BER-Driven Semantic Resilience and Receiver-Side Recovery for Text Communication

1st David Panebianco
Department of Electrical, Electronic and Computer Engineering
University of Catania
Catania, Italy
david.panebianco@phd.unict.it

2nd Ludovica Beritelli
Dep. of Mathematics and Computer Science
University of Catania
Catania, Italy
ludovica.beritelli@phd.unict.it

3rd Stefano Antonio Amico
Department of Electrical, Electronic and Computer Engineering
University of Catania
Catania, Italy
stefano.amico@infoamico.com

4th Roberta Avanzato
Dep. of Electrical, Electronic and Computer Engineering
University of Catania
Catania, Italy
roberta.avanzato@unict.it

5th Francesco Beritelli
Dep. of Electrical, Electronic and Computer Engineering
University of Catania
Catania, Italy
francesco.beritelli@unict.it

***Abstract*—Semantic communication shifts the design objective of communication systems from exact bit reconstruction toward the preservation of meaning and receiver-side utility. This shift is particularly critical in scenarios where small bit-level perturbations can alter numbers, negations, or operational commands, transforming a message while maintaining high lexical similarity. This paper proposes a Bit Error Rate (BER)-driven semantic resilience framework for the controlled evaluation of text-based communication. The framework models the transmission pipeline from binary encoding to channel corruption, decoding, and optional receiver-side semantic recovery using Large Language Models (LLMs). We introduce a multi-level evaluation methodology combining Word Error Rate (WER), embedding-based Semantic Textual Similarity (STS), and Semantic Textual Similarity Gain (STS Gain). Unlike conventional purely bit-oriented approaches, the proposed protocol explicitly evaluates the trade-off between exact lexical reconstruction and semantic robustness. The results provide a reproducible testbed for benchmarking semantic recovery strategies under varying BER conditions.**

***Index Terms*—Semantic Communication, Text Transmission, Bit Error Rate, Receiver-Side Recovery, Word Error Rate, Sentence-BERT, 6G.**

I. Introduction

Classical digital communication systems are primarily evaluated through physical- and link-layer indicators such as bit error rate (BER), packet error rate, signal-to-noise ratio, and throughput. Shannon's mathematical theory of communication provides the fundamental basis for reliable transmission over noisy channels, while intentionally separating engineering information from semantic interpretation [1]. This abstraction remains essential, but it is not sufficient for emerging task-oriented services in which the receiver must interpret the message and act on it. In several operational settings, exact bit recovery is less important than preserving the information required for a correct decision [2]–[5].

Semantic communication has therefore become a central research direction for beyond-5G and 6G systems. Recent surveys describe semantic representation, task-oriented transmission, knowledge-assisted interpretation, semantic metrics, and implementation as major open challenges [3]–[5], [20]. Deep learning has enabled practical semantic transceivers for text and speech. DeepSC introduced a Transformer-based semantic communication system for text transmission [6], while DeepSC-S extended the paradigm to speech transmission by emphasizing essential information [7]. These works show that meaning-oriented transmission can be more robust than purely bit-oriented baselines in some channel conditions.

However, evaluating semantic degradation remains difficult. A corrupted message such as "administer 100 mg" instead of "administer 10 mg" may retain high lexical similarity while conveying a completely different, and potentially dangerous, meaning. Likewise, the loss of "not" in "do not enter the electrical room" can invert the intended command. These examples show that standard metrics like WER, BLEU, or ROUGE alone cannot fully capture semantic correctness [11]–[16]. A communication system may appear acceptable under generic lexical evaluations while still failing on specific entities that matter most for the end-user.

This paper proposes a compact BER-driven framework for evaluating semantic resilience and receiver-side recovery in text communication. The framework does not claim to replace channel coding or end-to-end semantic transceivers. Instead, it provides a measurement layer that links controlled bit-level corruption to textual degradation, semantic similarity loss, and recovery gain. The contribution is methodological and experimental: it defines a reproducible way to test how meaning survives or fails under increasing BER, and how much a neural receiver-side recovery stage can improve interpretation.

The main contributions of this paper are as follows:

- We investigate the relationship between bit-level corruption and semantic degradation under controlled BER conditions.

- We propose a unified evaluation framework combining lexical fidelity (WER) and semantic preservation (STS and STS Gain).
- We introduce a text-domain receiver-side recovery paradigm and evaluate its capability to restore meaning without operating at the physical layer.
- We provide a reproducible experimental protocol highlighting the trade-off between syntactic degradation and semantic robustness.

II. Related Work

Semantic communication was anticipated by Weaver's discussion of technical, semantic, and effectiveness levels of communication [2]. Modern research revisits this idea using neural representation learning, knowledge bases, and task-oriented objectives [3]–[5]. DeepSC is a key text-oriented reference because it formulates sentence-level semantic transmission with neural encoders and decoders [6]. DeepSC-S is also relevant as it targets semantic speech transmission and uses attention to emphasize essential information [7]. Robust semantic communication and compensation methods further confirm the need to study degradation and recovery under channel impairment [8].

The present framework differs from end-to-end semantic transceiver design. Its target is text transmission over binary channels, allowing controlled analysis of bit-level corruption and receiver-side text repair. This choice is compatible with modern communication pipelines where information, regardless of its original source (e.g., audio, sensors, text), is processed as discrete token sequences [17], [18]. In this paper, textual data is directly exposed to channel-induced corruption to isolate and evaluate the semantic impact of bit errors.

Evaluation metrics are another critical component. WER remains a standard because it measures substitutions, deletions, and insertions [15]. BLEU and ROUGE are widely used for machine translation and summarization, but they rely strongly on token overlap and can penalize valid paraphrases [13], [14]. Sentence-BERT enables efficient sentence-level semantic similarity through dense embeddings [11], while BERTScore compares candidate and reference tokens using contextual representations [12]. Semantic Textual Similarity (STS) benchmarks provide a foundation for meaning-oriented comparison [16]. However, relying on a single metric is often insufficient. Real-world communication standards, such as 3GPP TS 22.179 for Mission Critical services [19], emphasize reliability and operational availability. This operational context motivates a stronger evaluation approach: measuring not only whether a message is readable, but whether it preserves its core meaning under severe channel degradation.

III. System Model and Experimental Setup

A. System Architecture

The proposed framework considers a text-based communication system operating over a noisy channel with a controllable Bit Error Rate (BER) p_b.

TABLE I
MAIN NOTATION USED IN THE PROPOSED FRAMEWORK.

Symbol	Meaning
t_0	Reference text message
b	Binary sequence obtained from the reference text
p_b	Target bit error rate used by the channel model
$\tilde{b}$	Corrupted binary sequence
t_c	Text decoded from the corrupted bitstream (*noisy* baseline)
t_r	Text generated after receiver-side semantic recovery (*fixed*)
Γ	Optional domain context and recovery constraints

Let t_0 denote the reference text message. The message is encoded into a binary sequence, transmitted through a noisy channel, and decoded at the receiver. An optional semantic recovery module may refine the corrupted output.

The processing chain is defined as:

$$b = \mathcal{E}(t_0), \tag{1}$$
$$\tilde{b} = \mathcal{C}(b, p_b), \tag{2}$$
$$t_c = \mathcal{D}(\tilde{b}), \tag{3}$$
$$t_r = \mathcal{R}(t_c, \Gamma), \tag{4}$$

where $\mathcal{E}$ denotes encoding, $\mathcal{C}$ the channel corruption process, $\mathcal{D}$ decoding, and $\mathcal{R}$ the optional semantic recovery module. The set Γ represents optional contextual constraints. Table I summarizes the main mathematical notation used in this work.

The recovery module is not assumed to be perfect and may introduce modifications to the decoded text. Therefore, the evaluation considers both lexical fidelity and semantic preservation.

B. Implementation Details

To quantitatively evaluate the effect of the semantic-aware receiver, an experimental study was conducted. The objective is to compare the baseline scenario, where the received text (t_c) is directly analyzed after channel corruption, with a semantic-aware scenario, where the neural compensation module generates a refined output (t_r).

a) Dataset and Evaluation Procedure: The experimental dataset consists of 50 English sentences selected to cover diverse syntactic structures and varying sentence lengths. This design aims to reduce bias toward specific linguistic patterns and improve the robustness of the evaluation. For each sentence and BER level, the system computes metrics over multiple random seeds to account for the stochastic nature of bit corruption. Results are reported as averaged performance across the dataset.

b) Channel Model: Channel corruption is simulated using a controlled BER model. Each sentence is encoded using ASCII representation and transmitted as a binary sequence. Bit-level corruption is introduced by flipping each bit with probability p_b. The evaluated BER values are $\{5 \cdot 10^{-5}, 5 \cdot 10^{-4}, 10^{-3}, 5 \cdot 10^{-3}, 8 \cdot 10^{-3}, 10^{-2}\}$, allowing evaluation under both low-noise and highly degraded conditions.

c) Semantic Recovery Module: The recovery module $\mathcal{R}$ is implemented using a pre-trained sequence-to-sequence model (`ai-forever/T5-large-spell`). Given a corrupted input t_c, the model generates a corrected output t_r by exploiting contextual information in the sentence. The model operates purely in the text domain and does not attempt to reconstruct the original bitstream. Its objective is to improve linguistic coherence under channel-induced corruption.

C. Evaluation Metrics

The proposed framework evaluates performance using lexical and semantic metrics to capture both surface-level errors and meaning preservation.

Word Error Rate (WER) measures the minimum number of edit operations required to transform the hypothesis into the reference sentence:

$$WER = \frac{S + D + I}{N} \tag{5}$$

where S, D, I, and N denote substitutions, deletions, insertions, and the number of reference words, respectively [15].

Semantic similarity between two sentences is computed using embedding-based representations. Let $Sim(t_a, t_b)$ denote the cosine similarity between Sentence-BERT embeddings [11], [12]:

$$Sim(t_a, t_b) = \cos(\mathbf{e}_a, \mathbf{e}_b) \tag{6}$$

The Semantic Textual Similarity Gain (STS Gain) quantifies the relative semantic improvement introduced by the receiver-side recovery module:

$$\text{STS Gain (\%)} = \frac{Sim(t_0, t_r) - Sim(t_0, t_c)}{Sim(t_0, t_c)} \times 100 \tag{7}$$

IV. Results

The experimental results are reported in Table II, which summarizes the performance of the noisy and semantic-aware (fixed) scenarios across different Bit Error Rate (BER) levels. All results are averaged over fifty sentences from the test dataset.

TABLE II
COMPARISON BETWEEN NOISY AND FIXED SCENARIOS AT VARYING BER LEVELS

BER	WER Noisy	WER Fixed	STS Noisy	STS Fixed	STS Gain (%)
$5 \cdot 10^{-5}$	0.001	0.001	0.999	1.000	0.102
$5 \cdot 10^{-4}$	0.033	0.016	0.963	0.987	2.480
10^{-3}	0.060	0.031	0.933	0.982	5.282
$5 \cdot 10^{-3}$	0.234	0.156	0.792	0.904	14.204
$8 \cdot 10^{-3}$	0.336	0.224	0.703	0.863	22.768
10^{-2}	0.386	0.264	0.675	0.822	21.753

Overall, channel noise has limited impact at low BER values. For BER $\leq 10^{-3}$, Semantic Textual Similarity (STS) remains above 0.93, indicating strong semantic preservation despite bit-level corruption. In this regime, the semantic-aware module provides only marginal improvements, as the received signal already preserves sufficient structure for correct interpretation.

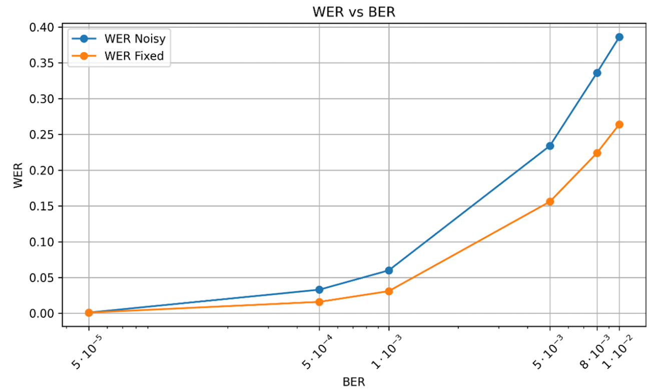

Fig. 1. WER vs BER comparison for Noisy and Fixed scenarios.

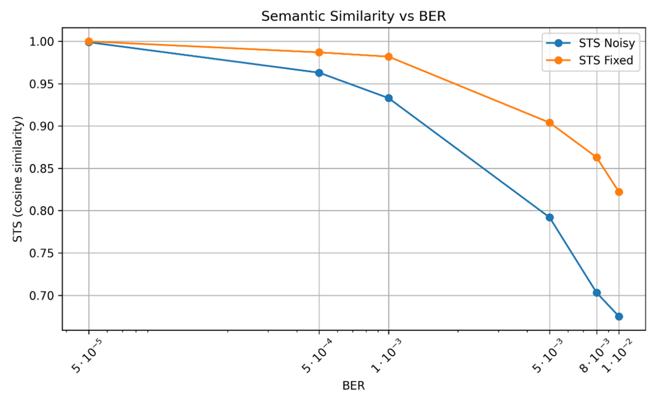

Fig. 2. STS vs BER comparison for Noisy and Fixed scenarios.

As BER increases, a clear divergence emerges between the two approaches. The noisy baseline exhibits severe degradation, including corrupted tokens and incomplete words, reflected in a sharp increase in Word Error Rate (WER). Figure 1 shows the evolution of WER as a function of BER for both noisy and semantic-aware scenarios.

In contrast, the semantic-aware model consistently reduces WER and improves STS, indicating partial reconstruction of corrupted content.

At high BER levels (e.g., $5{\cdot}10^{-3}$ to 10^{-2}), semantic benefits become more evident. The noisy signal reaches STS values as low as 0.675, while the semantic-aware output maintains values up to approximately 0.822, showing improved semantic preservation under severe channel impairment.

Figure 2 illustrates the behavior of Semantic Textual Similarity (STS) across different BER levels.

A consistent observation across all configurations is the trade-off between lexical accuracy and semantic preservation. Even when WER remains relatively high (e.g., 0.264 at 10^{-2}), STS remains comparatively stable, indicating that the recovery model prioritizes meaning preservation over exact word-level reconstruction.

The STS gain increases with BER, exceeding 20% in highly degraded conditions, confirming that semantic recovery be-

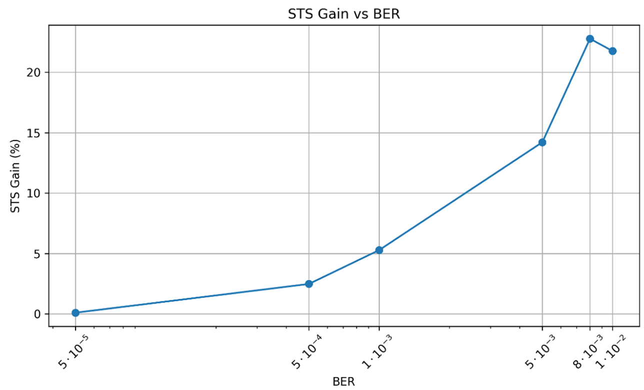

Fig. 3. STS Gain (%) vs BER.

comes increasingly beneficial as channel quality deteriorates.

Figure 3 highlights the semantic gain introduced by the recovery module as BER increases.

Overall, these results confirm that semantic-aware processing enhances robustness in noisy environments, particularly when traditional lexical reconstruction fails.

V. Discussion

The results highlight a fundamental property of semantic communication systems: Bit Error Rate (BER) and semantic degradation are related but not strictly equivalent. While BER measures physical corruption at the bit level, the semantic impact depends heavily on which specific tokens are affected.

Even under low BER conditions, errors might alter mission-critical terms (such as numbers, negations, or locations), leading to severe operational failures. Conversely, a higher Word Error Rate (WER) does not necessarily imply a complete loss of meaning if only non-essential words are corrupted. This mismatch strongly motivates the need for semantic-aware evaluation metrics beyond classical error rates.

The semantic recovery module demonstrates clear benefits in high-noise regimes, successfully bridging the gap between severe channel impairment and human or machine comprehensibility. However, it also introduces inherent risks. Language-model-based reconstruction may produce fluent yet incorrect outputs, including altered quantities, missing negations, or hallucinated entities. For this reason, in safety-critical contexts, explicitly signaling uncertainty may be preferable to producing a confident but incorrect reconstruction.

Several limitations remain in this study. The evaluation relies on a simplified independent bit-flip channel model and a relatively compact text dataset. Moreover, embedding-based similarity metrics (such as STS) are model-dependent and may not fully reflect human judgment in specialized operational domains, such as emergency response or healthcare. Finally, while this study focused on direct text transmission, validation on end-to-end speech-derived pipelines remains necessary.

VI. Conclusion

This paper presented a BER-driven semantic resilience framework for evaluating receiver-side recovery in text-based semantic communication systems.

The framework jointly analyzes lexical degradation, semantic similarity, and recovery gain under controlled channel impairments. Unlike traditional bit-level evaluation approaches, it explicitly focuses on the trade-off between exact lexical reconstruction and semantic robustness.

The experimental results show that semantic-aware recovery is highly effective under high-noise conditions, where traditional decoding fails to preserve meaning. However, the approach introduces potential risks related to hallucinated or inadvertently modified information, highlighting the future need for constrained and uncertainty-aware recovery mechanisms.

Future work will extend this study to real speech-to-text pipelines, incorporate more realistic channel models (including burst and correlated errors), and validate the proposed metrics through human expert evaluation in mission-critical scenarios.

References

[1] C. E. Shannon, "A Mathematical Theory of Communication," Bell System Technical Journal, vol. 27, no. 3, pp. 379-423, 1948; vol. 27, no. 4, pp. 623-656, 1948.

[2] W. Weaver, "Recent Contributions to the Mathematical Theory of Communication," in C. E. Shannon and W. Weaver, The Mathematical Theory of Communication. Urbana, IL, USA: University of Illinois Press, 1949.

[3] Z. Qin, X. Tao, J. Lu, W. Tong, and G. Y. Li, "Semantic Communications: Principles and Challenges," arXiv preprint arXiv:2201.01389, 2021.

[4] T. M. Getu, G. Kaddoum, and M. Bennis, "Semantic Communication: A Survey on Research Landscape, Challenges, and Future Directions," Proceedings of the IEEE, vol. 112, no. 11, pp. 1649-1685, Nov. 2024, doi: 10.1109/JPROC.2024.3520707.

[5] Z. Lu, R. Li, K. Lu, X. Chen, E. Hossain, Z.-F. Zhao, and H. Zhang, "Semantics-Empowered Communications: A Tutorial-cum-Survey," IEEE Communications Surveys & Tutorials, vol. 26, no. 1, pp. 41-79, 2024, doi: 10.1109/COMST.2023.3333342.

[6] H. Xie, Z. Qin, G. Y. Li, and B.-H. Juang, "Deep Learning Enabled Semantic Communication Systems," IEEE Transactions on Signal Processing, vol. 69, pp. 2663-2675, 2021, doi: 10.1109/TSP.2021.3071210.

[7] Z. Weng and Z. Qin, "Semantic Communication Systems for Speech Transmission," IEEE Journal on Selected Areas in Communications, vol. 39, no. 8, pp. 2434-2444, Aug. 2021, doi: 10.1109/JSAC.2021.3087240.

[8] Z. Weng and Z. Qin, "Robust Semantic Communications for Speech Transmission," arXiv preprint arXiv:2403.05187, 2024.

[9] A. Vaswani et al., "Attention Is All You Need," in Proc. Advances in Neural Information Processing Systems, Long Beach, CA, USA, 2017, pp. 5998-6008.

[10] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in Proc. NAACL-HLT, Minneapolis, MN, USA, 2019, pp. 4171-4186, doi: 10.18653/v1/N19-1423.

[11] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence Embeddings Using Siamese BERT-Networks," in Proc. EMNLP-IJCNLP, Hong Kong, China, 2019, pp. 3980-3990, doi: 10.18653/v1/D19-1410.

[12] T. Zhang, V. Kishore, F. Wu, K. Q. Weinberger, and Y. Artzi, "BERTScore: Evaluating Text Generation with BERT," in Proc. International Conference on Learning Representations, 2020.

[13] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, "BLEU: A Method for Automatic Evaluation of Machine Translation," in Proc. 40th Annual Meeting of the Association for Computational Linguistics, Philadelphia, PA, USA, 2002, pp. 311-318, doi: 10.3115/1073083.1073135.

[14] C.-Y. Lin, "ROUGE: A Package for Automatic Evaluation of Summaries," in Text Summarization Branches Out: Proc. ACL Workshop, Barcelona, Spain, 2004, pp. 74-81.

[15] A. C. Morris, V. Maier, and P. Green, "From WER and RIL to MER and WIL: Improved Evaluation Measures for Connected Speech Recognition," in Proc. Interspeech, Jeju Island, Korea, 2004, pp. 2765-2768, doi: 10.21437/Interspeech.2004-668.

[16] D. Cer, M. Diab, E. Agirre, I. Lopez-Gazpio, and L. Specia, "SemEval-2017 Task 1: Semantic Textual Similarity-Multilingual and Cross-lingual Focused Evaluation," in Proc. 11th International Workshop on Semantic Evaluation, Vancouver, Canada, 2017, pp. 1-14, doi: 10.18653/v1/S17-2001.

[17] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey, and I. Sutskever, "Robust Speech Recognition via Large-Scale Weak Supervision," in Proc. International Conference on Machine Learning, 2023.

[18] Z. Borsos et al., "AudioLM: A Language Modeling Approach to Audio Generation," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 31, pp. 2523-2533, 2023, doi: 10.1109/TASLP.2023.3288409.

[19] 3GPP, "Mission Critical Push To Talk (MCPTT); Stage 1," 3GPP TS 22.179, Technical Specification, 3rd Generation Partnership Project.

[20] G. Xin, P. Fan, and K. B. Letaief, "Semantic Communication: A Survey of Its Theoretical Development," Entropy, vol. 26, no. 2, Art. no. 102, 2024, doi: 10.3390/e26020102.