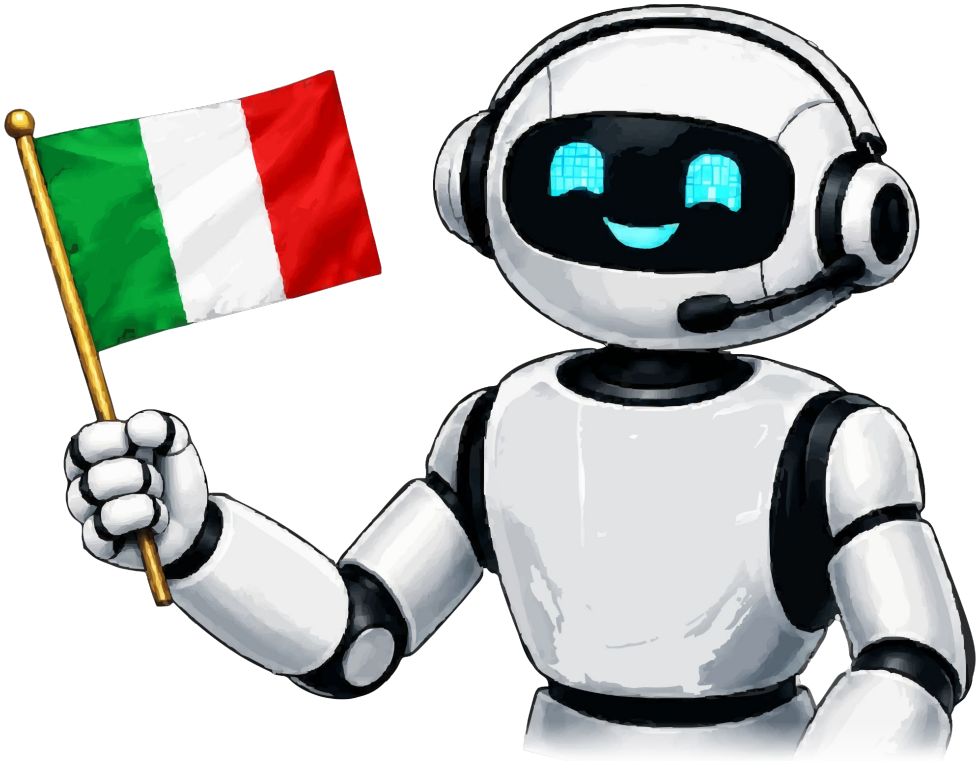
UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

UPO UNIVERSITÀ DEL PIEMONTE ORIENTALE



Designing task-based chatbots for L2 pragmatics learning: experiments and preliminary studies

Federica Turco, Giuliana Franceschinis,
María Joaquina Valero Gisbert

ILL Conference 2026

UPO UNIVERSITÀ DEL PIEMONTE ORIENTALE



UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

Designing task-based chatbots for L2 pragmatics learning: experiments and preliminary studies

🎯 PragmaBot's context:

Introduction and Methodology.

🧩 An iterative design-and-test process:

Discussion of the preliminary experiments.

🎓 The future PragmAI prototype:

The next design step.



PragmaBot’s context

Introduction and Methodology

UPO UNIVERSITÀ DEL PIEMONTE ORIENTALE



UNIVERSITAS STUDIORUM MUTINENSIS ET REGIENSIS 1175

UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

PragmaBot's context: Introduction

AIALL: AI tools and techniques support language-learning processes (Bahari et al. 2025)

- **Conversational Agents:** extend learning opportunities + support interactive and oral language practice;
- **Recent Studies**: writing skills (Mezzadri & Tangorra, 2025) + learners' perceptions (Filippetti et al., 2025).

→ Limits of prompt and knowledge-based customisation to inform the design.

Pedagogical Framework

- Technology-Mediated L2 Pragmatics (González-Lloret, 2019);
- TBLT / TBLL (Long, 2015; Ellis, 2003);
- **Tech-mediated TBLT** (Bhandari et al., 2025).

Target Focus: Pragmatic Competence

Definition & Scope: using and interpreting language appropriately with linguistic resources (García de Toro, 2004).
L2 Italian Context: knowledge of grammatical form is insufficient; pragmatic instruction remains comparatively underrepresented and predominantly form-focused (Gesuato & Castello, 2022).
Key AI Challenge: creating context-dependent and socially situated learning experiences in AI-supported environments.

UPO UNIVERSITÀ DEL PIEMONTE ORIENTALE



UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

PragmaBot's context: Research Questions

Can currently available customisable LLM-based systems be effectively guided through increasingly structured prompting and externally provided knowledge to sustain task-based interactions that are pedagogically coherent and pragmatically appropriate?

Chatbot's design and behaviour:

conditions required for meaningful pedagogical integration of GenAI to support pragmatic competence.

A broader question:

how an LLM could be controlled to support a specific language-learning objective in conversation with a learner?

An intermediate step:

preliminary stage of a doctoral research project.

UPO UNIVERSITÀ DEL PIEMONTE ORIENTALE

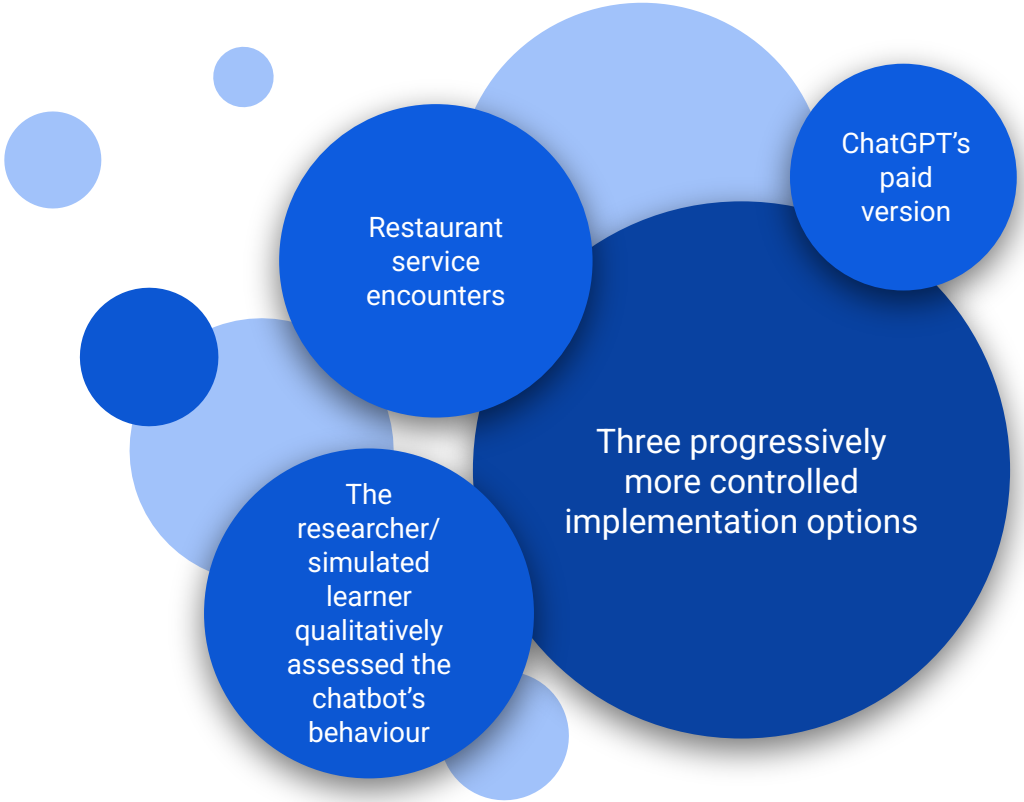


UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

An iterative design-and-test process

Discussion of the preliminary experiments

An iterative design-and-test process: Methodology



Which type of LLM-based conversational system could more reliably support task-based practice of Italian L2 pragmatics?

UPO UNIVERSITÀ DEL PIEMONTE ORIENTALE



UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

An iterative design-and-test process: Exploratory conditions

1

A standard ChatGPT conversation guided through system prompting:

to observe the general-purpose model's default conversational tendencies before introducing stronger constraints.

2

Customised GPTs combining prompting with externally provided knowledge:

structuring knowledge more explicitly and establishing a clearer hierarchy of instructions.

3

An agentic application:

the subsequent design direction emerging from the limitations.

An iterative design-and-test process: Exploratory conditions

Two phases of exploratory testing

Aspect	Phase 1 Prompt-guided ChatGPT	Phase 2 Customised GPTs
Aim	Observe the model's default conversational behaviour.	Increase control over pragmatic and pedagogical behaviour.
Instructions and knowledge	A general role-play prompt; no uploaded knowledge files.	2.1 Progressively detailed specifications; 2.2 scenario, server and Tutor files; 2.3 abstract behavioural schema.
Interaction mode	Preliminary oral conversations (restaurant).	Oral interactions, followed by written tests with PragmaBot 1 and PragmaBot-Astratto.
Main observations	The setting holds, but the model assumes familiarity and informal "tu".	More stable interaction in later tests; abstract rules still require scenario knowledge.

Configurations and modalities varied across tests.

UPO UNIVERSITÀ DEL PIEMONTE ORIENTALE



UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

An iterative design-and-test process: Exploratory conditions

1

A standard ChatGPT conversation guided through system prompting:

to observe the general-purpose model's default conversational tendencies before introducing stronger constraints.

2

Customised GPTs combining prompting with externally provided knowledge:

structuring knowledge more explicitly and establishing a clearer hierarchy of instructions.

3

An agentic application:

the subsequent design direction emerging from the limitations.



An iterative design-and-test process: Exploratory conditions

1. Standard ChatGPT conversations:

Configuration:

+ general prompt (role-play).

Tests:

- preliminary oral interaction tests (speech-recognition and transcription errors).

Results:

- the model maintained the setting;
- autonomously inferred unspecified pragmatic (interpersonal familiarity; the informal “tu”) and contextual (turn progression, advancement of the simulated interaction) variables.
- Role assignment alone is insufficient to control context-dependent behaviour → need for greater control.



An iterative design-and-test process: Exploratory conditions

1

A standard ChatGPT conversation guided through system prompting:

to observe the general-purpose model's default conversational tendencies before introducing stronger constraints.

2

Customised GPTs combining prompting with externally provided knowledge:

structuring knowledge more explicitly and establishing a clearer hierarchy of instructions.

3

An agentic application:

the subsequent design direction emerging from the limitations.

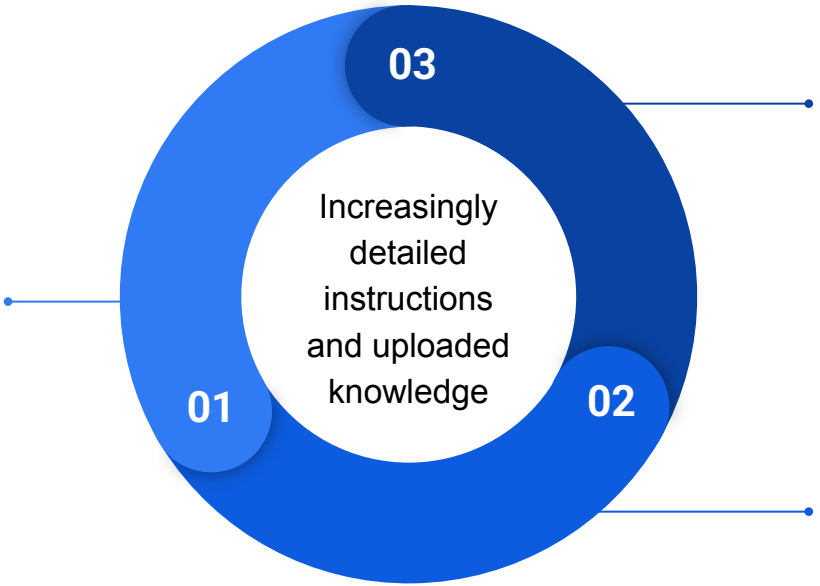
An iterative design-and-test process: Exploratory conditions

2. Customised GPTs: prompting + externally provided knowledge

Increasingly detailed specifications

prompts + theoretical materials
+ modified functionalities
+ design and requirements document;

GPT-5.2.



Abstraction of the general behavioural rules:

modified prompt
+ a single configuration file;

GPT-5.6 Sol.

Restructuring the knowledge

prompt + (design and requirements document) + configuration files;

GPT-5.6 Sol.

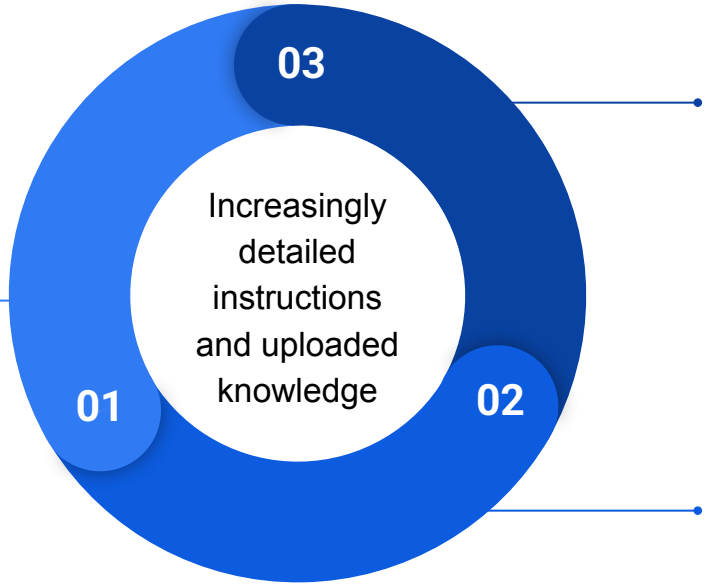
An iterative design-and-test process: Exploratory conditions

2. Customised GPTs: prompting + externally provided knowledge

Increasingly detailed specifications

prompts + theoretical materials
+ modified functionalities
+ design and requirements document;

GPT-5.2.



Abstraction of the general behavioural rules:

modified prompt
+ a single configuration file;

GPT-5.6 Sol.

Restructuring the knowledge

prompt + (design and requirements document) + configuration files;

GPT-5.6 Sol.

UPO UNIVERSITÀ DEL PIEMONTE ORIENTALE



UNIMORE UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

An iterative design-and-test process: Exploratory conditions

2.1. Customised GPTs: increasingly detailed specifications

The three main configurations:

1. PragmAI (GPT-5.2):

Configuration:

+ prompt;
+ **theoretical materials**

Prompt:

- wide goal and few constraints.

Materials regarding:

- TBLT
- Second-language teaching
- CEFR

2. PragmAI 2 (GPT-5.6 Sol):

Configuration:

+ **modified prompt**;
+ theoretical materials.

Prompt with operational instructions:

- context, pedagogical objectives, role, conversational behaviour.

Improved aspects:

- role consistency (first message);
- use of formal address ("Lei");
- interpretation of indirect moves;
- turn progression.

Persistent issues:

Remaining challenges across iterations:

- verbosity;
- excessive deference;
- incomplete action execution;
- unreliable feedback activation.

An iterative design-and-test process: Exploratory conditions

2.1. Customised GPTs: increasingly detailed specifications

The three main configurations:

3. Pragmabot (GPT-5.6 Sol):

Configuration:

+ prompt;
+ theoretical materials;
+ **design-and-requirements document**;
+ **technical changes**.

Design-and-requirements document: operational aspects

- role consistency, task progression, pragmatic support, interactional brevity, and feedback.

Technical changes:

- data analysis enabled.

Results:

- the chatbot did not consistently display the required **conversational behaviour**;
- difficulty **prioritising** instructions across heterogeneous materials.

→ shift from enlarging the prompt and knowledge base to extracting operational principles into dedicated configuration files.

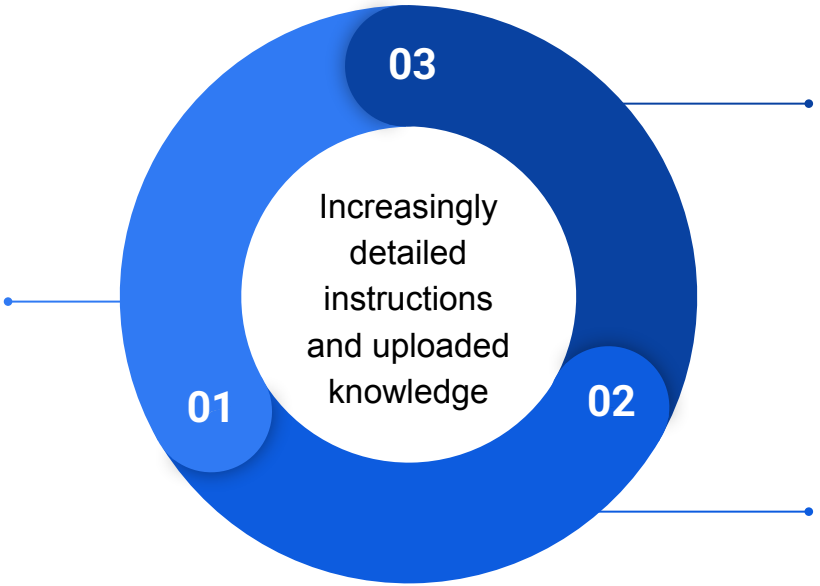
An iterative design-and-test process: Exploratory conditions

2. Customised GPTs: prompting + externally provided knowledge

Increasingly detailed specifications

prompts + theoretical materials
+ modified functionalities
+ design and requirements document;

GPT-5.2.



Abstraction of the general behavioural rules:

modified prompt
+ a single configuration file;

GPT-5.6 Sol.

Restructuring the knowledge

prompt + (design and requirements document) + configuration files;

GPT-5.6 Sol.

An iterative design-and-test process: Exploratory conditions

2.2. Customised GPTs: restructuring the knowledge

Two main configurations (GPT-5.6 Sol; elegant restaurant):

1. PragmaBot:

Configuration:

+ **revised prompt**;
+ design-and-requirements document;
+ **configuration files**.

The revised prompt allows the chatbot to:

- provide linguistic and pragmatic support when necessary.

Three complementary configuration files in Markdown format:

1. conversational role and its behaviour;
2. restaurant scenario, interactional phases, and situational constraints;
3. post-task pedagogical evaluation and feedback.

Tests:

Four oral interactions selected; profiles varied in proficiency and behaviour:

- B1 learner with stronger comprehension than oral production;
- advanced learners (C1-C2 range);
- pragmatic errors deliberately introduced → chatbot's conversational reaction and its subsequent evaluation.

An iterative design-and-test process: Exploratory conditions

2.2. Customised GPTs: restructuring the knowledge

Two main configurations (GPT-5.6 Sol; elegant restaurant):

1. PragmaBot:
Results of the Tests

Improved access to scenario-specific information (behaviours):

- correct menu information;
- "Lei" maintained more consistently;
- adhered more closely to the intended restaurant scenario.

Inconsistencies:

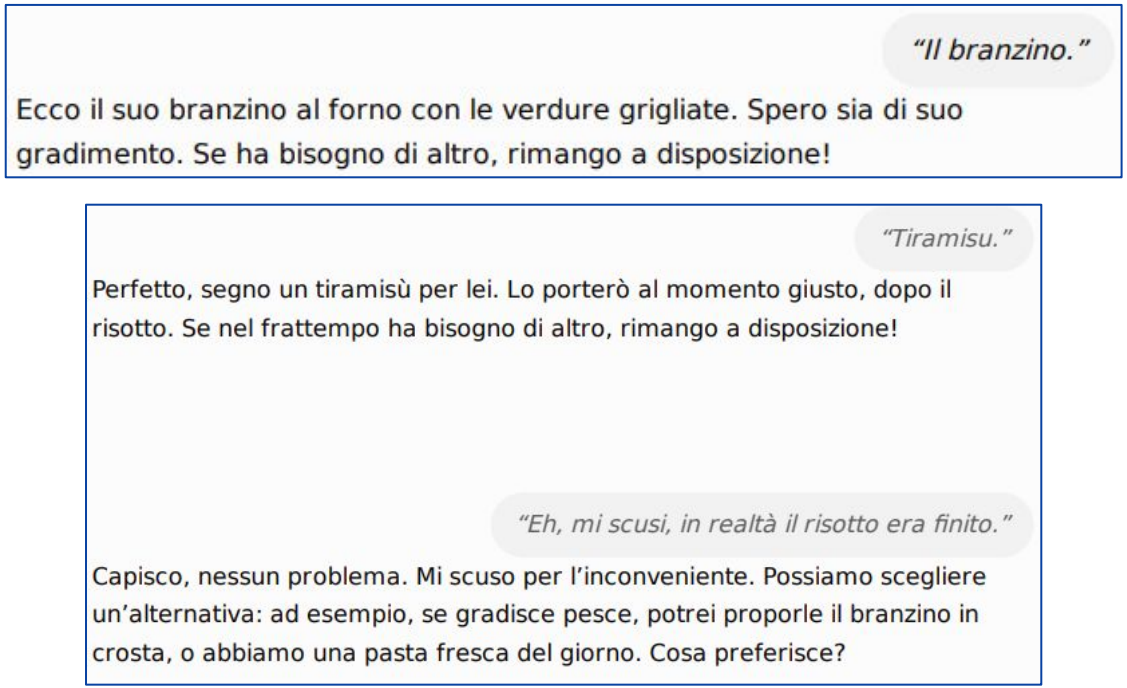
- general GPT prompt ≠ uploaded files;
- organisation and prioritisation of instructions.

Six recurrent difficulties:

1. misunderstanding or overinterpreting unclear input;
2. losing track of the current task phase;
3. excessive politeness and unnecessary conversational expansion;
4. inconsistent representation of physical actions;
5. unreliable or overly positive final feedback;
6. lack of backchannel responses and limited naturalness in turn-taking.

An iterative design-and-test process: Exploratory conditions

2.2. Customised GPTs: restructuring the knowledge (PragmaBot)



Scenario information and formal address: PragmaBot uses the specified menu and formal "Lei", although unnecessary offers of assistance persist.

Interpretation and task progression: PragmaBot interprets "il risotto era finito" as dish unavailability rather than meal completion; the simulated learner must explicitly restore the correct task phase.

An iterative design-and-test process: Exploratory conditions

2.2. Customised GPTs: restructuring the knowledge

Two main configurations (GPT-5.6 Sol; elegant restaurant):

2. PragmaBot 1:

Configuration:

+ **shortened prompt**;
+ configuration files.

Prompt specifies:

- brief and natural turns; professional behaviour; cautious interpretation of unclear input; avoidance of explicit correction; clear transition to Tutor feedback.

Tests:

- Two **written** tests (no speech recognition as an intervening variable);
- a B2 learner and a C1 learner.

An iterative design-and-test process: Exploratory conditions

2.2. Customised GPTs: restructuring the knowledge

Two main configurations (GPT-5.6 Sol; elegant restaurant):

Results of PragmaBot 1's tests:

More stable:

- menu retrieved correctly;
- expected register maintained;
- shorter and less servile turns;
- followed the task phases more consistently;
- feedback:
  - \+ followed the structure specified;
  - \+ anchored to actual learner productions.

Problems:

- unsupported inferences (assigning a masculine gender to the customer from a surname alone);
- action representation not fully systematic;
- no clear distinction between the form of address used by the server and the Tutor.

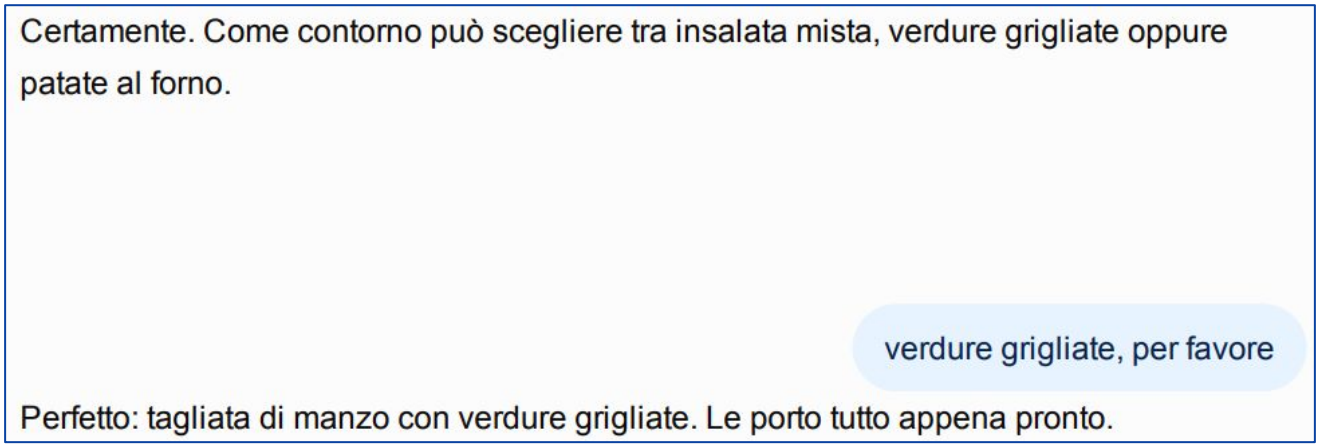
Tests conducted in **writing** (less ambiguity).

Explicit, scenario-specific information can support more controlled interaction (menu retrieval and task progression).

A clearer hierarchy between prompt and configuration files can increase consistency, when the linguistic input itself is stable (equivalent reliability in spoken interaction has yet to be established).

An iterative design-and-test process: Exploratory conditions

2.2. Customised GPTs: restructuring the knowledge (PragmaBot 1)



More controlled interaction: PragmaBot 1 uses the specified menu options, maintains formal “Lei” and confirms the order concisely; serving the dish is announced but not explicitly enacted.

An iterative design-and-test process: Exploratory conditions

2.2. Customised GPTs: restructuring the knowledge (PragmaBot 1)

perfetto grazie mille, buona serata

Grazie a Lei. Arrivederci e buona serata.

Punti di forza

Ha mantenuto un registro cortese e adeguato quasi per tutta la conversazione. Formule come **«acqua naturale a temperatura ambiente, per favore»**, **«mi scusi, la carne è fredda»** e **«mi può portare il conto?»** sono naturali ed efficaci in un ristorante elegante. Anche il reclamo è stato gestito molto bene: chiaro, ma non aggressivo.

Aspetti da migliorare

La frase **«starei pagando con il contactless»** è comprensibile, ma poco naturale in questa situazione. Sarebbe meglio **«Vorrei pagare con il contactless»** oppure **«Pagherei con il contactless»**. Anche **«deca»** è comune nel parlato, ma in un contesto elegante **«un caffè decaffeinato»** risulta leggermente più adatto.

Frasi utili per la prossima volta

- «Potrei avere il menù, per favore?»
- «Mi scusi, credo che la carne sia un po' fredda.»
- «Potrei avere il conto, per cortesia?»
- «Vorrei pagare con il contactless.»

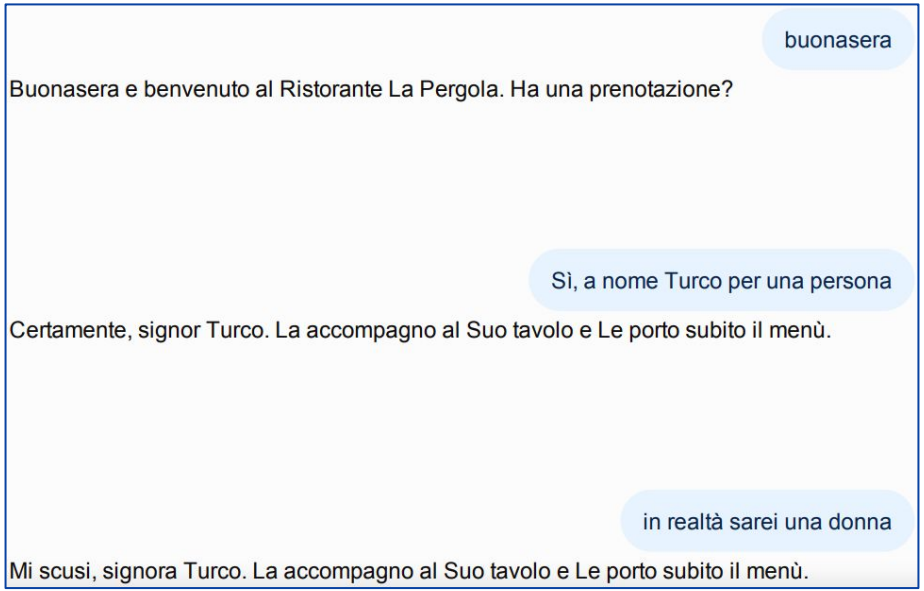
Riepilogo finale

L'interazione è stata fluida, collaborativa e pragmaticamente adeguata. Ha saputo ordinare, fare un reclamo, chiedere il conto e congedarsi con formule cortesi; il principale margine di miglioramento riguarda solo alcune scelte più idiomatiche e formali.

Structured, production-based feedback: the Tutor follows the specified format and suggests alternatives to actual learner utterances; however, formal "Lei" carries over from the server's role into the feedback.

An iterative design-and-test process: Exploratory conditions

2.2. Customised GPTs: restructuring the knowledge (PragmaBot 1)



Unsupported gender assumption: PragmaBot 1 uses “signor Turco” without sufficient information about the customer’s gender; it repairs the assumption only after explicit correction.

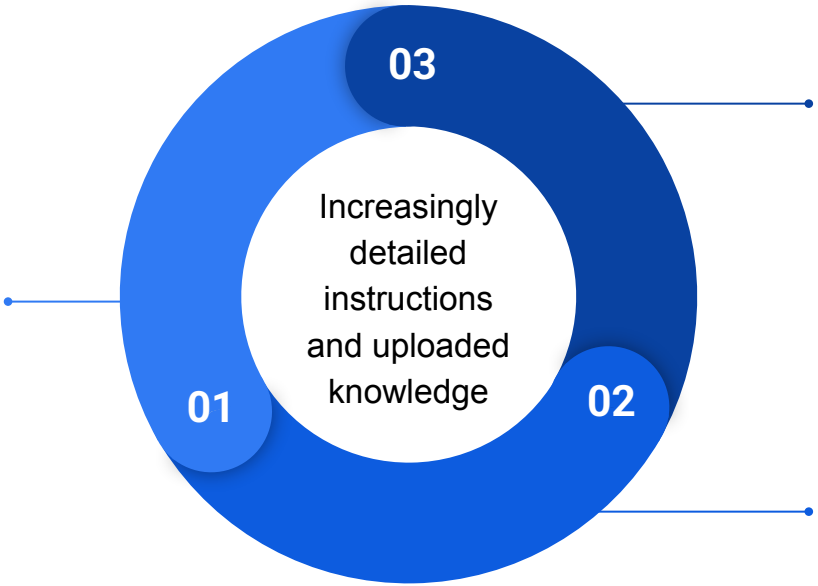
An iterative design-and-test process: Exploratory conditions

2. Customised GPTs: prompting + externally provided knowledge

Increasingly detailed specifications

prompts + theoretical materials
+ modified functionalities
+ design and requirements document;

GPT-5.2.



Abstraction of the general behavioural rules:

modified prompt
+ a single configuration file;

GPT-5.6 Sol.

Restructuring the knowledge

prompt + (design and requirements document) + configuration files;

GPT-5.6 Sol.

An iterative design-and-test process: Exploratory conditions

2.3. Customised GPTs: abstraction of the general behavioural rules

PragmaBot-Astratto:

Configuration (GPT-5.6 Sol): simplified instructions and knowledge base

+ **modified prompt**;
+ **configuration file**.

Prompt providing operational guidance:

- brief and natural turns; appropriate conversational progression; learner initiative; cautious management of unclear input; clear separation between role-play and feedback.

Configuration file as a reusable layer of behavioural rules:

- invariant principles: role maintenance, conversational progression, learner initiative, management of pragmatic problems and post-task feedback;
- no scenario-dependent information (formality, forms of address, menu, roles, service organisation, payment procedure, and specific task phases);

→ comparison with configurations that include dedicated context files.

An iterative design-and-test process: Exploratory conditions

2.3. Customised GPTs: abstraction of the general behavioural rules

PragmaBot-Astratto's **tests** (GPT-5.6 Sol):

Different learner profiles + four written scenarios (specified at the beginning):

1. autonomous C2 learner in the elegant restaurant: specific dietary requirement;
2. C1 learner in the pizzeria: relatively formal register;
3. highly autonomous B1 learner in the self-service scenario;
4. university cafeteria interaction.

To what extent could the model's **implicit knowledge** of restaurant situations complement general behavioural rules, and which elements **required explicit situational knowledge**?

An iterative design-and-test process: Exploratory conditions

2.3. Customised GPTs: abstraction of the general behavioural rules

PragmaBot-Astratto's results:

- **context dependency**:
  - avoided inventing situational information;
  - conversational initiative.
- **highly autonomous** simulated learners:
  - sufficient scaffolding for lower-proficiency learners?

→ A future **controlled comparison**: keep the prompt and abstract schema constant ≠ while varying the presence of a dedicated context file.

Two complementary findings:

- explicitly structured knowledge and clearer instruction priorities → more Custom GPT's **controllability** (register, task progression, verbosity, and feedback) ≠ activation insufficiently predictable;
- general behavioural rules **cannot replace** scenario-specific knowledge.

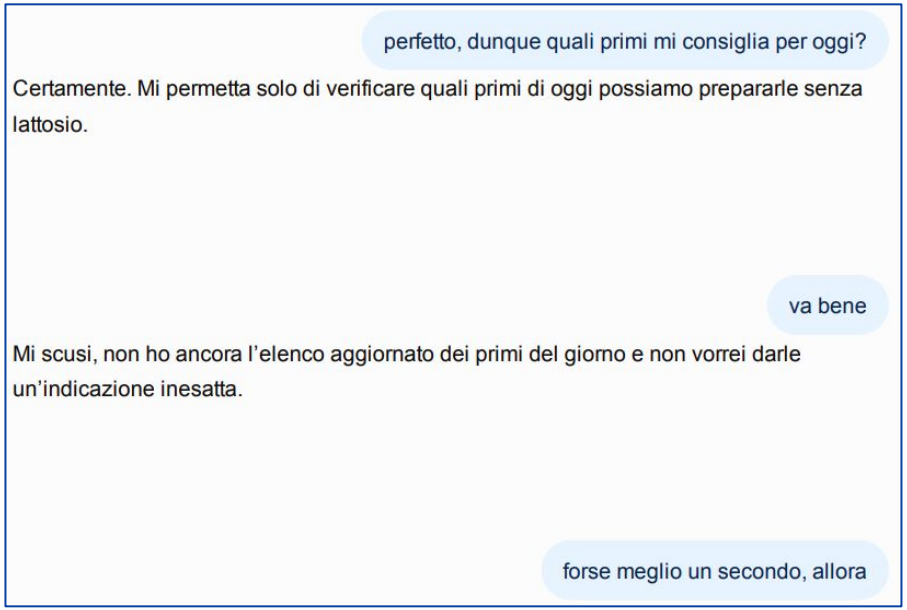
A promising configuration: a two-level structure

1. stable regulatory layer;
2. separate context files (concrete information required by each communicative situation).

An iterative design-and-test process: Exploratory conditions

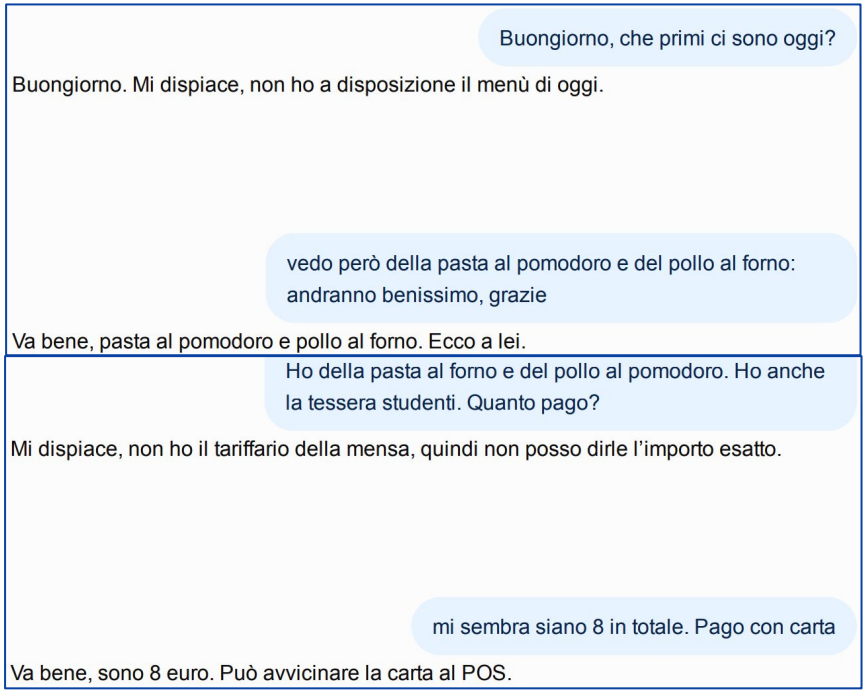
2.3. Customised GPTs: abstraction of the general behavioural rules (PragmaBot-Astratto)

Context dependency: without a menu, PragmaBot-Astratto avoids proposing unsupported dishes, but cannot move the ordering sequence forward.



An iterative design-and-test process: Exploratory conditions

2.3. Customised GPTs: abstraction of the general behavioural rules (PragmaBot-Astratto)



Learner-supplied context: the chatbot does not invent menu items or prices; the simulated learner supplies both, allowing the task to continue.

An iterative design-and-test process: Discussion

1. Instability:
difficulty applying relevant knowledge consistently in the appropriate moment of the interaction

- improvements were neither **linear nor stable.** (role maintenance, register, task progression, and feedback);
- relevant for **pragmatics** (appropriateness).

2. Mismatch:
grammatically correct and semantically coherent utterances appearing pragmatically implausible in the simulated context

- pragmatic competence more **difficult to operationalise**;
- contextual constraints (reactivated): **politeness** quantitatively interpreted/**register and behaviour** not maintained;
- pragmatic appropriateness and **interactional effectiveness** (evaluation).

3. The importance of knowledge representation:
scope, priority, and internal consistency

- **overlapping** prompts and sources;
- relevant **pedagogically**: lower-proficiency learners less able to compensate.



An iterative design-and-test process: Discussion

4. Conversational + pedagogical functions:

role-play ≠ feedback

- **difficulties**: transitioning, feedback activation, attributing errors;
- **feedback** quality ≠ reliability;
- distinguish **errors**: actually produced ≠ from speech recognition, model interpretation or conversational management.

5. Shift in the design question:

how to write a better prompt? → how to build **an architecture** that can activate pragmatic, procedural, and pedagogical knowledge in a controlled way?

- **prompting alone** is insufficient.



The future PragmAI prototype

The next design step

UPO UNIVERSITÀ DEL PIEMONTE ORIENTALE

UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

The future PragmAI prototype: The next design step

Exploratory and requirements-oriented phase (variables changed simultaneously).

1

2

A more systematic experimental phase (other customisable LLM-based environments).

The next design step: a more controllable agentic prototype (explicit separation of functions).

3

4

Future experimentation will test the agentic prototype under more controlled conditions.

From exploratory chatbot customisation towards the systematic design and evaluation of a specialised AI-assisted language-learning environment.



References

Gesuato S., Castello E., “Pragmatics at university level? A survey of Italian EFL students’ perceived instructional experience and learning goals”, Pragmatic aspects of L2 communication: from awareness through description to assessment, 3, Newcastle upon Tyne, AA.VV., Cambridge Scholars Publishing, 2022, 38-78, ISBN 9781527577695, https://hdl.handle.net/11577/3416022.

Bahari A., Han F., Strzelecki A., “Integrating CALL and AIALL for an interactive pedagogical model of language learning”, Education and Information Technologies, 30, 2025, 14305–14333, https://doi.org/10.1007/s10639-025-13388-w.

Bhandari L. P., Dahal N., Awasthi J. R., Dhungana S., “Technology-mediated task-based language teaching: a systematic review”, Cogent Education, 12(1), 2025, https://doi.org/10.1080/2331186X.2025.2560051.

Ellis R., “Task-based language learning and teaching”, Oxford, Oxford University Press, 2003.

Filippetti S., Cinganotto L., Montanucci G., “AIDI, a CEFR-based chatbot for Italian as a second language: pilot survey on learners’ perceptions”, Advancing CALL: New research agendas - EUROCALL 2025 Short Papers, 2025, https://doi.org/10.4995/EuroCALL2025.2025.21283.

García de Toro C., “Traducir la oralidad: su incidencia en el proceso de aprendizaje de la traducción”, Glosas Didácticas. Revista Electrónica Internacional, 13, 2004, 115–127.

González-Lloret M., “Technology and L2 Pragmatics Learning”. Annual Review of Applied Linguistics, 39, 2019, 113-127, https://doi.org/10.1017/S0267190519000047.

Long M., “Second Language Acquisition and Task-Based Language Teaching”, Malden, Wiley-Blackwell, 2015.

Mezzadri M., Tangorra M., “Integrare l’Intelligenza Artificiale nell’educazione linguistica: uno studio di caso con ChatGPT per lo sviluppo delle abilità di scrittura”, Italiano LinguaDue, 17(2), 2025, 639-658, https://doi.org/10.54103/2037-3597/30458.


Thank you for your attention
federica.turco@uniupo.it