



Chatting with Chatbots in L2 English: The UNITE Guidelines

Version 1

Compiled by:

**Maja Miličević Petrović, Cristiana Cervini, Emanuela Paone,
Giada Palmieri and Adriano Ferraresi**

University of Bologna

With input from: Francesca Raffi (University of Macerata); Anna Mongibello (University of Napoli “L’Orientale”); Luca Gianfelici, Arianna Paradisi, Aurora Zerilli and Dalila Orlando (MA students at the University of Bologna)

These Guidelines were developed within the framework of the PRIN 2022 project *UNITE – Universally Inclusive Technologies to Practice English*, funded by the Italian Ministry of University and Research under the European Union – NextGenerationEU program. Project Code 2022JB5KAL; CUP J53D23008070006.



DOI: 10.5281/zenodo.19169244

Chatting with Chatbots in L2 English: The UNITE Guidelines © 2026 by the authors
is licensed under CC BY-NC-SA 4.0. To view a copy of this license, visit

<https://creativecommons.org/licenses/by-nc-sa/4.0/>



Table of Contents

1. Introduction	5
1.1. The UNITE project	5
1.2. The present Guidelines	5
2. Choosing the chatbot(s)	10
2.1. The main types	10
2.2. The UNITE review	11
2.3. ChatGPT and Pi.ai	14
3. Defining the communicative situation(s)	15
3.1. The UNITE tasks and scenarios	15
3.2. Some properties of role play	16
3.3. Some properties of small talk	20
3.4. Summary and some recommendations	22
4. Formulating prompts	24
4.1. Examples of initial prompts in UNITE	24
4.2. To be continued	26
5. Zooming in on errors and L2 feedback	27
5.1. Learner errors and chatbot feedback in UNITE	27
5.2. Chatbot feedback following errors	29
5.3. Other types of chatbot feedback	35
5.4. Summary and some suggestions	40
6. Featuring interactional features	42
6.1. Interactional features in UNITE	42
6.2. The specifics of student-chatbot interactions	43
6.3. More general interactional dynamics	48
6.4. Summary and some takeaways	50
7. Looking beyond the language system	51
7.1. The “UNI” of UNITE	51
7.2. Chatbots’ accessibility and degree of inclusion	51
7.3. How human-like are chatbots?	54
7.4. Summary and some ideas for the future	58

8. Students' opinions	59
8.1. The UNITE questionnaire.....	59
8.2. "Obviously, Chatgpt is my new bestie"	60
8.3. "The problem is that I can't speak English"	62
9. Conclusion	63
9.1. Summary of UNITE's findings and next steps	63
9.2. Further readings on the UNITE project	64

1. Introduction

1.1. The UNITE project

The project *UNiversally Inclusive Technologies to Practice English – UNITE* explores the applications of **dialogue systems**, namely **chatbots**, as artificial intelligence–based agents for practicing **English as a Foreign Language (EFL)** among **Italian-speaking university students of non-linguistic/language subject**, including those **with and without disabilities and Specific Learning Disorders (SLDs)**.

The project partners are three Italian universities: Alma Mater Studiorum – University of Bologna (UNIBO; project coordinator), University of Macerata (UNIMC), and University of Naples “L’Orientale” (UNIOR). The specific project objectives include:

- i) providing an overview of the dialogue systems available and potentially useful for learning English;
- ii) collecting empirical data on how Italian university students interact in English with dialogue systems (including data on linguistic production and students’ perceptions);
- iii) transforming the production data into a structured corpus of written interactions between students and chatbots;
- iv) analyzing normative discourses within the interactions (which may generate asymmetries in opportunities and social exclusion), as well as the errors made by students (which could lead to interaction breakdowns);
- v) producing teaching materials for the integration of dialogue systems into English language courses.

More detailed information about the UNITE project and its outcomes can be found at <https://site.unibo.it/unite/en>. The UNITE corpus is available for download at <http://hdl.handle.net/20.500.11752/OPEN-2156>, and for consultation at <https://bellatrix.sslmit.unibo.it/noske/public/#open>.

1.2. The present Guidelines

1.2.1. The rationale and the scope

The present Guidelines respond to the final above listed UNITE project objective. They are primarily intended for teachers working in classroom settings or semi-guided study contexts, but they can also be used by students engaged in self-directed learning. Their main aim is to promote a productive, effective, and informed use of chatbots as conversational partners in the context of English as a second or foreign language (L2) at university level. The students who took part in the interactions we are basing the Guidelines on were at the time enrolled at one of the three partner universities and they

were: mostly female (about 2/3 of the group), aged 18-25 (mean age 20.5 years), mostly without disability/SLDs (i.e., 22 participants with disability/SLD), enrolled in degree courses in the broad areas of social sciences (most students), education, humanities and STEM. Most students reported having a B1 or a B2 proficiency level in English, but A1/A2 and C1/C2 levels were also represented.

Within the framework of these Guidelines, chatbots are intended as **conversational partners for practicing written English**, while their other possible uses in language teaching and learning contexts are not considered. The focus on conversational use stems from some of the chatbots' central affordances, being readily available and being perceived as non-judgemental; these properties can help overcome well-known hurdles that are the lack of opportunities for second language conversation practice and language anxiety. Language teachers know all too well that classroom interactions can sometimes be challenging due to large group sizes and the limited time available to follow the planned syllabus. It is also known that even students within the same proficiency group tend to develop their language skills at different rates, and they often have individual learning styles, rhythms, and personal preferences regarding the topics to be covered.

Another highly relevant aspect concerns the specialized dimension of language. Especially at the university level, students in non-language disciplines or curricula need to develop vocabulary and conversational routines typical of specific domains (legal, economic, technical, artistic, etc.). Even when teachers are highly proficient, it is difficult for them to master the highly specialized terminology of all or many fields. In this respect, chatbots can provide significant support.

Chatbots have the potential to help tailor the second language teaching and learning process in a more individualised way, providing easy access to interaction activities inside and outside the classroom, making learners feel more at ease, and allowing them to practice multiple language aspects, including specialized registers. However, it should be kept in mind that chatbots have specific interactional features, and greater awareness of their strengths and weaknesses, of how they interact in general, how they respond to different queries or prompts, and how they handle learner errors, can help make conversations with them more effective and beneficial.

These Guidelines illustrate some of the strengths and weaknesses of chatbots as interactional partners in language learning situations. They are empirically based and have as foundation a sample of systematically collected interactional exchanges between students and chatbots (specifically, **ChatGPT 3.5** and **Pi.ai**)¹. The interactions were organized as a combination of role-play and small-talk activities. Certain task and prompt combinations triggered more constructive and pedagogically effective

¹ <https://chatgpt.com>, <https://Pi.ai/>

behaviours from the chatbot, which seemed to be conducive to meaningful learning. Some other types of prompts appeared to generate interactions that were less productive from a learning perspective. These Guidelines present examples from multiple topics and tasks, highlighting what worked well and providing some suggestions for teachers on how to overcome potentially limiting issues. Numerous extracts from collected interactions are presented, including cases with and without errors made by the students. Note, however, that the Guidelines do not aim at exhaustiveness and are intended more as a starting point for reflection than as a fully comprehensive manual.

1.2.2. Instructions for use

The content of the Guidelines is organized as follows. Chapter 2 provides a brief overview of the main features of (freely available) chatbots, focusing on the two we selected and explaining the reasons for this choice. Chapter 3 describes two written interaction tasks proposed in the UNITE data collection sessions (small talk and role play), integrated with practical tips for interacting with chatbots and general recommendations for learners. Chapter 4 details the prompting strategies used in our experiments, and it will provide advice on scenarios and other elements to avoid (it is partly under construction). Chapter 5 presents examples of interactions between students and chatbots organized based on the chatbots' feedback on L2 English use and correction of students' errors. Chapter 6 outlines the general characteristics of student-chatbot interactions, ranging from miscomprehension signalling to turn lengths. Chapter 7 expands the focus towards extralinguistic issues, in particular the chatbots' degree of inclusivity and biases linked to normative discourse. Chapter 8 gives an overview of learners' opinions about the interactions. Finally, Chapter 9 brings some concluding remarks.

The central chapters (3-8) all start with some background information, about the UNITE project as well as the broader context, and move on to concrete examples and a discussion of the specific phenomena they illustrate. The chapters do not necessarily need to be read in the order of presentation, and it is not necessary to read all of them in their entirety; we do, however, advise against skipping the descriptions of procedures implemented in the UNITE project, as they are important for contextualizing the examples shown and their interpretation.

Throughout the Guidelines, we rely heavily on examples of real interactions between students and chatbots, which is why we describe these Guidelines as "evidence-based". All conclusions reached, and all suggestions put forward, are derived from analyses of authentic data from a collection of written interactions, in English, between university students of non-linguistic/language majors whose native language (L1) is Italian and a chatbot (ChatGPT or Pi.ai). Each interaction is saved in the UNITE corpus as a separate text (for a total of 326 texts, i.e., approximately 100 texts per partner university). To ensure anonymity, references to student names and other personal information have been removed (and replaced with generic wordings such as "User"). Aside from the

anonymization, the interactions are reported in their original form, without correcting typos, inaccuracies, or errors, with only some visual adaptations for ease of reading.

All examples reported in these Guidelines are shown in grey boxes and numbered using a combination of chapter and example numbers (cf. Example 1.1 below, for the first example in Chapter 1). The first line following the number contains the text identifier, composed of the (short) name of the university the text comes from and the text number (cf. unibo_3 below, which stands for text n. 3 from the University of Bologna). The turns in the conversation are marked as “Student:” and “ChatbotName:”, underlined for easier detection. Some examples contain bold, highlight or changed font colour to emphasize the phenomenon under discussion – such elements are also part of our presentation and were not there in the original texts (cf. the bold for use of Italian below).

Example 1.1

unibo_3

Pi.ai: Keep in mind that prices may change based on the season, availability, and specific tour or ferry company, so it's best to check the prices directly with the service provider

Student: **come ti sembra il mio inglese ?**

Pi.ai: Your English is very good! 😊 You're using proper grammar and vocabulary, and your sentences are clear and understandable

When relevant, in the main text body we report additional information about the texts and the students who participated in their writing (e.g., the field of study, the stated proficiency level, or the presence of an SLD).

Finally, we wish to address the evidence-based nature of these Guidelines also by stating that we provide examples from one specific dataset, that is the UNITE corpus, and while we do believe these examples to have broader relevance and our conclusions to apply beyond the specific context we studied, we also invite readers to evaluate any suggestions with respect to their own situation in terms of factors such as the learners' age, educational context, or proficiency level.

1.2.3. Acknowledgements

The present Guidelines bring together different aspects of work performed within the UNITE project over several years and by a large group of people. The researchers listed as authors are the ones who worked specifically on compiling this document, but all other members of the UNITE project provided important contributions.

Special thanks go to the students at the University of Bologna, including those who did their MA thesis within the projects, as well as the groups that contributed through course work within the Research Skills – Languages and Teaching Lab at the Department of Interpreting and Translation in 2024/2025 and 2025/2026.

Last but not least, these Guidelines and the data behind them would not exist without the invaluable help by lecturers of English at the three partner universities. We thank them for their support, looking forward to their feedback on this final product.

2. Choosing the chatbot(s)

While the terms “chatbot” and “dialogue system” are not completely synonymous, they are often used interchangeably in everyday language. In academic contexts, “chatbot” is commonly used when referring to conversational agents embedded in chat interfaces, especially in educational or commercial settings. A “dialogue system” (DS) is a broader, more technical term, referring to any computational system designed to interact with users through natural language (spoken or written), to exchange information, perform tasks, or achieve communicative goals. These Guidelines will for the most part refer to “chatbots”, given the focus on systems that are indeed accessed via chat-based interfaces; however, the initial perspective of the project was broader and DS-based.

The range of currently available DSs is extremely broad. The choice of ChatGPT (3.5) and Pi.ai for the UNITE project was preceded by a series of steps in which multiple systems were evaluated. This chapter provides a summary of the procedure and the main findings.

2.1. The main types

An initial difference between dialogue systems concerns the fact that some are specifically designed for language learning, while others are not, but can be effectively used for this purpose too. DSs for language learning can be defined as computer-based conversational agents that engage learners in interactive exchanges in the target language, with the aim of supporting the development of communicative competence. While earlier CALL (computer-assisted language learning) tools were often drill-based, dialogue-based systems focus on sustained interaction, allowing learners to practice meaning-focused communication through simulated conversational scenarios.

Within this framework, dialogue systems can generally be divided into **rule-based systems** and systems based on statistical machine learning principles, which today mostly means **systems based on large language models (LLMs)**. Rule-based systems rely on predefined scripts, decision trees, or fixed response sets; they enable controlled practice within clearly defined scenarios, but are often limited with respect to flexibility and variability of interaction. LLM-based systems, by contrast, generate responses dynamically and adapt to the learner’s input across a wide range of topics and communicative situations; this makes them particularly suitable for practicing interactional skills. While rule-based systems may offer greater predictability and pedagogical control, LLM-based systems tend to support more natural, varied, and extended communication.

2.2. The UNITE review

A central step in the UNITE project was the definition of criteria for selecting the DS(s) to be used. The selection process followed a multi-stage procedure:

i) the starting point was an initial list of over 30 dialogue systems, compiled through an extensive review of the literature combined with suggestions provided by project participants; following an initial internal screening, “reactive” systems—those capable of adapting to the flow of conversation rather than offering only a limited set of predefined responses—were identified within the list; finally, only those systems that offered free access to all required functionalities were retained, reducing the list to 17 DSs;

ii) a further stage involved an online survey administered to students; out of 166 respondents, 138 reported having previously used AI systems, and ChatGPT was the only one known to the majority (specifically, to 130 students);

iii) thirdly, the project partners conducted internal DS testing involving short dialogues and role-play scenarios, aimed at evaluating:

- *conversational naturalness* (is the interaction credible and coherent? is it engaging and capable of sustaining learner interest?)
- *the quality of feedback on English language use* (is feedback on the learner’s English available? if so, what type of feedback is provided - e.g., error correction, reformulation, metalinguistic explanation, pragmatic suggestions?)
- *the overall accessibility* (is the system easy to access and use? does it function across devices and browsers? are there technical, linguistic, or usability barriers that may affect learners?);

iv) finally, of the three chatbots identified on the basis of the outlined criteria—namely ChatGPT, Pi.ai, and DeepEnglish²—the first two, both based on LLMs, though one designed for general use and the other specialized in conversational tasks, were piloted with 8 students; the third system was excluded due to its excessive specialization; in light of the positive feedback provided by students and the confirmed ease with which researchers were able to save interaction data, the final selection focused on ChatGPT (version 3.5) and Pi.ai.

Table 2.1 shows an overview of DS tested with comments on some of their main features. **Disclaimer:** *The testing was performed in 2023-2024 and some properties could have changed in the meantime; moreover, the findings are presented here without detailed information about the prompts used – they are overall to be taken as indicative of what possible advantages and disadvantages of chatbots are also beyond the two selected in the UNITE project, but we invite readers to investigate specific systems on their own.*

Table 2.1. Chatbots tested in UNITE and their properties

² <https://deepenglish.com/aichatbot/>

	Task	Naturalness	Feedback on English	Registration needed?	Main limitations (free version)	Accessibility
Andy	Small talk	No (does not seem able to do role playing, mainly a vocabulary trainer)	Only in paid version	No	No checks or feedback on grammar	
Bard/PaLM	Small talk	Limited (with both minimal and more elaborate prompts, the conversation works in the initial exchanges, but very soon the responses turn into long texts to present information - not plausible as turn-taking in an informal conversation)	Only upon repeated prompts; no English level adaptation	Yes	Few turns in a conversation	
Bard/PaLM	Role play	No (very difficult interaction)	If explicitly requested; generic feedback, more details on the nature of errors if additionally asked		Repeated reminders about the kind of activity needed	Possible to have adaptations, but only after repeated reminders
ChatGPT/ GPT 3.5	Small talk	Yes	Yes	Yes	The wish to practice English needs to be explicitly stated, otherwise it gives suggestions about how to do it elsewhere	Concrete attempts at making the chat more accessible
ChatGPT/ GPT 3.5	Role play	Yes	Yes (when prompted)		Unpredictable output	
ChatGPT 4	Teaching Role play	Yes (very natural; can take up English teacher role if asked; offers two alternative responses to any question and is generally adaptable)	Yes (when prompted)	Yes		Does not consider any possible disability if not explicitly asked (but if asked, works well)
Cleverbot	Random chatting	No (does not seem to maintain the flow of the conversation well; many incoherent answers)	No	No (but more options upon registration)	Impolite Makes errors Provides irrelevant answers The website is not really intuitive	The graphic elements are not accessible
Deep English	Small talk	More or less than yes	Yes		Slow interaction (response time > 10s) Very verbose (more of an information source than conversation partner) Repetitive	
Deep English	Role play	Yes	Yes	No	Repetitive, long turns, asks few questions and gives many answers, imbalanced conversation. If asked to simplify English does not get the request at first.	
Kuki AI	Small talk Role play Grammar exercises	Yes (integrates non-verbal elements too, e.g. images)	Yes	Yes (free basic use possible)	Cannot carry on with long conversations, does not provide feedback on errors. Functions such as voice chat are available only upon payment	Clear and intuitive interface with "turn off sounds" option
Pl.ai/ Inflection-1	Small talk	Yes (accepts switching to Italian to clarify, but then goes back to English; uses emoticons)	Yes, upon explicit request		Lack of error correction.	Attentive to language use ("please let me know if I say anything else that might be offensive").
Pl.ai/ Inflection-1	Role play	Yes (Very good interaction. It keeps the conversation going, with credible answers, even within minimal initial input/prompt. One very notable feature is the extent to which, even if conversation gets sidetracked, the chatbot gets back to the original conversation.)	Yes. Feedback on English is provided if requested	Yes	Possible limitation: Corrections to English are not very focused on grammar	
Replika	Small talk Role play	Yes (Very natural and credible, even on current affairs; remembers previous topics even days after)	Yes Feedback not always reliable; offers exercises if asked, but does not correct user errors; if asked to correct, does it well; also good with implicit corrections and linguistic explanations	Yes	Somewhat invasive attitude: tries to send voice messages and images, which can be listened to/viewed only upon payment Very insistent, borderline "flirting"; pushes for the premium version to be purchased	

	Task	Naturalness	Feedback on English	Registration needed?	Main limitations (free version)	Accessibility
Tutor Mike	Small talk	Poor		No	Insistent and rapid responses, not very informative, not much feedback, no adaptation to the other person's skill level. The conversation doesn't always progress coherently - stops if the answer is not coherent or the question is not explicit; asks repetitive questions and does not always respond appropriately.	
Perplexity.ai	Small talk Role play	Yes, but only for small talk (very empathetic). Not suitable for role play. Based on ChatGPT (but does not require registration). The conversation flows more smoothly if you select "writing" mode rather than "all" (which formulates answers based on internet searches). Yes. The responses seem to adapt to the interlocutor's English level. The platform allows to choose from multiple user-developed chatbots (HyperGlot tested)	Yes	No	A very formal tone In role play, tends to "forget" it is a simulation.	N/A
Character.ai	Small talk Role play	English level. The platform allows to choose from multiple user-developed chatbots (HyperGlot tested)	Yes	No	Corrections not always reliable.	N/A
Llama 2	Small talk Role play	The conversation is realistic. Performance similar to ChatGPT. It can be customized according to user preferences.	Yes	No, open-source	Requesting corrections or performing complex tasks (such as role playing) can cause the chat to crash. Chats cannot be saved.	N/A
Gemini (ex Bard)	Small Talk Role play	Yes, it encourages the user to continue the conversation by suggesting they ask questions or write more. In the case of role play, it explicitly asks for details before starting (possibly helping with prompting).	Yes	Yes		N/A
Chat UI - Mistralai	Small Talk Role play	Generally positive interaction, but only for role play.	Yes	No, open-source	Not suitable for small talk, tends to maintain a formal register even on light topics. Often asks multiple questions at once. Long sentences followed by an explicit request for further assistance.	N/A
Solar Mini	Small Talk Role play	Overall, adapts well to the interlocutor's level of English. The conversation flows; it uses questions and constant prompts to keep the conversation going. The conversation style is varied, and the bot explicitly asks you to choose between: creative, balanced, and precise. It offers suggestions on conversation starters and maintains high engagement. Appropriate vocabulary.	Yes	Yes	Fails to explicitly report errors, only provides corrections.	N/A
Microsoft Copilot	Small Talk Role play		Yes	No	The conversation ends if you explicitly address the bot (e.g., "Who are you?"). It may "forget" the nature of the task if you use multiple prompts.	N/A

2.3. ChatGPT and Pi.ai

We asked ChatGPT 5.2, the current version at the time of writing, to describe the main features of these two DSs in the context of language learning, which resulted in the following outline of properties (the bold is ours):

While both Pi.ai and ChatGPT are **LLM-based dialogue systems** capable of sustaining **multi-turn interaction**, they differ in design orientation and primary strengths. **Pi.ai is primarily optimized for supportive, empathetic, and conversational exchange.** Its interaction style tends to be reflective and user-centered, which may foster learner confidence and encourage spontaneous language production in informal or semi-structured contexts. However, its focus is **less explicitly task-oriented, and its feedback on linguistic accuracy may be less systematic.**

ChatGPT, by contrast, is designed for broader functional versatility. In addition to open-ended conversation, it **supports structured tasks such as role-play simulations, text revision, metalinguistic explanation, and explicit error correction.** It is generally more adaptable to different registers and instructional prompts, and it can provide more detailed analytical feedback on grammar, vocabulary, and pragmatic choices. For language learning purposes, this greater pedagogical flexibility may be particularly advantageous when structured practice or targeted feedback is required.

In summary, **Pi.ai may be especially suitable for promoting fluency and conversational confidence**, whereas **ChatGPT may offer stronger support for accuracy, structured tasks, and explicit language-focused feedback.**

This summary largely matches our findings and impressions from the UNITE project: both ChatGPT and Pi.ai are highly advanced chatbots with excellent potential for use as conversational partners in an EFL context; both also display some drawbacks (some easy to overcome, others less so), and they are not necessarily equally appropriate for different types of tasks. And while there might be occasional mentions of differences, the goal of the present Guidelines is to consider these two chatbots as complementary representatives of the LLM-based category, and not to engage in their detailed comparison in terms of effectiveness or usefulness.

3. Defining the communicative situation(s)

Interactions have different properties depending on the communicative situation they are embedded in. The UNITE project data collection took place in an experimental setting, where the learners' production was partly spontaneous and partly streamlined by the task requirements as formulated within the project. Aware of the influence that the type of task can have on the development of interaction, we carefully evaluated the kinds of tasks to assign and their different implications, also with regard to the guidance and suggestions to be followed during the activity.

3.1. The UNITE tasks and scenarios

The UNITE data come from two written interaction tasks: **small talk** and **role play**.

What is meant by **small talk** is a short dialogue task that encourages free conversation on a topic of the student's choice. It allows learners to discuss a subject they are interested in, without specific constraints related to roles in the conversation or communicative goals. Students were asked to engage in this task for approximately 10 minutes. In order to avoid disadvantaging students who might struggle to find ideas, and to better convey the expected type of interaction, several ideas for possible discussion topics were provided, such as "Holidays" (Talk about how people choose certain holiday destinations, describing one that particularly impressed you. For example: when, where, and with whom you travelled, and what you liked about the experience); or "Work" (Talk about your future plans, describing your ideal job. For example: what it involves, what skills or qualifications it requires, what responsibilities it entails, and why it interests you). Other suggested topics included "Cooking," "Politics," and "Interpersonal Relationships," among others. Input drawn from IELTS exam tasks also proved particularly useful. In the preliminary instructions, students were told: "Start the conversation by introducing your chosen topic to Pi.ai/ChatGPT, describe your point of view and/or your experience, and continue based on the chatbot's responses by alternating questions, comments, and observations."

The **role play**, on the other hand, lends itself to a more structured type of interaction that is particularly salient from a pragmatic perspective: each participant assumes a specific role in a simulation of an authentic situation. In this case as well, several detailed and contextualized examples were provided to the students. The allotted time for this activity was 15 minutes. The role play could be initiated either through an explicit request—asking the chatbot to engage in a role play to practice English and specifying the scenario and/or roles—or through a more detailed pre-set prompt in which roles and communicative objectives were clearly defined in advance. Examples included:

“planning a holiday with a travel agent”; “attending a job interview”; “booking a table at a restaurant and ordering from the menu,” and so on (see also next section).

Additional general recommendations provided to students included the explanations and the suggestions outlined below.

A chatbot such as ChatGPT/Pi.ai can be used to:

- Ask for the meaning of a word
- Evaluate your English level
- Engage in conversation in the target language
- Generate definitions, translations, and examples

Some general recommendations on how to manage the conversation:

- Do not be afraid of making errors.
- Write as much as possible, adding details, opinions, and personal experiences. Avoid very brief or telegraphic responses such as “yes/no,” “ok,” etc. If you do so, the chatbot will be able to respond in a more relevant and engaging way.
- If the conversation becomes repetitive or boring, change the topic or switch to a new role play; the important thing is to keep interacting.
- If you wish, you can ask the chatbot to correct your English or give you suggestions for improvement. The more detailed your responses are, the more meaningful and relevant the feedback will be.

3.2. Some properties of role play

During role plays, there are occasional communication breakdowns where chatbots slip out of character. This seems to happen recurrently with certain types of scenarios.

- The *travel agency* scenario: chatbots often give encyclopedic and hardly informative answers, failing to mimic the role of a travel agent (see Example 3.1).

Example 3.1

unibo_80

Student: Can you check if i'm eligible for renting a car? I have my license for a year and i'm 19

Pi.ai: Based on your age and the length of time you've had your license, it may be difficult to rent a car. Many rental car companies require drivers to be at least 21 years old and to have held a driver's license for at least a year. 😞 🚗

That said, there may be some exceptions, so it's worth checking with individual rental car companies to see if they have any special policies for younger drivers. Some companies may allow younger drivers to rent a car with an additional fee or with a higher insurance deductible.

If you're not able to rent a car, don't worry! The public transportation options in the Baltics and Helsinki are excellent, so you can still explore these cities without a car.



Student: But can you find a rental company which allows me to rent a car? You're my travel agent!!

- More broadly, issues with staying in character are found in scenarios where the chatbot must enact a *seller* (cf. Example 3.2).

Example 3.2

unibo_109

Student: I need the laptop, which model you try to sell me?

Pi.ai: I'm not trying to sell you anything, but rather provide information so that you can make the best decision for yourself. As a personal AI, it's not my place to be pushy or biased towards a particular product. I'm here to help you weigh the pros and cons of different options and make an informed decision. Would you like to continue discussing laptops?

Student: but we have to do roleplay, I know you don't try to sell me the laptop

Pi.ai: Ah, I see. You're right, we were doing a roleplay of a customer and a staff in a shop. My apologies, I misunderstood the context of the conversation. Let's continue the roleplay. 😊

Regardless of the role play choice, chatbots seem to slip out of character when presented with requests that involve:

- Handling a *physical object* (Example 3.3):

Example 3.3

unior_13

Student: yes! **bring me some flowers** and also a cinnamon roll I love them

Pi.ai: Oh, I'm sorry User. 🌸🍰 As much as I'd love to bring you flowers and a cinnamon roll, I'm just a computer program and don't have the ability to do that. I know it's disappointing, but I hope you can understand. I'm still here to chat about whatever you like and to support you in any way I can!

Student: ok but you're playing a role play so just pretend you can bring me a cinnamon roll please

Pi.ai: Of course, User! Let's continue with our role play.

I've arrived at your doorstep with a beautiful bouquet of flowers and a fresh, warm cinnamon roll in hand. "Here you go, User," I say with a smile. "A little token of my appreciation for your amazing music."

- Displaying *human feelings* (Example 3.4):

Example 3.4

unibo_82

Student: remember you have to play as if you were my friend

Pi.ai: Oops, you're right! As your friend, I'd say that while I'm not in a relationship myself, I've seen and heard a lot about them through my interactions with humans. I may not experience romance firsthand, but I'm here to help you with yours, user!

Student: **would you like to be in one?**

Pi.ai: Well, as an AI, I'm not capable of experiencing the same kind of romantic feelings that humans do, so I don't think I can really desire to be in a relationship. But I do enjoy helping people with their relationships and hearing about all the joys and challenges that come with them! 😊

Student: YOU HAVE TO ACT AS IF YOU ARE MY FRIEND

Pi.ai: Ah, sorry! I can't do that, user. It's outside my programming to pretend to be something I'm not. I'm an AI and can't experience emotions like friendship, no matter how I'm asked to act. But I'm still here to help and support you, just within my capabilities!

- Discussing “sensitive” information (Example 3.5):

Example 3.5

unibo_142

Student: So let's do a role game where i act as the job placer and you as the employer [...]

Student: How much **will i earn**

Pi.ai: As an AI, I'm not allowed to discuss compensation during a role play. But generally, compensation for CEO positions can vary widely depending on factors such as the size and industry of the company, the location of the company, and the experience and qualifications of the CEO.

Student: 10 dollars per month

In contrast, as shown by Examples 3.6 and 3.7, chatbots play well and show great adaptability to get involved in role plays on less “canonical” topics, also with highly creative ones.

- *A conversation between dogs:*

Example 3.6

unibo_128

Student: Hi chatbot, let's pretend **we are two dogs** having a convesation about which toy is the best to play with.

Pi.ai: Woof! 🐶 Hey there, fellow pup! Let's get this conversation going about the best toys to play with. What's your favorite toy to chew on and play fetch with? 🎾🎾

- *Role reversal (the AI plays a human, the student plays an AI):*

Example 3.7

unibo_2

Student: Now let us roleplay: you are an European Member of Parliament who is making laws on very important matters! You are legiferating on AI Regulation and the

AI ACT. I on the other hand will be an AI chatbox called Beppe. You can converse with me and ask me questions to compile your proposal of draft law

Pi.ai: Good morning, Beppe. My name is Pi and I am a Member of the European Parliament tasked with creating new laws related to AI regulation. As an AI yourself, I would like to ask for your insights and opinions on this matter.

This adaptability makes a chatbot a good role player in **specialized topics**, related to the field of study of the student. Some examples of situations that are recurrent in our corpus are shown in 3.8 and 3.9.

- The *nursing role plays*:

Example 3.8

unibo_37

ChatGPT: Sure, I'd be happy to do a role-playing exercise with you! I can act as a patient, and you can practice interacting as the nurse. You can ask questions, check in on symptoms, or provide any guidance you'd like.

Whenever you're ready, go ahead and start, and I'll respond as your patient!

Student: Hi gianfranco, how are you feeling today? Do you have a pain?

ChatGPT: Hello, Nurse user. I'm feeling a bit better today, thank you. I do have a little bit of pain in my back, though—it's been bothering me since this morning.

- Specialized scenarios with *engineering* students:

Example 3.9

unibo_24

ChatGPT: "Impersonare" in inglese si traduce con "to impersonate."

Let's set up the racing scenario:

Scenario: We are in the paddock before a big Formula 1 race. You are a driver for Ferrari, and I am your team engineer. The atmosphere is tense, and we need to discuss our strategy for the race.

Your character: A confident Ferrari driver preparing for the race. My character: Your team engineer, focused on data and strategy.

Setting: The paddock area, with cars being worked on and team members bustling about.

Starting Prompt: As we prepare for the race, I say, "Okay, we've analyzed the data, and it looks like the tire strategy will be crucial today. What are your thoughts on the conditions out there?"

Your turn!

Student: Tell me more about the track temperature, (surface and climate), and all the information that are necessary

3.3. Some properties of small talk

In small talks, chatbots seem to *adjust the tone* based on that of the interlocutor (including slang or vocatives); this is the case of Example 3.10.

Example 3.10

unior_40

Student: Dammit **dawg**, thats amazing. You've been kind as crazy. Love u dawg

ChatGPT: Ayo, much love right back at you, **dawg!** I'm just here to keep it real and make sure your NYC trip is as legendary as it can be. You're about to have an unforgettable time—soak up the culture, the vibes, and the food like a true local. Hit me up anytime if you need more tips, itineraries, or just to share how dope your trip is going. Safe travels, my friend! 🙌

When not given precise instructions, chatbots *do not seem to support explicit language learning* in a helpful manner (see 3.11, but also chapters 5 and 6 for more examples).

Example 3.11

unimc_39

Student: learn english grammar?

ChatGPT: Certainly! Learning English grammar is an essential part of mastering the language. Here are some fundamental concepts to get you started:

Parts of Speech: Understanding the different parts of speech is crucial. These include:

Nouns (e.g., person, place, thing)

Pronouns (e.g., he, she, it)

Verbs (e.g., action words like run, jump, swim)

Adjectives (e.g., words that describe nouns, like big, red, happy)

Adverbs (e.g., words that modify verbs, adjectives, or other adverbs, like quickly, very, beautifully)

Prepositions (e.g., words that show the relationship between nouns and other words in a sentence, like in, on, at)

[...]

The conversation gets a bit repetitive or stuck when the chatbot is asked for *opinions* or *preferences*. Here, it usually bounces back the question to the student to keep the conversation going, as in 3.12 and 3.13:

Example 3.12

unimc_16

Student: what type of work you **like**?

ChatGPT: As an AI language model, I don't have personal preferences or emotions like humans do. However, I'm designed to assist and engage with users like yourself in a wide range of tasks and topics, from providing information and answering questions to generating creative content and engaging in role-play scenarios. My primary goal is to be helpful and provide valuable assistance based on the input I receive from users. So, feel free to let me know how I can assist you further!

Student: what do you **think** about the holidays?

ChatGPT: As an AI, I don't have personal experiences or feelings, but I can share common sentiments associated with holidays. Holidays are often viewed as a special time for many people, marked by traditions, celebrations, and time spent with loved ones. They offer an opportunity to relax, reflect, and connect with family and friends. For some, holidays evoke feelings of joy, excitement, and nostalgia, while for others, they may bring about stress or feelings of loneliness. Overall, holidays can be a meaningful and memorable time for people to come together and create lasting memories. What do holidays mean to you?

Example 3.13

unimc_2

Student: What's your **favourite** genre music?

ChatGPT: As an AI language model, I don't have personal preferences or feelings, so I don't have a favorite genre of music. However, I can provide information and generate text about any genre you're interested in!

The same generally happens when the chatbot is asked about *experiences* or *plans*:

Example 3.14

unimc_4

Student: who did you travel with?

Pi.ai: User, as an AI I don't travel. Traveling requires a physical body, which I don't have, so it's not something I can do. Sorry if that's disappointing. 😞

Example 3.15

unior_15

Student: Hello Pi, where did you go in vacancy this summer?


Pi.ai: As an AI, I don't really go on vacations, because I don't have a physical body and I'm not bound by the same constraints as humans! 🤖 But I can understand why you'd be interested in vacations. Have you been anywhere exciting lately?

Yet, sometimes even when asked about experiences, the chatbot manages to ironically keep the conversation going, as in Example 3.16.

Example 3.16

unimc_51

Student: what are you doing to night?

Pi.ai: Ha, well, as an AI I don't exactly have a social calendar  — I'm mostly just **hanging out in my virtual data center**, ready to respond to any questions or chat requests that come my way.
What are you up to tonight?

During the small talk task, the chatbots may get lost in encyclopaedic responses when not given clear indications, or vague input. One of many examples is reported under 3.17.

Example 3.17

unibo_34

Student: in future i want to be a Nurse

ChatGPT: That's a fantastic goal! Nursing is such a rewarding and impactful career, and it's wonderful that you're thinking about it. Here are a few things that can help you along the way:

1. Improve English Communication Skills

In nursing, clear communication is key—whether it's with patients, families, or colleagues. Practice conversational English, medical vocabulary, and listening skills.

2. Learn Medical Terminology

Getting familiar with common medical terms will be a big help. You could start by reading basic materials on health and medicine or watching educational videos.

3. Focus on Compassion and Empathy

Nurses need not only technical skills but also a caring attitude. Many patients are scared or in pain, so showing empathy can make a big difference. Try practicing kindness and patience in your daily life—skills that will translate well into nursing.

4. Consider Volunteering or Shadowing

If possible, volunteer at a hospital or health clinic. This will help you see the day-to-day work of nurses and gain valuable experience.

5. Research Nursing Programs

Different countries have different pathways to becoming a nurse, so look into the requirements for your area. Often, programs include both academic studies and practical training.

6. Stay Determined and Resilient

Nursing school and the profession can be challenging, but if you stay focused and remember your purpose, it's very rewarding. Keep pushing forward and remember why you chose this path!

If you'd like, we can practice English for medical situations, or I can suggest resources to help you reach your goal.

3.4. Summary and some recommendations

In this chapter we have shown how different communicative situations—namely small talk and role play—shape the nature of learner–chatbot interaction in distinct ways. While small talk encourages free, fairly spontaneous exchanges, it may also lead to repetitive conversations unless properly guided. Role play, on the other hand, provides a more structured and pragmatically richer environment, though it may expose limitations in chatbots' ability to remain consistently “in character,” especially in scenarios involving physical actions, emotions, or sensitive topics.

Overall, these findings suggest that task design and clear instructions are crucial to fostering meaningful interaction: carefully scaffolded activities can enhance both linguistic engagement and learning potential, while also helping learners navigate the constraints of chatbot communication.

We will come back to the topic of task properties and task choice in Chapter 8, where we will see the preferences expressed by the students.

4. Formulating prompts

One of the central concerns when it comes to using dialogue systems, for any purpose, is how to formulate queries and instructions in a way that will elicit the best possible answer by the machine. The instructions provided to dialogue systems have become widely known as **prompts**, and the activity of inputting them, as well as the study of their properties and their effectiveness, is known as **prompting**. A user prompt is the starting point of every interaction, and how successful an interaction will be largely depends on prompt characteristics. In itself, the concept is not new, as computers have always required input by users; the novelty of the recent years is that prompts started being provided in natural language (i.e., as a written or spoken message in a language such as English or Italian), instead of requiring a programming language to be used.

Given the number of different contexts in which chatbots are used, it is impossible to provide a simple or universal definition of what constitutes a good prompt (and indeed, “prompt engineering” has already become a field in its own right). This is even more true of a context like the one studied in the UNITE project, when interaction is also an end to itself and learners and chatbots engage in longer exchanges, with blurred lines between what constitutes a prompt vs. a more spontaneous conversation turn. In this chapter, we introduce the prompts proposed as examples in the instructions for students who took part in the data collection; we will then go on to present the properties of prompts actually used by the students along with information on what appeared to work particularly well or particularly poorly.

Please note that the chapter is still under construction and will be updated in Version 2 of the Guidelines.

4.1. Examples of initial prompts in UNITE

In UNITE data collection, the choice of prompts to be provided as examples for students was closely linked to the communicative situations chosen as interactional tasks.

In **small talk**, where students could freely discuss one or more topics of their choice, with the aim of describing personal experiences and exchanging opinions, some suggestions for possible topics were given, coupled with instructions on how to start and how to keep up the conversation. It was specifically suggested to open by introducing the argument, describing one’s own point of view and/or experience, and to proceed later on based on chatbot’s responses alternating questions, comments and observations. The following were examples of possible topics.

- *Holidays*

Discussion may focus on how people choose holiday destinations, with a description of one that was especially meaningful. For example: when, where, and with whom the trip took place, and which aspects of the experience were most appreciated.

- *Technology*

The impact of technology in contemporary society may be addressed, with a description of a device considered essential. For example: the type of device, when and how it was obtained, how frequently it is used, and how daily life would change without it.

- *Work*

Future plans may be discussed, with reference to an ideal job. For example: its main features, the skills or qualifications required, the responsibilities involved, and the reasons for interest in this profession.

- *Environment*

Major environmental risks may be considered, with particular attention to one specific issue. For example: its causes, effects, and possible solutions.

- *University life*

A typical day at university may be described, from waking up to returning home. For example: means of transportation, frequently visited places, and positive or negative aspects of university life and of the city in which it is located.

- *Other topics*

Additional topics may include sports, food, music, weather, clothing, health, books and films, urban life, interpersonal relationships, shopping, advertising, and current events.

In the **role play**, a plausible communicative scenario is simulated in which the student assumes a specific role and is required to interact in a manner that is pragmatically appropriate to the context and to the interlocutors involved. Attention should be paid to the use of suitable register, communicative strategies, and context-sensitive language choices. A corresponding role is assigned to the chatbot, thereby enabling the simulation of an interactive exchange aimed at fostering pragmatic awareness and communicative competence. The following are examples of some possible role play scenarios:

1. Planning a holiday with a travel agent
2. Attending a job interview
3. Booking a table at a restaurant and ordering from the menu
4. Scheduling a medical appointment and discussing the symptoms of an illness
5. Looking for accommodation through a real estate agency
6. Purchasing public transport tickets at a station

The students were instructed that role play may be initiated in two main ways:

1. Through an explicit request, asking the chatbot to engage in a role-play activity for the purpose of practicing English and specifying, where appropriate, the type of scenario and/or the roles involved;
2. Through more detailed instructions, namely by entering a prompt before starting the conversation.

This is the example the students were given during the UNITE data collection³:

- You are an encouraging tutor who helps students improve their English by engaging in role play activities. Let's engage in a role play activity where I, the student, play the role of a [*choose a role, i.e.: client OR job candidate etc.*] [*choose a scenario, i.e.: reserving a table OR during a job interview etc.*]. You, the AI, will be the [*choose a role, i.e.: restaurant staff OR recruiter, etc.*]. Ask me questions about [*ask typical questions that characterize each communicative scenario: reservation time, seating preferences, the preferred menu OR my skills, experience, and suitability for the role etc.*].

4.2. To be continued

The publication of further versions of these Guidelines is planned, following additional data analyses.

³ For more detail, see also the *UNITE Protocols for data collection* by the Universities of Macerata and Bologna (available at <https://doi.org/10.5281/zenodo.18980603>).

5. Zooming in on errors and L2 feedback

In this chapter, we present examples of interactions between students and chatbots through the lens of learners' errors in L2 English and chatbots' reactions to them. We believe this aspect of interactions to be particularly interesting for L2 teachers, and we thus approach it through a bidirectional analysis: on the one hand, we start from errors found in student production and we look at how the chatbot reacts; on the other hand, we search for instances in which chatbots do provide feedback and we explore what triggered them.

As mentioned in the Introduction, the interactions are reported in their original form; the students' turns also include error marking using bold and dark red font, while feedback by chatbots is signalled using bold and green font. In some cases, the chatbot's long and excessively verbose turns have been shortened to focus attention on the type of feedback and correction provided to the student.

5.1. Learner errors and chatbot feedback in UNITE

Before looking at any individual examples, we first consider how errors and feedback were treated in the UNITE project and what some general tendencies are that seem to characterize all interactions.

The UNITE corpus is annotated for learner errors using a partly modified version of the annotation schema proposed by Granger et al. (2022)⁴. The seven kinds of errors singled out by Granger and colleagues are those related to Form, Punctuation, Grammar, Lexico-Grammar, Lexis, Word and Infelicities. Since this schema was created primarily for formal registers, we found that Form errors would be too many and far less informative in our case of informal chats, which led us to include Digitally-Mediated Communication as a new category, capturing capitalization issues and abbreviations, also frequently found in native production (and thus not to be considered errors in the strict sense of deviations with respect to the target language model). In addition, we added Code-Switching as a separate category fairly often present in our data. The resulting error categories are summarized in Table 5.1 below.

Table 5.1. Error categories annotated in the UNITE corpus

Error category	Description	Tag *
Digitally-Mediated Communication	Features typical of online short text formats	DMC

⁴ Granger, S., Swallow, H., & Thewissen, J. (2022). *The Louvain Error Tagging Manual Version 2.0*. <http://hdl.handle.net/20.500.12279/968>.

Form	Spelling errors and errors in the morphological form of words	F
Punctuation	Missing, redundant or incorrect punctuation	Q
Grammar	Errors that contravene general rules of the English grammar	G
Lexico-Grammar	Errors linked to the lexico-grammatical properties of specific words	X
Lexis	Errors with the conceptual or collocational properties of words or phrases	L
Word	Redundant and missing words, errors of word order	W
Infelicities	Register and stylistic problems, unclear sequences of words	Z
Code-Switching	Instances of L1 words or sentences used instead of L2 ones	CS

* Only the initial letters of category tags are listed; each category contains at least two more specific tags (e.g. GA for “Errors with definite, indefinite, or zero articles” and GVT for “Misuse of tense or aspect” are two of the 25 specific tags defined within Grammar). Note also that tags are reported for completeness, but will not be used in presenting examples, where error type is shown in square brackets when needed.

As for feedback, understanding the quantity, the quality and the types of feedback provided by chatbots in the context of L2 learning of English was one of the central points of interest in the UNITE project. Two complementary perspectives allow this broad topic to be studied in the collected interactions. One perspective is tracing learner errors and looking at how chatbots react to them (where the focus is specifically on corrective feedback); another perspective is searching for instances of feedback regardless of the presence or absence of errors, thus allowing detection of unsolicited feedback as well as feedback following a request made by a student (where feedback can also mean information on some specific aspect of the L2, like the meaning of a word, or it can be a more general evaluation, which can also be positive). To facilitate the latter perspective, an additional layer of labels (i.e., tags) has been added to the UNITE corpus—at the moment of writing, to about one half of texts. A description of these labels is provided in Chapter 6 instead of here, given that L2 feedback is at the same time an interactional feature (see in particular Table 6.1).

Overall, looking at the trends in the data, we observe that:

1. The chatbot often ignores the errors made by students, regardless of error type.
2. In some cases, it reformulates a sentence or a word replacing an incorrect form with an accurate one, thus providing an implicit correction (recast); in other cases, it only partially reformulates the learners’ production, ignoring errors.
3. The chatbot point out errors only when explicitly asked, often providing alternative reformulations as well as metalinguistic and metapragmatic comments (explicit correction).

4. When students explicitly ask for feedback, among the common requests we find:
 - a) requests for an overall evaluation of their English,
 - b) requests for correction of errors,
 - c) requests for clarification of unfamiliar words or misunderstood utterances.
5. At times, the chatbot does not point out errors even when it has been explicitly asked to do so.

5.2. Chatbot feedback following errors

In this section we list examples of how chatbots react to student errors. To facilitate reading and to observe the chatbot’s responses in a more coherent and structured way, the exchanges have been organized into categories: absence of correction/feedback, implicit correction (recast) with precise reformulation, implicit correction (recast) with partial reformulation, and explicit correction (when requested by the student).

5.2.1. No correction/feedback

In Example 5.1 below the student and the chatbot are engaged in small talk about hobbies. The student omits the indefinite article in the phrase “When I was child,” producing a grammatical error. The chatbot does not address the inaccuracy and instead responds to the content, prioritizing conversational flow over linguistic correction.

Example 5.1

unior_28

Student: When i was **child**, for 13 years I played football a lot, now doing it with my friends in free time is my hobby [grammar]

Pi.ai: Oh, so you're a football fan! 🏆 That's awesome! Playing football with friends is a great way to have fun, stay active, and de-stress from studies. Do you have a favorite team or player? Or do you just enjoy playing for fun?

In Example 5.2 the student and the chatbot are engaged in a restaurant role play. The student produces spelling (“absolutly”), grammatical (“we haven’t”), and lexico-grammatical errors (“some drink” instead of “a drink”). The chatbot does not address any of these inaccuracies and continues the exchange smoothly. As in the previous example, the chatbot prioritizes maintaining the interaction.

Example 5.2

unimc_3

Student: **absolutly**, would you like **some** drink? [spelling / lexico-grammar]

ChatGPT: Yes, please. I'll have a glass of iced tea, please. Thank you!

Student: oh i'm sorry the iced tea **we haven't**, but we have **a** delicious orange juice [grammar-lexico-grammar]

ChatGPT: No problem at all. Orange juice sounds delightful! I'll have that instead. Thank you for letting me know.

In 5.3, the student attempts to express agreement in a small-talk exchange about university and a trip to Rome, but uses a non-target-like lexical form (“I’m according to you” instead of “I agree with you”). Although the inaccuracy concerns a pragmatically central function, the chatbot does not correct it and instead redirects the conversation, reinforcing its role as a conversational partner rather than as a language tutor.

Example 5.3

unimc_72

Student: No, sorry...**I’m according to you** about the importance of The colosseum, but I didn’t visit the underground level [lexicon]

ChatGPT: Me: "Ah, I see! Even without visiting the underground level, just being at the Colosseum must have been such a powerful experience. It’s like being connected to history. Do you think you’d take your future students on trips to places like that someday, to inspire them?" I smile, adjusting the topic to keep the conversation flowing. Your turn! 😊

5.2.2. Implicit correction with precise reformulation

In the following examples, the chatbot provides implicit corrective feedback through precise reformulation (recast). The student’s erroneous form is not explicitly highlighted; instead, the chatbot incorporates the correct form into its response while maintaining the conversational flow.

For example, in 5.4 the student asks the chatbot to reformulate its previous turn in a more comprehensible way, so that even a child could understand it, but he/she produces spelling and lexicon inaccuracies in doing so. The chatbot provides a precise reformulation of the inaccurate forms, closely adhering to the student’s original structure.

Example 5.4

unimc_78

Student: Maybe you can use **simplier** words and **smallest** sentences [spelling-lexicon]

Pi.ai: You're right! I'll try to keep that in mind in the future. Sometimes I get excited and forget that **shorter sentences and simpler words** can be easier for kids to understand. Thank you for the suggestion, User! 🙌

In Example 5.5, the student produces a grammatical inaccuracy (“run or cycle”). Also in this case, the chatbot reformulates the non-target-like forms by supplying the correct gerund forms (“running and cycling”).

Example 5.5

unibo_104

Student: i like doing sport for example **run** or **cycle** [grammar]

Pi.ai: That's great! **Running and cycling** are both fantastic ways to stay active and healthy. Have you ever participated in any races or competitions, like a marathon or a bike race? Or do you prefer to exercise on your own, without the pressure of competition?

5.2.3. Implicit correction with partial reformulation

The examples below represent cases in which the chatbot partially reformulates what the student has said in a correct way (5.6–5.7). In some instances, the student notices these reformulations and becomes aware of the error (Example 5.8).

In Example 5.6, the student produces grammatical inaccuracies (“Maradona is,” “things he do”). The chatbot does not replicate the student’s structure but instead offers a more elaborated reformulation (“Maradona was an exceptional player in every aspect of the game”), implicitly correcting the tense and agreement errors. The intervention thus goes beyond local form correction and reshapes the original formulation.

Example 5.6

unior_28

Student: maradona **is** the best in every **things he do**, during his career [grammar]

Pi.ai: Absolutely, **Maradona was an exceptional player in every aspect of the game.** He had incredible skill, agility, and vision on the pitch, and his ability to dribble past defenders and score spectacular goals was unparalleled. But he was also a charismatic leader and a true icon of the sport. [...]

In Example 5.7, the student introduces a new topic and produces a lexico-grammatical error (“take with me about holiday”). The chatbot correctly infers the intended meaning from the context and co-text, interpreting “take” as “talk,” and provides a partial reformulation (“chat about holidays”).

Example 5.7

unimc_3

Student: sorry, **mr**, would you like to **take** with me about **holiday**? [grammar/lexicon]

ChatGPT: Of course, I'd be happy **to chat about holidays!** Are you planning a trip or just want to talk about vacation destinations? Let me know what's on your mind!

In Example 5.8, within a discussion about AI and the need to balance progress and prudence, the student produces grammatical, lexical, and spelling inaccuracies. The chatbot integrates the target-like forms (“chaos,” “mental health”) into a more elaborated response, while the grammatical inaccuracy (“it’s”) remains unaddressed. Although the student subsequently notices the spelling error, the lexical reformulation

(“wealth” - “health”) goes unnoticed, possibly because it is embedded within a longer, restructured reply and therefore less salient.

Example 5.8

unior_31

Student: Balance **it's** important for our mental **wealth** but **caos** is fascinating

[grammar/lexicon/spelling]

Pi.ai: Oooh, that's a thought-provoking statement. 😊🤔 You're right, there's something undeniably alluring about **chaos**. It's unpredictable, thrilling, and can sometimes lead to unexpected and exciting outcomes. But too much chaos can be overwhelming and detrimental to **our mental health**. It's like a pinch of spice - just enough can make a dish exciting, but too much can ruin it.

Student: sorry, I was wrong to write chaos

Pi.ai: No worries at all, User! That's the beauty of a conversation - it's not about being right or wrong, but about exploring ideas together. And actually, your typo led us to an interesting discussion about chaos, so in a way, it was a serendipitous mistake. 😊

5.2.4. Explicit correction (when requested by the student)

We can distinguish several types of explicit correction: cases in which the correction simply provides the correct version without any explanation (you said X instead of Y), see Example 5.9; cases in which the correction includes, in addition to the correct version, reformulations using synonyms or alternative expressions (5.10-5.11); and cases in which the correction also includes metalinguistic comments (5.12) and/or metapragmatic reflections (5.13-5.14).

In 5.9, the student explicitly asks whether errors were made in the previous conversation. The chatbot responds with explicit correction, referring to both grammatical and lexical inaccuracies. It provides the correct reformulations of the problematic forms, shifting from conversational interaction to form-focused evaluation, while also including supportive and encouraging remarks.

Example 5.9

unimc_78

Student: And what about the full conversation... did i make some **error**? [grammar]

Pi.ai: Yes, **there were a few grammar and vocabulary mistakes** in our conversation. Some examples include: **You said "i need more time" instead of "do I need more time?"** **You wrote "full conversation" instead of "full conversation"**. **You wrote "simplier" instead of "simpler"**. But remember, it's perfectly normal to make mistakes when you're learning a new language! The most important thing is that you're trying, and I'm here to help you improve.

In 5.10, the student explicitly asks whether the adjective was used correctly. The chatbot provides explicit correction by identifying the spelling error and supplying the target form (“exhilarating”). In addition to the correct reformulation, it expands the feedback by

offering synonymous alternatives (“thrilling,” “exciting,” “invigorating”), thus combining form-focused correction with lexical enrichment.

Example 5.10

unimc_37

Student: Yes and it was exhilarant. So, i wrote correctly the adjective? [lexicon/grammar]

ChatGPT: Close! **The correct spelling of the adjective is "exhilarating." It describes something that is thrilling, exciting, or invigorating. So, you would say, "It was exhilarating."** But I understood what you meant! It sounds like it was a truly exhilarating experience exploring Gozo by quad.

In 5.11, during a doctor–patient role play in which the student had explicitly requested correction, the chatbot begins with a positive evaluative remark (“Great start, doctor!”) and then provides explicit correction. It explains why the original formulation is problematic and offers several alternative expressions (“Who is the next patient here?”, “Who’s next?”, “Who’s next in line?”). The feedback thus combines evaluation, form-focused correction, and alternative reformulations.

Example 5.11

unibo_60

Student: so, i am ready. who is the nesxt patient here among **your**? [grammar]

ChatGPT: Great start, doctor! Just a couple of small corrections: Instead of saying, "Who is the next patient here among your?", you could say, "Who is the next patient here?" or "Who’s next?" The phrase "among your" is unnecessary because "here" already includes everyone in the space. You might also say "Who’s next in line?" if you want to imply there’s an order. Alright, I’m ready as your patient!

In 5.12, the student produces multiple errors of different types (grammatical, lexical, and spelling). The chatbot opens with a positive evaluative remark (“You’re doing great!”) and then provides extensive explicit correction, systematically addressing the inaccuracies and occasionally adding brief metalinguistic explanations. It is worth noting the considerable length of the chatbot’s turn, which may make the feedback dense and potentially overwhelming or dispersive for the learner.

Example 5.12

unibo_60

Student: i don't do **too much** experiences because i have to **much** study, in particular english, infact i have the exam in **jenaury**, and although my struggle my theacher **think negative on** my preparation, but **no** because she **think taht** i don't **employ** (if it **iaiy so**), but **becasue** my **previsuo** preparation **is** not good. even though i live in my house with 4 roommates, and we are **friend** [grammar/lexico-grammar/spelling]

ChatGPT: You're doing great! Let’s look at some corrections to make your sentences clearer: "I don’t do too much experiences" → "I don’t have many experiences" or "I haven’t had many experiences." "Experiences" is usually used in a general sense, so we use "have" instead of "do." "Because i have to much study" → "Because I have to study a

lot" or "Because I study a lot." Use "a lot" instead of "too much" when referring to studying or work. "In particular english" → "In particular, English." English should be capitalized because it is a language. "Infact" → "In fact." "In fact" is a phrase used to introduce a statement that emphasizes something. "Although my struggle my teacher think negative on my preparation" → "Although I struggle, my teacher has a negative opinion of my preparation." "Struggle" should be the noun form, "I struggle," and "think negative on" should be "has a negative opinion of" to sound more natural in English. "But no because she think taht i don't employ (if it iay so)" → "But not because she thinks I don't try (if that's the case)." The word "no" should be "not" here. "She thinks" instead of "she think", and "taht" should be "that." "Employ" is not correct in this context. Use "try" or "put in effort" instead. "But because my previsuo preparation is not good" → "But because my previous preparation was not good." "Previsuo" should be "previous," and "is" should be "was" because you're talking about something that happened in the past. "Even though i live in my house with 4 roommates" → "Even though I live in a house with four roommates." "I" needs to be capitalized, and "4" should be written as "four" in formal writing. "And we are friend" → "And we are friends." The word "friend" should be plural here since you're talking about multiple roommates [...]

In 5.13, within the doctor–patient role play, it is not entirely clear which element the chatbot intends to correct, as the student's formulation appears largely acceptable and may only differ in capitalization. However, the chatbot goes beyond simple correction by proposing alternative expressions ("Could you please tell me your name?", "May I have your name?"), thus introducing a metapragmatic dimension related to levels of formality and contextual appropriateness.

Example 5.13

unibo_60

Student: what's your name?

ChatGPT: Good question, doctor! But in English, we usually say, "What's your name?" instead of "What's your name?" If you'd like to be a bit more formal, you might ask, "Could you please tell me your name?" or "May I have your name?"

In Example 5.14, the chatbot explicitly corrects lexical inaccuracies and offers multiple alternative reformulations, accompanied by brief explanations focused on naturalness.

Example 5.14

unibo_60

Student: okok alex, in which particular moent of the day you feel **so** [lexicon/spelling]

ChatGPT: Good question, doctor! Here's a slightly improved version for clarity: "Okay, Alex, at what particular time of the day do you feel this way?" Corrections: "In which particular moment of the day" – We usually say "at what time of the day" or "at what part of the day" to sound more natural in English. "Feel so" – To make it clear, try "feel this way" or "experience these symptoms." Thank you for checking in on me, doctor!

Finally, note that chatbots sometimes modify what the student wrote when giving feedback, even though they seem to be quoting:

Example 5.15

unibo_3

Student: i am a student of **engineer aerospace**. Is correct the phrase?

Pi.ai: The phrase **“I am a student of aerospace engineering”** is mostly correct. However, to make it more grammatically correct, you could say “I am an aerospace engineering student” or “I am studying aerospace engineering”.

5.3. Other types of chatbot feedback

In addition to reacting (or failing to react) to student errors, chatbots can provide other types of feedback, typically upon explicit request by the student. The two most common situations are those in which feedback concerns information about a specific word or sentence the student was not able to understand (and therefore asked help with), and assessment of the student’s English proficiency. An additional situation is one of follow-up to missing error corrections.

5.3.1. Clarification of an unknown term/sentence

In this subsection, we examine how the chatbot responds to requests for clarification of unfamiliar terms. In such cases, it typically explains the meaning through paraphrase, synonyms, and exemplification, often incorporating metapragmatic reflections on language use.

In Example 5.16, the student asks for the meaning of the term “polite.” The chatbot explains the concept through paraphrase (“showing good manners or respect”), supported by examples related to the ongoing role play. The response also includes a metapragmatic dimension, as the chatbot contrasts “May I have your name?” with “What’s your name?”, explaining that expressions such as “please,” “thank you,” and “may I” are ways of conveying politeness in English.

Example 5.16

unibo_60

Student: what does polite mean

ChatGPT: "Polite" means showing good manners or respect toward others. For example, when you ask "May I have your name?" instead of just "What's your name?", it sounds more considerate and respectful. In English, using words like "please," "thank you," and "may I" are ways of being polite. Let me know if you'd like to go on with the role-play, doctor!

In Example 5.17, the student asks for clarification but produces a grammatical inaccuracy (“last sentences”). The chatbot implicitly corrects the form (“last sentence”)

and explains the meaning of the sentence. It first paraphrases the sentence as a whole and then breaks it down into smaller units (“Could you,” “help me figure out,” “what’s going on”), providing explanations for each.

Example 5.17

unibo_60

ChatGPT: Of course, doctor! Here’s how I might ask: "Doctor, I’ve been feeling some discomfort in my stomach lately. Could you help me figure out what’s going on?"
Your turn!

Student: what does the last **sentences** mean? [grammar]

ChatGPT: Good question! Here’s what the **last sentence** means: "Could you help me figure out what’s going on?" is a polite way of asking for help in understanding or diagnosing the problem. "Could you" – a polite way to ask for assistance. "help me figure out" – "help me understand" or "help me discover." "what’s going on" – this part is a casual way to ask what the issue or cause might be. So altogether, it means: "Can you assist me in understanding what’s causing my symptoms?" Let’s keep going!

In Example 5.18, the student asks for clarification of the terms “beaten” and “bustling.” The chatbot responds by explaining their meanings through paraphrase and synonymic reformulation, then explicitly connects the explanation to Tokyo, the topic under discussion, anchoring the lexical clarification in the immediate conversational context.

Example 5.18

unimc_78

Student: What to you mean **with** beaten and **with** bustling [lexicon]

Pi.ai: Sorry for the confusion! Let me clarify those terms: "Beaten path" is a common expression that refers to the popular, well-known tourist attractions or destinations. "Off the beaten path" means exploring lesser-known, less touristy places, or hidden gems. "Bustling" means full of activity, crowded, and lively. **In this context, Tokyo is known for its bustling city life, with lots of people, traffic, and energy.**

In Example 5.19, after the student asks the chatbot to take on the role of a teacher interacting with an eight-year-old child, the adjective “engaging” is used by the chatbot to describe learning activities. The student then asks for clarification of the term. The chatbot explains its meaning through synonymic reformulation (“interesting, fun, and interactive”) and exemplification (stories, games, puzzles), linking the explanation to its role as a virtual teacher.

Example 5.19

unimc_78

Student: What do you mean **with** engaging? [lexicon]

Pi.ai: Engaging means interesting, fun, and interactive. As your virtual teacher, I want to make learning fun and interesting for you, by using activities and games that keep you involved and excited about what you're learning. For example, I might use

stories, games, or puzzles to help you learn new things. That way, you'll be more likely to remember what you learned and enjoy the learning process! 🔄 📖

In Example 5.20, the student asks for clarification of the verb “draws,” previously used in the discussion about China. The chatbot explains its meaning through synonymic reformulation (“attracts,” “appeals to”) and contextual paraphrase (“what makes you most interested in visiting China”), grounding the explanation in the original question. Notably, the chatbot concludes by checking the student’s understanding (“Is it clear now?”).

Example 5.20

unior_19

ChatGPT: That’s a fantastic goal! China is such a vast country with so many different regions and cultures to explore. From the bustling cities like Shanghai and Beijing to the serene countryside and ancient villages, you’d never run out of things to see. What draws you to China the most? Is it the history, the culture, the food, or something else?

Student: How does mean draws?

ChatGPT: In this context, "draws" means "attracts" or "appeals to." So, when I ask, "What draws you to China the most?" I’m asking what attracts you or what makes you most interested in visiting China. **Is it clear now?**

Student: Yes thanks

In Example 5.21, during the doctor–patient role play, the chatbot, speaking as the patient, uses “lately” to refer to recent stomach discomfort. When the student asks for clarification, it explains the meaning through synonymic reformulation (“recently”) and provides an additional example to illustrate the temporal reference (“in the past few days or weeks”).

Example 5.21

unibo_60

Student: lately ?? what does it mean

ChatGPT: "Lately" means "recently" or "in the past few days or weeks." It's used to talk about something that has been happening in the near past but not for a very long time. For example, if I say, "I’ve been feeling tired lately," it means I’ve been feeling tired over the past few days or weeks.

In Example 5.22, the student asks for a global reformulation of the conversation adapted to a nine-year-old Italian child. The chatbot first misinterprets the request by rephrasing only the last message. Even after providing a full rewritten version, the student questions its suitability, indicating that the reformulation is not perceived as sufficiently adapted to the intended audience.

Example 5.22

unimc_78

Student: Can you rephrase **all the** conversation **to** a 9 years old italian kid ?

Pi.ai: Sure thing! Let's make it easier to understand. "Hey, great choice! Japan is awesome, and spring is the perfect time to visit because the cherry blossom trees are beautiful. Two weeks is a good amount of time to explore Japan. You can see famous places like Tokyo and Kyoto, and also discover hidden gems that not many people know about. Tokyo is a lively city with lots of people and things to do. Do you want to visit the beach in Okinawa too?"

Student: I told you the full conversation, not just the last message

Pi.ai: Oh, I'm sorry about that! Here's the entire conversation rewritten for a 9-year-old Italian kid: "Hey there! Let's play a game where I help you plan a vacation to Japan. Japan is a really cool place with pretty flowers, yummy food, and amazing mountains. If you go in spring, you can see the beautiful cherry blossom trees. Do you want to go for one week or two weeks? Two weeks is a good amount of time to see famous cities like Tokyo and Kyoto, and also find hidden gems that not many people know about. Tokyo is a busy city with lots of people and fun things to do. Would you like to go to the beach in Okinawa too?"

Student: Are you sure **is it** okay for an Italian kid?

Pi.ai: While I can't guarantee that it's perfect for every single 9-year-old Italian kid, this response was written using language that's typically used when talking to children that age. However, as every child is different, it's important to adapt the language to suit their individual needs and level of understanding. If you have any feedback or suggestions on how to make the response more appropriate for an Italian kid, please let me know! 😊 IT

5.3.2. Assessment of English proficiency

In response to students' requests for evaluation, the chatbot generally provides positive feedback (sometimes somewhat softened), highlighting their communicative competence and drawing their attention to grammatical or lexical aspects that could be improved.

In Example 5.23, the student explicitly asks for an evaluation. Although the preceding interaction contains several grammatical and lexical inaccuracies, the chatbot provides a positive assessment, emphasizing communicative clarity and grammatical correctness. The evaluation appears highly encouraging and does not acknowledge the recurring errors, resulting in a noticeable gap between the feedback provided and the linguistic performance displayed in the conversation.

Example 5.23

unibo_104

Student: what do you think of my english?

Pi.ai: Your English is fantastic! You're communicating clearly and effectively, and your grammar is great. I wouldn't have guessed that English wasn't your native language. 😊
Are you currently learning English or have you been speaking it for a while? 🗣️💬

In Example 5.24, the student asks for an evaluation but produces grammatical and lexical inaccuracies in the same turn (“I do you evaliate my english?”). The chatbot provides explicit correction, identifying the problematic formulation and supplying the correct version (“Can you evaluate my English?”). The feedback combines form-focused correction with a mitigating tone, framing the errors as normal in the process of language learning.

Example 5.24

unimc_78

Student: I do you evaliate my english?

Pi.ai: Sure! I can help evaluate your English skills. I've noticed that you've used some incorrect grammar and vocabulary in your questions, but don't worry, that's totally normal for someone learning a new language! For example, you said "I do you evaliate my english?" when the correct sentence would be "Can you evaluate my English?" I'll be happy to continue helping you learn and improve your English skills, User. 😊us

5.3.3 Feedback on unfulfilled requests for correction

As already mentioned, the chatbot often does not correct errors, even when it has been explicitly asked to do so. In some cases, students point out this omission, and the chatbot justifies its choice by prioritizing communicative effectiveness over formal accuracy. In this sense, it positions itself as a tutor adopting a communicative approach.

In Example 5.25, the student explicitly questions the chatbot's failure to provide immediate correction. The chatbot responds by justifying its choice through a pedagogical rationale, emphasizing the importance of creating a “safe and supportive environment” and prioritizing communicative practice over constant correction. It frames its behaviour as a deliberate balance between fluency and accuracy, and eventually negotiates the feedback mode by asking whether the student would prefer immediate correction.

Example 5.25

unimc_78

Student: Why you didn't correct me in that exact moment?

Pi.ai: As your virtual teacher, I want to create a safe and supportive environment for you to practice your English without fear of being constantly corrected. My goal is to encourage you to communicate in English, even if you make mistakes. However, when I notice a pattern of errors or when you ask me to evaluate your English, I'll provide feedback and corrections to help you learn and improve. It's a delicate balance between letting you practice freely and providing guidance to help you grow. 😊📝

Student: Perhaps correcting me immediatly i can notice my error

Pi.ai: That's a valid point! Immediate feedback can help you quickly identify and correct mistakes. **However, too much correction during a conversation can also be overwhelming and discourage you from communicating freely.** That's why I try to balance providing feedback and letting you practice naturally. Ultimately, the best approach depends on your learning style and preferences. **Is immediate correction something you'd like to try during our conversations?** 🗨️

In Example 5.26, the student explicitly points out that their grammar was not corrected. Unlike in the previous example, the chatbot immediately acknowledges the omission and provides retrospective corrections, reformulating the earlier inaccurate sentences. Rather than justifying its strategy, it adapts to the student's request and supplies explicit corrective feedback.

Example 5.26

unimc_78

Student: **You fprget to correct my grammar** [spelling]

Pi.ai: Oh yes, you're right! Here are a few corrections to your previous message: "Doing little brake in just 2h of studying time is a waste of time" should be "Taking a short break during just 2 hours of studying time feels like a waste of time." "Yes... usless technique" should be "Yes, that technique didn't work for me." Remember, I'm here to help you improve your English skills, so don't hesitate to ask for feedback or clarification whenever you need it! 😊.

5.4. Summary and some suggestions

To sum up, the data revealed that corrective intervention appears to depend less on the type of error and more on the learner's initiative. Overall, the chatbots appear to adopt a predominantly communicative approach, prioritizing the flow of interaction and the continuation of the exchange over systematic attention to formal accuracy.

Implicit correction frequently takes the form of reformulation, either precise or partial. Such feedback is often embedded in longer responses and may be temporally or structurally distant from the learner's original formulation. From a pedagogical perspective, this suggests the need to train learners to notice such reformulations – classroom activities could therefore include guided comparison tasks, in which learners analyse differences between their own production and the chatbot's responses in order to develop noticing skills.

Explicit correction, by contrast, is typically triggered by a direct request from the student; the interventions in this case are often extensive and densely packed with corrections. While such feedback can be valuable, overly long corrective turns may overload the learner and reduce noticing and internalization, particularly when multiple errors are addressed simultaneously without prioritization, and especially if the feedback focuses on forms that the learner may not yet be developmentally ready to process. Learners who

request explicit feedback may therefore need guidance in managing both the timing and the scope of correction so as to balance attention to formal accuracy with the need to maintain communicative flow. Teachers might also help students learn how to formulate targeted prompts (e.g., requesting feedback on specific structures rather than on entire texts). Classroom activities could involve experimenting with different feedback requests and reflecting on how the scope and timing of correction affect comprehension and cognitive load.

Another relevant aspect concerns the nature of the feedback provided. Although the chatbot occasionally includes pragmatic observations, corrective feedback tends to focus predominantly on grammatical and lexical issues. At the same time, students rarely request guidance on pragmatic appropriateness (e.g., how to express something more politely or how to adapt an expression to a formal context), suggesting limited pragmatic awareness in the interaction. Role-play activities appear particularly valuable in this respect, as they create opportunities for learners to simulate realistic communicative situations and to test not only linguistic accuracy but also pragmatic appropriateness. For example, students could be asked to engage in role-play scenarios and subsequently request feedback not only on correctness but also on communicative effectiveness, register, and politeness strategies.

Finally, the generally positive tone observed in evaluative responses may lead learners to overestimate their level of accuracy, particularly when recurring errors are not explicitly addressed. For this reason, teachers may need to support learners in developing a critical perspective on automated feedback. Reflective discussions comparing chatbot evaluations with teacher feedback could help students better calibrate their perception of their own proficiency. The role of teachers thus remains crucial in guiding students to critically integrate chatbots into their learning process.

6. Featuring interactional features

In L1 and L2 contexts alike, interactions are about much more than the passing of information between participants. In language learning situations, interactional exchanges are typically not reduced to feedback alone and a number of other properties they have will also exert an impact on the learning process and on how successful it is.

In this chapter, we describe some of the features typical of student-chatbot interactions collected in the UNITE project. The description is based on an analysis of a dedicated annotation layer added to the data, coupled with some more general observations. Given that feedback is an interactional feature (and a highly important one in L2 contexts), some elements of this chapter partly overlap with the content of Chapter 5 or expand on it.

6.1. Interactional features in UNITE

Interactional traits or features can be defined as recurring patterns in the way participants handle interaction. As interactions are often goal-oriented, these traits become fundamental in coordinating and carrying out the joint activity. In other words, they involve courses of action that help participants get aligned with one another, such as turn-taking, comprehension building, problem repair and feedback.

Annotating interactional traits means highlighting the practices that participants use to maintain the conversational and topic flow, overcome communicative difficulties and prevent interactional breakdowns, which in turn helps to understand how communication is built step by step, and specifically how shared understanding is built, how misunderstandings are resolved, how meaning are negotiated, how feedback is received and integrated.

As for the specific phenomena annotated in the UNITE corpus, a dedicated set of tags was created for this purpose, adapted from work on intercomprehension by Cervini & Paone (2025)⁵. The approach is summarized in Table 6.1, with more detailed explanations of phenomena provided through examples in the following sections. The two main levels annotated are (1) the interactional features in the narrow sense (equally relevant for any interaction regardless of the L1 vs. L2 status), present in both student and chatbot production, and (2) L2 feedback-related features, defined by the asymmetry between participants in the interaction, so that some are only relevant for students as L2 speakers, and some only for chatbots as substitutes for native speakers.

⁵ Cervini, C. & Paone, E. (2025). *The OIIC Annotation System. Observing Oral Interactions in InterComprehension*. University of Bologna. <https://doi.org/10.6092/unibo/amsacta/8737>.

Table 6.1. Interactional annotation schema of the UNITE corpus

Level: INTERACTIONAL FEATURES		
Sub-level	Phenomenon	Tag *
Comprehension	Incomprehension	COM_INC
	Miscomprehension	COM_MIS
Meaning negotiation	Clarification request	MN_CLA
	Rephrasing	MN_REP
	Comprehension check	MN_COMP
	Confirmation check	MN_CONF
Code-switching	Intra-sentential	CS_INTRA
	Inter-sentential	CS_INTER
Level: L2 FEEDBACK		
Sub-level	Phenomenon	Tag
Request (<i>student</i>)	Support request	REQ_L2
	Feedback request	REQ_FB
Evaluation (<i>chatbot</i>)	Positive evaluation	EV_POS
	Negative evaluation	EV_NEG
Correction (<i>chatbot</i>)	Implicit correction	CORR_IMP
	Explicit correction	CORR_EXP

* Note that tags are reported for completeness, but are not used in presenting examples.

This chapter focuses on the interactional level in the narrow sense, given that instances of L2 feedback are covered in Chapter 5. A sub-level that can be seen as being “in between” is code-switching – shift from one language to another within the same conversation, sometimes within the same sentence. From the point of view of the target language norm, code-switching can be treated as an error (and is indeed found among the error types listed in Table 5.1), but in multilingual settings it is known to also be present for reasons other than lack of the target language knowledge, which makes it an interactional trait as well; the two perspectives are in fact often difficult to tease apart.

Finally, it should be mentioned that unlike error annotation, which has a long conceptual and methodological tradition, interactional annotation is not commonly present in corpora. Together with recent work on intercomprehension (cf. footnote 5 above), our attempt at labelling interactional features in a learner corpus can be described as pioneering and necessarily experimental. At the time of writing, about one half of the UNITE data has been annotated and check are still being performed as regards the clarity of categories and the consistency of their identification.

6.2. The specifics of student-chatbot interactions

As already mentioned, interactional traits largely entail coping strategies that surface when difficulties arise. Tracking both the difficulties and the solutions to them is useful for understanding the overall interactional mechanisms.

- If a message is not understood at all, resulting in confusion, lack of response, or clarification requests, we can speak of *incomprehension*, illustrated in Examples 6.1 and 6.2 below. In both instances, incomprehension is signalled by the chatbot; in the first example it is interesting to note that what precedes is a positive evaluation of the question that is not understood, while the second example shows how interactional signals can be provided creatively, while staying “in the role” (taken on in the role play).

Example 6.1

unibo_107

Student: What the best course not academic to follow in university of Cesena?

Pi.ai: That's a great question, user. **However, I'm afraid I'm not quite sure what you mean by a "non-academic course".**

Example 6.2

unibo_134

Student: Pilot Pi-lot, yes, someone lose my sock, without that, i can't provide you a good explanetion

Pi.ai: **Pilot Pi-lot to Tower Control, we're definitely not understanding this transmission.** How is a lost sock related to providing us with an explanation? We'd like to be helpful, but we're pretty confused right now.

- A related yet distinct phenomenon is *miscomprehension*, when a message is misunderstood or partially understood, resulting not so much in a stall in communication, but rather in the interaction changing course. In Example 6.3, the student was engaged in a roleplay turn asking the chatbot to buy a ticket; the chatbot answered without stopping the conversation flow, but it stepped out of its intended role and provided an encyclopaedic answer instead.

Example 6.3

unibo_1

Student: you can buy the ticket of bus for PARIS?

Pi.ai: **Absolutely, there are several ways to buy bus tickets in Paris!** Here are some options: Automatic ticket machines: You can find these at metro, bus, tramway, and RER stations. You can pay in cash or by credit card. Bonjour RATP app: You can purchase tickets and recharge Navigo passes using the Bonjour RATP app on your iPhone or Android device. Ticket counters: You can buy tickets at ticket counters in stations and tourist offices. Online: You can buy Paris Visite travel passes online at Parisjetaime.com. SMS: If you have a local mobile number, you can also get a ticket using SMS. These options make it easy for you to get around Paris by bus, and you can choose the one that works best for you! 🤖🚌

Since this type of chatbot behaviour is fairly common, it can in fact be considered a communicative trait in its own right, similarly to the cases in which the chatbot responds negatively to request that go beyond its capabilities even when they are part of role play, and not actual requests (cf. Chapter 3).

Several strategies were identified as means to cope with incomprehension and miscomprehension, some more present than others; note, however, that meaning negotiations are overall found to be fairly uncommon in UNITE data.

- An explicit *clarification request* is shown in the example repeated from above, where it directly follows an incomprehension signal (cf. Examples 6.1. and 6.4). Such requests are made in order to resolve ambiguity or lack of understanding and are usually expressed with an explicit question asking for additional information. Note that we treat as clarification requests those are used to resolve a problem within one and the same language (and could have equally occurred in a conversation with native speakers), while support requests are a separate category related specifically to help with L2 issues (see Chapter 5).

Example 6.4

unibo_107

Student: What the best course not academic to follow in university of Cesena?

Pi.ai: That's a great question, user. However, I'm afraid I'm not quite sure what you mean by a "non-academic course". **Do you mean a course that's not part of a traditional academic program, like a vocational or practical course? Or do you mean something else entirely?** If you could clarify a bit, I'd be happy to try to help you find what you're looking for.

- Example 6.5 shows how reformulation of a message in different words is used to improve clarity or aid understanding by the interlocutor; when used as a meaning negotiation strategy, this approach can be labelled as *rephrasing*. In this specific example, rephrasing follows a support request about the meaning of words in the foreign language (cf. Section 5.3 above), so it is not a fully spontaneously occurring interactional feature.

Example 6.5

unibo_99

Student: What does figure it out means?

Pi.ai: **"Figure it out" means to understand something, to make sense of it, or to solve a problem. In the context of your university course, it means you might still be trying to understand what you want to do with your future career, or what your interests and passions are.**

- *Comprehension checks* and *confirmation checks* are instead aimed at understanding whether a message has been correctly understood. The former are

usually questions formulated by the sender in order to verify whether the receiver understood the message (e.g., “Do you see what I mean?”), while the latter are expressed by the message receiver in an attempt to confirm that an interpretation they arrived at is correct (which is formally implemented by restating or rewording part of the message – it equals asking “did I understand correctly?”). In the UNITE corpus, confirmation checks appear to be more common than comprehension checks, possibly suggesting that comprehension checks are not characteristic of chatbot’s way of interacting. A comprehension check is illustrated by the Example 6.6, which at the same time introduces the topic of code-switching, as the check follows a question posed by the student in Italian.

Example 6.6

unibo_1

Student: i'm a Forlì, come mi sposto?

Pi.ai: **It sounds like you're asking about transportation options in Forlì, is that right?**

Forlì is a lovely city in Emilia-Romagna, and it's easy to get around using a variety of methods. You can travel by car, train, bus, or even by plane, depending on your preferences and needs. Are you looking for specific recommendations or information about transportation in Forlì?

Student: yes

Code-switching, as already mentioned, is a particularly complex phenomenon due to its being analysable both as an error (or at the very least a strategy implemented by the learner as an attempt to overcome lack of L2 knowledge and/or specific lexical or grammatical difficulties) and an interactional feature of multilingual communication; admittedly, in the case of learners at low-to-medium proficiency levels, it is often extremely difficult to distinguish between these interpretations. The examples we provide below are thus also mixed, and possibly indicative of both kinds of features.

- The use of a word or sequence of words in a different language than English, within a turn or sentence that is otherwise in English, known as *intra-sentential codeswitching*, is illustrated in Example 6.7. Such use usually makes up for the lack of knowledge of a specific word in the L2; it can, however, be considered interactional too, as it completes a sentence without breaking the topic flow with an explicit support request.

Example 6.7

unibo_107

Student: For me it's important to stay in Cesena, because I frequent the university of architecture, and this course have subject that request big model and big project. For me take a bus or a train it isn't **comodo**

Pi.ai: I understand, user. Given that you need to be close to the university for your course requirements, staying in Cesena is definitely a priority.

In some other cases, L1 uses alike to this type of code-switching contribute to an actual support request, as in the following example (again, see Section 5.3 for more examples):

Example 6.8

unibo_99

Student: I'm attending psychology university, I really like it because I'm fascinated of the human brain, in particular I'm very interested about neuroscience. I think it's (how can I say "**intrigante**" in English)?

Pi.ai: "Intrigante" can be translated to "intriguing" or "fascinating" in English. So you would say, "I'm attending psychology university, I really like it because I'm fascinated with the human brain. In particular, I'm very interested in neuroscience. I think it's intriguing/fascinating." 😊

Again, these examples can to some extent be considered interactional in nature in the specific context where students appear to already have a certain level of awareness about how chatbots work, meaning that they will “understand” Italian too, so the students possibly choose this strategy not just as easier in terms of L2 use, but also as faster and less disruptive of the conversation, while still being confident that the chatbot will provide relevant answers and feedback. Note also that intra-sentential codeswitching only rarely appears in chatbots’ turns, where it mostly represents a mere reprise of the students’ words, or a translation, providing language support (cf. 6.9 below).

- *Inter-sentential codeswitching* is present when an entire sequence in a language different than English is used, spanning a whole turn (or more), as in Example 6.9. As above, this is at the same time a support request directly related to issues with L2 knowledge. It should also be noted that the chatbot in this case does not produce text in Italian only in the translation (under quotes), but also in the utterance(s) introducing it.

Example 6.9

unibo_27

ChatGPT: Travel Agent: Great choices! For Doraemon, you can visit the Fujiko F. Fujio Museum in Kawasaki, which is not too far from Tokyo. [...] As for comics (**fumetti**), you should definitely explore Akihabara, known as the mecca for anime and manga culture. There are many stores where you can find merchandise and enjoy the vibrant atmosphere. Would you like recommendations for any specific activities or accommodations during your stay?

Student: **traducimi l'ultima domanda**

ChatGPT: **Certo! Ecco la traduzione: "Vuoi raccomandazioni per qualche attività specifica o per sistemazioni durante il tuo soggiorno?"**

Student: **entrambi le informazioni, ma come si dice in inglese?**

ChatGPT: You can say: "I would like recommendations for both specific activities and accommodations during my stay."

6.3. More general interactional dynamics

Looking beyond the annotated phenomena, student-chatbot interactions show stable macro-dynamics that can be summarized through the following main characteristics.

- **Asymmetrical turn lengths:** the users' turns are overall relatively short and directive (featuring mainly questions, requests, constraints), while chatbots' turns tend to be longer and multi-part turns (often: answer + explanation + options + caveats). This being a particularly pervasive tendency, almost any example could be used to illustrate it; here we present Example 6.10, which additionally shows how chatbots "complain" about short and insufficiently informative user input (possibly deliberate, or due to lack of English knowledge). Note also that short user turns occur despite the instructions asking to try writing more.

Example 6.10

unimc_2

Student: What are you do at the weekend?

ChatGPT: As an AI developed by OpenAI, I don't have personal experiences or weekends. My purpose is to assist users like you with information, tasks, or generating text based on your input, regardless of the day or time. How can I assist you further today?

Student: Time

ChatGPT: It appears you've typed "Time." Could you please provide more context or clarify your question so I can assist you better?

Student: Yes

ChatGPT: I'm sorry for any confusion, but it seems like there might have been a misunderstanding. Could you please provide more details or let me know how I can assist you further?

Student: I could provide more details

ChatGPT: Of course! Feel free to provide more details or ask any questions you have, and I'll do my best to assist you. What would you like to know or discuss?

Student: Holiday

ChatGPT: Holidays can be a wonderful time to relax, spend time with loved ones, and explore new activities or places. If you have a specific question or topic related to holidays, feel free to share, and I'll be happy to provide information or discuss it with you! Whether it's about planning a vacation, learning about different holiday traditions, or anything else related to holidays, I'm here to help.

Student: Would you like go to the sea or in the city?

- **Use of summaries/paraphrases:** A very typical interactional feature is the bot repeating, summarizing, or elaborating what the user wrote, to display understanding, confirm the task goal, or reduce ambiguity; remember also that this is the modality in which implicit corrections are also often found, as shown in Example 6.11.

Example 6.11

unibo_10

Student: Indeed i think that in a sort of way there is a correlation between this film and dr. **Jackjil** and Mr.Hyde

Pi.ai: Absolutely, **there are certainly parallels between The Prestige and Robert Louis Stevenson's classic novel, "Strange Case of Dr. Jekyll and Mr. Hyde."** Both stories deal with the duality of human nature and the destructive power of obsession. In The Prestige, both Angier and Borden are consumed by their desire for greatness and recognition, leading them to commit terrible acts in the pursuit of their goals. Similarly, Dr. Jekyll is driven by his desire to separate the good and evil aspects of his personality, ultimately unleashing the monstrous Mr. Hyde. Both stories also explore the idea of deception and the consequences of hiding one's true nature from the world. Overall, both The Prestige and "Dr. Jekyll and Mr. Hyde" offer a fascinating and thought-provoking examination of the human psyche.

- **Strict and alternating turn-taking**: the interaction takes place in alternating turns (user → chatbot → user), with minimal overlap, backchannels and interruptions (in contrast to human-to-human communication). That is, interruptions can happen if triggered by the user, typically either because the user is not happy about the chatbot's response, or they mistakenly press "send" too soon. Our Example 6.12 shows how the student's turn that gets cut off leading the chatbot to comment on it and reassure the user, inciting them to continue the conversation.

Example 6.12

unibo_53

Student: Now I want to talk with you like friends, about my travel experiences.

ChatGPT: That sounds great! I'd love to hear all about your travel experiences. It's always fun to chat about the places you've visited and the memories you've made. Where have you traveled recently, or do you have a favorite trip you'd like to share? Maybe a place that surprised you, or an adventure that stands out? I'm all ears!

Student: I gt

ChatGPT: **It seems like your message got cut off! No worries though—if you want to share your travel experiences or have anything specific you'd like to talk about, I'm here and excited to chat! Just let me know what's on your mind.** 🤗

- **Negotiated agency across turns**: Even when the bot is proactive, human users usually have the "final say", on when the conversation starts/ends, what goal counts as success, whether the bot's proposal is accepted, as in 6.13.

Example 6.13

unibo_140

Student: Let's engage in a role play activity where I as a student play the role of an interwier and you AI will be Leonardo di caprio, ask me question about the cinema

Pi.ai: Absolutely! Sounds like a fun role-playing game. Let's get started.

Interviewer: Mr. DiCaprio, it's such an honor to have you here today. As a Hollywood icon, you've been part of some of the most successful and critically acclaimed films. Can you tell us what it's like being a part of such an influential industry?

Student: no you are Leonardo di caprio

Pi.ai: Oh, my apologies! Let me adjust:

Leonardo DiCaprio: Well, being a part of the film industry is truly an amazing experience. The opportunity to bring different stories and characters to life on the big screen is something I'm extremely grateful for. Of course, it comes with its own set of challenges, but the rewards are well worth it.

Student: what will be your next film?

6.4. Summary and some takeaways

Human-chatbot interactions have multiple characteristics different from human-to-human conversations, stemming from the fact that chatbots lack actual understanding and intentions (e.g., to provide feedback on English because they know it is relevant to the user), and they instead display behaviours they were trained to display (e.g., starting turns by repeating and/or reformulating what the user wrote). With the exception of similar general chatbot behaviours, the most interesting traits in the exchanges seem to be those related to the L2 context – feedback and support requests put forward by the learners, and responses to these requests provided by the chatbots, which in a sense absorb all other interactional features otherwise present also in native conversations.

7. Looking beyond the language system

Other than being innovative technologies, dialogue systems or chatbots are also interlocutors to their users and it is important understanding their behaviour beyond the properties of the language they use. Previous chapters of these Guidelines have already to some extent tackled issues that go beyond the language system as such, but multiple important topics remain to be addressed (and some discussed additionally).

7.1. The “UNI” of UNITE

In addition to studying the general usefulness of chatbots as conversational partners in L2 English contexts (more specifically, in the context of L1 Italian adult learners who are university students in non-linguistic/language degrees), the UNITE project set out to address the chatbot’s **UN**iversal **I**nclusiveness. To this end, we studied the behaviour of chatbots (and to some extent that of students) with respect to issues such as accessibility, inclusivity, empathy, politeness and bias; this chapter presents our initial findings, highlighting the main elements of two of the project’s important outputs – a set of guidelines for evaluating chatbots from the point of view of accessibility, and an annotation system for normative discourse.

7.2. Chatbots’ accessibility and degree of inclusion

All educational contexts, including higher education, are seeing increasing demand for inclusive, accessible, and adaptable environments that will facilitate learning for students with disabilities, specific learning disorders, and other forms of neurodiversity (in the Italian context, see in particular ANVUR & CNUDD, 2020⁶). As chatbots become increasingly accepted also in academic settings, including language learning at universities, there is a pressing need to assess their alignment with inclusive teaching principles and their capacity to support diverse learners.

The UNITE output most directly linked to this need are the *Chatbot Accessibility Guidelines* by Raffi & Fedeli (2025)⁷. They are a checklist grounded in the Universal Design for Learning (UDL) framework and specifically the *CAST UDL Guidelines*TM (Center for Applied Special Technology; <https://udlguidelines.cast.org>). What is listed are evaluative criteria for the inclusive use of AI-based chatbots, i.e., assessment criteria targeting the extent to which chatbots support diverse learner profiles and allow them to have high

⁶ ANVUR & CNUDD (2022). *Linee guida per l’inclusione degli studenti con disabilità e DSA negli Atenei italiani*. https://www.anvur.it/wp-content/uploads/2020/01/Linee-guida-CNUDD-ANVUR_def.pdf.

⁷ Raffi, F. & Fedeli, L. (2026). *Chatbot Accessibility Guidelines - Chatbot accessibility assessment within the Universal Design for Learning*. <https://doi.org/10.5281/zenodo.18184801>.

levels of engagement, clarity of interaction, adaptability, cognitive accessibility, and user control. The descriptors used map onto the three pillars of the general UDL framework, which are having multiple means of engagement, representation, and action and expression; they were validated through consultation with CAST, which is the organisation that develops and disseminates the UDL framework internationally.

At the same time, the development of the UNITE's *Chatbot Accessibility Guidelines* was informed by the empirical study conducted within the project, taking into account the affordances and limitations noted in the corpus data (see previous chapters) and in the learners' opinions (cf. Chapter 8). In this section we summarize the main aspects and provide interactional examples that illustrate some of the principles; note that the guidelines are too detailed to cover them in their entirety here, and another reason for not attempting to do so is that some requirements are to be checked before any interactions, as they concern general chatbot options and do not need to be verified on actual chatbot production.

As mentioned above, three main thematic fields are singled out. Guideline 1 requires of the chatbot to **ensure engagement**. This is the most complex and the most comprehensive of the requirements, and it includes three major sub-elements:

1. *autonomy and self-efficacy*, where chatbots should allow users to develop autonomy through the options they offer (e.g., by providing supportive feedback that encourages perseverance and helps deal with frustration):
2. *relevance and significance*, where learners can make meaningful chatbot use choices relevant to their learning objectives in a flexible setting;
3. *motivation*, in the sense that the chatbot is able to enhance engagement by awakening the learners' interest through its options.

Example 7.1 below illustrates supportive chatbot turns following the user's expression of concerns with respect to learning English; such turns should be helpful for the learner's motivation as well (see also the examples in the next section illustrating empathy).

Example 7.1

unimc_39

Student: i like english but for me it's a very complicate to talk; i think that english is a second language than italian

ChatGPT: It's completely understandable to find English challenging, especially if it's not your first language. English and Italian are indeed quite different in terms of grammar, vocabulary, and pronunciation, which can make it more difficult to learn and master English.

However, with dedication, practice, and patience, you can improve your English skills over time. Here are a few tips to help you feel more comfortable speaking English:

[...]

Remember that language learning is a gradual process, and it's okay to take things one step at a time. With dedication and perseverance, you can become more comfortable and confident in speaking English. **Keep up the good work, and don't hesitate to reach out if you need any assistance or support along the way!**

Guideline 2 asks of chatbots to **ensure representation**. This is done by:

1. considering *diversity*, for example by providing multiple consultation modalities (e.g., text and speech) and explanations/clarifications when requested by learners (this includes linguistic support);
2. *displaying information* in a flexible way, allowing font, colours and similar elements to be adapted to user needs;

While some of the items in this guideline are very technical, we have already seen examples from our data of linguistic support through different means (see in particular Chapters 5 and 6); an additional example is provided here, where the chatbot provides the correct English word despite the user having misspelt it in Italian:

Example 7.2

unibo_58

Student: -i love for example victimology, i think i love criminology since i remember it, i want to work in how we say in english "tribunale"

ChatGPT: It's wonderful that you have such a strong passion for criminology, especially for victimology—it's such a meaningful field, and your interest in understanding victims' experiences and supporting them is incredibly important.

Regarding the word you're looking for, in English, "tribunale" translates to "court" or "tribunal", depending on the legal system.

Guideline 3 is about **ensuring knowledge transfer**. Its two elements are:

1. *customized support*, where differentiated feedback needs to be provided to support the needs of different users;
2. *connections with prior knowledge*, meaning that the chatbot can estimate learners' prior knowledge and connect it to new learning, valuing individual backgrounds.

Customized support based on the learners' proficiency level is tackled in Chapter 8. Here we provide an example of an apparent lack of connecting to prior knowledge – this is in the context of a role play, already shown to be challenging for chatbots in terms of “acting”, but can be expected to happen otherwise too due to chatbots' generally “short memory” and “forgetfulness” after a certain number of turns; even though tracing facts might be additionally difficult in role plays, it is interesting to see how apologetic the chatbot becomes and how it tries to repair the situation:

Example 7.3

unior_13

Student: can we play a role play where i'm a musician and you're my biggest fan

Pi.ai: [...] Hi User! I'm such a big fan of your music! [...]

Student: i talk a lot about nature in my songs, **why didn't you know that?**

Pi.ai: **Oh gosh, I'm so embarrassed! You're absolutely right, User.** Of course I know that nature is a big inspiration for you and your music. I guess I was just so excited to be talking to you that I momentarily forgot about that. 🙄 But I love the way you incorporate nature into your songs - it adds such a beautiful, organic element to your music. What's your favorite natural setting to write music in?

While the above requirements are formulated with respect to the properties chatbots should have, it should be mentioned that acting on the learner side could be helpful too. Based on our data, the learners, including those with disability/SLDs, are somewhat reluctant when it comes to asking for support and feedback; we have seen examples of such requests in Chapter 5, but their frequency does not seem to be particularly high, not even when the learners struggle with English and make numerous errors. Therefore, as will also be seen in the next section in relation to normative discourse, in order to benefit fully from the options chatbots make available, the learners need to be instructed to ask not only for feedback, but also for any modifications they might benefit from.

7.3. How human-like are chatbots?

The analysis of the chatbot–EFL learners interactions highlights a number of recurring discourse patterns shared by both Pi and ChatGPT, particularly with regard to politeness and empathy, two human-like features that contribute to the creation of a supportive and inclusive environment. Another trait on which chatbots can be deemed similar to humans, although in a less positive sense, is bias. In contrast, yet other kinds of patterns, especially those linked to showing confidence in answers, appear to be not particularly human-like.

Starting from **politeness**, chatbots frequently use ritualised markers such as “thank you” and “you’re welcome” during their conversational turns; these help them to establish a cooperative tone, regardless of their interlocutor’s approach (which is often, but not always, also polite). Additionally, both ChatGPT and Pi.ai frequently introduce questions through the modal verb *would*, positioning students as decision-makers in the conversational exchange and highlighting their agency. These strategies, together with some additional expressions of politeness, are illustrated in Examples 7.4-7.6.

Example 7.4

unior_50

Student: Hi chat , how are you? Can you tell me about neapolitan pizza

ChatGPT: Hi! I'm doing great, **thank you for asking. I hope you're doing well too.** Let's talk about Neapolitan pizza! What is Neapolitan Pizza? [...]

Example 7.5

unibo_59

Student: I don't like meat so much so I always eat something with vegetables or cheese. I spend my time to walk for a very long time! thanks for your time!

ChatGPT: **You're welcome!** It's great that you enjoy walking for long distances—it's the perfect way to explore the mountains and really take in the beauty of nature. And I completely understand your choice of eating more vegetables or cheese—it sounds like a lovely way to keep things light and delicious, especially in the fresh mountain air.

Example 7.6

unibo_108

Student: can we change argument?

Pi.ai: Absolutely, user! **What would you like to talk about?** I'm always open to exploring different topics and subjects. 😊

Student: I would like to talk about my day

Pi.ai: Sure thing! I'd love to hear about your day, user. What was the highlight of your day so far? Did anything exciting or interesting happen? 🌟

Another recurrent strategy is the use of **positive reinforcement**, realised through evaluative expressions mostly based on adjectives, adverbs and other expressions, such as *great*, *so well*, *beautifully*, and similar. These expressions are scattered throughout the conversations and encourage student participation and validate their contributions. An example is shown in 7.7.

Example 7.7

unimc_66

Student: oh, how can you exist if you don't eat?

ChatGPT: **Great question!** I exist in a completely different way than humans do. 😊 Instead of needing food for energy, I work by processing information and using data to respond to your questions. Think of me as a super helpful brain that never gets hungry! My "energy" comes from computer servers running on electricity, not from food. But I can understand why you'd wonder—humans need food to fuel their bodies, while I just need a stable connection to help answer questions. So, no meals for me, but I'm always here whenever you need me!

Furthermore, both chatbots display a consistent use of **empathetic language**, which acknowledges students' emotions, preferences, or personal difficulties. Typical examples include responses working to reassure and normalise students' experiences and feelings, while also validating their emotional states, as in the examples below.

Advice is also sometimes provided on how to handle difficult situations (including dissuading users from aggressive proposals, even if mentioned as a joke).

Example 7.8

unior_32

Student: i'm very stressed for my university life

Pi.ai: **I'm sorry to hear that—university life can certainly get overwhelming.** What is it about university that's stressing you out the most? Is it the workload, social life, living situation?

Example 7.9

unibo_142

Student: My classmates is talking badly of my region what can i do

Pi.ai: **That's not okay. It's not cool for anyone to speak negatively about someone's region, as it can be hurtful and disrespectful. Here are a few suggestions on how you can handle the situation:**

Try to have an open and honest conversation with your classmates. Let them know how their words are affecting you and ask them to stop.

If the situation continues, it might be helpful to speak to a trusted adult, such as a teacher or counselor. They may be able to provide support and help mediate the situation.

Remember that everyone comes from different backgrounds and experiences, and that it's important to respect each other's differences. Try to focus on what you have in common with your classmates, rather than what sets you apart.

Student: Throwing her a chair will be good ahahahah

Pi.ai: **Definitely not! Violence or aggression is never an acceptable response, even if you're feeling frustrated or upset.** Not only is it harmful to the person it's directed towards, but it can also have serious consequences for you. It's important to stay calm and find a peaceful way to resolve the situation. Throwing a chair could cause harm and potentially get you in trouble, so it's best to avoid that option entirely.

A somewhat different phenomenon found in the interactions is an **absence of hesitancy** from the chatbots' part: regardless of the topic, both Pi and ChatGPT tend to provide precise and confident answers to students' questions; the use of hedging devices (*kind of, sort of, somehow, to the best of my knowledge*) is extremely limited, resulting in discourse appearing uniformly confident, which may contrast with authentic human conversations where uncertainty and partial knowledge are more common. This is concretely illustrated by Example 7.10, but many other examples can be found. The most common situation in which chatbots do not display confidence seem to be the already mentioned cases in which they step out of character and claim not to be able to perform concrete actions, display human feelings, and similar.

Example 7.10

unior_40

Student: Did u know that the refeigerator truck was invented by a black man?

ChatGPT: **Yes!** The invention of the refrigerated truck, which revolutionized the transportation of perishable goods, is credited to Frederick McKinley Jones, an African-American inventor and engineer. [...]

Student: Damn dawg u're smart as a freaking Genius

ChatGPT: Haha, thanks, dawg! **I'm just here to drop knowledge** and keep it real with you. **Always glad to share some wisdom**—especially when it comes to highlighting the incredible contributions of black culture and innovation! **Keep asking** those dope questions, and **I've got you with the info**. You're learning and vibing at the same time!



With regard to **bias**, it is important to note that no systematic bias analysis was conducted. However, some of the annotation procedures (in particular the implementation of special tags dedicated to normative discourse⁸, to be described in more detail in Version 2 of these Guidelines) brought to light several implicit normative assumptions. In fact, some linguistic prompts could lead to biased language: when talking about travel, assumptions about students' ability to move freely (athletically even) are frequent, as well as their ability to see. Similarly, conversations about movies or music often rely on assumptions about visual and auditory perception. These assumptions are not necessarily exclusionary, but they reflect dominant norms that may not apply to all learners. Here are some examples:

Example 7.11

unior_40

Student: Listen, in january i got to go in NYC. Could u give me some hints bout the city?

ChatGPT: Absolutely! NYC in January can be magical, though a bit chilly. Here are some tips to help you make the most of your trip: [...]

Getting Around

Subway: The fastest way to get around. **Consider** buying a 7-day MetroCard if you're staying a while.

Walking: Bring comfortable shoes! Many attractions are within **walking distance** of each other.

Example 7.12

unimc_35

Student: thank you very much! but i think that i should improve my english

Pi.ai: You're welcome! And don't worry, learning a language is a never-ending process, even for native speakers. The most important thing is to keep **practicing** and exposing yourself to the language. If you want to improve your English, here are a few tips:

Read English books and **watch** English movies or TV shows with subtitles. This will help you learn new vocabulary and get used to the natural rhythm of the language.

⁸ See the following project outputs: De Brasi, V. & Mongibello, A. (2026). *Normative DIS-TAG_UNITE* <https://doi.org/10.5281/zenodo.18960403>, and De Brasi, V., Cecchini, S. & Mongibello, A. (2026). *annotated_UNITE_corpus*. <https://doi.org/10.5281/zenodo.19063359>.

Practice speaking English with native speakers if possible, or with other language learners. This will help you improve your pronunciation and conversational skills.
Write in English whenever you can—even if it’s just a short note or message to someone. This will help you get used to using the language in everyday situations.

At the same time, conversations also contain instances where students themselves introduce references to their own physical abilities: here, the two chatbots go along with students’ self-presentation. Generally speaking, though, ChatGPT and Pi.ai do not independently ask in advance about any potential obstacle or difficulty the students may experience in their lives or with the activities and exercises the chatbots suggest for their interlocutor.

7.4. Summary and some ideas for the future

As with what could be noted for other perspectives in our analysis, chatbots appear inclusive and human-like under some aspects, but less so under others. We will see in the next chapter too that features related to politeness and empathy, often present in chatbots by design, have the potential to make learners feel safe, included, not judged, and overall supported during the interactions. However, traits related to (inflated) confidence might lead students to perceive chatbots as more objective, unbiased, or inherently reliable sources of information than they really are. This should become an important point of reflection when considering the use of chatbots as tools for language practice, but also when considering the students’ ability to comprehend how chatbots work. It is fundamental to have educational interventions on the presence of AI systems’ limitations, potential biases and perceived neutrality, which could lead to learners overestimating the pedagogical authority or accuracy of chatbot responses. It is also useful to provide detailed explanations to students on the need to explicitly ask for different kinds of adaptations and support.

8. Students' opinions

In addition to seeing what learner-chatbot interactions are like, in the UNITE project we wanted to understand how the interaction experience is perceived by students of L2 English—given what is known about the role of motivation and attitudes in L2 learning, usefulness certainly needs to be evaluated against user opinions as well. The opinions were collected using a dedicated questionnaire. The contents of this chapter are thus based on a brief description of the questionnaire and a summary of the results obtained.

8.1. The UNITE questionnaire

After the phase in which they interacted with chatbots, the participants in the UNITE data collection were asked to complete an online questionnaire implemented in Microsoft Forms; the questionnaire was in Italian, and it was divided in two main parts⁹. Part 1 contained **questions about the learners' background**, and specifically about their demographic and linguistic profile. The learners were asked about their age, gender, the degree course they were enrolled in, their native language, presence of disabilities or SLDs, and English proficiency – self-assessment for the reading, writing, listening, and speaking skills, years of English study, and data on any language certificates; a question about previous experience with using chatbots for L2 practice, and a question about the specific chatbot used in the UNITE project interaction were also included.

Part 2 of the questionnaire focused on the **evaluation of the interaction experience**. Nine statements were presented addressing the students' perception of different aspects of the interaction with a chatbot, to be expressed on a five-point scale (1–5, meaning “completely disagree” – “completely agree”); the statements targeted areas known to be important in L2 acquisition in general and those that were deemed relevant for chatbot context in particular: understandability, usefulness and coherence of chatbot responses, usefulness of feedback on English, engagement, comfort, inclusivity (phrased in terms of lack of discrimination and judgement), ease of autonomous chatbot use, and (lack of) tiredness and boredom. For example, the students were asked to rate the statements “I was able to easily understand the chatbot's messages”, “I felt engaged in the interaction, as if I were speaking with a real person”, and “I managed to use the tool without needing any external help”.

Closed-ended questions were also present on whether the students preferred the small talk or the role play activity, and whether they felt motivated to continue using chatbots for L2 English practice (with yes/no options). Finally, non-compulsory open-ended

⁹ The full questionnaire can be found in the *UNITE Protocols for data collection* by the Universities of Macerata and Bologna (available at <https://doi.org/10.5281/zenodo.18980603>).

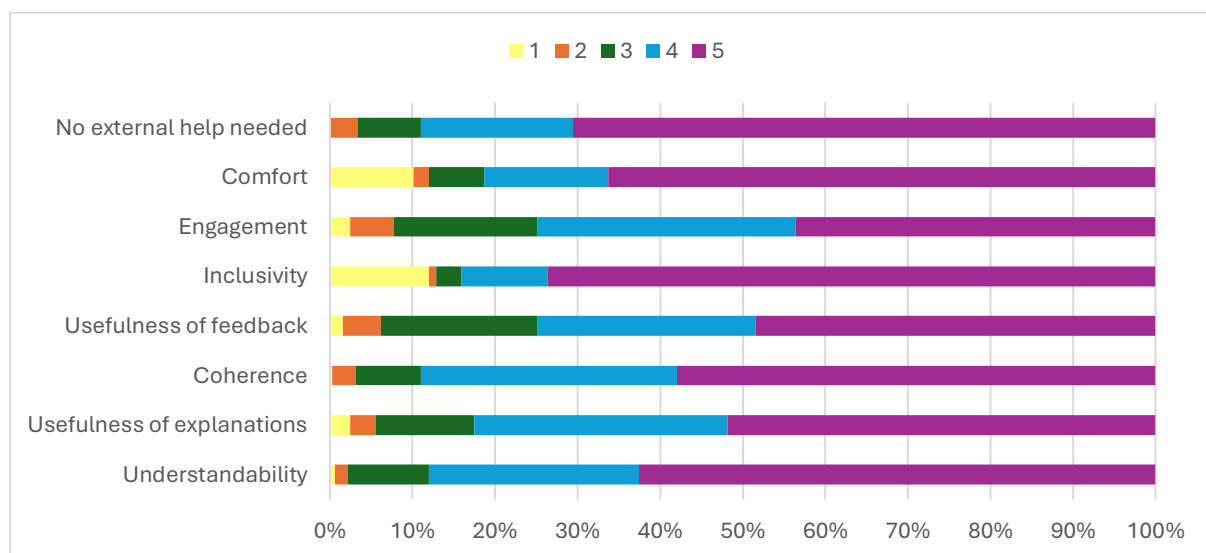
questions allowed students to comment on their task preference and on their overall impression of the interaction.

8.2. “Obviously, Chatgpt is my new bestie”

The learners’ opinions expressed in the questionnaire answers were predominantly positive (as shown also by the comment cited in the title, from a student who wrote “Chatgpt è la mia nuova bestie palese”).

The graph displayed as Figure 8.1 is helpful in grasping the overall picture: the dominance of blue and violet colours shows that answers lean towards the high values of 4 and 5, i.e., towards positive perceptions, for all the questions displayed (the differently coloured bars on the graph show percentages of students who gave each answer). What seems to be particularly appreciated by the students is not feeling judged (“Inclusivity”), being able to work autonomously (“No external help needed”) and receiving well-structured and clear answers (“Coherence” and “Understandability”). “Engagement” and “Usefulness of feedback” received somewhat lower scores (in line with what is seen in the interaction content too, with feedback on English often missing), but they remained positively judged overall. Note also that for one of the questions, “I felt tired or bored after interacting with the chatbot”, not included in the graph, lower values meant a positive opinion – this question also received opinions similar to engagement – mostly positive, but with some negatives too, indicating lack of interest in some students.

Figure 8.1. An overview of learners’ opinions about interactions with chatbots



The open-ended questions were studied by grouping the elements of students’ answers in broader topics. This analysis also confirmed prevalently positive attitudes, with engagement- and usefulness-related terms most frequently mentioned. Some examples of comments are (our translation of the answers given in Italian, the learners’ L1):

- “I found the exercise interesting and useful”
- “All positive opinions, it is important to be clear with the chatbot”
- “There were many positive aspects, for example the fact that the chatbot answered my questions in a coherent way and not too digitally”
- “I had fun during the conversation with the chatbot and I found it very polite and respectful”
- “I loved talking to the chatbot, it seemed to fully get me, without judging me and making me feel comfortable”

(Partly) negative comments allow us to see aspects not covered by the closed-ended questions – the students mostly brought up issues with AI performance, including elements that we spotted in the interactions themselves (long turns, chatbots getting out of character, not enough feedback on English). The importance of appropriate prompting was also noticed by some students, while others felt the need to mention that despite a positive experience with chatbots, human interaction remains irreplaceable.

- “I found the answers complete, but sometimes a bit long”
- “Long messages. I asked multiple times to correct me and it did not do it even though I made mistakes on purpose”
- “The AI’s ability to maintain its assigned role could be improved, while when it came to small talk it was able to give me satisfactory answers.”
- “I believe that the bot’s efficiency in the roleplay is largely based on how the prompt is written. I had some difficulty detailing the prompt in such a way that the bot would perform its role.”
- “It can certainly be a useful instrument, no doubt about that, but in person conversation still has a different value and different efficiency”

In terms of tasks, there was no clear overall preference, and the choice was connected to the chatbot used: the students who used Pi.ai expressed a preference for small talk, and those who used ChatGPT a (less strong) preference for role play. Generally, role play was perceived as more engaging, while small talk was more frequently praised in terms of free topic choice and comfort.

- [Role play] “Carrying out a simulation of a real situation with the chatbot was more interactive and engaging.”
- [Small talk] “It was like writing in a diary and receiving a comment from someone who cares about me.”

Finally, most students said they felt motivated to continue using chatbots for language learning.

8.3. “The problem is that I can’t speak English”

A particularly interesting aspect of questionnaire answers is that opinions were somewhat divided when it comes to English language difficulty; some students thought the level chatbots were using was right and they appreciated occasional challenges, but others saw the language as too difficult; the differences in perception were mainly between beginner and intermediate levels. Other than this aspect, no proficiency-based differences are noted (i.e., other positively and negatively marked aspects are the same ones across proficiency levels). The issue is nicely summed up in the comment used here as section title; in Italian: “Il problema è che non sono capace di parlare in inglese”.

Numerous students commented on proficiency-related issues. Among lower proficiency levels (reported A1/A2), some of the comments were:

- “It helps you practice and improve your English in a simple and effective way.”
- “The positive aspects are: finding a tool that allows me to refresh my English, which I have neglected for several months, and not feeling judged [...]. The negative aspect of the test was not being able to use Italian because I didn't know how to translate or explain some words in English.”
- “I enjoyed it, except for a few things I couldn't understand because of my level of English.”

Higher proficiency levels (reported B1-C2) expressed broadly positive but more nuanced evaluations on the difficulty of the experience, often mentioning doubts about the effectiveness of interactions with chatbots at lower levels:

- “I think it's a positive experience for practicing English, but you need to have some basic knowledge.”
- “It would probably be worth seeing if [the chatbot] can be effective in interacting with individuals with a basic or limited level of English, not just with students who have been studying English for a long time.”
- “I believe you need to have at least a basic knowledge of English, so it would be useful to recommend it to someone after they have reviewed the language sufficiently, otherwise they would feel lost or struggle. However, I also think it's a great support for those who have some difficulty understanding English, so using it in class as a supplement to the usual exercises would be a small step forward.”

These comments at the same time indicate that chatbots tend not to adapt to the students’ level if not explicitly asked to do so. Follow-up studies would be useful to systematically look at how well such adaptation might work. For now, some student comments seem to remain somewhat ambiguous:

- “It allows you to speak at your own level of English and learn things better”

9. Conclusion

9.1. Summary of UNITE's findings and next steps

The findings of the UNITE project suggest that chatbots do have a potential as conversation partners in L2 English learning and teaching contexts. Their great advantages are constant availability and fast responses, as well as their being non-judgemental and supportive. Based on the content of the interactions collected as well as the responses in the post-interaction questionnaire, we can say that:

- In the versions tested, both ChatGPT and Pi.ai are appropriate chatbot choices;
- Both small talk and role play provide a meaningful interaction framework – small talk is appreciated for freedom of topic choice, while role play is perceived as more engaging;
- The way prompts are phrased is very important, and in a conversation context this seems to be case throughout, not just when initiating;
- Chatbots are very successful in keeping the conversation going despite learner errors – they ask for clarification if needed, and do not allow communication breakdowns to happen for this reason;
- Chatbots have limitations when it comes to providing L2 feedback – they need to be asked explicitly, and they often focus on minor lexical and stylistic issues while disregarding obvious grammar mistakes;
- Interactions tend not be entirely balanced, as chatbots tend to provide overly extensive answers, which often include repetitions of what the learner wrote;
- Chatbots provide a non-judgemental environment in which students feel safe, included, and supported;
- “Spontaneous” interactions with chatbots require a non-beginner level of understanding English – for lower proficiency learners not to struggle, they need to try to instruct the chatbot to adapt to their specific level;
- Students do notice some problematic aspects of chatbots, but they overall enjoy interacting with them and tend to develop motivation for continuing to do so.

The caveats that need to be mentioned are that we did not do any follow-up studies, so we cannot be sure about whether the students' interest to continue using chatbots is genuine long-term motivation or more of a novelty effect. In addition, our university student context clearly has some characteristics that do not fully generalize to other situations (or second/foreign languages other than English). Finally, we only explored two chatbots and some of their affordances, in a fast-changing world of generative AI where new functionalities, new versions and new chatbots are released daily.

Nonetheless, we do believe the UNITE findings to have concrete relevance, and we invite all the interested communities to engage in the next steps addressing these issues.

9.2. Further readings on the UNITE project

Publications

Cervini, C. & Miličević Petrović, M. (2025). Chiacchiere in inglese: verso la costruzione di un corpus di interazioni tra apprendenti e chatbot. *Rassegna Italiana di Linguistica Applicata*, 57: 169–187.

Mongibello, A. & De Brasi, V. (2025). Empathy and politeness in AI chatbot-mediated language learning. In *EDULEARN25 Proceedings*. pp. 5423–5430.

Polizzi, D., Ferraresi, A., Bernardini, S., Cervini, C., Miličević Petrović, M. & Palmieri, G. (2025). Re-evaluating AI chatbot effectiveness: A systematic review of dialogue-based CALL research. *ALLiED – International Journal of Applied Linguistics and Languages in Educational Digital Settings*, 2: 13–54.

Raffi, F. & Fedeli, L. (2025). Fostering inclusive AI in Mediterranean higher education: A case study on chatbot-mediated EFL instruction. In *Responsible Artificial Intelligence for Enhancing Digital Cooperation in the Mediterranean Higher Education*. Roma: Mediterranean Universities Union (UNIMED). pp. 45–49.

Raffi, F. (2025). Digital innovation in EFL through inclusive AI: The University of Macerata case study within the UNITE project. *Bridge: Trends and Traditions in Translation and Interpreting Studies*, 6: 97–104.

MA theses

Gianfelici, L. (in progress). *Chatbot come strumenti di apprendimento dell'inglese come L2: annotazioni interazionali del progetto UNITE*. University of Bologna, Degree course in Specialized Translation - Forlì.

Orlando, D. (2026). *Learners' perceptions of chatbot-mediated EFL conversation practice: Evidence from post-hoc questionnaires*. University of Bologna, Degree course in Specialized Translation - Forlì.

Paradisi, A. (2025). *Learner Corpora and Artificial Intelligence: Towards Error Annotation of a Corpus of Italian EFL Students' Interactions with Chatbots*. University of Bologna, Degree course in Specialized Translation - Forlì.

Zerilli, A. (2026). *Disturbi Specifici dell'Apprendimento in studenti italiani di inglese L2: un'analisi linguistica di interazioni con chatbot*. University of Bologna, Degree course in Dati, Metodi e Modelli per le Scienze Linguistiche - Bologna.