, SD = 0.27$, $t(17) = 1.84, p = .083, \Delta = 0.06$, 95% CI [−0.01, 0.13], $dav = 0.22$. Error-tag rates corroborated these patterns: relative to human-only outputs, non-target collocation declined from 13.9 to 8.8 per 100 term instances (≈37% reduction) and register drift from 8.1 to 5.6 per 100 (≈31% reduction), while under-/over-specification and semantic calque showed modest, non-systematic changes (≤10% relative). Process measures favoured the corpus-assisted workflow. Time-on-task was lower for corpus-assisted PE (median = 19.5 min) than for human-only translation (median = 24.0 min), Wilcoxon signed-rank $Z = -2.17, p = .030, r = .36$. Acceptability ratings indicated high perceived usefulness (median = 4, IQR = 1) and moderate perceived effort (median = 3, IQR = 1); open-ended comments most frequently cited easier checking of multiword terms and clearer phraseological models. Order effects were not detected for the primary outcome (both $ps \geq .50$), and a confirmatory mixed-effects model with random intercepts for participant and term reproduced the condition effect ($\beta = 0.27, SE = 0.07, t = 3.98, p < .001$). Collectively, these findings meet the stated aim by demonstrating that a brief corpus-assisted post-editing activity improves terminological accuracy relative to human-only translation in this single-cohort pilot, primarily by strengthening phraseological fit while reducing effort.

As implications for teaching and assessment, the results of our study indicate that phraseology – not concept selection – is the binding constraint for student translators of chemistry texts. Accordingly, instruction should prioritise collocational/phraseological fit around multiword terms (MWTs): teaching with KWIC concordances, discipline-typical frames (e.g., prepare / was determined / was maintained at), and targeted drills on modifier order and noun–noun compounds. The concise two-dimension rubric (conceptual exactness; collocational fit) proved reliable and classroom-feasible, supporting formative assessment at the term level and enabling feedback that is diagnostically specific (e.g., under-/over-specification vs. non-target collocation). Because the corpus-assisted post-editing (PE) workflow improved accuracy while reducing time-on-task, instructors can integrate short PE sessions as low-cost "quality checkpoints" in ESP/EAP courses without adding undue workload.

As implications for MT/GenAI use in LSP settings, the findings support a human-in-the-loop pipeline – MT→PE with corpus/termbase support – for specialised terminology. Gains concentrated in collocational fit, suggesting that MT may supply a semantically plausible baseline, while corpus evidence and PE guidance supply the discipline-specific phraseology students lack. Practically, programs should maintain small, curated termbases (IUPAC vs. conventional mappings, morphological families, typical collocates) and provide quick-access concordancing to anchor decisions on multiword units. For institutional QA, the rubric and error tags can be embedded in editorial checklists (e.g., for theses or lab manuals) to standardise expectations across courses.

At the curriculum level, we recommend integrating short micro-modules into chemistry-ESP sequences that (a) explicitly connect morphology to phraseology – moving from derivational families to discipline-preferred frames, (b) provide deliberate practice with modifier stacking and compound-noun parsing, and (c) institutionalize brief post-editing-with-corpus checkpoints at key assessment moments. To ensure portability across cohorts and instructors, departments should maintain small, shared repositories that include mini-corpora, vetted term lists (aligned with IUPAC and conventional nomenclature), and exemplar concordances illustrating stable procedure and reporting frames.

Conclusions and prospects for further research. This pilot study set out to (a) map recurrent equivalence failures in student translations of multiword chemical terms and (b) pilot-test a brief corpus-assisted post-editing (PE) activity against human-only translation within a single cohort ($n = 18$). Stage 1 established a reliable, concise, two-dimensional rubric for term-level assessment – covering conceptual exactness and collocational/phraseological fit – and showed that about two in five annotated instances contained at least one equivalence failure, with non-target collocation, modifier-order issues, and register drift as dominant patterns. Stage 2 demonstrated that corpus-assisted PE produced a statistically significant improvement in terminological accuracy over human-only translation (a medium–large within-subject effect), with the largest gains in collocational/phraseological fit; conceptual exactness showed only a small, non-significant increase. Importantly, corpus-assisted PE also reduced time-on-task and was rated as pedagogically acceptable, suggesting a favourable trade-off between quality and effort. Taken together, the findings indicate that for chemistry-focused ESP/EAP tasks, the binding constraint for novice translators is discipline-preferred phraseology around multiword terms, rather than concept selection per se. A small, targeted toolkit – KWIC concordances, a curated term base aligned with IUPAC and conventional usage, and the two-dimensional rubric – can be feasibly integrated into coursework to provide diagnostic feedback and low-cost quality checkpoints. These results warrant scaled replications and programmatic integration (e.g., brief PE-with-corpus checkpoints at key assessment moments). In corpus-based translation studies (CBTS), the study shows how small, domain-specific corpora and KWIC evidence can be operationalized not only for analysis but also for assessment and targeted intervention with student translators. In MT/PE for specialized domains (LSP), it isolates where PE with corpus support yields the largest marginal gains – phraseology around multiword terms – and quantifies those gains within subjects. In ESP/EAP for STEM, it links disciplinary genre frames (e.g., procedure verbs; methods/results reporting) to measurable improvements and supplies a scalable rubric suitable for classroom use. In assessment and rater reliability for LSP contexts, it demonstrates that term-level scoring can achieve acceptable reliability with concise instruments, supporting program-level quality assurance. Future work should expand to multi-site cohorts, add an MT-only arm to isolate the contributions of PE and corpus support, test retention/transfer on delayed tasks, and integrate representational translation training (verbal ↔ symbolic ↔ particulate) to examine potential crossover effects on both conceptual accuracy and phraseology.

Bibliography:

1. Kübler N., Mestivier A., Pecman M. Error annotation and analysis of a (semi-) specialised English-French learner translation corpus. *Meta.* 2024a. V. 69, № 2. P. 331–354. DOI: 10.7202/1118381ar.
2. Naveen P., Trojovský P. Overview and challenges of machine translation for contextually appropriate translations. *iScience.* 2024. V. 27, № 10. Article 110878. DOI: 10.1016/j.isci.2024.110878.
3. Vukićević-Đorđević L. Chemistry vocabulary teaching: A short guide for ESP teachers. *The Journal of Teaching English for Specific and Academic Purposes.* 2025. V. 13, № 1. P. 151–158. DOI: 10.22190/JTESAP250114013V.
4. Dam H. V., Zethsen K. K. Professionals' views on the concepts of translation: the challenge of categorisation. *The Translator.* 2024. V. 30, № 3. P. 388–406. DOI: 10.1080/13556509.2024.2320953.
5. Rędzioch-Korkuz A. Revisiting the concepts of translation studies: equivalence in linguistic translation from the point of view of Peircean universal categories. *Language and Semiotic Studies*. 2023. V. 9, № 1. P. 33–53. DOI: 10.1515/lass-2022-0008.
6. Ursin J., Hyytinen H., Silvennoinen K., Toom A. Linguistic, Contextual, and Experiential Equivalence Issues in the Adaptation of a Performance-Based Assessment of Generic Skills in Higher Education. *Frontiers in Education.* 2022. V. 7. Article 885825. DOI: 10.3389/feduc.2022.885825.
7. Guerberof A., Moorkens J. Machine translation and post-editing training as part of a master's programme. *The Journal of Specialised Translation.* 2019. Issue 31. P. 217–238. DOI: 10.26034/cm.jostrans.2019.184.
8. Wang G., Xin Y. An analytical framework for corpus-based translation studies. *Humanities and Social Sciences Communications.* 2024. V. 11. Article 1709. DOI: 10.1057/s41599-024-04250-4.
9. Wu K., Lei V., Li D. Charting the Trajectory of Corpus Translation Studies: Exploring Future Avenues for Advancement. *Corpus-based Studies across Humanities.* 2024. V. 2, № 1. P. 51–77. DOI: 10.1515/csh-2024-0001.
10. Deng J. M., Flynn A. B. "I am working 24/7, but I can't translate that to you": The barriers, strategies, and needed supports reported by chemistry trainees from English-as-an-additional language backgrounds. *Journal of Chemical Education.* 2023. V. 100, № 4. P. 1523–1536. DOI: 10.1021/acs.jchemed.2c01063.
11. Hu B., Zhu L., Bi H. Effect of computer simulations on student ability to translate chemical representations when learning the particulate nature of matter concept. *Journal of Chemical Education.* 2024. V. 101, № 10. P. 4149–4160. DOI: 10.1021/acs.jchemed.4c00964.
12. Lu H., Hao X., Zhang Y. Popularization and scientization in terminology translation: A case study of interlingual terminological shifts in the Chinese-English translation of San Ti (Three Body). *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication*. 2024. V. 30, № 2. P. 250–284. DOI: 10.1075/term.23015.lu.
13. Robinson J. Multi-language chemistry dictionary to be created to avoid confusion. Chemistry World. 2024. URL: https://www.chemistryworld.com/news/chemistry-body-to-create-multi-language-chemistry-dictionary-to-avoid-confusion/4020071.article (date of access: 12.08.2025).
14. Cabezas-García M., León-Araúz P. Machine versus corpus-based translation of multiword terms. *Digital Scholarship in the Humanities.* 2023. T. 38, Suppl. 1. P. i6–i16. DOI: 10.1093/llc/fqad026.
15. Polcz K., Hamsovszki S., Huszár E., Jámbor E., Szigetváry N., Válóczi M. Translation procedures in secondary term formation across languages in online glossaries: Universal or language-specific? *Across Languages and Cultures.* 2023. V. 24, № 2. P. 239–256. DOI: 10.1556/084.2023.00400.
16. Kübler N., Martikainen H., Mestivier A., Pecman M. Post-editing neural machine translation in specialised languages: the role of corpora in the translation of phraseological structures. У: Monti J., Pastor G. C., Mitkov R., Hidalgo-Ternero C. M. (eds.). *Recent Advances in Multiword Units in Machine Translation and Translation Technology*. Amsterdam; Philadelphia: John Benjamins Publishing Company, 2024b. P. 57–78. DOI: 10.1075/cilt.366.04kub.
17. Li X. Instrumentalising foreign language pedagogy in translator and interpreter training: methods, goals and perspectives: by O. I. Seel, S. Roiss & P. Z. González, Amsterdam, John Benjamins, 2023, x + 291 pp. *The Interpreter and Translator Trainer.* 2025. V. 19, № 2. P. 218–225. DOI: 10.1080/1750399X.2025.2463756.
18. Martin P. P., Graulich N. Beyond language barriers: Allowing multiple languages in postsecondary chemistry classes through multilingual machine learning. *Journal of Science Education and Technology*. 2024. V. 33. P. 333–348. DOI: 10.1007/s10956-023-10087-4.
19. Kwok H. L., Shi Y., Xu H., Li D., Liu K. GenAI as a translation assistant? A corpus-based study on lexical and syntactic complexity of GPT-post-edited learner translation. *System.* 2025. V. 130. Article 103618. DOI: 10.1016/j.system.2025.103618.

Тарасова В. Виклики перекладу та питання еквівалентності хімічної термінології у студентському науковому дискурсі

Анотація. У пілотному дослідженні проаналізовано, як студенти-перекладачі опрацьовують багатослівну хімічну термінологію в англійській для спеціальних/академічних цілей, а також чи підвищує корпусно-асистоване післяредагування точність термінологічного відтворення порівняно з «людським» перекладом без інструментів. Участь взяли 18 студентів другого курсу хімічного профілю Київського національного університету імені Тараса Шевченка. Етап 1 (описовий/корпусний) передбачав укладання маломасштабного доменно-орієнтованого корпусу: студентського підкорпусу (12 054 токени; 83 тексти) та експертно-референтного підкорпусу (16 265 токенів; 54 фрагменти зі статей, лабораторних протоколів і підручників). Виокремлено 96 цільових багатослівних хімічних термінологічних блоків і анотовано 518 терміновживань за двовимірною рубрикою (концептуальна точність; колокаційно-фразеологічна відповідність). Міжоцінювальна узгодженість відповідала апріорним порогам: зважений $\kappa = .79$, 95% ДІ [.70, .88] для концептуальної точності; зважений $\kappa = .73$, 95% ДІ [.61, .84] для колокаційної відповідності; ICC(2,2) = .82, 95% ДІ [.70, .90], $p < .001$ для загального бала, що свідчить про добру абсолютну узгодженість. Майже у двох із п'яти випадків фіксувалися порушення еквівалентності, серед яких домінували нетипові колокації, збій порядку модифікаторів та зсув регістру. Етап 2 (внутрішньосуб'єктне порівняння) зіставляв «людський» переклад і корпусно-асистоване післяредагування машинного перекладу з доступом до кураторованого конкордансера і термбази на двох контрзбалансованих текстах (≈250–300 слів) із погодженими наборами багатослівної хімічної термінології. Термінологічна точність (0–4 за термін) була вищою за умови корпусно-асистованого післяредагування, $M = 2.90, SD = 0.38$, порівняно з «human-only», $M = 2.62, SD = 0.42; t(17) = 4.43, p < .001$, середня різниця = 0.28, 95% ДІ [0.15, 0.41], dav = 0.73. Приріст зосереджувався у колокаційно-фразеологічній відповідності ($\Delta = 0.22; t(17) = 3.12, p = .006$), тоді як концептуальна точність збільшилася незначно і статистично незначуще ($\Delta = 0.06; t(17) = 1.84, p = .083$). Час виконання зменшився (медіана 19.5 проти 24.0 хв; Wilcoxon $Z = -2.17, p = .030$), студенти високо оцінили корисність підходу (медіана = 4/5). Ефект порядку не виявлено; модель зі змішаними ефектами підтвердила вплив умови ($\beta = 0.27, SE = 0.07, t = 3.98, p < .001$). Результати свідчать, що дисциплінарно типова фразеологія навколо багатослівної хімічної термінології, а не власне вибір концепту, є головним обмеженням для початківців-перекладачів у хімії. Дослідження пропонує: (а) надійну й придатну до аудиторного використання рубрику термінорівневого оцінювання; (б) емпірично окреслений ландшафт помилок для багатослівної хімічної термінології; (в) докази того, що короткі KWIC-керовані сеанси корпусно-асистованого післяредагування з кураторованими базами термінів (узгодженими з IUPAC і усталеною практикою) забезпечують вимірюване підвищення якості за меншого навантаження. Практичне застосування включає інтеграцію контрольних пунктів корпусно-асистованого післяредагування-з-корпусом і цілеспрямованих вправ із фразеології в курси англійській для спеціальних/академічних цілей хімічного спрямування.

Ключові слова: корпусно-асистоване післяредагування; корпусні дослідження перекладу; англійська для спеціальних/академічних цілей; студентський переклад; машинний переклад; багатослівна хімічна термінологія; студентський науковий дискурс у ЗВО.

Дата першого надходження рукопису
до видання: 24.10.2025
Дата прийнятого до друку рукопису
після рецензування: 25.11.2025
Дата публікації: 30.12.2025