Kajian Linguistik dan Sastra
https://journals2.ums.ac.id/index.php/kls

Quantum Translation: A New Heuristic for Cognitive Uncertainty in the AI Era *(Terjemahan Kuantum: Pendekatan Heuristik Baru dalam Ketidakpastian Kognitif di Era AI)*

Nurul Widiatul Ihsan[1], Jiehan Ashara Halim[2], Claudya Fritsca[3], Refika Andriani[4]

[1] English Language Education Program, University of Lancang Kuning
Surel: nurulwidiatul29@gmail.com

[2] English Language Education Program, University of Lancang Kuning
Surel: jiehanahara@gmail.com

[3]English Language Education Program, University of Lancang Kuning
email: claudyafritsca18@gmail.com

[4]English Language Education Program, University of Lancang Kuning
email: refika@unilak.ac.id

DOI: 10.23917/kls.v10i2.13587
Diterima: 31 Oktober 2025. Revisi: 1 November 2025. Disetujui: 5 Desember 2025
Tersedia secara elektronik: 18 Desember 2025. Terbit: 30 Desember 2025
***Sitasi*:**
N. W. Ihsan, J. AsharaHalim, C. Fritsca, and R. Andriani, "Quantum Translation: A New Heuristic for Cognitive Uncertainty in the AI Era," Kaji. Linguist. dan Sastra, vol. 10, no. 2, pp. 234–258, 2025, doi: DOI: 10.23917/kls.v10i2.13587.

Abstract

This study reframes uncertainty in translator cognition by proposing a Quantum Translation (QT) heuristic superposition, collapse, and entanglement as a probabilistic lexicon for process analysis. Using a PRISMA-consistent systematic literature review, we screened records from Scopus, Crossref, and Google Scholar (2020–2025) via database queries and citation chasing, yielding 22 empirical studies. Data extraction targeted instruments used in primary studies (e.g., eye tracking, key logging, screen capture) and findings were synthesized thematically. Across the corpus, uncertainty is acknowledged as central yet treated implicitly as ambiguity, difficulty, or risk. Product-focused evaluation routinely obscures process-level signals such as cognitive load, recursive drafting, and attentional control. QT addresses this gap by modeling (i) superposition as coexisting candidate renderings, (ii) collapse as context-triggered resolution constrained by skopos, register, and pragmatics, and (iii) entanglement as cross-level dependencies linking lexical, syntactic, and discourse decisions. The review also charts convergences between human process traces and computational predictors (e.g., surprisal), informing risk-aware human AI workflows. We contribute a testable heuristic and implications: integrate QT-informed diagnostics in translator education; report AI use transparently; and adopt evaluation models that combine process and product. Together, these steps strengthen accountability and professional preparedness for human AI collaboration.

Penulis Korespondensi: Nurul Widiatul Ihsan
Nurul Widiatul Ihsan, Jiehan Ashara Halim, Claudya Fritsca, Refika Andriani, University of Lancang Kuning
Surel: nurulwidiatul29@gmail.com

***Keywords**: Heuristic decision-making, LLM-assisted translation, quantum-like cognition, translator cognition, uncertainty*

Introduction

Cognitive Translation Studies (CTS) has redirected translation studies from end products to the processes that generate them. In this view, translators are not conduits but situated decision-makers who weigh competing options under constraints of purpose, time, and domain knowledge. The central challenge in that decision space is uncertainty. It arises from lexical polysemy and syntactic complexity, from gaps in information and competing pragmatic demands, and from the need to coordinate local textual choices with global communicative aims. From this standpoint, context is less a static backdrop than a control parameter that, when activated at the right moment, shrinks the set of admissible interpretations and steers the emerging target text toward situational fit [1]. Process research has made this challenge empirically tractable. Think-aloud protocols, eye tracking, and keyboard logging produce temporally fine-grained traces that illuminate how cognitive load is distributed and how text complexity modulates consultation and problem solving [2]. These traces have demystified parts of the “black box,” while also exposing the limits of evaluation schemes that judge only the final text. As Vanroy, Schaeffer, and Macken [3] argue, product-oriented metrics systematically overlook the cognitive costs of recursive drafting, sentence restructuring, and attentional control that accumulate during the act of translation.

The rapid diffusion of Large Language Models (LLMs) intensifies, rather than diminishes, the relevance of a process perspective. Translators increasingly work within mixed cognitive ecologies in which tasks, risks, and burdens are distributed across human and machine actors. Educational priorities are shifting from basic post-editing toward metacognitive capability: the critical appraisal of system outputs, the calibration of risk, and the design of strategic interventions [4], [5]. Empirical frontiers are advancing as researchers align human traces with computational predictors; difficulty and effort are being linked to measures such as surprisal and model-internal attention, enabling more detailed accounts of process dynamics [6]. Complementary studies delineate contexts in which human intuition and situational awareness still surpass algorithmic processing, especially when non-local constraints or pragmatic subtleties are salient [7]. Work on prompting further shows how translators use iterative reasoning to steer LLMs, enacting a human-in-the-loop architecture that keeps professional responsibility with the human agent [8].

Despite these advances, the field's ordinary vocabulary ambiguity, difficulty, risk does not fully capture two recurring features of contemporary workflows: the simultaneous presence of multiple viable renderings and the abrupt, context-triggered resolution by which one option becomes preferred. What is needed is a compact, testable lexicon that treats uncertainty not as a flaw to be eliminated but as an intrinsic state of potentiality from which decisions are selected. To address this gap, the present article advances a Quantum Translation (QT) heuristic. The term is figurative rather than physical: it names a way of speaking about probabilistic states and context-driven resolution without presupposing any literal quantum mechanism. The heuristic centers on three linked notions. First, superposition designates the coexistence of candidate interpretations or renderings represented, implicitly or explicitly, as a distribution conditioned by co-text, genre, and domain knowledge. Second, collapse names the moment when task constraints skopos, register, and pragmatic fit stabilize one rendering and suppress alternatives, often following the activation of a "context packet" comprising briefs, style guides, terminology, or exemplars. Third, entanglement describes cross-level dependencies: a local lexical choice can reconfigure syntactic packaging and ripple into discourse patterns, creating non-local effects that matter for cohesion and style. Conceptually, the dynamic interplay between a superposed state being pulled by various contextual constraints toward a final collapse can be visualized in Figure 1.

Kajian Linguistik dan Sastra Vol. 10 (2) (2025) 234–258

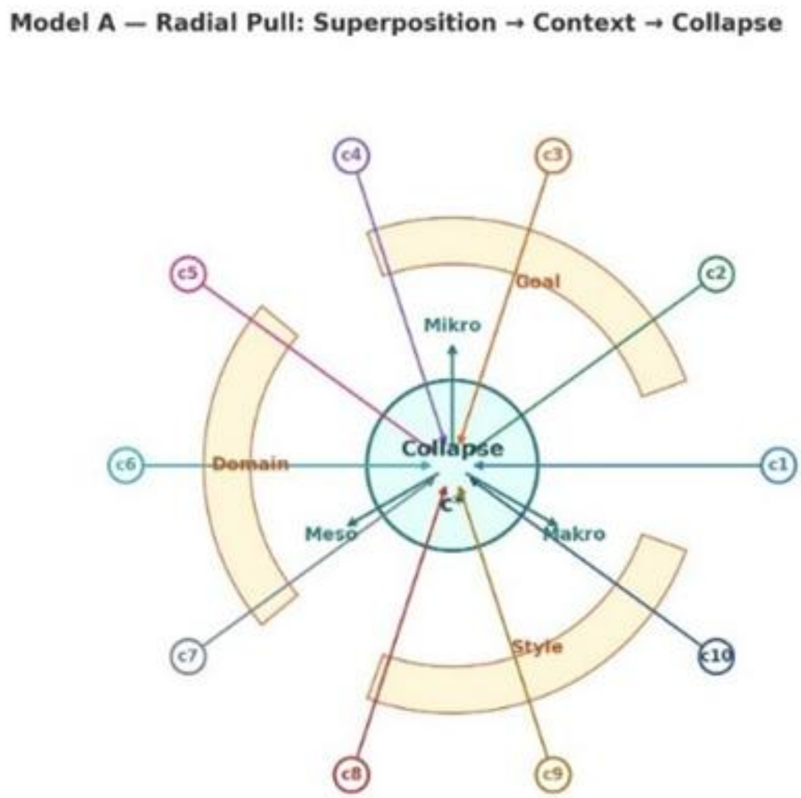

Figure 1. A Conceptual Model of the Quantum Translation (QT) Heuristic

The conceptual "Radial Pull" model of the Quantum Translation (QT) heuristic. The superposition state, which contains multiple viable candidate renderings (c1-c10), is influenced by various contextual constraints (e.g., Goal, Style, Domain). These constraints act as a "pull" that triggers a collapse from a high-entropy state to a single, stabilized choice (c), resolving the cognitive uncertainty. These contextual constraints operate at different levels: micro (e.g., co-text), meso (e.g., genre conventions), and macro (e.g., the overall communicative goal). QT matters because it invites operationalization. If relevant context is the trigger for collapse, we should observe entropy reduction within candidate sets and shorter time-to-decision among experts who mobilize context efficiently. If decisions are entangled across levels, controlled lexical manipulations should induce measurable shifts in syntactic structure and discourse coherence. And if risk concentrates where models register high surprisal, those zones should correlate with human effort signals such as pauses, regressions, and bursts of revision. Accordingly, this article translates QT into measurements. Candidate entropy indexes the dispersion of alternatives during superposition. Time-to-collapse captures decision dynamics from the emergence of credible options to the stabilization of a choice. Cross-level coupling metrics quantify entanglement by tracking how local edits propagate into syntax and discourse. These variables can be estimated from combinations of key-logging timelines, eye-movement indicators, screen-capture annotations, and computational signals, yielding converging evidence about where uncertainty concentrates and how context reduces it. The contribution is threefold. Conceptually, the article proposes QT as a compact language for reasoning about uncertainty in translator cognition. Methodologically, it shows how to render that language testable with observable variables that scale across languages, directions, and expertise levels. Practically, it derives implications for translator education diagnosing superposition segments, activating context packets, and reflecting on cascading effects alongside recommendations for transparent reporting of AI assistance and for evaluation models that register process costs as well as products. The remainder of the paper proceeds as follows. We situate QT within prior work and describe the review protocol and inclusion criteria that ground the synthesis. We then map the instrument signal landscape and present thematic findings on how uncertainty manifests in process data across recent studies. Next, we specify operational definitions and sketch study designs for testing QT's predictions, before turning to implications for curricula, professional standards, and policy. We close by outlining limitations and proposing directions for future research on human AI translation ecologies. Throughout, our aim is to consolidate insights into a coherent, testable vocabulary that supports cumulative progress across tools, settings, languages, and research communities.

Uncertainty in CTS: from proxy to construct

In CTS, uncertainty is typically indexed by proxies ambiguity, perceived difficulty, and risk rather than treated as an explicit latent construct with operational definitions. Consultation studies demonstrate how translators mobilize both linguistic (terminology, collocations) and extralinguistic (world knowledge, situational frames) resources to reduce uncertainty at critical junctures [2], [9]. Product-focused metrics, while indispensable for quality assurance, struggle to register process-level costs [3]. Reception studies underscore that these costs are not neutral: poor process control can externalize burden onto readers [10], [11].

Human machine divergences and diagnostic predictors

Where human and machine outputs diverge, the fault lines often lie in morphosyntax and pragmatics [7]. Parallel advances link computational predictors to human effort: surprisal correlates with reading time and production difficulty, offering a bridge between model-internal probabilities and human processing [6], [12]. Pedagogically, this supports AI literacy as diagnostic reading: recognizing risk zones, hypothesizing failure modes, and planning interventions [4], [13], [5]. Industry-oriented work emphasizes expectation management and risk-aware deployment [14].

Toward a probabilistic lexicon: QT

QT provides a minimal vocabulary for process reasoning:

- Superposition: coexistence of viable renderings, implicitly represented as a distribution conditioned by co-text, genre, and domain knowledge.
- Collapse: context-triggered stabilization of one option, often following activation of a context packet (briefs, style guides, terminology, exemplars, audience models).
- Entanglement: cross-level dependencies in which a local lexical choice reconfigures syntax and propagates into discourse patterns, with measurable effects on cohesion and style.

What are the conceptualizations, operationalizations, and measurements of uncertainty in translator-cognition studies from 2020 to 2025?

To what extent do prevailing cognitive models of translation account for systematic human–LLM divergences, and in what ways could a Quantum Translation (QT) heuristic offer a complementary framework to address any identified gaps?

Method

Review Design and Protocol

This review followed a pre-specified protocol and was reported in accordance with PRISMA 2020. The protocol defined the information sources, finalized search strings, set inclusion and exclusion criteria, and described a quality-appraisal rubric together with screening steps (identification, screening, eligibility, inclusion) and a structured data-extraction template to ensure transparency and replicability. Searches covered Scopus, Crossref (for metadata enrichment and forward citation checks), and Google Scholar (for supplementary recall); Web of Science was not queried because of institutional policy. The coverage window was 1 January 2020 to 1 October 2025, and the last search was executed on 13 October 2025 (UTC); no automated alerts were monitored after that date. The search strategy was piloted and then fixed prior to the full run. Queries combined controlled terms and free-text synonyms for three concepts: translation process and cognition; uncertainty, ambiguity, and risk; and process-tracing instruments. Truncation and proximity or field operators were used where supported. An example Scopus query was:

1. Scopus (Title/Abstract/Keywords)
   (("translator cognition" OR "translation process") AND (uncertainty OR ambiguity OR "decision-making" OR risk) AND ("large language model" OR "LLM" OR "machine translation" OR neural)) AND 2020–2025
2. Crossref (metadata)
   translator cognition; uncertainty; translation; decision-making; large language model; filters: 2020–2025; journal article.
3. Google Scholar (phrase + Boolean)
   "translator cognition" (uncertainty OR ambiguity OR "decision-making") ("LLM" OR "machine translation") 2020-2025

and the first approximately 200 results per query were screened after de-duplication. Crossref was used to resolve DOIs and titles, to harvest reference lists for backward citation-chasing, and to retrieve citing records for forward citation-chasing. The eligibility criteria (summarized in Table 3) admitted peer-reviewed empirical studies on human translation or post-editingthat reported process-level data (such as eye movements, keystrokes, screen captures, or think-aloud protocols) and that addressed uncertainty, ambiguity, or risk explicitly or implicitly. Reviews, theoretical essays without data, theses, preprints without peer review, tool demonstrations without human data, and records outside the 2020–2025 window were excluded. Records were de-duplicated by DOI, title, and author. Two reviewers independently screened titles and abstracts and then full texts against the criteria; disagreements were resolved by discussion, with recourse to a third reviewer when needed. A PRISMA flow diagram documents counts at each

Kajian Linguistik dan Sastra Vol. 10 (2) (2025) 234–258

stage.

Data were extracted with a structured template capturing bibliometrics, sample characteristics (for example, expertise and language pair), task type, instruments and signals (pauses, regressions, revisions), operationalizations of uncertainty, and main findings relevant to process-level inference. Quality appraisal used a five-item rubric covering design clarity, instrument validity, data transparency, analytic adequacy, and bias control. Twenty percent of studies were double-coded; inter-rater agreement (Cohen's kappa) was targeted at 0.80 or higher, with consensus reconciliation for discrepancies. Findings were synthesized thematically with attention to instrument-signal mappings (see Table 1) and to operational proxies of uncertainty.

Table 1. Instrument-to-Signal Mapping

Instrument	**Primary Signals (Proxy for Cognitive Effort)**	**Time Resolution**	**Key Advantages**	**Limitations**
Eye Tracking (ET)	Fixations, Fixation Duration, Regressions, Pupil Dilation	Very High	Measures attention and initial processing on screen; captures reading behavior and cognitive load (via pupil size)	Sensitive to calibration; subject to gaze loss; does not capture off-screen or internal thought processes
Keystroke Logging (KT)	Pauses, Insertions, Deletions, *Undo/Redo* Sequences, Lexical Substitutions	High	Measures production time, revisions, and technical effort; essential for identifying candidate renderings (Superposition)	Cannot track reading/orientation time without typing; results are influenced by the translator's typing skill (non-cognitive factors)

Kajian Linguistik dan Sastra Vol. 10 (2) (2025) 234–258

Instrument	Primary Signals (Proxy for Cognitive Effort)	Time Resolution	Key Advantages	Limitations
Screen Capture/Recording	Opening of Termbases, Glosseries, Concordances, Search Queries, Screen Navigation	Medium	Records context mobilization and information-seeking behavior; crucial for identifying the activation of a "context packet"	Does not provide continuous or fine-grained cognitive data (only discrete events)

Where measures were comparable, the synthesis noted convergence between human process traces and computational predictors such as surprisal. Heterogeneity in designs and measures precluded meta-analysis; therefore, effect directions and consistencies are reported narratively, supported by structured tables and figures.

PRISMA counts and flow diagram

Searches yielded 1,261 records from databases plus 5 from snowballing, with a detailed breakdown provided in Table 2. After de-duplication (by DOI/title/author), 1,261 unique records were screened at title/abstract; 1,239 were excluded. Twenty-two (22) full texts were assessed and included; none were excluded for quality. A PRISMA-style flow diagram (Figure 2) documents counts at each stage.

Table 2 Database Queries and Yields (2020–2025)

Database / Source	Search Strings (as listed)	Retrieved	Duplicates Removed	Records Screened	Records Excluded	Full Texts Assessed	Studies Included
Scopus	TITLE-ABS-	56	–	56	54	2	2

Crossref	KEY(...) metadata filters 2020–2025	1,000	–	1,000	996	4	4
Google Scholar	phrase + Boolean	200	–	200	189	11	11
Other (Snowballing)	backward/forward chasing	5	5	0	0	5	5
Total	–	1,261 + 5 = 1,266	5	1,261	1,239	22	22

Table 3 Inclusion and Exclusion Criteria (per scaffold)

Criterion	**Inclusion**	**Exclusion**
Publication type	Peer-reviewed journal article	Conference/workshop papers; theses; book chapters; reviews/editorials/opinion
Focus	Human translator cognition; human–LLM workflows	Purely engineering/system papers without translator focus
Window	2020–2025	<2020 or >2025
Language	English	Other languages without accessible full text
Availability	Abstract/full text retrievable	Unretrievable

Kajian Linguistik dan Sastra Vol. 10 (2) (2025) 234–258

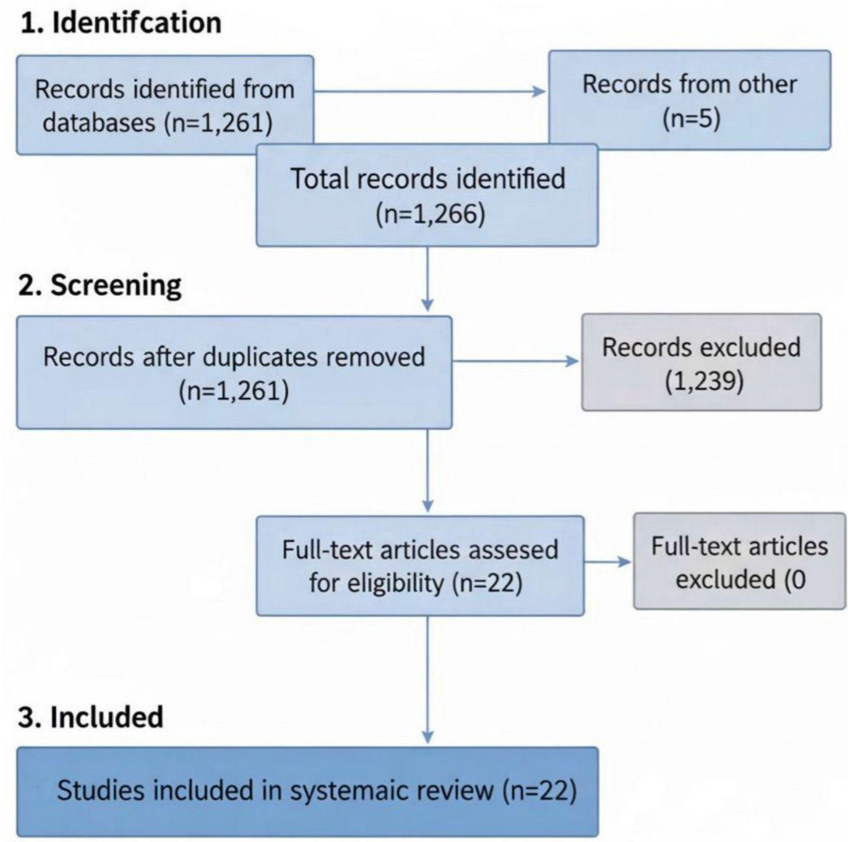

Figure 2 Prisma-Style Flow (Final n = 22)

Results and Discussion

This study's Results are organized around five interconnected themes that collectively address the research questions. The discussion synthesizes evidence from the reviewed corpus to build a coherent argument, moving from the conceptualization of uncertainty to its methodological, pedagogical, and professional implications, and culminating in a proposal for a new theoretical framework.

Theme 1: Conceptualizing and Measuring Uncertainty as a Core Cognitive Construct

Our analysis verifies that uncertainty functions as a fundamental, if frequently implicitly articulated, concept in modern process-oriented translation studies (RQ1). This is not represented by explicit probabilistic frameworks. Instead, it is indexed by its observable correlates: lexical and syntactic ambiguities, task-perceived difficulty, and expected risk. The influence on the translation process is well documented in the literature. Integrating methodologies of higher resolution, especially those that combine eye-tracking and screen-capture video, shows the extent to which localized constraints affect translators' information-seeking behavior. These studies illustrate distinct cognitive profiles, detailing the flexible, responsive nature of translators in their consulting patterns, where they alternate between specific queries on terminology and broader questions about the core ideas to resolve emerging cognitive gaps [2], [9]. A prominent result highlighted in numerous studies is the methodological inadequacy of solely analyzing the product to understand this phenomenon. [3] Persuasively contend that the seeming fluency of a finished translation may conceal a laborious and unpredictable production process.

This mismatch highlights the need to triangulate data from many process traces (e.g., keystrokes, eye movements, temporal pauses) to create a more precise model of cognitive effort. Furthermore, the potential consequences of unchecked ambiguity are considerable. Extending the study to the other end of the communication chain, reader-side evidence demonstrates that poorly aligned or pragmatically erroneous translations significantly hinder information absorption for the end user [10], [11]. This important finding reconceptualizes uncertainty as not merely a burden to the translator; it also represents a cognitive strain that can either be alleviated during the task or transferred to the reader, with noteworthy implications for communicative effectiveness. The occurrence of multiple reasonable frameworks and information requests based on a single point of uncertainty illustrates superposition as proposed in quantum theory of communication.

Theme 2: Evaluative innovation and methodological rigor in an uncertain environment

Measuring uncertainty requires simultaneous attention to methodological and evaluative rigor. Our findings underscore an increasing agreement that research outcomes are significantly influenced by experimental design. Variables such as sample size, text typology, and translation directionality (L1 vs. L2) significantly influence cognitive effort and output quality, necessitating caution against over-generalization from limited studies [15], [16], [17]. This highlights the significance of design-conscious assessment in both human and machine translation studies. The integration of human cognition with machine learning is catalyzing important methodological innovations. [6] Observed that some of the computational features, such as surprisal and attention scores embedded within Neural Machine Translation (NMT) systems and Large Language Models (LLMs), can predict the difficulty level of information for users quite accurately. This establishes a crucial link, facilitating the association of cognitive effort a normally inaccessible internal human state with external computable, probabilistic data. This connection possesses instant diagnostic value. For example, analyses of human machine translation divergences in particular linguistic domains such as morphosyntax have evolved beyond mere description; they can identify systematic and predictable "risk zones" where machine outputs are prone to errors or non-idiomatic expressions, [7] facilitating more focused post-editing and quality assurance strategies. The ability to predict and mitigate the challenges encountered by a human processor by particular computer signals provides an empirical validation to the notion of collapse, where a piece of data, deriving from either a machine or human source, resolves an uncertain cognitive state.

Theme 3: Imperatives for Teaching Fostering Ai Literacy and Metacognitive Tendencies

The differences between humans and AI, as well as the dangers each machine poses, affects translation education. There is a notable shift in the pedagogy from tool-agnostic instruction to advanced AI literacy. This encompasses far more than the capacity to perform basic functions; it emphasizes the importance of learning the nuances of diagnostic and reflective assessment tools [4], [5]. The objective is to train translators to examine AI-generated texts critically, identify minor errors, and evaluate AI-constructed arguments to determine their validity, usefulness, or potential for augmentation. The current research shows that successful technology integration requires more than just technical skills; it demands certain basic psychological qualities as well. According to [18], critical thinking, cognitive flexibility, and professional self-efficacy are important factors that influence a translator’s ability to manage human-AI relations. Therefore, instruction should train students not only to wield certain tools but to understand the rationale behind their application.

This means that training should not only educate students how to use tools, but also why they should use them. This will help them build the strong and critical thinking abilities they need to stay sharp in a world where work is becoming more automated. Translators need to keep track of the whole human-AI workflow as one system, because one small change can change the whole text. This shows how important entanglement is by illustrating that translation choices are rarely made in a vacuum; instead, they are part of a complicated text structure.

Theme 4: Theoretical integration in the development of a probabilistic translation lexicon (QT)

Based on the previous issues, the research indicates the viability of a more organized, probabilistic lexicon to elucidate the translation process. We propose that metaphorical constructs originating from quantum theory offer a stimulating yet constructive framework for what we designate as Quantum Translation (QT). This is supported by three findings that all point in the same way. The ability of LLMs to simulate human-like staged decision-making by strategic prompting [8] suggests that the translation process can be seen as a sequence of controlled state reductions. Second, the finding that surprisal and attention measures significantly influence variability in human output and reading times [6] demonstrates a direct relationship between probabilistic information and cognitive effort. Third, the long-established finding of context-triggered collapse, where a single piece of information can clarify several ambiguities [1] is analogous to the measurement effect in quantum mechanics.

The QT framework thinks about the process in terms of three main ideas:

1. Superposition: An unclear source segment is a superposition of several possible translation possibilities, which shows the process-level uncertainty that is not obvious in the final product.
2. Collapse: The cognitive process of decision-making, initiated by contextual knowledge or consultation, "collapses" this superposition into a singular, definitive representation.
3. Entanglement: Decisions about translation aren't independent; choices made at one level (like lexical) affect other levels (like syntactic and stylistic) in ways that aren't obvious, making an entangled, interdependent system.

Theme 5: Field-Level Translation Integrating Research into Professional Practice and Policy

Finally, the discussion turns to the actual application of these findings at the industry level. There is a substantial agreement between academic research and suggestions arising from industry-focused investigations. Professional perspectives strongly advocate for risk-aware MT integration and disciplined expectation management [14]. This call for pragmatism directly reflects the empirical findings of systematic "risk zones" and human-MT divergences. The industry's need for nuanced quality control aligns perfectly with the academic argument for a holistic, process+product evaluation model that is sensitive to contextual factors like text type and directionality. Consequently, the research not only enhances academic theory but also establishes an evidence-based framework for creating more intelligent, sustainable, and human-centered translation workflows in the professional domain.

Based on these thematic findings, which collectively highlight the need for a more structured, probabilistic lexicon to account for cognitive uncertainty, the following section moves from the conceptual to the concrete. It instantiates the Quantum Translation (QT) heuristic by operationalizing its core components superposition, collapse, and entanglement into specific, measurable constructs. This operationalization serves to render the QT framework empirically testable and directly applicable to future process-oriented research. This section translates the Quantum Translation (QT) framework into a testable empirical model by defining three core, measurable constructs. We operationalize superposition as candidate entropy, collapse as time-to-collapse, and entanglement as cross-level coupling, with formal definitions provided in Table 4. For each construct, we provide a formal definition, a clear procedure for estimation, and specific, testable predictions. The section concludes by outlining study designs and analysis plans to validate this heuristic.

Table 4 Operational Definitions for QT Metrics

Variable	**Operational definition**	**Unit / Index**	**Data source(s)**	**Estimation summary**	**Primary predictions / hypotheses**
Candidate entropy	Dispersion of concurrently viable translation options at time *t* for a unit (token/phrase/segment); formalized as Shannon entropy	Bits ($\log_2$) or nats (ln); also report derivative/ slope (per s) and AUC (bit·s).	Key-logging (keystrokes, revisions), eye-tracking (fixation clusters	(1) Define decision window (e.g., segment start → stable commitment). (2) If using	Enriched context packets (brief/termbase/style guide) and higher expertise → lower peak, faster half-life, steeper

Kajian Linguistik dan Sastra Vol. 10 (2) (2025) 234–258

Variable	Operational definition	Unit / Index	Data source(s)	Estimation summary	Primary predictions / hypotheses
	over candidate probabilities conditioned on source + active context packet. Report dynamic features: peak entropy, half-life after context activation, post-context slope, AUC over the decision window.		indicating alternative planning), screen capture; optional model probes (LLM top-k with probs) aligned to the same timeline.	model probes, obtain top-k candidate distribution at each *t*; else approximate option weights from human traces (e.g., competing partial strings + pause-weighted likelihood). (3) Compute H(t); extract peak, half-life after context insertion, slope, and AUC.	negative slope. Systematic profile differences across HT / PE / LLM; novices show higher peaks and slower decay.
Time-to-collapse (TtC)	Latency from segment onset or from context-packet activation to first stable commitment on	Seconds; survival/hazard metrics (e.g., median	Key-logging (onset → final commit timestam	(1) Mark t_0 (segment start) and *t_ctx* (context activation).	Richer context and higher expertise → shorter TtC (HR>1 vs baseline). PE

Variable	Operational definition	Unit / Index	Data source(s)	Estimation summary	Primary predictions / hypotheses
	a rendering (no reversals beyond Δ≥τ ms and no competing edits for a grace window). Model at the trial level.	TtC, HR with 95% CI).	ps, backspace bursts), eye-tracking (final fixation dwell before commit), on-screen markers for context activation.	(2) Define stability rule (e.g., no edits ≥800 ms + no alternative switch). (3) Estimate Kaplan–Meier curves and Cox models with expertise, mode (HT/PE/LLM), and context richness as factors; report HR, CI, diagnostics.	typically shortens TtC relative to HT for routine segments; LLM shows very short “apparent” TtC but may carry downstream edits (flag in notes).
Cross-level coupling	Strength of dependence between local lexical commitments and higher-level changes (syntax/discourse/genre register) within/between segments; captures “entanglement”	Bits (conditional mutual information) or standardized regression coefficients (β); optionally report odds ratios	Aligned sequences: committed tokens/phrases + syntactic parses, discourse coherence metrics, register/	(1) Construct time-aligned features (lexical choice events; subsequent syntactic/discourse shifts). (2) Estimate CMI(lex → syntax/disco	controls) and/or multilevel regressions with random effects (segment/participant). (3) Report effect sizes, CIs, and variance components.

Kajian Linguistik dan Sastra Vol. 10 (2) (2025) 234–258

Variable	Operational definition	Unit / Index	Data source(s)	Estimation summary	Primary predictions / hypotheses
	across linguistic levels.	for categorical shifts.	style features; condition on mode (HT/PE/LLM) and context richness.	urse	

Candidate Entropy as Superposition

Candidate entropy is the dispersion of viable renderings for a unit uat time t. Greater dispersion indicates a broader superposed state in which multiple alternatives remain credible. The estimation proceeds by constructing a candidate set for each unit on a shared time grid. Three complementary evidence streams contribute to this set. First, keystroke logging yields near synonymous substitutions, back and forth lexical changes, and undo redo sequences that reveal candidates under consideration. Second, screen capture records the opening of termbases, glossaries, and concordanced examples that introduce additional candidates even when they are not typed. Third, model probes query a machine translation system or a large language model for context conditioned alternatives so that plausible but untyped options are represented. Each candidate iat time treceives a normalized probability p_i (t)derived from observed frequencies, from model scores converted to probabilities through a calibrated softmax, or from a validated fusion of observation and model evidence. Candidate entropy is then computed as Shannon entropy, candidates. In addition to pointwise entropy, the analysis summarizes the peak value, the time to half peak, the post context slope after activation of a context packet, and the area under the entropy curve for the unit. Expert translators show lower peaks, faster halflife, and steeper post context declines once a context packet is available (as visualized in Figure 3). Trajectories differ by text type and by direction of translation, consistent with prior evidence in the literature referred to in the reference list.

$$\mathrm{H}(X) := -\sum_{x\in\mathcal{X}} p(x)\log p(x),$$

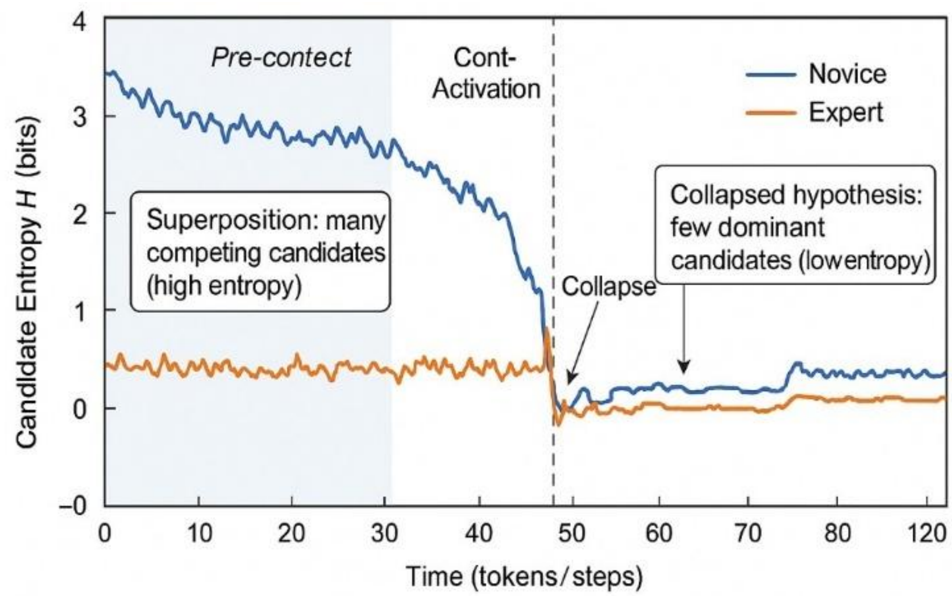

Figure 3 Dynamic Trajectory of Candidate Entropy for Novice and Expert

Time to Collapse as Decision Dynamics

Time-to-collapse (TtC) measures the latency of the decision-making process. It is the duration from when at least two viable translation options first appear for a unit until a final choice is stabilized. A choice is considered stable when it is no longer revised and the translator has moved on to subsequent parts of the text. The measurement procedure involves two key steps. First, we detect decision onset, defined as the moment entropy first rises above a baseline or when at least two candidates surpass a probability threshold. Second, we detect commitment, defined as the final edit made to the unit before a sustained period of forward progress. To ensure accuracy, data from eye-tracking, keystroke logging, and screen capture are aligned on a common timeline. The distribution of TtC values can be analyzed using survival and hazard models. Our primary hypothesis is that the availability of enriched context (e.g., style guides, termbases) will significantly shorten TtC. Conversely, factors that increase cognitive load, such as high interface clutter, are predicted to lengthen TtC.

Cross Level Coupling as Entanglement

Cross-level coupling quantifies the "entanglement" of translation decisions. It measures the strength of the dependency between local, word-level choices (lexical decisions) and subsequent adjustments at higher levels, such as sentence structure (syntactic packaging) and overall text flow (discourse organization).

To measure this, our analysis tracks how controlled lexical substitutions propagate into structural shifts across the text. We employ two complementary analytical frames:

1. Structural Analysis: This frame uses edit distance, syntactic parsing, and discourse cohesion metrics (e.g., connectives, referential chains) to track changes in form and structure.
2. Information-Theoretic Analysis: This second frame quantifies the dependency using conditional mutual information (CMI) and multilevel regression. This approach measures the relationship between a lexical decision and a structural outcome while statistically controlling for contextual factors.

We hypothesize that human translators exhibit task-sensitive coupling that is finely tuned to register and communicative goals (skopos). In contrast, we predict that LLM outputs will show systematic divergences, particularly in morphosyntax, when non-local constraints are present. These divergences should manifest as weaker coupling effects and reduced conditional mutual information compared to human translation.

Study Designs and Analysis Plans

To empirically validate the Quantum Translation (QT) framework, we propose several experimental designs. These designs are structured to manipulate key factors believed to influence cognitive processes and uncertainty in translation:

1. Context Richness Manipulation: This study would compare a minimal context condition against an enriched context condition, where participants are provided with a comprehensive brief, termbase, and style guide. The primary outcomes to be measured are Time-to-collapse (TtC), the trajectory of candidate entropy (specifically its peak, half-life, and post-context

slope), and final product quality (e.g., MQM/HTER).

2. Expertise Comparison: A second design would compare participant groups based on their level of expertise: Novice, Intermediate, and Professional translators. This study aims to map how expertise modulates TtC, entropy peak and decay, the number of edits per segment, and the strength of cross-level coupling.
3. Translation Mode Analysis: This design would investigate differences across three common working modes: Human Translation (HT) from scratch, traditional Post-Editing (PE), and LLM-assisted Post-Editing (LLM+PE). The focus would be on how these modes affect TtC, total editing effort, cross-level coupling, and discourse coherence metrics.
4. Translation Directionality: This study would compare translation processes between L1→L2 (forward) and L2→L1 (inverse) directions. Key outcomes would include differences in TtC, entropy patterns (peak and half-life), and the distribution of error types based on a predefined taxonomy.
5. Text Genre and Register: To assess task-based variation, this design would compare processes across diverse text genres, such as Technical, Legal, Marketing, and Literary texts. The primary investigation would focus on how cross-level coupling (lexis→syntax/discourse), coherence, consistency, and style adherence vary according to specific genre constraints.
6. Context-Packet Timing: This study would manipulate when contextual information is provided: either pre-activated (available from the start) or via mid-segment insertion (introduced at a moment of high uncertainty). This design directly tests the "collapse" mechanism by measuring the entropy half-life post-context and identifying any hazard shift in TtC.
7. Risk Band Triage: Finally, a study would pre-assess segments into Low, Medium, and High uncertainty bands (based on pre-assessed entropy levels). This tests the predictive validity of the QT heuristic by measuring whether high-entropy segments correlate with increased editing effort, higher residual entropy at commitment, and a greater number of downstream revisions.

Kajian Linguistik dan Sastra Vol. 10 (2) (2025) 234–258

Implementation Specifications Suitable for Immediate Execution

The All modalities share a single timestamp source. A simple sensor fusion procedure with linear interpolation maps keystrokes, gaze events, and screen capture markers to a common time grid. All events are stored in a tall table with columns for time, event type, unit identity, translator identity, and context metadata. Each unit uses a moving window that reflects the most recent input events. Near synonymous substitutions are grouped with an embedding based cosine threshold such as 0.8. Entries from termbases and glossaries that are open in the same screen region are added to the set. Model probes use identical context windows to maintain informational equivalence between human and model derived candidates. Observational evidence and model evidence are combined by a weighted average and then normalized. Weights are selected by cross validation on a labeled subset. Reliability diagrams assess calibration and isotonic regression or Platt style correction is applied if needed. Onset is the first time entropy exceeds the unit baseline by a prespecified delta or two candidates surpass a probability threshold τ. Commitment is the final edit before stable downstream progression within a grace window such as 30 seconds or two subsequent units completed without backtracking.

Baselines, thresholds, and window lengths are set in a pilot study and locked prior to the main data collection. Structural change is quantified with a syntactic parser and with cohesion measures. Conditional mutual information is computed between lexical decisions and structural outcomes while conditioning on context. A parallel multilevel regression predicts structural change from lexical decisions with random intercepts for translators and segments. Entropy is analyzed with mixed models that represent time using splines and include interactions among time, context richness, expertise, and direction. Time to collapse is analyzed with proportional hazards models accompanied by checks of model assumptions and by accelerated alternatives if needed. Coupling is analyzed with the combination of conditional mutual information and multilevel regression. All hypotheses are preregistered and all tuning choices are recorded before inspecting the main outcomes. The project releases a preprocessing manifest and fully reproducible scripts. A session level quality dashboard is provided with indicators for proportion of lost gaze, synchronization deviation, proportion of excluded segments, and any departures from protocol.

Conclusion

This study places uncertainty at the heart of translational action, elevating it from a mere proxy such as ambiguity or difficulty to a powerful, if figurative, quantum-like heuristic. We demonstrate how an excessive focus on the final product often obscures critical signals in the cognitive process, such as cognitive load and recursive drafting. To bridge this gap, we propose Quantum Translation (QT): a conceptual framework that views translation as a superposition of multiple candidate meanings, which then "collapses" into a single final choice triggered by context, where every choice is "entangled" with other decisions. This framework is more than a metaphor; it is measurable through concrete metrics like candidate entropy, time-to-collapse, and cross-level coupling. Ultimately, QT offers a bridge from theory to practice. It equips educators to train diagnostically "AI-literate" translators and encourages professionals to build human-AI collaborative workflows that are more transparent, accountable, and aware of the underlying cognitive processes.

Reference

[1] Z. Zu, “The Right Contextual Information Determining the Success of Communication on Translation,” *Appl. Transl.*, vol. 15, no. 2, pp. 12–16, 2021, doi: 10.51708/apptrans.v15n1.1423.

[2] Y. Cui and B. Zhen, “English-Chinese Translation: an Eye-Tracking, Screen-Recording and Retrospective Study,” *Perspect. Stud. Transl. Theory Pract.*, vol. 29, no. 5, pp. 740–760, 2021, doi: https://doi.org/10.1080/0907676X.2020.1760899.

[3] B. Vanroy, M. Schaeffer, and L. Macken, “Comparing the Effect of Product-Based Metrics on the Translation Process,” *Front. Psychol.*, vol. 12, no. August, pp. 1–16, 2021, doi: 10.3389/fpsyg.2021.681945.

[4] H. Mei and H. Chen, “Assessing Students’ Translation Competence: Integrating China’s Standards of English With Cognitive Diagnostic Assessment Approaches,” *Front. Psychol.*, vol. 13, no. March, pp. 1–13, 2022, doi: 10.3389/fpsyg.2022.872025.

[5] J. Zhang and S. Doherty, “Investigating Novice Translation Students’ AI Literacy in Translation Education,” *Interpret. Transl. Train.*, vol. 19, no. 3–4, pp. 234–253, 2025, doi: https://doi.org/10.1080/1750399X.2025.2541478.

[6] Z. W. Lim, E. Vylomova, C. Kemp, and T. Cohn, "Predicting Human Translation Difficulty with Neural Machine Translation," *Trans. Assoc. Comput. Linguist.*, vol. 12, pp. 1479–1496, 2024, doi: 10.1162/tacl_a_00714.

[7] J. Luo, C. Cherry, and G. Foster, "To Diverge or Not to Diverge: A Morphosyntactic Perspective on Machine Translation vs Human Translation," *Trans. Assoc. Comput. Linguist.*, vol. 12, pp. 355–371, 2024, doi: 10.1162/tacl_a_00645.

[8] Z. He *et al.*, "Exploring Human-Like Translation Strategy with Large Language Models," *Trans. Assoc. Comput. Linguist.*, vol. 12, pp. 229–246, 2024, doi: 10.1162/tacl_a_00642.

[9] Y. Cui and B. Zheng, "Extralinguistic Consultation in English–Chinese Translation: A Study Drawing on Eye-Tracking and Screen-Recording Data," *Front. Psychol.*, vol. 13, pp. 1–15, 2022, doi: 10.3389/fpsyg.2022.891997.

[10] B. Whyatt, E. Tomczak-Łukaszewska, O. Witczak, and O. Lehka-Paul, "Readers Have to Work Harder to Understand a Badly Translated Text: an Eye-Tracking Study into the Effects of Translation Errors," *Perspect. Stud. Transl. Theory Pract.*, vol. 33, no. 5, pp. 1085–1105, 2024, doi: 10.1080/0907676X.2024.2418016.

[11] B. Whyatt, O. Witczak, E. Tomczak-Łukaszewska, and O. Lehka-Paul, "The Proof of the Translation Process is in the Reading of the Target text: An Eyetracking Reception Study," *Ampersand*, vol. 11, 2023, doi: https://doi.org/10.1016/j.amper.2023.100149.

[12] N. J. Smith and R. Levy, "The effect of word predictability on reading time is logarithmic," *Cognition*, vol. 128, no. 3, pp. 302–319, 2013, doi: 10.1016/j.cognition.2013.02.013.

[13] S. O'Brien and M. Ehrensberger-Dow, "MT literacy—A Cognitive View," *Transl. Cogn. Behav.*, vol. 3, no. 2, pp. 145–164, 2020, doi: https://doi.org/10.1075/tcb.00038.obr.

[14] L. Nunes Vieira and E. Alonso, "Translating Perceptions and Managing Expectations: an Analysis of Management and Production Perspectives on Machine Translation," *Perspect. Stud. Transl. Theory Pract.*, vol. 28, no. 2, pp. 163–184, 2020, doi: 10.1080/0907676X.2019.1646776.

[15] Y. Wang and A. J. Daghigh, "Cognitive Effort in Human Translation and Machine Translation Post-Editing Processes A Holistic and Phased View," *FORUM. Rev. Int. d'interprétation traduction/International J. Interpret. Transl.*, vol. 21, no. 1, pp. 139–162, 2023, doi: https://doi.org/10.1075/forum.22009.wan.

[16] Y. Wang and A. J. Daghigh, "Effect of Text Type on Translation Effort in Human Translation and Neural Machine Translation Post-Editing Processes: Evidence from Eye-Tracking and Keyboard-Logging," *Perspect. Stud. Transl. Theory Pract.*, vol. 32, no. 5, pp. 961–976, 2024, doi: https://doi.org/10.1080/0907676X.2023.2219850.

[17] Y. Wang, S. Li, and Y. Z. Rasmussen, "Translators' Allocation of Cognitive Resources in Two Translation Directions: A Study Using Eye-Tracking and Keystroke Logging," *Appl. Sci.*, vol. 15, no. 8, pp. 1–21, 2025, doi: 10.3390/app15084401.

[18] X. Li, Z. Gao, and H. Liao, "The Effect of Critical Thinking on Translation Technology Competence among College Students: the Chain Mediating Role of Academic Self-Efficacy and Cultural Intelligence," *Psychol. Res. Behav. Manag.*, vol. 16, no. April, pp. 1233–1256, 2023, doi: 10.2147/PRBM.S408477.