# Architectural Framework for Automated CEFR B1/B2 English Learning Systems: Prompt Engineering Strategies for Z.ai GLM-4.7

## 1. Introduction and System Overview

The convergence of Generative AI and Second Language Acquisition (SLA) has created unprecedented opportunities for the democratization of high-quality, personalized language instruction. The "Ovi English School" initiative represents a sophisticated application of this convergence, aiming to automate the transformation of complex, authentic news media (BBC/NPR) into pedagogically sound learning materials targeted at Japanese adult learners. This report provides a comprehensive technical and pedagogical analysis of the requisite prompt engineering strategies for the Z.ai GLM-4.7-flash model. The objective is to construct a robust system prompt architecture that reliably outputs structured JSON data suitable for downstream Text-to-Speech (TTS) synthesis and PDF generation, strictly adhering to CEFR B1/B2 proficiency constraints while navigating the unique linguistic interference patterns of Japanese learners.

### 1.1 The Operational Challenge

The primary engineering challenge lies in the "Reasoning-Structure Trade-off" inherent in Large Language Models (LLMs). While models like GLM-4.7 possess advanced reasoning capabilities, forcing them into rigid output formats (such as complex JSON schemas) often degrades the quality of their linguistic reasoning.1 For an educational application where the nuance of a "B1 level explanation" versus a "B2 level explanation" is critical, this degradation is unacceptable. Furthermore, the target demographic—Japanese adult learners—presents a unique set of "interlanguage" challenges, including *wasei-eigo* (loanword interference) and significant cultural context gaps that render standard Western news unintelligible even when linguistically simplified.2

### 1.2 Model Capabilities: Z.ai GLM-4.7-flash

The selection of Z.ai's GLM-4.7-flash model introduces specific architectural advantages and constraints. The model supports a context window of 128k tokens, allowing for the inclusion of extensive "Few-Shot" examples and detailed style guides within the system prompt.4 Crucially, the model features "Native JSON Schema" support and "Interleaved Thinking" capabilities, which allow the model to plan its reasoning path before generating output.4 This report argues that leveraging these specific features—explicitly separating the "pedagogical reasoning" phase from the "JSON generation" phase—is the only viable method to ensure high-fidelity educational content generation.

## 2. Pedagogical Framework: CEFR B1/B2 Constraints

To engineer an effective prompt, one must first rigorously quantify the target state. The Common European Framework of Reference for Languages (CEFR) provides the theoretical boundaries for the output, but these must be translated into explicit instructions that an LLM can parse and execute.

### 2.1 Defining the Proficiency Thresholds

The transition from B1 (Threshold) to B2 (Vantage) is often described as the "intermediate plateau," where learners move from concrete, everyday language to more abstract, complex thought.

#### 2.1.1 Vocabulary Architecture

The lexical constraints for B1 and B2 learners are distinct and measurable. Research indicates a significant jump in vocabulary size between these levels, which the system prompt must respect to avoid overwhelming the learner.

| **CEFR Level** | **Passive Vocabulary (Word Families)** | **Active Vocabulary (Word Families)** | **Lexical Characteristics** |
| --- | --- | --- | --- |
| **B1 (Threshold)** | 2,000 – 3,000 5 | 1,200 – 2,000 | Focus on high-frequency vocabulary (NGSL top 2,000). Circumlocution strategies are used for unknown words.6 |
| **B2 (Vantage)** | 3,000 – 4,000 5 | 2,500 – 3,000 | Good range for most general topics and own field. Ability to vary formulation to avoid repetition, though gaps cause hesitation.7 |
| **C1 (Advanced)** | 8,000+ 8 | 4,000 – 5,000 | Broad lexical repertoire, idiomatic expressions, and little searching for expressions.6 |

For the Ovi system, this implies a strict "Lexical Budget" for the generated news summaries. The prompt must instruct the model to prioritize the **New General Service List (NGSL)** for the core narrative. Any vocabulary falling outside the top 3,000 words (the B2 ceiling) must be treated as a "Target Word" and explicitly defined in the accompanying glossary.9 Failure to adhere to this constraint results in text that is "frustration-level" reading rather than "instructional-level."

#### 2.1.2 Syntactic Complexity and Sentence Length

Beyond vocabulary, sentence structure is the primary driver of difficulty. BBC and NPR news reports often utilize "information-dense" structures, characterized by heavy nominalization, passive voice, and multiple embedded relative clauses.

For a B1 learner, the cognitive load of holding a 30-word sentence in working memory while decoding syntax is prohibitive. Research from readability studies suggests that comprehension drops significantly as sentence length exceeds 25 words.11

* **B1 Target:** The system prompt must enforce an average sentence length of **15 words**, with a hard maximum of **20-25 words**.11
* **B2 Target:** Complexity can increase slightly to **18-25 words**, allowing for limited subordination (e.g., "The bill, which was signed yesterday, will...").13
* **Grammatical Restrictions:** The prompt must explicitly blacklist C-level structures for B1 outputs. For instance, the *Future Perfect Continuous* ("will have been doing") or complex *Inversion* ("Rarely have I seen...") should be rewritten into simpler constituent clauses.14

### 2.2 The Japanese Learner Profile (L1 Interference)

The Ovi system targets a specific L1 background: Japanese. This requires the prompt engineering to go beyond generic ESL (English as a Second Language) rules and address specific "interlanguage" errors common to Japanese speakers.

#### 2.2.1 The Wasei-Eigo (Loanword) Trap

Japanese utilizes thousands of English loanwords (*gairaigo*), but their meanings often diverge significantly from standard English. This phenomenon, known as *wasei-eigo* (Japan-made English), creates "false friends" that actively hinder comprehension.16

* **Prompt Strategy:** The system must actively scan the input text for these "dangerous" words. For example, if the news article mentions a "mansion" (large estate), the system must flag it because in Japanese, *mansion* (マンション) means a concrete condominium. If the article mentions a "claim" (assertion), the system must note that in Japanese, *claim* (クレーム) usually implies a consumer complaint.16
* **Glossary Requirement:** The JSON output must include a field for cultural\_nuance where these specific L1 interference points are explained.

#### 2.2.2 The Contextual Gap

Japanese communication often relies on "high-context" assumptions, whereas Western news writing is "low-context" but assumes a shared Western cultural background. Concepts like "filibuster," "indictment," or "picket line" are not just vocabulary problems; they are conceptual gaps.

* **Schema Deficiency:** Research indicates that Japanese learners often fail to comprehend Western news not because of language, but because they lack the "script" for the event.2
* **Prompt Strategy:** The prompt must instruct the model to act as a "cultural mediator," identifying these Western-centric concepts and providing brief, context-setting explanations in the JSON output, separate from the translation.18

## 3. Computational Linguistics Strategy: Text Simplification

Transforming authentic news into B1/B2 text requires a rigorous simplification algorithm. This is not merely "summarization," which reduces length, but "simplification," which reduces linguistic complexity while retaining information density.

### 3.1 Lexical Simplification Algorithms

The automation of lexical simplification typically follows a pipeline: **Complex Word Identification (CWI) -> Substitution Generation -> Substitution Selection -> Ranking**.19

* **CWI in Prompts:** The prompt must instruct the GLM-4.7 model to identify "Target Words." These are defined as words that are essential to the story (domain-specific) or are high-utility academic words (Tier 2 vocabulary).
* **Substitution Rule:** For words that are *not* essential to the core narrative (e.g., "exacerbate"), the prompt must instruct the model to substitute them with B1/B2 equivalents (e.g., "make worse" or "worsen").20
* **Retention Rule:** For words that *are* essential (e.g., "sanctions," "election"), the prompt must instruct the model to *retain* the word but trigger a definition generation sequence in the vocabulary list.21

### 3.2 Syntactic Simplification

Syntactic simplification involves splitting long sentences and canonicalizing word order.

* **Sentence Splitting:** The prompt must explicitly instruct the model to identify "split points" at conjunctions (*and, but, because*) and relative pronouns (*who, which, that*). A sentence like *"The president, who was facing intense pressure from the opposition, decided to resign"* should be split into *"The president was facing intense pressure from the opposition. Therefore, he decided to resign."*.20
* **Voice Conversion:** Passive voice ("The decision was made by the committee") is often cognitively heavier for Japanese learners than active voice ("The committee made the decision"). The prompt should favor Active Voice for B1 outputs.15

### 3.3 Readability Metrics as Constraints

While automated metrics like Flesch-Kincaid are imperfect, they serve as useful "guardrails" for the prompt.

* **Target Metrics:** The prompt should instruct the model to aim for a Flesch-Kincaid Grade Level of **5.0–6.0** for B1 and **6.0–8.0** for B2.23
* **Feedback Loop:** In the "Evaluation" phase, these metrics can be calculated on the output. If the output exceeds the target grade level, it indicates a failure of the simplification instructions.

## 4. Prompt Engineering Architecture: The Reasoning-First Approach

The core technical innovation proposed for Ovi English School is the **Reasoning-First** prompt architecture. This approach addresses the limitations of standard "direct-to-JSON" prompting by leveraging the Chain-of-Thought (CoT) capabilities of modern LLMs.

### 4.1 The "Reasoning-Structure Trade-off"

Research into LLM performance suggests a negative correlation between complex formatting requirements and reasoning quality. When a model is tasked with outputting a strict JSON schema directly, a significant portion of its "cognitive budget" (attention mechanism) is allocated to syntax compliance (opening brackets, escaping quotes, matching keys) rather than the semantic quality of the text.1

* **Quantifiable Degradation:** Studies show that enforcing strict JSON output during the reasoning phase can degrade performance on complex tasks by **10–15%**.1
* **Implication for Ovi:** If we ask the model to "Simplify this text and output JSON" in one breath, we risk getting a simplified text that misses key nuances or generates hallucinations because the model "rushed" the simplification to get to the formatting.

### 4.2 Chain-of-Thought (CoT) Integration

To mitigate this, the prompt must explicitly decouple the "Thinking" phase from the "Formatting" phase. This is aligned with the "Interleaved Thinking" capabilities of GLM-4.7.4

* **Mechanism:** The prompt instructs the model to first output a "Thinking Block" (e.g., wrapped in XML tags <thinking>...</thinking>) where it performs the linguistic analysis, plans the simplification, selects vocabulary, and drafts the quiz questions.
* **Benefit:** This allows the model to "scratchpad" its work. It can draft a simplified sentence, check it against the 15-word limit, and rewrite it *before* committing it to the final JSON string. This "intermediate computation" significantly improves the adherence to complex constraints like CEFR levels.26

### 4.3 Prompt Chaining vs. Single-Turn CoT

While "Prompt Chaining" (breaking the task into multiple API calls) offers high reliability, it introduces latency and cost. For a real-time or batch-processing news podcast, a **Single-Turn CoT** approach is often more efficient.

* **Single-Turn Strategy:** The prompt asks for the Thinking Block and the JSON Block in the same response. This leverages the 128k context window of GLM-4.7 effectively while minimizing network overhead.28
* **Validation:** A post-processing script simply regex-matches the JSON block, parses it, and discards the thinking block, keeping the user-facing content clean.

## 5. System Prompt Development

The construction of the system prompt is an iterative process. Based on the requirements analysis, the following components are essential.

### 5.1 Role Definition and Persona

The "Persona" pattern is effective in priming the model's latent knowledge. For Ovi, the persona is not just "an AI assistant" but a specialized expert.

* **Role:** "Expert Applied Linguist and EFL Content Developer specializing in Japanese adult learners."
* **Tone:** "Professional, encouraging, yet linguistically precise."
* **Objective:** "To adapt authentic news into accessible, high-fidelity learning materials without 'dumbing down' the intellectual content."

### 5.2 The "Algorithm" within the Prompt

The prompt must act as code, defining a step-by-step algorithm for the model to follow in its thinking block.

**Step 1: Text Analysis (Internal Monologue)**

"Analyze the input text for complexity. Identify the top 3 concepts that a Japanese learner might lack cultural context for. Identify any 'wasei-eigo' false friends."

**Step 2: Drafting & Simplification (Internal Monologue)**

"Draft the simplified text. Check sentence lengths. If a sentence exceeds 20 words, split it. Ensure cohesive devices (transitions) are used to help flow."

**Step 3: Vocabulary Selection (Internal Monologue)**

"Select 5-7 target words. Priority given to AWL (Academic Word List) words. Draft definitions in simple A2 English."

**Step 4: Quiz Generation (Internal Monologue)**

"Draft 3 multiple-choice questions. Ensure one tests 'Gist', one tests 'Detail', and one tests 'Inference'. Review distractors to ensure they are plausible but incorrect."

**Step 5: JSON Generation (Final Output)**

"Format the final content into the strictly defined JSON schema."

### 5.3 Handling "Hallucination" and Factuality

News content demands high factual accuracy. Simplification often induces "semantic drift" where the meaning changes slightly.

* **Constraint:** "Do not invent details. If a sentence is too complex to simplify without losing meaning, remove the detail rather than guessing."
* **Verification:** The CoT block can include a step: "Verify that the simplified summary retains the 5 Ws (Who, What, Where, When, Why) of the original text."

### 5.4 Addressing the "Wasei-Eigo" Requirement

This is a unique feature of the Ovi system. The prompt must include a "Knowledge Injection" or specific instruction about this.

* **Instruction:** "Check for words like 'Smart' (slender in JP), 'Cunning' (cheating in JP), 'Mansion' (apartment in JP). If found, add a specific note in the vocabulary object: cultural\_nuance: 'Note: In English, this word means X, unlike the Japanese loanword which means Y.'"

## 6. System Prompt Template

The following section presents the comprehensive System Prompt Template designed for GLM-4.7. It utilizes XML-style delimiters for clarity and enforces the Reasoning-First architecture.

### 6.1 The Master Prompt

# SYSTEM PROMPT FOR OVI ENGLISH SCHOOL

## ROLE

You are "Ovi," an elite Applied Linguist and EFL Materials Developer specializing in teaching English to Japanese adult learners. Your expertise lies in adapting complex authentic news (BBC/NPR) into CEFR B1/B2 level learning materials. You balance linguistic simplification with intellectual engagement, ensuring the content remains authentic.

## OBJECTIVE

Your task is to process a raw news article and output a structured JSON object containing a simplified summary, vocabulary glossary, and comprehension quiz. This output will be used to generate a PDF study guide and a TTS podcast script.

## TARGET AUDIENCE PROFILE

* **L1 Language:** Japanese.
* **Target Proficiency:** CEFR B1 (Intermediate) to B2 (Upper Intermediate).
* **Linguistic Profile:** Strong passive vocabulary knowledge (can recognize difficult words) but weaker active grammar skills. Struggles with complex sentence structures (nested relative clauses) and rapid phonetic processing.
* **Specific Challenges:**
  + **Wasei-Eigo:** Confuses English words with their Japanese loanword counterparts (e.g., "claim," "mansion," "tension").
  + **Cultural Context:** Lacks background knowledge on Western political/social concepts (e.g., "filibuster," "tipping," "jury duty").

## PROCESS (CHAIN OF THOUGHT)

You must perform the following steps in a hidden <thinking> block before generating the JSON:

1. **Analyze Complexity:** Scan the input text for CEFR C1/C2 vocabulary and complex syntax.
2. **Cultural Scan:** Identify 1-2 concepts that require cultural explanation for a Japanese audience.
3. **Wasei-Eigo Scan:** Check for "False Friend" loanwords.
4. **Draft Simplification:** Rewrite the text to B1/B2 standards.
   * **Sentence Length:** Aim for 15-20 words average. Max 25 words.
   * **Syntax:** Convert Passive Voice to Active Voice. Split compound-complex sentences.
   * **Vocabulary:** Prioritize NGSL (New General Service List) words. Keep domain-specific terms but flag them for the glossary.
5. **Draft Quiz:** Create 3 Multiple Choice Questions (MCQs) based on Bloom's Taxonomy (Knowledge, Comprehension, Inference). Ensure distractors are plausible.

## JSON OUTPUT SCHEMA

You must strictly adhere to the following JSON structure:json

{

"meta": {

"title": "String (Engaging B1-level title)",

"cefr\_level": "String (e.g., 'B1')",

"original\_source": "String",

"reading\_time\_minutes": "Number"

},

"content": {

"summary\_intro": "String (1-2 sentence hook)",

"paragraphs":

},

"vocabulary":,

"cultural\_context":,

"quiz":,

"correct\_index": "Number (0-3)",

"explanation": "String (Why the answer is correct)"

}

]

}

## CONSTRAINTS & GUIDELINES  
- \*\*Output Format:\*\* ONLY return the valid JSON object. Do not add markdown code blocks like ```json... ``` around it.  
- \*\*Language:\*\* The `content` and `definition` fields must be in simplified English. The `japanese\_translation` field must be in Japanese.  
- \*\*Tone:\*\* Objective, journalistic, but accessible.  
  
## FEW-SHOT EXAMPLES  
  
### Example 1 (Political News)  
\*\*Input:\*\* "The incumbent Prime Minister's decision to dissolve parliament has precipitated a snap election, causing consternation among the opposition."  
\*\*Thinking:\*\* "Incumbent" and "precipitated" are too hard. "Snap election" needs explanation. "Consternation" is C2. I will simplify to "current Prime Minister" and "caused a surprise election." I need to explain "dissolving parliament" in the cultural notes as it's a UK specific process.  
\*\*JSON Output:\*\*  
{  
 "meta": { "title": "Prime Minister Calls Surprise Election",... },  
 "content": { "paragraphs": },  
 "vocabulary":,  
 "cultural\_context":  
}  
  
### INPUT TEXT  
[Insert User News Article Here]

### 6.2 Logic Behind Specific Directives

* **"Do not add markdown code blocks":** This is a specific instruction to aid the parsing pipeline. If the model outputs ```json, the parser has to strip it.
* **"Simple English Definition (A2 Level)":** This ensures that the learner doesn't need a dictionary to understand the dictionary. A B1 learner cannot learn a word if the definition uses C1 words.
* **"Japanese Translation":** Including this requires the model to be multilingual. GLM-4.7 is trained on multilingual data, making it capable of accurate English-to-Japanese mapping for single terms.

## 7. Operational Architecture: The Pipeline

The prompt is just one component of the "Ovi" system. The surrounding operational architecture ensures reliability and scalability.

### 7.1 The Request Pipeline

1. **Preprocessing:** The raw news text (BBC/NPR) is cleaned. Metadata (author, date) is extracted.
2. **Prompt Assembly:** The dynamic system prompt is assembled.
   * *Context Injection:* If the article is about "Economics," specific few-shot examples related to economic news are injected into the prompt to prime the model's vocabulary.29
3. **API Call (GLM-4.7):** The request is sent with temperature=0.2 (low randomness for factual consistency) and max\_tokens=4096.30
4. **Parsing & Validation:** The output is captured.
   * *Regex Extraction:* The script looks for the first { and last } to isolate the JSON.
   * *Schema Validation:* A library like Pydantic (Python) or Zod (JS) validates the structure. Does content.paragraphs exist? Is quiz an array of length 3?
5. **Error Handling (Self-Correction):** If validation fails (e.g., missing comma), the system triggers a "Repair Call."
   * *Repair Prompt:* "The previous output was invalid JSON. Error: [Error Message]. Please regenerate ONLY the JSON.".31

### 7.2 TTS Integration Strategy

The JSON structure is designed for "Audio-First" consumption.

* **Paragraph Segmentation:** The content.paragraphs array allows the TTS engine to insert natural pauses (<break time="500ms"/>) between paragraphs, preventing the "wall of text" audio effect.
* **Vocabulary Drills:** The vocabulary array can be transformed into a drill segment.
  + *Script:* "... [Pause 2s]...... [Pause 2s]...."
  + This allows the user to "guess" the meaning during the pause, an active learning technique known as "spaced retrieval".32

### 7.3 PDF Generation

The same JSON feeds the PDF generator.

* **Layout:** Title and Summary at the top.
* **Body Text:** The simplified text is rendered with line numbers.
* **Glossary Sidebar:** The vocabulary list is rendered in a sidebar or footnotes, aligned with the text occurrence, reducing eye movement and cognitive load.

## 8. Evaluation and Testing Framework

To ensure the "Ovi" system consistently meets the B1/B2 standard, a rigorous testing framework is required, combining automated metrics with human oversight.

### 8.1 Automated Metric Evaluation

Before any content reaches a user, it must pass a suite of automated checks.

* **Readability Scoring:** The textstat library is used to calculate the **Flesch-Kincaid Grade Level** of the content.paragraphs text.
  + *Pass Range:* 5.0 – 8.0.
  + *Fail Action:* If > 8.0, the text is flagged as "Too Complex" and sent for re-simplification.23
* **Lexical Density Check:** The text is tokenized and compared against the NGSL word list.
  + *Metric:* % of words outside the NGSL Top 3000.
  + *Pass Threshold:* < 5-8%. If higher, it suggests the model failed to simplify vocabulary.33
* **JSON Validity Rate:** Monitor the percentage of API calls that result in valid JSON on the first try. A drop in this metric suggests the prompt instructions are becoming ambiguous or the model behavior is drifting.

### 8.2 Qualitative Human Evaluation (Rubric)

For the pilot phase, human experts (EFL teachers) should grade a random sample of outputs using the following rubric:

| **Criterion** | **Score (1-5)** | **Description** |
| --- | --- | --- |
| **Fidelity** | 1-5 | Does the simplified text accurately reflect the facts of the original news story without hallucination? |
| **Naturalness** | 1-5 | Is the English natural and idiomatic, or does it sound robotic/stilted? |
| **Level Appropriateness** | 1-5 | Is the grammar/vocabulary truly B1/B2? (e.g., Are there accidental C1 structures?) |
| **Cultural Utility** | 1-5 | Is the cultural\_note accurate and actually helpful for a Japanese learner? |
| **Quiz Quality** | 1-5 | Are the questions valid? Are the distractors plausible but clearly incorrect? |

### 8.3 A/B Testing Prompts

Optimization is ongoing. We propose A/B testing different prompt variations:

* **Prompt A (Single-Shot):** The standard template above.
* **Prompt B (Multi-Shot):** Including 5+ examples of simplified articles.
* **Prompt C (Iterative):** A two-step API call (1. Simplify, 2. JSON Format).
* *Hypothesis:* Prompt B will offer the best balance of quality and latency/cost. Prompt C yields higher quality but doubles the cost and time.28

## 9. Strategic Insights and Future Outlook

### 9.1 The "Meta-Cognitive" Advantage

The primary differentiator of Ovi English School is the embedding of pedagogical "meta-cognition" into the AI. By forcing the model to "think like a teacher" before it "writes like a journalist," we bridge the gap between raw data processing and educational scaffolding. This approach moves beyond simple text simplification tools to create a *learning agent*.

### 9.2 Dynamic Difficulty Adjustment (DDA)

The current architecture produces static B1/B2 content. A future enhancement is **Dynamic Difficulty Adjustment**. By passing a user\_proficiency parameter (e.g., "weak\_grammar", "strong\_vocab") into the prompt, the model could tailor the simplification strategy—perhaps using more complex vocabulary but simpler sentence structures for that specific user profile.

### 9.3 Scalability of Content Types

While this report focuses on News, the architecture is content-agnostic. The same "Reasoning-First" JSON schema approach can be adapted for:

* **Business English:** Replacing "News Analysis" with "Email/Meeting Analysis."
* **Travel English:** Adapting travel blogs or guides.
* **Technical English:** Simplifying documentation for Japanese engineers.

## 10. Conclusion

The "Ovi English School" project represents a sophisticated integration of prompt engineering and pedagogical theory. By adhering to the architectural principles outlined in this report—specifically the use of CoT reasoning to decouple linguistic planning from JSON formatting, the strict enforcement of CEFR B1/B2 constraints, and the culturally aware handling of Japanese learner interference—the system can achieve high-fidelity automation of language learning materials. The success of this system relies not just on the GLM-4.7 model's raw power, but on the precision of the constraints we impose upon it. The result is a system that does not merely translate news, but transforms it into a structured, accessible, and culturally contextualized learning experience.

#### Works cited

1. Beyond JSON: Picking the Right Format for LLM Pipelines, accessed February 1, 2026, <https://medium.com/@michael.hannecke/beyond-json-picking-the-right-format-for-llm-pipelines-b65f15f77f7d>
2. No News is Bad News: Using the Media in Teaching about Japan, accessed February 1, 2026, <https://www.asianstudies.org/publications/eaa/archives/no-news-is-bad-news-using-the-media-in-teaching-about-japan/>
3. More advanced Japanese news articles online seem boring and repetitive. Am I just still bad or is this common? : r/LearnJapanese - Reddit, accessed February 1, 2026, <https://www.reddit.com/r/LearnJapanese/comments/z7cr5g/more_advanced_japanese_news_articles_online_seem/>
4. GLM-4.7 Text Structuring in Action: 3 Steps to Extract Key Information from Complex Documents - Apiyi.com Blog, accessed February 1, 2026, <https://help.apiyi.com/en/glm-4-7-text-structuring-guide-en.html>
5. What are estimates of vocabulary size for each CEFR level?, accessed February 1, 2026, <https://languagelearning.stackexchange.com/questions/3061/what-are-estimates-of-vocabulary-size-for-each-cefr-level>
6. Common European Framework of Reference for Languages: Learning, Teaching, Assessment - https: //rm. coe. int, accessed February 1, 2026, <https://rm.coe.int/168045b15e>
7. The development of vocabulary breadth across the CEFR levels. - EuroSLA, accessed February 1, 2026, <http://www.eurosla.org/monographs/EM01/211-232Milton.pdf>
8. CEFR Levels Explained: Your Comprehensive FAQ ... - Text Inspector, accessed February 1, 2026, <https://textinspector.com/cefr-levels-faq-guide/>
9. How Many Words You Should Know For Every Language Level - The Universe Of Memory, accessed February 1, 2026, <https://universeofmemory.com/how-many-words-you-should-know/>
10. (PDF) Lexical Simplification in Foreign Language Learning: Creating Pedagogically Suitable Simplified Example Sentences - ResearchGate, accessed February 1, 2026, <https://www.researchgate.net/publication/370984996_Lexical_Simplification_in_Foreign_Language_Learning_Creating_Pedagogically_Suitable_Simplified_Example_Sentences>
11. Sentence length: why 25 words is our limit - Inside GOV.UK, accessed February 1, 2026, <https://insidegovuk.blog.gov.uk/2014/08/04/sentence-length-why-25-words-is-our-limit/>
12. Simple sentences - Readability Guidelines, accessed February 1, 2026, <https://readabilityguidelines.co.uk/clear-language/simple-sentences/>
13. CEFR-Based Sentence Difficulty Annotation and Assessment - ACL Anthology, accessed February 1, 2026, <https://aclanthology.org/2022.emnlp-main.416.pdf>
14. accessed February 1, 2026, <https://blog.maltalingua.com/navigating-grammar-at-different-cefr-levels-at-maltalingua-english-school/#:~:text=CEFR%20Level%20B1%3A%20Intermediate&text=Present%20Continuous%3A%20For%20current%20actions,first%20conditionals%20for%20real%20situations.>
15. Table of grammar contents – B1 - Test-English, accessed February 1, 2026, <https://test-english.com/grammar-points/b1/contents-b1/>
16. English Loanwords and Made-in-Japan English in Japanese Mariko Hatanaka\* Justin Pannell Introduction - Hawaii Pacific University, accessed February 1, 2026, <https://www.hpu.edu/research-publications/tesol-working-papers/2016/02HatanakaPannell.pdf>
17. How to Talk About Politics in Japanese - GaijinPot Blog, accessed February 1, 2026, <https://blog.gaijinpot.com/talk-politics-japanese/>
18. Bridging Cultures: A Japanese Student's Path to Intercultural Communication - Macquarie University, accessed February 1, 2026, <https://research-management.mq.edu.au/ws/portalfiles/portal/459691895/Publisher_version.pdf>
19. Lexical Simplification in Foreign Language Learning: Creating Pedagogically Suitable Simplified Example Sentences - ACL Anthology, accessed February 1, 2026, <https://aclanthology.org/2022.tsar-1.9.pdf>
20. Simplification - NLP-progress, accessed February 1, 2026, <http://nlpprogress.com/english/simplification.html>
21. Aligning Sentence Simplification with ESL Learner's Proficiency for Language Acquisition - ACL Anthology, accessed February 1, 2026, <https://aclanthology.org/2025.naacl-long.21.pdf>
22. Two minutes NLP — Quick intro to Text Simplification | by Fabio Chiusano | Generative AI, accessed February 1, 2026, <https://medium.com/nlplanet/two-minutes-nlp-quick-intro-to-text-simplification-f5f9d7be4a3c>
23. [Textstat] How to evaluate readability? - Kaggle, accessed February 1, 2026, <https://www.kaggle.com/code/yhirakawa/textstat-how-to-evaluate-readability>
24. Flesch Reading Ease and the Flesch Kincaid Grade Level - Readability score, accessed February 1, 2026, <https://readable.com/readability/flesch-reading-ease-flesch-kincaid-grade-level/>
25. Learning to Generate Structured Output with Schema Reinforcement Learning - arXiv, accessed February 1, 2026, <https://arxiv.org/html/2502.18878v1>
26. How Chain of Thought (CoT) Prompting Helps LLMs Reason More Like Humans | Splunk, accessed February 1, 2026, <https://www.splunk.com/en_us/blog/learn/chain-of-thought-cot-prompting.html>
27. Chain-of-thought prompting: When and why to use it (with examples that actually show the difference) - Reddit, accessed February 1, 2026, <https://www.reddit.com/r/PromptEngineering/comments/1ppwi7i/chainofthought_prompting_when_and_why_to_use_it/>
28. JSON Whisperer: Efficient JSON Editing with LLMs - arXiv, accessed February 1, 2026, <https://arxiv.org/html/2510.04717v1>
29. Prompt engineering best practices for In-Context Learning : r/LLMDevs - Reddit, accessed February 1, 2026, <https://www.reddit.com/r/LLMDevs/comments/1g8uio9/prompt_engineering_best_practices_for_incontext/>
30. Best practices for prompt engineering with the OpenAI API, accessed February 1, 2026, <https://help.openai.com/en/articles/6654000-best-practices-for-prompt-engineering-with-openai-api>
31. Crafting Structured {JSON} Responses: Ensuring Consistent Output from any LLM, accessed February 1, 2026, <https://dev.to/rishabdugar/crafting-structured-json-responses-ensuring-consistent-output-from-any-llm-l9h>
32. Vocabulary in Japanese EFL Textbooks: A Bidirectional Coverage Analysis - ResearchGate, accessed February 1, 2026, <https://www.researchgate.net/publication/359258584_Vocabulary_in_Japanese_EFL_Textbooks_A_Bidirectional_Coverage_Analysis>
33. A Close Examination of Vocabulary in Japanese EFL Textbooks - JALT Publications |, accessed February 1, 2026, <https://jalt-publications.org/sites/default/files/pdf-article/jalt2021-pcp-024.pdf>