Beyond Statistical Mapping: Bridging Large Language Models and Human Cognition through Conceptual Intermediate Representation for Enhanced Machine Translation

JUNG HO PARK*

Abstract

Large Language Models (LLMs) have achieved remarkable fluency in machine translation (MT) through statistical pattern recognition and mapping of vast linguistic data. [cite: 7] However, current LLM-based translation often falls short in capturing the deep semantic nuances, cultural implications, and speaker's cognitive perspectives inherent in human language, leading to "translationese," cultural misinterpretations, and occasionally hallucinations. [cite: 8] This paper proposes a novel paradigm for LLM-based machine translation that explicitly integrates principles from Cognitive Grammar (CG). [cite: 9] We hypothesize that translation quality can be significantly enhanced if LLMs are trained to encode source language expressions into a **language-independent Conceptual Intermediate Representation (CIR)**, and subsequently decode this CIR back into the target language. [cite: 10] This approach shifts the role of LLMs from direct language mappers to **'cognitive re-conceptualizers,'** enabling a more robust transfer of meaning, intention, and cultural context. [cite: 11] We argue that by grounding LLM translation in a cognitively plausible framework, we can achieve more human-like, natural, and reliable translations, especially for complex linguistic phenomena such as metaphors, emotions, and culturally bound expressions. [cite: 12] This paper outlines the theoretical foundations, proposes a conceptual model for integrating CIR into LLM architectures, and discusses its potential to mitigate common translation errors and improve explainability. [cite: 13] **AI Usage Notice:** Large Language Models (Google Gemini) were utilized in the ideation, drafting, and structuring processes of this paper. [cite: 14] The ultimate review and responsibility for all content rest with the author. [cite: 15]

*Affiliation: KOREA NATIONAL OPEN UNIVERSITY, Email: sorihanul@gmail.com, Project GitHub: https://github.com/sorihanul/CognoTranslate-Gem

1 Introduction

1.1 Capabilities and Limitations of Current LLM-based Machine Translation

In recent years, Large Language Models (LLMs) have made remarkable advancements in the field of Natural Language Processing (NLP), particularly in Machine Translation (MT). [cite: 17] Transformer architecture-based Neural Machine Translation (NMT) systems, along with pre-trained LLMs like GPT and BERT, have learned from vast amounts of multilingual text data, providing fluent and natural translations previously unimaginable with rule-based or statistical translation systems. [cite: 18] This has significantly contributed to lowering global communication barriers. [cite: 19]

However, despite these impressive advancements, current LLM-based translation still exhibits several inherent limitations. [cite: 20] First, most LLMs rely on statistical pattern recognition and direct superficial mapping between languages in large datasets. [cite: 21] This can lead to literal translations that disregard context or result in unnatural "translationese." Second, a lack of deep understanding of the cultural implications, subtle speaker intentions, and the cognitive ways in which specific concepts are constructed often leads to culturally inappropriate or completely missed translations. [cite: 22] Third, LLMs sometimes exhibit "hallucinations," arbitrarily generating information not present in the source text, which degrades the reliability of translations. [cite: 23] These limitations suggest that current LLM translation has not yet reached a true "understanding" and reconstruction of the fundamental "meaning" embedded in language, going beyond mere linguistic form transformation. [cite: 24] Therefore, a new paradigm shift beyond current statistical pattern matching is needed. [cite: 25]

1.2 The Perspective of Cognitive Grammar on Language and Meaning

This paper proposes Ronald Langacker's **Cognitive Grammar (CG)** as a theoretical foundation to overcome the limitations of LLM translation [3, 4, 5]. [cite: 25] Cognitive Grammar views language as the product of "conceptualization," based on human cognitive abilities and experience. [cite: 26] That is, language does not merely reflect objective reality but is rather a way for speakers to understand, structure, and express the world from a particular perspective. [cite: 27] Key concepts of CG include: [cite: 28]

- **Image Schemas:** Universal, abstract cognitive structures derived from human bodily experiences (e.g., space, movement, force), which form the basis of linguistic expressions [7]. [cite: 28, 29] (e.g., CONTAINER, PATH, FORCE, UP-DOWN schemas). [cite: 29]
- **Profiling and Perspective:** Even for the same event or concept, linguistic expressions vary depending on which aspects (parts) the speaker 'pro-

files' and which are backgrounded. [cite: 29] They also reflect the speaker's perspective in conceptualizing the event (e.g., active/passive voice, proximal/distal viewpoint). [cite: 30]

- **Force Dynamics:** Schemas representing the relationships of interacting forces between two or more entities (e.g., resistance, facilitation, inhibition, permission, compulsion), crucial for understanding causality and state changes. [cite: 31]

CG emphasizes that the meaning of language is not merely an arbitrary combination of symbols but a conceptual structure formed through human cognitive processes. [cite: 32] Research by Lakoff and Johnson [2], which demonstrated that metaphors are not just rhetorical devices but fundamental aspects of human cognition, further supports this view. [cite: 33] This cognitive perspective provides crucial insights necessary for LLMs to "understand" and "reconstruct" deep semantic structures beyond the superficial forms of language. [cite: 34]

1.3 Hypothesis: Connecting LLMs and Cognition through Conceptual Intermediate Representation (CIR)

The core hypothesis of this paper is that LLMs should learn and utilize a **language-independent Conceptual Intermediate Representation (CIR)** based on Cognitive Grammar principles during the translation process. [cite: 35] Traditional LLMs adopt an End-to-End learning approach, directly mapping source language (SL) sentences to target language (TL) sentences. [cite: 36] We propose a two-stage approach instead: [cite: 37]

1. **Source Language -> CIR Encoding:** The LLM receives source text as input and transforms the embedded cognitive meaning (image schemas, speaker's perspective, emotion, force dynamics, etc.) into a structured CIR. [cite: 37, 38]

2. **CIR -> Target Language Decoding:** The LLM receives this CIR as input and 're-conceptualizes' the meaning in the most natural and appropriate way according to the target language's grammar, vocabulary, and cultural context, thereby generating the target language sentence. [cite: 38, 39]

This approach shifts the role of LLMs from mere language mapping tools to **'cognitive re-conceptualizers,'** enabling a more robust transfer of the fundamental concepts, intentions, and cultural contexts that language aims to convey, allowing for an "artistic transference of meaning" beyond simple word-for-word or sentence-for-sentence translation. [cite: 39]

1.4 Contributions of This Paper

This paper makes the following key contributions: [cite: 40]

- **Proposing a theoretical framework for CG-based LLM translation:** Demonstrating theoretically how LLMs can improve translation quality by mimicking human cognitive language processing. [cite: 40]
- **Suggesting a conceptual architecture for CIR integration:** Proposing a conceptual model for how CIR can be structurally integrated and utilized within LLM architectures. [cite: 41]
- **Discussing translation quality improvement:** Discussing how this framework can enhance translation quality in terms of naturalness, semantic fidelity, and accurate transfer of cultural nuances. [cite: 42]
- **Implying reduced hallucination and increased explainability:** Emphasizing the potential for CIR to partially reveal the internal reasoning processes of LLMs, thereby reducing translation hallucinations and increasing the explainability of 'black-box' models. [cite: 43]

2 Background and Related Work

2.1 Traditional Machine Translation Approaches

Machine translation research began in the 1950s with Rule-Based Machine Translation (RBMT). [cite: 44] RBMT performed translations based on grammatical rules and dictionaries manually crafted by linguists. [cite: 45] While offering high consistency and control over translations, it faced limitations in scalability and flexibility due to the complexity of language and its exceptions. [cite: 46] From the 1990s onward, Statistical Machine Translation (SMT) emerged as the dominant paradigm. [cite: 47] SMT learned statistical patterns from large parallel corpora to perform translations, outperforming RBMT in fluency, but still exhibiting limitations in contextual understanding or subtle meaning transfer. [cite: 48] Example-Based Machine Translation (EBMT) built databases of existing translation examples to find and translate similar sentences, which was effective in specific domains but difficult to apply to general language. [cite: 49]

2.2 Neural Machine Translation (NMT) and Large Language Models (LLMs)

From the mid-2010s, Neural Machine Translation (NMT) brought about a revolutionary change in machine translation research. [cite: 50] Sutskever et al. [6] first demonstrated the potential of NMT with sequence-to-sequence models, and Bahdanau et al. [cite: 51] [1] significantly improved translation quality by introducing the Attention mechanism. Particularly, Vaswani et al. [cite: 52] [8] introduced the Transformer architecture, leading to a groundbreaking shift that forms the basis of current LLMs. [cite: 53] In recent years, Large Language Models (LLMs) such as GPT-3, PaLM, and Gemini have further boosted NMT performance. [cite: 54] These LLMs acquire various language understanding and generation capabilities through pre-training on vast text data and fine-tuning for specific tasks, with translation being one of their powerful applications. [cite: 55] LLMs have significantly improved contextual fluency and consistency, resolving many issues that previous generation translation systems struggled with. [cite: 56]

2.3 Cognitive Linguistics and Machine Translation

Cognitive Linguistics is a field of linguistics that attempts to explain linguistic phenomena based on human cognitive structures and experiences. [cite: 57] In the past, there have been continuous attempts to integrate cognitive linguistic theories into Natural Language Processing (NLP) or Machine Translation. [cite: 58] For example, research on semantic representation using image schemas or knowledge representation through conceptual graphs. [cite: 59] However, these attempts were mainly limited to rule-based systems or knowledge engineering approaches and struggled to handle the complexity of large linguistic data and generate flexible translations. [cite: 60] It was difficult to integrate complex aspects of human cognition with the computing power and model architectures available at the time. [cite: 61] The advent of LLMs has opened up new possibilities for integrating these cognitive linguistic insights into large-scale language processing systems. [cite: 62] LLMs excel at learning vast amounts of knowledge and linguistic patterns on their own, and it is expected that explicitly injecting cognitive grammar principles can improve the quality of their learning. [cite: 63]

2.4 Limitations of Current LLM Translation from a Cognitive Perspective

Current LLM-based translation primarily operates through statistical associations and pattern matching. [cite: 64] This leads to the following cognitive limitations: [cite: 65]

- **Failure to Capture Speaker's Perspective and Cognitive Schemas:** LLMs often transform superficial words without a deep understanding of how the speaker profiled objects (e.g., active/passive voice, proximal/distal viewpoint) or what basic cognitive schemas (e.g., CONTAINER, PATH) were used to construct concepts in the source text. [cite: 65] This can result in awkward or altered meanings in the target language. [cite: 66]

- **Difficulty in Conveying Emotional Nuances and Cultural Implications:** Language is rich with emotional tones and cultural implications that are not explicitly stated. [cite: 67] Concepts like Korean 'Han' (한) or 'Jeong' (정), or specific metaphors and proverbs, are difficult to translate fully through simple word mapping. [cite: 68] LLMs may translate them superficially or miss them entirely, failing to convey the intended emotional/cultural experience to the target language reader. [cite: 69]

- **Hallucination and Semantic Distortion:** LLMs sometimes generate information not present in the source text or misinterpret complex contexts or subtle meanings, leading to hallucinations that distort the original meaning. [cite: 70] This can occur because they rely on superficial patterns rather than clearly understanding the 'conceptual skeleton' of the language. [cite: 71]

- **'Black Box' Nature:** It is difficult to explain the internal reasoning processes of current LLMs for why they produce a particular translation. [cite: 72] This poses challenges in identifying the causes of translation errors and improving the system. [cite: 73]

For example, consider a passage from Edgar Allan Poe's "THE TELL-TALE HEART": "Whenever it fell upon me, my blood ran cold." This sentence conveys more than simple meaning; [cite: 74] it vividly describes the narrator's (murderer's) paranoid fear and psychological reaction. [cite: 75] Current LLMs might translate this as "Whenever it fell upon me, my blood became cold," but this often fails to convey the cognitive and emotional meaning of "my blood ran cold," which implies intense horror. [cite: 76] A more natural and psychologically accurate translation in Korean would be " " (my blood ran cold). [cite: 77] This illustrates the need to re-conceptualize the **cognitive schema** of 'a sensation of freezing due to fear' in both languages, beyond direct linguistic mapping. [cite: 78]

3 Cognitive Grammar-based Conceptual Intermediate Representation (CIR) Framework

3.1 Defining CIR

To overcome the limitations of LLM-based machine translation, we propose a **Conceptual Intermediate Representation (CIR)** framework. [cite: 79] CIR is a language-independent semantic representation that goes beyond the surface form of language, grounded in how humans perceive and conceptualize the world. [cite: 80] It is structured to integrate concepts from Cognitive Grammar, such as image schemas, profiling, force dynamics, and subjectification, and should be able to express the same conceptual content regardless of the specific language. [cite: 81] CIR goes beyond mere syntactic parse trees or semantic role labeling. [cite: 82] It captures the deeper cognitive processes underlying language, such as how the speaker 'views' a particular entity, what 'path' was taken to conceptualize an event, or what 'forces' are at play. [cite: 83] This differs from traditional interlinguas, which primarily focus on logical/propositional meaning, by emphasizing **human embodied meaning construction based on experience.** [cite: 84]

3.2 Key Components of CIR

CIR includes the following cognitive grammar elements to represent conceptual meaning: [cite: 84]

- **Image Schemas:**
  - **CONTAINER:** Concept of an enclosed space (e.g., inclusion, inside/outside, crossing boundaries). [cite: 84, 85]
  - **PATH & SOURCE-PATH-GOAL:** Concepts of movement, process, or transition (starting point, path, goal). [cite: 85, 86]
  - **FORCE DYNAMICS:** Interaction of forces (resistance, facilitation, permission, compulsion) and resulting state changes. [cite: 86, 87]
  - **UP-DOWN:** Vertical orientation (increase/decrease, ascent/descent, change of state). [cite: 87]
  - **TR/LM (Trajector/Landmark):** Relationship between the focused entity (TR) and the reference point (LM). [cite: 88]
  - **PROFILING / PERSPECTIVE:** How the speaker emphasizes certain entities and from what perspective (e.g., active/passive) an event is viewed. [cite: 89]
  - Others include BALANCE, LINK, PART-WHOLE, CENTER-PERIPHERY, FRONT-BACK, NEAR-FAR, etc. [cite: 89]
- **Subjectification:**
  - The way a speaker or experiencer embeds their emotions, attitudes, cognitive states, or epistemic modality into linguistic expressions (e.g., Korean expressions like 일 것 같다 (seems like), 인 것 같다 (appears to be), 네요 (exclamatory ending), 군요 (realization ending), which convey emotion/certainty). [cite: 90]
- **Cultural Mapping Points:**
  - Concepts or emotions unique to a specific culture (e.g., Korean Han (한), Jeong (정)), or metaphorical expressions and customary phrases understood only within that culture. [cite: 91] These are marked with special 'tags' or 'nodes' within the CIR, indicating the need for reconceptualization appropriate to the target language culture. [cite: 92]

3.3 Examples of CIR Representation (Source Language: Korean, Target Language: English)

Let's illustrate how CIR operates with specific examples using actual sentences. [cite: 93] **Example 1: Simple Sentence (Basic CIR Structure)**

- **Korean Source:** 그녀는 방 안에 있다. (She is in the room.) [cite: 93]
- **Surface Analysis:** 그녀 (she, subject), 방 (room, place), 안에 (in, location) [cite: 93]
- **Cognitive Analysis:** The speaker conceptualizes she (TR) as being located inside room (LM), which is a CONTAINER schema. [cite: 94]
- **Conceptual Intermediate Representation (CIR):**

```
[EVENT: 'BE_LOCATED',
 TR: {TYPE: 'HUMAN', ID: 'FEMALE_1'},
 LM: {TYPE: 'SPACE', ID: 'ROOM_1', SCHEMA: 'CONTAINER'},
 RELATION: {SCHEMA: 'CONTAINER_INTERNAL'},
 TIME: 'PRESENT',
 PERSPECTIVE: 'NEUTRAL_EXTERNAL']
```

- **English Target (Re-conceptualization):** `She is in the room.` [cite: 94]

Example 2: Emotion and Figurative Expression (Deep Cognitive Analysis)

- **Korean Source:** 그 눈이 나를 바라볼 때마다 내 피는 차갑게 식었다. (Whenever that eye looked at me, my blood grew cold.) [cite: 94] *(From the Korean translation of Edgar Allan Poe's "THE TELL-TALE HEART")* [cite: 95]
- **Surface Analysis:** 눈 (eye, subject), 바라보다 (look at, verb), 내 피 (my blood, subject), 차갑게 식다 (grow cold, verb phrase) [cite: 95]
- **Cognitive Analysis (Considering the context of the Korean source):** [cite: 95]
  - **그 눈이 나를 바라볼 때마다 (Whenever that eye looked at me):** [cite: 95]
    - 눈 (eye): The object (LM) and an abnormal cognitive stimulus source. [cite: 96] From the speaker's perspective (TR), it is a 'perceived object'. [cite: 96]
    - 바라볼 때마다 (whenever looked at): Indicates a repetitive and continuous 'path of perception' of the PATH schema. [cite: 97] It also includes the temporal aspect of 'repetition'. [cite: 98]
    - **Subjectification:** The phrase 나를 (me) profiles the speaker's direct experience and sense of threat. [cite: 99]
  - **내 피는 차갑게 식었다 (My blood grew cold):** [cite: 99]
    - 내 피 (my blood): The entity (TR) representing a physiological change within the speaker. [cite: 100]

    * 차갑게 식었다 (grew cold): A combination of the UP-DOWN schema (temperature drop) and FORCE DYNAMICS schema (strong external cognitive stimulus compels an internal physiological change) that are complexly applied. [cite: 101] This is not merely a temperature change but strongly implies the metaphorical meaning of **'physical paralysis/chilling sensation due to extreme fear.'** [cite: 101]
    * **Subjectification:** The speaker's extreme fear and paranoid psychological state are expressed. [cite: 102] This emotion is not merely stated but metaphorically conceptualized through the physical phenomenon of 'blood growing cold.' [cite: 103]

- **Conceptual Intermediate Representation (CIR):**

```
[EVENT: 'REPEATEDLY_PERCEIVED_BY_EYE',
 PERCEIVER: {TYPE: 'HUMAN', ID: 'NARRATOR', ROLE:
     'EXPERIENCER', PERSPECTIVE: 'INTERNAL_TRAJECTOR'},
 OBJECT_OF_PERCEPTION: {TYPE: 'BODY_PART', ID:
     'OLD_MAN_EYE', SCHEMA: 'POINT_OF_PERCEPTION',
     QUALIFIER: 'VULTURE_LIKE_PALE_BLUE_FILM_OVER'},
 TRIGGER_CONDITION: 'ON_EACH_OCCURRENCE_OF_PERCEPTION',
 CONSEQUENCE_EVENT:
     'PHYSIOLOGICAL_CHANGE_OF_NARRATOR_BLOOD',
 CHANGE_DIRECTION: {SCHEMA: 'UP_DOWN', DIRECTION:
     'DOWNWARD', ASPECT: 'TEMPERATURE'},
 FORCE_DYNAMICS: {AGENT_OF_FORCE: 'OBJECT_OF_PERCEPTION',
     PATIENT_OF_FORCE: 'NARRATOR_BLOOD', FORCE_TYPE:
     'COMPULSION/OVERWHELMING_STIMULUS', RESULT:
     'FEAR_INDUCED_PHYSIOLOGICAL_RESPONSE'},
 EMOTION: 'INTENSE_FEAR_HORROR_PARANOIA',
 SUBJECTIFICATION:
     'NARRATOR_EXPERIENCE_FEELING_OF_COLD_PARALYSIS_DUE_TO_FEAR',
 SCHEMA_EXTENSION:
     'METAPHORICAL_RELATION_BETWEEN_COLD_BLOOD_AND_FEAR']
```

- **English Target (Re-conceptualization via CIR):** `Whenever it fell upon me, my blood ran cold.` [cite: 104]

  - CIR's `FORCE DYNAMICS` (compelling change due to overwhelming stimulus) and `EMOTION` (intense fear) information indicate that the translation should not be a mere "grew cold" but rather a more idiomatic expression like "ran cold," which effectively conveys the psychological nuance of the original English text (and the Korean translation accurately captured this). [cite: 105]

Example 3: Korean Honorifics/Non-Honorifics (Social Relations and Subjectification)

- **Korean Source:** [cite: 105]
  - 수고하셨습니다. (Formal, typically used for superiors or elders) [cite: 106]
  - 수고했어. (Informal, typically used for subordinates or peers) [cite: 106]
- **Cognitive Analysis:** Both sentences share the core concept of 'having exerted effort' or 'having gone through hardship,' but they express the speaker's **social relationship (hierarchy)** and **attitude (respectful vs. informal)** towards the listener through Subjectification. [cite: 106] This is an essential cognitive element for Korean speakers. [cite: 106]
- **Conceptual Intermediate Representation (CIR):** [cite: 106]
  - 수고하셨습니다. (CIR-FORMAL):

```
[EVENT: 'EFFORT_EXERTED',
 AGENT: {TYPE: 'HUMAN', ID: 'HEARER'},
 SUBJECTIFICATION: {
     SPEAKER_STANCE: 'RESPECTFUL_FORMAL',
     HEARER_STATUS_RELATION:
         'HIGHER_OR_EQUAL_RESPECT_REQUIRED',
     SPEAKER_INTENTION: 'EXPRESS_APPRECIATION_FORMAL'
 }]
```

  - 수고했어. (CIR-INFORMAL):

```
[EVENT: 'EFFORT_EXERTED',
 AGENT: {TYPE: 'HUMAN', ID: 'HEARER'},
 SUBJECTIFICATION: {
     SPEAKER_STANCE: 'CASUAL_INFORMAL',
     HEARER_STATUS_RELATION: 'LOWER_OR_EQUAL_CASUAL',
     SPEAKER_INTENTION: 'EXPRESS_APPRECIATION_INFORMAL'
 }]
```

- **English Target (Re-conceptualization via CIR):** [cite: 108]
  - For 수고하셨습니다.: `Well done, sir/ma'am.` or `Thank you for your hard work.` (re-conceptualized into more formal and respectful expressions) [cite: 109]
  - For 수고했어.: `Good job!` or `You did well.` (re-conceptualized into more informal and friendly expressions) [cite: 109]
  - This example demonstrates that CIR contributes to selecting appropriate tone and expression in the target language by capturing subtle social and emotional relationships between speaker and listener embedded in the language, beyond mere propositional meaning. [cite: 110]

Example 4: Korean Idiom/Proverb (Cultural Mapping Point)

- **Korean Source:** 식은 죽 먹기였다. (It was as easy as eating cold porridge.) - meaning it was very easy. [cite: 111]
- **Surface Analysis:** 식은 죽 (cold porridge), 먹기 (eating act) [cite: 111]
- **Cognitive Analysis:** The act of 'eating cold porridge' is conceptualized in Korean culture as an experience that is 'very easy and requires little effort.' [cite: 112] This is a Metaphorical_Extension that cannot be understood without specific cultural background knowledge. CIR must recognize such Cultural_Mapping_Points. [cite: 113]
- **Conceptual Intermediate Representation (CIR):**

```
[CONCEPT: 'EASY_TASK',
 SCHEMA_SOURCE: 'CULINARY_EASE',
 METAPHORICAL_EXTENSION: {
     SOURCE_DOMAIN:
         'PHYSICAL_ACTIVITY_OF_EATING_COLD_PORRIDGE',
     TARGET_DOMAIN: 'ABSTRACT_TASK_EASE',
     MAPPING_RULES:
         'LOW_EFFORT_IN_SOURCE_MAPS_TO_LOW_EFFORT_IN_TARGET'
 },
 CULTURAL_MAPPING_POINT: 'KOREAN_IDIOM_COLD_PORRIDGE',
 TIME: 'PAST',
 SUBJECTIFICATION: 'NEUTRAL_EVALUATION']
```

- **English Target (Re-conceptualization via CIR):** `It was a piece of cake.` or `It was a breeze.` [cite: 114]
  - The CULTURAL_MAPPING_POINT and METAPHORICAL_EXTENSION information included in the CIR guides the LLM to re-conceptualize '식은죽 먹기' not as a literal translation but as a similar metaphorical expression (e.g., 'piece of cake') that conveys 'an easy task' in Anglo-American culture. [cite: 115] This enables true meaning transfer across cultural barriers. [cite: 115]

4 Proposed LLM Architecture: Model for CIR-based Machine Translation (Conceptual Stage)

4.1 Two-Stage / Hybrid LLM Model

Our proposed model primarily consists of two core modules: [cite: 115]

- **Encoder (Source Language -> CIR Transformation Module):**

  - This LLM module takes source language text (e.g., Korean) as input and identifies the embedded cognitive grammar elements (image schemas, profiling, force dynamics, subjectification, etc.) to transform them into a structured CIR. [cite: 116]
  - This requires the ability to understand the deep conceptual structure and the speaker's cognitive intention of the text, beyond mere syntactic parsing. [cite: 117]
  - Training can be done by fine-tuning using a dataset of (source language sentence, CIR) pairs. [cite: 118] The CIR is expected to be systematically represented using predefined schemas and attributes. [cite: 119]

- **Decoder (CIR -> Target Language Generation Module):**
  - This LLM module receives the CIR generated by the encoder as input and re-conceptualizes it in the most natural and fluent way, adhering to the target language's (e.g., English) grammar, vocabulary, and cultural context, to generate the target language sentence. [cite: 120]
  - Here, based on the conceptual information in the CIR, the decoder determines the most appropriate and idiomatic expressions, culturally nuanced lexical choices, and sentence structures that preserve the emotional tone of the original text. [cite: 121]
  - Training can be fine-tuned using a dataset of (CIR, target language sentence) pairs. [cite: 122] During decoder training, it is crucial to design a loss function that evaluates whether all information within the CIR (especially subjectification, cultural mapping points, and emotion-related schemas) is appropriately reflected in the target language expression. [cite: 123]

4.2 Training Methodology (Conceptual Proposal)

To effectively train such a two-stage model, the following methodologies can be considered: [cite: 123]

- **Hypothetical Dataset Construction Strategy:** [cite: 123]
  - **Phase 1: Building a Small-Scale Cognitive Grammar Annotated Dataset:** Initially, we propose a process where a small group of experts with deep linguistic knowledge and understanding of Cognitive Grammar manually annotate key sentences from carefully selected Korean-English parallel corpora (e.g., literary works, news articles, daily conversations, including various genres) to create CIRs. [cite: 124] This annotation process would rigorously follow clearly defined guidelines for using the CIR components outlined in

Section 3.2 (image schemas, profiling, force dynamics, subjectification, cultural mapping points). [cite: 124, 125] This small, high-quality dataset would serve as a crucial 'gold standard' for the LLM to learn the structure and meaning of CIR. [cite: 125]

  - **Phase 2: Large-Scale CIR Generation via LLM and Human Review:** After fine-tuning an initial LLM (encoder) with the small-scale dataset from Phase 1, it would be used to automatically generate preliminary CIRs for a large volume of Korean text. [cite: 126] Subsequently, a relatively smaller number of human reviewers would examine and correct the generated CIR drafts, thereby increasing the efficiency of data construction. [cite: 127] This process ensures the consistency and accuracy of CIR while presenting a scalable data collection method. [cite: 128]

- **Hypothetical Training Process:** [cite: 128]

  - **Encoder Training:** A pre-trained large language model (e.g., the encoder part of a Korean-English NMT model or a multilingual embedding model) would be used as the base model. [cite: 129] Fine-tuning would be performed on this model using (source language sentence, CIR) pair datasets. [cite: 130] The loss function could include a multi-task loss to improve prediction accuracy for each CIR component (schema, TR, LM, relations, subjectification, cultural tags, etc.). [cite: 131] For instance, a combination of classification and regression losses could be used for predicting each slot in the CIR. [cite: 132]

  - **Decoder Training:** The decoder, which generates target language sentences from CIR input, would be fine-tuned with (CIR, target language sentence) pair datasets. [cite: 133] During decoder training, it is crucial to design a loss function that evaluates whether all information within the CIR (especially subjectification, cultural mapping points, and emotion-related schemas) is appropriately reflected in the target language expression. [cite: 134]

- **Multi-task Learning:** [cite: 135]

  - When training the encoder LLM, related cognitive tasks such as source text summarization, key concept extraction, and sentiment analysis could be learned alongside CIR generation to enhance the representational power of the CIR. [cite: 135, 136] This helps the LLM better understand the deeper meaning of language. [cite: 136]

  - The decoder LLM could also improve its understanding of CIR by performing auxiliary tasks, such as generating explanations for how specific CIR elements are realized in the target language, in addition to translation. [cite: 137]

- **Potential for Human Feedback-based Reinforcement Learning (RLHF) of CIR:** [cite: 138]
  - To evaluate the naturalness of translations and the faithful transfer of cognitive nuances from the original text, especially for subjective judgments required in literary translation, humor, and emotional expression, human feedback from translators or experts would be collected. [cite: 138, 139] This feedback could be used as a reward signal to fine-tune the CIR-to-target language decoder using Reinforcement Learning, further enhancing translation quality. [cite: 139]

4.3 Advantages over End-to-End LLM Translation (Addressing Existing LLM Limitations)

The CIR-based LLM architecture offers significant advantages over existing End-to-End LLM translation models: [cite: 140]

- **Explicit Meaning Control and Reduced Ambiguity:** The presence of CIR as an intermediate stage allows for clear understanding and control over the 'content' that the LLM is translating. [cite: 141] This reduces errors that can occur when interpreting ambiguous expressions in the source text. [cite: 142]
- **Improved Explainability of Translation Choices:** By providing a cognitive rationale (CIR) for the LLM's translation process, which previously operated as a 'black box,' it opens up the possibility of explaining why specific translation results were generated. [cite: 142, 143] This increases trustworthiness. [cite: 143]
- **Reduced Probability of Hallucination:** The approach of clearly encoding the conceptual gist of the source text into CIR before re-conceptualizing it significantly reduces the phenomenon of hallucination, where LLMs arbitrarily generate information not present in the source or distort its meaning. [cite: 143] That is, by 'understanding' before 'generating,' the process becomes more robust. [cite: 144]

5 Expected Outcomes and Benefits

The CIR-based LLM translation framework proposed in this paper is expected to yield the following significant outcomes and benefits. [cite: 145]

5.1 Enhanced Translation Quality

- **More Natural and Idiomatic Target Language Expressions:** By reconceptualizing based on cognitive concepts beyond the surface structure of language, the framework can reduce literal translation artifacts and produce idiomatic expressions and writing styles that sound natural to native target language speakers. [cite: 146]

- **Accurate Transfer of Cultural Nuances, Metaphors, and Emotional Tones:** Including cultural mapping points and subjectification elements within the CIR will minimize misinterpretations caused by cultural differences between languages, and effectively transferring the subtle emotional tones and figurative expressions of the original text to the target language. [cite: 147]

- **Reduction of 'Translationese' and Literal Translation:** Cognitive Grammar-based re-conceptualization is not just about word substitution; [cite: 148] it involves re-expressing the original concepts to fit the cognitive structures of the target language, resulting in translations that possess the natural fluency of texts originally written by native speakers. [cite: 149]

5.2 Mitigation of Hallucination and Semantic Errors

- **Information Manipulation Prevention:** Since the CIR explicitly captures the 'conceptual truth' of the source text, the LLM will be much less likely to invent facts not present in the original or distort information during the translation process. [cite: 150] Translations will become more 'factually' grounded. [cite: 151]

- **Ensuring Semantic Consistency:** As the core concepts and relationships of the source text are clearly represented in the CIR, the translation results will maintain consistency with the original meaning, and important information or nuances will not be omitted. [cite: 151, 152]

5.3 Improved Explainability and Controllability

- **Transparency in LLM Interpretation:** The CIR as an intermediate representation allows for partial verification of how the LLM 'interprets' the source text and what cognitive judgments it makes. [cite: 153] This enhances the explainability of traditional 'black box' models, providing insights for both researchers and users. [cite: 154]

- **Human Intervention/Correction at a Conceptual Level:** If a translation error occurs, instead of merely correcting the output translation, there is a possibility of directly modifying specific elements of the CIR that caused the error, thereby achieving the desired translation. [cite: 154, 155]

5.4 Cross-Lingual Generalization

- **N-to-N Language Translation Efficiency:** Due to the language-independent nature of CIR, when new language pairs are added, there is no need to retrain the entire model; [cite: 155] only the mapping between that language and the CIR needs to be learned. [cite: 156] This can significantly improve the efficiency of building and maintaining multilingual translation systems. [cite: 157] That is, instead of creating separate Korean-English and English-Japanese models, various language pair translations can be achieved with only Korean-to-CIR, CIR-to-English, and CIR-to-Japanese modules. [cite: 158]

6 Challenges and Future Work

While the CIR-based LLM translation framework proposed in this paper holds great potential, addressing the following significant challenges will be necessary for its practical implementation. [cite: 159]

6.1 CIR Definition and Standardization

Human cognitive processes are highly complex and subtle, making it a formidable task to define and standardize them into a clear and formalized CIR. [cite: 160] Even when based on Cognitive Grammar theory, research is needed to systematically design and reach consensus on CIR schemas and attributes that can comprehensively capture the cognitive implications of all linguistic expressions. [cite: 161] This demands close collaboration among linguists, cognitive scientists, and computer scientists. [cite: 162]

6.2 Data Annotation

For CIR-based LLM training, a large-scale parallel dataset consisting of (source language sentence, corresponding CIR, target language sentence) is essential. [cite: 163] CIR annotation is a highly specialized and time-consuming task, making the construction of such a dataset one of the biggest bottlenecks. [cite: 164] Research into innovative data construction methodologies is needed, such as efficient annotation tool development, semi-supervised learning techniques, or hybrid approaches where initial CIRs are generated with the help of LLMs and then reviewed by humans. [cite: 165]

6.3 Model Complexity

A two-stage LLM architecture can be more complex in its internal structure than an end-to-end model, which implies increased computational requirements during model training and inference. [cite: 166] Research is needed on architecture and algorithms for optimizing the performance of each module, managing efficient information flow, and improving computational resource efficiency. [cite: 167]

6.4 Evaluation Metrics

Current machine translation evaluation primarily relies on superficial similarity or fluency metrics such as BLEU, ROUGE, and BERTScore. [cite: 168] To properly evaluate the success of CIR-based translation, the development of new evaluation metrics capable of measuring 'cognitive fidelity' or 'accuracy of conceptual transfer' is essential. [cite: 169] These metrics must be able to quantitatively assess how well the translated text re-conceptualized the original's cognitive intent and cultural nuances. [cite: 170]

6.5 Multimodal Integration

Human cognition occurs through various sensory modalities, such as sight and hearing. [cite: 171] Therefore, in the long term, this research could be extended to include multimodal inputs like images, video, and speech, transforming them into integrated CIRs for translation. [cite: 172] This would contribute to artificial intelligence understanding and communicating with the world in a holistic way, similar to humans. [cite: 173]

7 Conclusion

This paper proposed a new paradigm for overcoming the limitations of current Large Language Model (LLM)-based machine translation by utilizing a **Conceptual Intermediate Representation (CIR)** based on the principles of Cognitive Grammar. [cite: 174] We argued that LLMs can significantly enhance translation quality by moving beyond surface-level pattern matching to 'understand' the conceptual meaning formed through human cognitive processes and 're-conceptualize' it in the target language. [cite: 175] The proposed two-stage LLM architecture is expected to enable more natural and idiomatic translations, along with the accurate transfer of cultural nuances and emotional tones, through encoding from source language to CIR and decoding from CIR to target language. [cite: 176] Furthermore, this approach holds the potential to reduce translation hallucinations and improve the explainability of LLMs. [cite: 177] Certainly, challenges remain, such as the definition and standardization of CIR, the construction of large-scale datasets, and model complexity. [cite: 178] However, this research represents an important first step toward opening new horizons in the future of machine translation through interdisciplinary collaboration in computational linguistics, cognitive science, and artificial intelligence research. [cite: 179] Ultimately, we aim for LLMs to evolve from mere 'language translation machines' to 'intelligences that understand and communicate concepts.' [cite: 180]

8 References

References

[1] Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. [cite: 181] *arXiv preprint arXiv:1409.0473*. [cite: 182]

[2] Lakoff, G., & Johnson, M. (1980). *Metaphors We Live By*. University of Chicago Press. [cite: 182, 183]

[3] Langacker, R. W. (1987). *Foundations of Cognitive Grammar, Volume 1: Theoretical Prerequisites*. Stanford University Press. [cite: 183]

[4] Langacker, R. W. (1991). [cite: 184] *Foundations of Cognitive Grammar, Volume 2: Descriptive Application*. Stanford University Press. [cite: 184]

[5] Langacker, R. W. (2008). *Cognitive Grammar: A Basic Introduction*. [cite: 185] Oxford University Press. [cite: 185]

[6] Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to Sequence Learning with Neural Networks. [cite: 185, 186] *arXiv preprint arXiv:1409.3215*. [cite: 186]

[7] Talmy, L. (2000). *Toward a Cognitive Semantics*. MIT Press. [cite: 186, 187]

[8] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). [cite: 187] Attention Is All Need. *Advances in Neural Information Processing Systems, 30*. [cite: 188]