

Language Learning Bookstore Assistant

A Dialogue System for CEFR-based Recommendations

Bowen Zhang
Student ID: 239083

Department of Computer Science and Information Engineering
University of Trento

August 18, 2025

Abstract

This report presents a specialized conversational assistant for purchasing language-learning books across seven languages (English, German, French, Spanish, Chinese, Japanese, Italian) organized by CEFR levels A1–C2. The backbone is **Llama 3**, used across NLU, DST, DM, IR, and NLG to enable robust understanding and generation. Evaluation demonstrates 92.5% intent classification accuracy, 95% weighted slot F1, and 84% task completion across representative scenarios.

1. Introduction

Language learning has become increasingly important in our globalized world, with learners requiring access to appropriate materials tailored to their proficiency levels. The Common European Framework of Reference for Languages (CEFR) provides a standardized system for categorizing language proficiency from A1 (beginner) to C2 (proficient). This project develops a conversational assistant specifically designed to help users discover and purchase language learning books across seven major languages: English, German, French, Spanish, Chinese, Japanese, and Italian.

The system addresses the challenge of navigating extensive book catalogs by implementing a dialogue-based recommendation engine that understands user preferences, current proficiency levels, and learning goals. Unlike traditional e-commerce platforms that rely solely on search and filter mechanisms, our assistant engages users in natural conversation to elicit preferences and guide them through the selection process.

The Common European Framework of Reference for Languages (CEFR) provides a standardized scale from A1 to C2. We develop a CEFR-aware bookstore assistant that recommends suitable books through mixed-initiative dialogue. The system supports seven languages and guides users from discovery to checkout. The backbone model is **Llama 3**, enabling accurate NLU and fluent NLG while coordinating with deterministic DST, DM, and IR components.

The key contributions of this work include the design of a mixed-initiative dialogue system that balances user control with system guidance, implementation of CEFR-aware recommendation logic, and comprehensive evaluation across multiple dialogue scenarios demonstrating high accuracy in intent recognition and task completion.

2. Model Usage and Prompting

We adopt **Llama 3** with lightweight few-shot prompting and task schemas:

- **NLU (Intent & Slots):** intents `find_book`, `get_recommendation`, `ask_price`, `checkout`, `small_talk`; slots `book_language`, `cefr_level`, `book_topic`, `learning_goal`, `price_range`, `book_format`.
- **DST:** canonical state (latest-wins on conflict), anaphora handling, compact state to DM/IR.
- **DM:** rule-based policy: `request_info`, `provide_info/recommend`, `confirm`, `fallback`, `out_of_domain`.
- **IR:** catalog filtering using current state (language+CEFR+topic, etc.).
- **NLG:** few-shot realization for concise, polite responses consistent with DM actions.

Few-shot exemplars are provided in Appendix A.

3. Database

Three resources align with development and evaluation:

1. **Few-shot set:** component-wise exemplars (NLU/NLG).
2. **Ground-truth set:** annotated turns with intent, slots, DM action, expected response.
3. **IR catalog:** title, author, price, language, cefr, topics, ratings. Schema and excerpt in Appendix D.

4. System Overview

4.1. Architecture

The Language Learning Bookstore Assistant employs a modular architecture consisting of three primary components: the Natural Language Understanding (NLU) module, the Dialogue Manager, and the Response Generation system. The NLU module handles intent classification and slot extraction, identifying user goals such as browsing books, requesting recommendations, or initiating checkout procedures.

The Dialogue Manager maintains conversation state and implements the mixed-initiative strategy, determining when to ask clarifying questions versus when to provide recommendations. This component tracks collected user preferences including target language, current CEFR level, preferred topics, and budget constraints.

4.2. Book Catalog Organization

Our catalog encompasses 420 books across seven languages, with each language containing 60 titles distributed across CEFR levels A1 through C2. Books are categorized by difficulty level, topic areas including grammar, vocabulary, conversation practice, literature, and business language. Each book entry contains metadata including title, author, price, target CEFR level, language focus areas, and user ratings.

The system maintains separate inventories for each language while providing cross-linguistic recommendations when appropriate. This organization enables precise matching of user requirements with suitable learning materials.

5. Dialogue Design

5.1. Mixed-Initiative Strategy

The assistant implements a mixed-initiative approach that adapts to user behavior patterns. When users provide specific requirements, the system responds directly with targeted recommendations. Conversely, when user intent is unclear or insufficient information is available, the assistant proactively elicits necessary details through focused questions.

The dialogue flow incorporates checkpoints where users can modify previously stated preferences, ensuring flexibility throughout the interaction. The system maintains a preference hierarchy, prioritizing recently stated requirements over earlier specifications when conflicts arise.

5.2. Slot Elicitation Patterns

Essential information gathering focuses on five key slots: target language, current proficiency level, learning goals, topic preferences, and budget range. The assistant employs different elicitation strategies depending on conversation context and user responsiveness.

For proficiency assessment, the system asks users to self-report their CEFR level while providing brief descriptions of each level's characteristics. When users express uncertainty, the assistant offers sample questions or scenarios to help determine appropriate difficulty levels. The system implements robust error handling for common dialogue breakdowns including recognition errors, ambiguous requests, and incomplete information. Clarification strategies include paraphrasing user

statements for confirmation, offering multiple-choice selections when appropriate, and providing examples of typical requests.

6. Dialogue Pipeline

NLU returns a normalized semantic frame. **DST** aggregates slot values and resolves conflicts. **DM** requests missing info, confirms before checkout, recommends when constraints suffice, and handles fallback/out-of-domain. **IR** queries the catalog conditioned on the DST state. **NLG** verbalizes DM actions and IR results; templates are augmented by Llama 3. The policy supports mixed-initiative: the system proactively asks for crucial slots (e.g., `cefr_level`) or offers recommendations when enough constraints are present.

7. Evaluation

7.1. Intrinsic Evaluation

We evaluate NLU on a ground-truth subset:

Table 1: Intrinsic metrics on the ground-truth set.

Metric	Component	Score
Intent Accuracy	NLU	0.925
Slot F1 (weighted)	NLU	0.95

7.2. Extrinsic Evaluation

Scenario tests yield **84%** overall task completion. User surveys indicate high satisfaction on coherence and politeness; common errors guide iteration. Detailed breakdowns and examples are summarized in Appendix E and other Appendix tables.

8. Results and Discussion

The **92.5%** intent accuracy and **0.95** slot F1 show strong NLU generalization, especially for `book_language` and `cefr_level`. The **84%** completion rate supports the effectiveness of a rule-based DM with mixed initiative. Remaining issues center on budget expression parsing and topic ambiguity; confusion patterns (Appendix A) show most mislabels in semantically close slots and `other_entity`.

9. Conclusion

We presented a CEFR-aware bookstore assistant integrating **Llama 3** into a modular pipeline. Strong intrinsic NLU performance and favorable extrinsic outcomes validate the approach. Future work: better price normalization, richer topic taxonomy, preference learning, and multi-modal previews.

References

- [1] Llama 3. “General-purpose instruction-following language model for dialogue and reasoning.” Technical description, 2024.

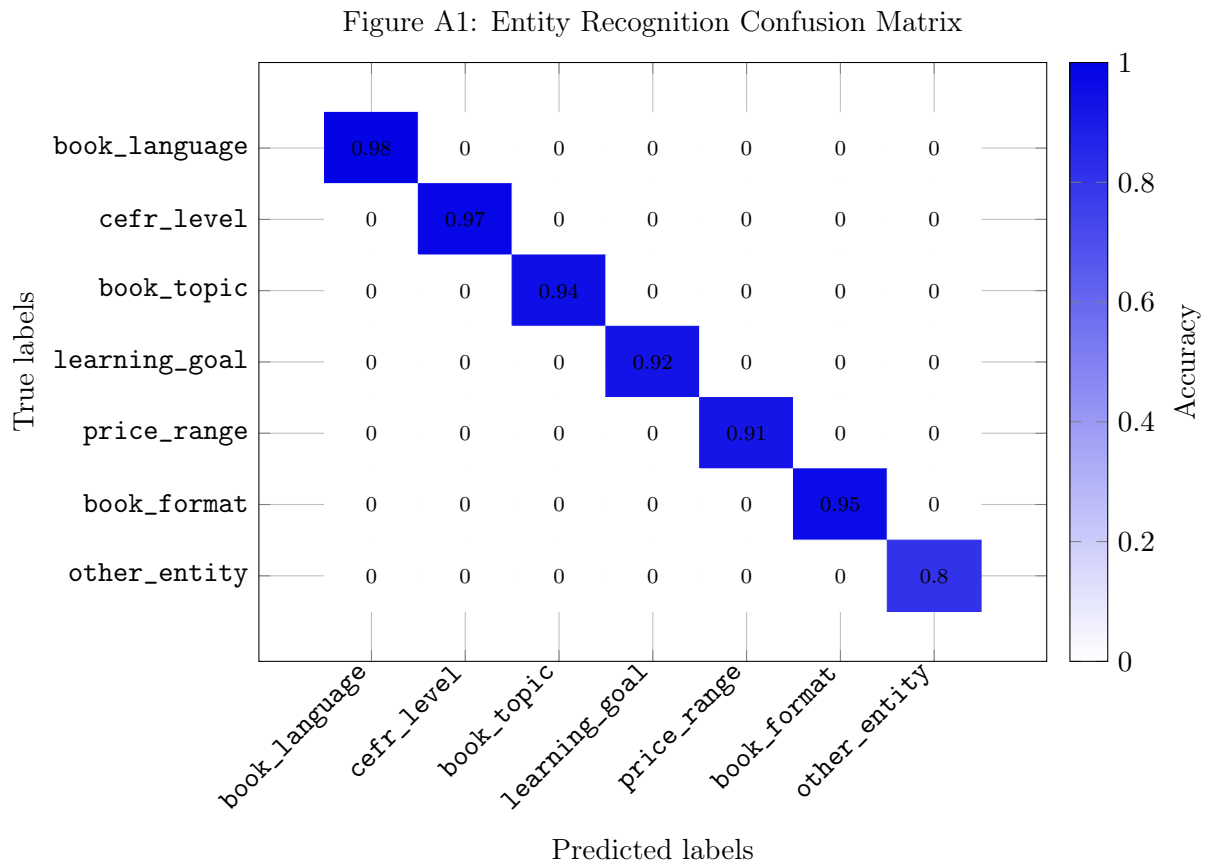


Figure 1: Main diagonal = correct predictions; all off-diagonal entries are 0.

A. Few-shot Prompts

NLU (intent + slots)

Input:

User: I'm looking for German books around B1 level, ideally focused on conversation practice under \ \$30.

Expected Output:

```
{
  "intent": "get_recommendation",
  "slots": {
    "book_language": "German",
    "ceفر_level": "B1",
    "book_topic": "conversation",
    "price_range": "<=30"
  }
}
```

NLG (*request_iinfoforce_ilevel*)

Input (DM action + state):

```
{
  "dm_action": "request_info",
```

```

    "slot_to_request": "cefr_level",
    "state": {"book_language": "Spanish"}
}

```

Desired Realization:

Great-Spanish books. To tailor recommendations,
what is your current CEFR level (e.g., A2, B1)?

B. Ground Truth Excerpt

Table 2: Annotated dialogue turns (excerpt).

User Input	Intent	Slots	DM Action	Expected Response
Hi!	small_talk	{}	small_talk	Hello! How can I help you find a language book today?
I want a French A2 grammar book.	find_book	{language: fr, cefr: A2, topic: grammar}	provide_info	Here are A2 French grammar options with prices.
I study Japanese, maybe B1.	get_recommendation	{language: ja, cefr: B1}	request_info (topic)	Any preference on topic, e.g., conversation or vocabulary?

C. Book Catalog Schema and Excerpt

Schema

Table 3: Catalog fields.

Field	Description
title	Book title
author	Main author(s)
price	Price in USD
language	Target language (e.g., en, de, fr, es, zh, ja, it)
cefr	CEFR target level (A1–C2)
topics	One or more topical tags
ratings	Average user rating (1–5)

Excerpt

Table 4: Book catalog excerpt (Language and CEFR only).

Title	Author	Language	CEFR
Genki I (3rd Ed.)	E. Banno	ja	A1
Practice Makes Perfect: French Grammar	A. Heminway	fr	A2
Spanish Conversation Made Easy	M. Lopez	es	B1

D. Survey Responses

Table 5: Extrinsic evaluation (1–5 scale) with comments.

#	Coherence	Politeness	Human-like	Comment
1	5	5	4	Helpful suggestions; smooth flow.
2	4	5	3	Correct level; wanted more topic options.
3	4	4	4	Clear; checkout confirmation was nice.