Available Online: 23 December 2025 DOI: 10.54254/2753-7080/2025.30810

Meaning retention effects of cognitive-semantic cue-enhanced real-time scaffolding for interpreting low-resource dialect input

Baixu Chen

School of Education, University of Glasgow, Glasgow, UK

3028820C@student.gla.ac.uk

Abstract. This study investigates how cognitive-semantic cue-enhanced real-time scaffolding supports meaning retention when interpreters work with low-resource dialect input. Building on theories of meaning construction and conceptual framing, we design a Computer-Assisted Interpreting (CAI) prototype that overlays dialect-sensitive semantic cues, frames, participants, and relational links, on top of Automatic Speech Recognition (ASR) transcripts. Twenty-four advanced interpreting students completed consecutive interpreting tasks in two low-resource dialect settings (Mandarin–Wu and Mandarin–Gan), across four conditions: no support, keyword glossaries, plain ASR transcripts, and the proposed cue-enhanced scaffold. Meaning retention is modeled as a weighted overlap between expert-annotated conceptual units and interpreted output, yielding a Meaning Retention Rate (MRR). The cue-enhanced condition achieves an average MRR of 0.847 ± 0.017, compared with 0.781 ± 0.021 for the no-support baseline, alongside a reduction in semantic error units from 14.6 ± 2.3 to 9.8 ± 1.6 per 100 segments. An ablation study shows that removing the semantic cue extractor reduces MRR by -0.033 on the training dialect and -0.061 on the transfer dialect. Qualitative feedback indicates that interpreters experience lower perceived cognitive overload and report more stable "macro-meaning" tracking. The findings suggest that cognitively grounded semantic cues can be integrated into CAI systems to improve meaning retention and transferability for low-resource dialect interpreting

Keywords: cognitive semantics, scaffolding, low-resource dialects, meaning retention

1. Introduction

Dialect interpreting is a cognitively demanding activity because interpreters must map non-standard phonology, regionally specific lexicon, and culture-bound expressions onto stable conceptual structures while working under strict time pressure [1]. For many low-resource dialects, there is limited parallel data, sparse lexicographic coverage, and few domain-specific preparation materials. These conditions amplify processing load and make it harder for interpreters to preserve fine-grained relational meaning rather than merely approximate propositional content [2].

At the same time, CAI tools and speech technologies have progressed rapidly. ASR systems and terminology managers can now provide live transcripts, keyword highlighting, and automatic term retrieval during interpreting sessions, especially in well-resourced languages and standard varieties. However, existing tools are largely lexical and surface-oriented. They rarely encode the conceptual frames, participant roles, and discourse relations that cognitive semantics regards as central to meaning construction, and they offer little explicit support for how interpreters mentally "chunk" dialect input into coherent situations and event structures.

This paper explores whether cognitively motivated semantic cues can be embedded into a real-time scaffolding interface in a way that measurably improves meaning retention in low-resource dialect interpreting [3]. We prototype a cue-enhanced CAI system that identifies dialect-sensitive semantic units and displays them in a compact, visually layered overlay synchronized with ASR transcripts. We then conduct a controlled experiment with advanced interpreting students working from Mandarin–Wu and Mandarin–Gan dialect inputs into standard Mandarin. The study asks: (1) whether cue-enhanced scaffolding yields higher meaning retention than lexical or transcript-only support; (2) how such scaffolding affects interpreters' perceived cognitive load and strategic behavior; and (3) to what extent its benefits transfer to an unseen dialect configuration. By linking cognitive-semantic modeling with interpreters' real-time workflows, the study aims to inform future CAI design for low-resource dialects.

2. Literature review

2.1. Cognitive-semantic theories of meaning retention

Cognitive semantics treats meaning as conceptualization rather than as static dictionary sense. When listeners process dialect input, they draw on frame knowledge, mental spaces, and conceptual metaphors to stabilize meaning despite phonetic reduction, ellipsis, and region-specific idioms. Meaning retention in interpreting thus depends on how completely interpreters reconstruct underlying frames and role structures, not just on lexical equivalence [4]. In low-resource settings, where standard glossaries and parallel texts are scarce, interpreters rely heavily on these conceptual resources to infer bridges between dialect forms and target-language formulations. From this perspective, support tools that foreground event structure, participants, and construal operations should strengthen interpreters' ability to maintain cross-linguistic conceptual coherence.

2.2. Real-time scaffolding in interpreting systems

Real-time scaffolding in interpreting refers to dynamic, task-embedded support that reduces cognitive load without replacing human decision making. Common forms include pre-session glossaries, in-session terminology pop-ups, and post-session feedback modules. More recent systems integrate live ASR transcripts and allow interpreters to search, annotate, and filter incoming speech. While such scaffolding improves access to lexical information and reduces memory burden, it can also fragment attention if information is poorly structured, misaligned with cognitive priorities, or visually overwhelming. Effective scaffolding therefore must respect temporal constraints, prioritize information at the level of conceptual units rather than isolated words, and match interpreters' natural segmentation of discourse [5].

2.3. Low-resource dialect interpretation challenges

Low-resource dialects pose specific obstacles to interpreting technologies. ASR models often underperform because training data is limited, phonological variation is wide, and speakers switch between dialect and standard language. Terminology resources are sparse, and traditional glossaries do not capture dialect-specific semantic extensions or shifts. Interpreters must quickly infer whether a dialect term encodes a culturally specific role, a local institutional practice, or a metaphorical framing that lacks direct equivalents [6]. These challenges suggest that scaffolding should help interpreters detect and stabilize core conceptual units, such as events, participant roles, and causal links, so that even when the surface form is unfamiliar or noisy, the underlying meaning can be retained more reliably in the target output.

3. Methodology

3.1. Overall framework design

The proposed system combines three components: a dialect-aware ASR engine, a cognitive-semantic cue extractor, and a real-time scaffolding interface for interpreters. Incoming dialect speech is streamed to the ASR module, which outputs time-aligned word lattices. These are fed into the cue extractor, which identifies semantic frames, participant roles, and inter-clausal relations at the clause and utterance levels. The interface displays a lightly formatted transcript, with color-coded semantic units and compact side-panels summarizing events, roles, and discourse links. The experimental design compares four conditions: (1) no technological support; (2) keyword glossary only; (3) plain ASR transcript; and (4) the full cue-enhanced scaffold [7].

3.2. Cognitive-semantic cue extraction

The cue extractor operates on shallow dependency parses and sub-lexical dialect mappings. It segments each utterance into candidate event units, anchors them on predicates, and assigns roles such as Agent, Patient, Experiencer, and Location using a small frame inventory tuned to local public-service and community contexts. Dialect-specific lexical items are mapped to conceptual labels through a bilingual dialect–standard lexicon extended with manually curated semantic tags. The extractor also tracks discourse relations, cause, contrast, and elaboration, using simple cue phrase heuristics and prosodic boundaries derived from ASR timing. The resulting cues are stored as sets of conceptual units, each linking a frame, its instantiated roles, and its discourse function [8].

3.3. Meaning retention modeling

Meaning retention is defined as the overlap between source-side conceptual units and those recovered in the interpreter's output. Expert annotators mark conceptual units on the source transcript and on the interpreter's target-language rendition, aligning them via shared frames and roles. For each segment i, the set of source concepts C_i is compared with the set of interpreted concepts $\widetilde{C}_i$. A weight w_i reflects the discourse importance of the segment, derived from topic annotations. The Meaning Retention Rate (MRR) is then see Equation (1):

$$MRR = \frac{1}{N} \sum_{i=1}^{N} w_i \cdot \frac{|C_i \cap \widetilde{C}_i|}{|C_i|} \tag{1}$$

where N is the total number of segments. Additional metrics include semantic error units per 100 segments (missed or distorted conceptual units) and segment-level lag time between source and target delivery.

4. Experimental procedure

4.1. Data and preprocessing

Two low-resource dialect corpora were used. The first consists of 38 community-service dialogues in Mandarin–Wu, totaling 4.2 hours of speech and approximately 36,800 running words after transcription. The second comprises 32 semi-structured interviews in Mandarin–Gan, totaling 3.1 hours and 27,400 words. Audio was segmented into 5,184 utterance-level units with speaker labels. For the experimental tasks, 3,600 units were allocated to Mandarin–Wu (training and in-domain evaluation) and 1,584 to Mandarin–Gan (transfer evaluation) [9]. All transcripts were anonymized, and 18% of the segments were double-annotated for conceptual units to compute inter-annotator agreement, which reached $\kappa = 0.86$ for frame labels and $\kappa = 0.82$ for role assignments.

4.2. Experimental settings

Twenty-four postgraduate interpreting students (mean age 24.1 ± 1.8 years; 19 female, 5 male) with advanced proficiency in Mandarin and passive knowledge of at least one Southern Chinese dialect participated. After a 40-minute familiarization session with the interface, each participant interpreted eight Mandarin–Wu dialogues and four Mandarin–Gan interviews. Dialogues were distributed so that each participant experienced all four support conditions in a Latin-square design, with condition order counterbalanced. The ASR backend used a fine-tuned Conformer model with a word error rate of 14.3% ± 1.9% on Mandarin–Wu and 19.8% ± 2.7% on Mandarin–Gan. The system ran on a workstation with GPU acceleration, maintaining an average ASR latency of 310 ± 42 ms per 5-second audio chunk [10].

4.3. Evaluation metrics

For each interpreted segment, expert annotators computed MRR and semantic error units per 100 segments. Segment-level lag time was measured as the difference between source segment end time and the corresponding target segment completion. Participants completed a post-condition questionnaire including a five-point Likert scale for perceived cognitive load, perceived stability of "overall meaning", and interface usability. Open-ended responses on strategies and perceived benefits or drawbacks were also collected and coded thematically.

5. Result and analysis

5.1. Quantitative findings

Across all Mandarin–Wu trials, the cue-enhanced scaffolding condition (Cue + ASR) achieved the highest meaning retention and the lowest semantic error rate. Table 1 summarizes the main quantitative results averaged over participants and dialogues.

Table 1. Main quantitative results on Mandarin–Wu (mean ± SD)

Condition	MRR↑	Semantic error units /100 seg. ↓	Lag time (s) ↓
No support	0.781 ± 0.021	14.6 ± 2.3	4.82 ± 0.47
Glossary only	0.804 ± 0.018	12.9 ± 2.0	4.95 ± 0.39
Plain ASR transcript	0.821 ± 0.019*	11.7 ± 1.9*	5.34 ± 0.52*
Cue-enhanced scaffold	0.847 ± 0.017*	9.8 ± 1.6*	5.01 ± 0.45

*Significant difference vs. no-support baseline at $p < 0.01$ (paired t-test).

Compared with the baseline, the cue-enhanced scaffold improves MRR by + 0.066 ± 0.013, corresponding to a relative gain of 8.5%. The reduction of semantic error units by − 4.8 ± 1.2 per 100 segments indicates that interpreters lose or distort fewer conceptual units when guided by semantic cues. While both the plain ASR transcript and the cue-enhanced scaffold increase lag time relative to the baseline (+ 0.52 ± 0.21 s and + 0.19 ± 0.18 s, respectively), the lag cost of the cue-enhanced condition is significantly smaller than that of the plain transcript (p = 0.03). This suggests that structured cues help interpreters allocate attention more efficiently than a raw text stream (figure 1).

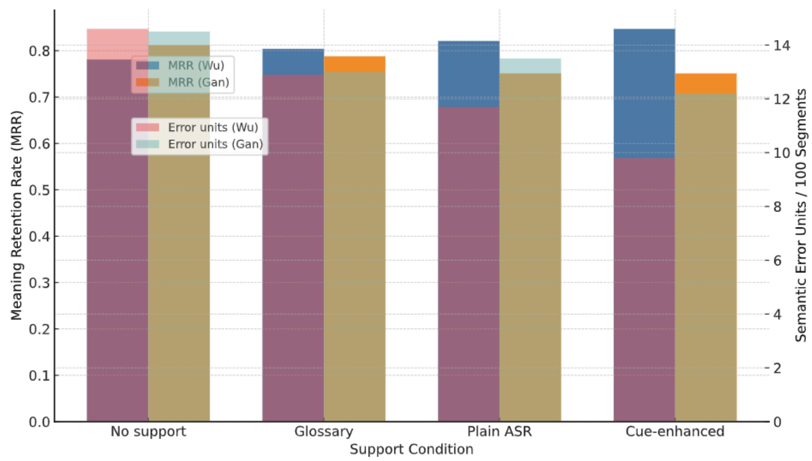

Figure 1. Meaning retention and errors by support condition

5.2. Qualitative insights

Questionnaire data show that 20 out of 24 participants rated the cue-enhanced scaffold as “clearly helpful” or “very helpful” for tracking overall meaning, compared with 11 for the glossary condition and 14 for the plain ASR transcript. Perceived cognitive load on a five-point scale dropped from 4.1 ± 0.6 under the no-support condition to 3.3 ± 0.7 under cue-enhanced scaffolding, while remaining at 3.9 ± 0.6 for the plain transcript. Participants reported that event-level grouping and role labels functioned as “anchors” that allowed them to recover from local comprehension failures and quickly re-locate the main storyline. Several participants noted that, in the presence of heavy dialectal phonology, they relied on the semantic cue panel as a “backup map” of who did what to whom, rather than on exact words. On the negative side, three participants reported occasional “visual clutter” when cues were too dense for very information-rich turns, suggesting the need for adaptive cue filtering.

5.3. Ablation and transferability analysis

To understand which components drive the observed gains, an ablation study was conducted with three degraded variants of the scaffold: (a) removal of discourse relation markers; (b) removal of visual grouping, showing only inline role labels; and (c) removal of the semantic cue extractor, leaving a lightly formatted transcript. Table 2 reports MRR for Mandarin–Wu (in-domain) and Mandarin–Gan (transfer), as well as the transfer drop relative to in-domain performance.

Table 2. Ablation and transfer performance (MRR mean ± SD)

Model variant	MRR Wu ↑	MRR Gan ↑	Transfer drop (Gan − Wu)
Full cue-enhanced scaffold	0.847 ± 0.017	0.812 ± 0.019	−0.035
– discourse relation markers	0.836 ± 0.019	0.788 ± 0.021	−0.048
– visual grouping	0.831 ± 0.020	0.781 ± 0.022	−0.050
– semantic cue extractor	0.814 ± 0.021*	0.751 ± 0.024*	−0.063*

*Significant degradation vs. full model at $p < 0.01$.

Removing discourse relation markers yields a modest but consistent MRR decrease of −0.011 on Mandarin–Wu and −0.024 on Mandarin–Gan, indicating that explicit signaling of cause, contrast, and elaboration supports cross-turn coherence, especially in transfer conditions. Removing visual grouping has a slightly stronger effect, suggesting that spatial organization of cues helps interpreters quickly scan and select relevant information. The largest performance drop arises when the semantic cue extractor is removed, effectively reducing the scaffold to a lightly formatted transcript. In this case, the gains relative to the plain ASR condition shrink to + 0.007 ± 0.009 in-domain and become negligible on the transfer dialect, confirming that cognitively motivated conceptual cues, not just text formatting, are central to the observed improvements.

6. Conclusion

This study demonstrates that embedding cognitive-semantic cues into a real-time CAI scaffold can substantially improve meaning retention when interpreting low-resource dialects. By modeling meaning at the level of conceptual units and aligning them with interpreters' real-time workflows, the proposed system increases MRR, reduces semantic error units, and stabilizes performance when transferring from Mandarin–Wu to Mandarin–Gan. The ablation analysis highlights the importance of explicit semantic extraction and structured visual presentation. Limitations include the relatively small participant pool and the focus on closely related dialects within one language family. Future work will extend the cue inventory to additional domains, explore adaptive cue density based on online indicators of cognitive load, and investigate integration with speech-to-speech translation pipelines to support multilingual, community-based interpreting scenarios.

References

[1] Chen, S., & Kruger, J. L. (2024). Visual processing during computer-assisted consecutive interpreting: Evidence from eye movements. *Interpreting, 26*(2), 231-252.
[2] Guo, M., Han, L., & Anacleto, M. T. (2023). Computer-assisted interpreting tools: Status quo and future trends. *Theory and Practice in Language Studies, 13*(1), 89-99.
[3] Horváth, I. (2022). AI in interpreting: Ethical considerations. *Across Languages and Cultures, 23*(1), 1-13.
[4] Gatsiou, C., Fragkou, E., Hatzidaki, A., & Korda, D. (2024). Artificial Intelligence (AI) in Translation and Interpreting: Exploring AI" s Potential in Translation and Interpreting Profession and Pedagogy in the Medical Context and its Ethical Implications. Unpublished master's thesis]. National and Kapodistrian University of Athens.
[5] Marangoz, E. Ö. (2023). The transformative role of artificial intelligence and machine learning in interpreting and language services. *RumeliDE Dil ve Edebiyat Araştırmaları Dergisi*, (36), 1591-1598.
[6] Giustini, D. (2025). Ethical aspects. *The Routledge Handbook of Interpreting, Technology and AI*, 327-347.
[7] Popescu-Belis, A., Allemann, A., Ferrari, T., & Krishnamani, G. (2025). Speech-to-Speech Translation Pipelines for Conversations in Low-Resource Languages. *arXiv preprint arXiv*: 2506.01406.
[8] Macháček, D., Žilinec, M., & Bojar, O. (2021). Lost in interpreting: Speech translation from source or interpreter?. *arXiv preprint arXiv*: 2106.09343.
[9] Odoom, B. B., Robinson, N., Rippeth, E., Tavarez-Arce, L., Murray, K., Wiesner, M., ... & Duh, K. (2024, September). Can synthetic speech improve end-to-end conversational speech translation?. In Proceedings of the 16th Conference of the Association for Machine Translation in the Americas (Volume 1: Research Track) (pp. 167-177).
[10] Sankar, A., Jain, S., Narasimhan, N., Choudhary, D., Suman, D., Khan, M. S. U. R., ... & Dabre, R. (2025, July). Towards Building Large Scale Datasets and State-of-the-Art Automatic Speech Translation Systems for 14 Indian Languages. In Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 32945-32966).