INTERNATIONAL CONFERENCE **Dehli, India, 2026**
https://eijmr.org/conferences/

SEMANTIC ERRORS IN MACHINE TRANSLATION SYSTEMS AND THEIR CAUSES

Damir Shermatov
Bachelor Student, Faculty of English Philology and Translation Studies

(Ingliz Filologiyasi va Tarjimashunoslik Fakulteti),
Samarkand State Institute of Foreign Languages, Samarkand 140100, Uzbekistan

Abstract:

Semantic errors in machine translation (MT) occur when the output is fluent but fails to preserve meaning—by mistranslating word senses, omitting or adding content, or generating "hallucinated" information not supported by the source. Such errors are especially problematic because they can look grammatically perfect while being semantically wrong, making them hard to detect during post-editing. Research on neural machine translation (NMT) highlights that adequacy problems such as omissions and additions can appear in otherwise fluent output, masking meaning loss. Studies on hallucinations show that NMT can produce translations "untethered" from the input, sometimes triggered by rare tokens or distribution shifts. This article explains the main types of semantic MT errors, links them to underlying causes (lexical ambiguity, domain shift, data noise, decoding behavior, and model uncertainty), and illustrates them with short examples followed by clarifying commentary. It also summarizes evaluation practices (MQM-style categories and adequacy-focused learned metrics) and argues for targeted quality checks beyond surface fluency.

Keywords:

machine translation; semantic errors; adequacy; hallucination; word sense disambiguation; omissions; additions; domain shift; evaluation metrics; post-editing

Introduction

Machine translation has improved dramatically in fluency, especially with neural models, but fluency is not the same as meaning preservation. A translation can sound natural and still be semantically incorrect, which creates a risk in professional use: readers may trust a polished sentence even when it contains a wrong sense, missing information, or invented details. Studies on NMT adequacy errors report that omissions and additions are particularly prominent and may not be signaled by fluency problems—meaning the output can be "perfectly fluent" while semantically wrong. This is why semantic errors are often more dangerous than grammatical errors: they are less visible and can propagate misinformation in business, legal, medical, and academic settings. Research on hallucinations further reinforces this concern, showing that NMT systems can generate pathological translations that are not faithful to the source. The goal of this article is to explain what semantic errors in MT look like, why they happen, and how translators and users can detect them more reliably.

What counts as a "semantic error" in MT

In MT evaluation, semantic errors are usually discussed under "adequacy" or "accuracy": whether the translation preserves the meaning of the source. Modern error frameworks and studies commonly break adequacy problems into mistranslation, omission, and addition (and related subtypes). Semantic errors include (1) wrong word sense (lexical ambiguity failure), (2) wrong roles or relations (who did what to whom), (3) wrong polarity or negation, (4) omissions/under-translation (missing content), (5) additions/over-translation (extra content), and (6) hallucinations (content untethered from the source). Hallucinations are widely defined as unfaithful outputs that fabricate or introduce content not supported by the source. A key practical insight is that semantic errors can be invisible if you only read the target text; detecting them often requires comparing against the source or using adequacy-focused checks.

Core semantic error types with examples

Wrong word sense (lexical ambiguity). Example: Source: "He went to the bank after work." MT Output A: "U ishidan keyin bankka bordi." MT Output B: "U ishidan keyin daryo qirg‘og‘iga bordi." Both are grammatically plausible in Uzbek, but only one matches the intended sense. The supportive point is that the source word "bank" is polysemous, and MT must resolve the correct sense using context (money-related vs river-related cues). Even strong NMT systems struggle with word sense disambiguation for uncommon senses and low-context sentences, and reference sources note that lexical ambiguity remains a major challenge for MT.

Omission (under-translation). Example: Source: "The patient denied chest pain but reported dizziness." MT Output: "Bemor bosh aylanishini aytdi." Here the denial of chest pain disappears. The supportive point is that omission can flip clinical interpretation: "denied chest pain" is important negative information, and its loss can change decision-making. Research on adequacy errors in NMT emphasizes that omissions can appear inside fluent output and require careful source comparison to detect.

Addition (over-translation). Example: Source: "She visited the museum." MT Output: "U muzeyga sayohat qilib, gid bilan ekskursiya qildi." The system adds "with a guide" and "tour," which were not stated. The supportive point is that additions often sound plausible but introduce factual claims; in many domains this becomes misinformation rather than style. Adequacy studies treat additions as a central error class because they alter propositional content.

Hallucination (untethered output). Example: Source: "The meeting was postponed." MT Output: "Yig‘ilish ertaga soat 10:00 ga ko‘chirildi, chunki ob-havo yomon." Time and cause are invented. The supportive point is that hallucinations can look informative and confident, which makes them especially risky for users who do not cross-check with the source. Research from Google and MIT Press (TACL) documents that NMT can produce pathological, source-untethered translations and studies hallucination mechanisms and detection.

Why semantic errors happen: main causes

Lexical ambiguity and insufficient context. Many source sentences are short, underspecified, or ambiguous, so the model must guess a sense from limited cues. Reference overviews of MT note that lexical ambiguity and word sense disambiguation remain difficult for NMT, especially for rare senses or sparse context. This is one reason short segments (headlines, UI strings, chat messages) can yield more semantic errors: the model lacks signals to disambiguate.

Domain shift and out-of-distribution inputs. MT systems learn from training data distributions. When

the test domain differs—medical, legal, dialectal, informal chat—models can produce fluent but semantically off translations. Studies on translation difficulty and MT quality using WMT data discuss how source-text features relate to MT performance variation across languages and domains. Domain shift increases wrong-sense choices, terminology drift, and hallucination risk.

Training data noise and misalignment. Parallel corpora may contain misaligned segments, paraphrases, or inconsistent translations. Models can learn "good-sounding" target patterns that are only loosely tied to the source. This contributes to additions, omissions, and stylistic smoothing that harms adequacy. Research on MT error causes highlights that different evaluation criteria reveal different error types and that adequacy problems may be overlooked when fluency dominates perception.

Decoding behavior and uncertainty. Beam search and length biases can encourage shorter outputs (increasing omissions) or produce overly generic, "safe" phrasing (semantic smoothing). Over/under-translation is recognized as an adequacy issue worth detecting explicitly in evaluation pipelines. When the model is uncertain (rare terms, long-distance dependencies, unusual syntax), it may default to common patterns, which can silently drop or distort meaning.

Hallucination triggers. Research shows hallucinations can be triggered by rare tokens or perturbations and that NMT can generate outputs barely related to source inputs in pathological cases. More recent work investigates hallucination detection and interpretability, reinforcing that hallucination is a known failure mode rather than a rare anomaly.

How to detect semantic errors more reliably

Because semantic errors hide behind fluency, detection should prioritize adequacy checks. Human evaluation frameworks like MQM categorize errors under accuracy/adequacy (mistranslation, omission, addition), which aligns with what post-editors look for in practice. On the automatic side, learned metrics such as COMET are designed to correlate with human judgments and focus more on adequacy than older n-gram metrics, though metrics still have blind spots and often need complementary checks. The practical point is: do not rely on "looks fluent" or on one metric. Instead, combine (1) source-vs-target spot checks, (2) targeted term checks (especially names, numbers, negation, units), and (3) adequacy-focused evaluation or error annotation when stakes are high. Explainable evaluation research (e.g., xCOMET) also argues for moving from single scores to error insights, which is directly useful for diagnosing semantic problems.

Practical prevention strategies in real workflows

First, segment awareness: MT outputs should be reviewed in context, not only sentence-by-sentence, because many semantic errors involve cross-sentence reference (pronouns, ellipsis, discourse relations). Second, "high-risk trigger list": always verify negation, quantities, dates, named entities, and domain terms; these are common locations for semantic failure. Third, controlled post-editing: adequacy checks should include a quick scan for omissions/additions by aligning key content words between source and target. This is motivated by evidence that omissions/additions can occur in fluent NMT output. Finally, domain adaptation and terminology constraints: where possible, use domain-specific MT or glossaries to reduce ambiguity and terminological drift, because domain shift is a documented driver of quality variability.

Conclusion

Semantic errors in machine translation are not disappearing; they are evolving. As MT becomes more fluent, errors of meaning—wrong sense selection, omissions, additions, and hallucinations—become harder to detect and potentially more harmful. Research on adequacy errors shows that fluent NMT output can still omit or add content, requiring careful source-based checking. Hallucination research demonstrates that systems can generate unfaithful translations untethered from the source, sometimes triggered by rare tokens or distribution shifts. The most reliable response is a workflow that treats adequacy as primary: combine targeted human checks, error taxonomies (mistranslation/omission/addition), and evaluation tools that provide insight into semantic quality rather than only surface similarity scores.

References:

Agarwal, A., et al. (n.d.). Hallucinations in neural machine translation. Google Research.
Gupta, P., et al. (2021). Detecting over/under-translation errors for determining translation adequacy. arXiv.
Guerreiro, N. M., et al. (2024). xCOMET: Transparent machine translation evaluation. TACL/ACL Anthology.
Lee, K., et al. (2018). Hallucinations in neural machine translation. OpenReview (PDF).
Popović, M. (2021). On nature and causes of observed MT errors. ACL Anthology (MT Summit).
ScienceDirect Topics. (n.d.). Machine translation—overview (lexical ambiguity and WSD challenge).
Ustaszewski, M. (2019). Exploring adequacy errors in neural machine translation. ACL Anthology (PDF). Vardaro, J., et al. (2019). Translation quality and error recognition in professional contexts (MQM categories). MDPI.
Wan, Y., et al. (2022). Challenges of neural machine translation for short texts. Computational Linguistics (MIT Press).
Xu, W., et al. (2023). Understanding and detecting hallucinations in neural machine translation. TACL (MIT Press).