

“Communication” in Learners’ Interaction with Chatbots in Undergraduate Linguistic Courses in Egypt

Hanaa Youssef Shaarawy Ismail
Associate Professor of Linguistics
Badr University in Cairo (BUC)

Abstract

People all over the world communicate with Artificial Intelligence (AI) to facilitate job tasks. Learners also communicate with different chatbots to facilitate their learning process. However, learners sometimes suffer from irrelevant and inaccurate responses given by chatbots. The purpose of this article is to report on a preliminary experience of different interactions performed in Fall 2023 between beginner AI users in two linguistic courses taught by the researcher to first- and fourth-year students of the English Department in one of the Egyptian private universities and different chatbots and analyze these interactions within the framework of the pragmatic relevance theory. Two research questions were posed in the study; one related to the user’s prompts and the other related to the AI responses: (1) How far do learners manifest relevant prompts while communicating with chatbots, and (2) To what extent do chatbots provide relevant responses? These questions were answered qualitatively to account for (ir)relevance of prompts and responses and quantitatively using the F1 score formula. Findings revealed that to get relevant responses from chatbots, the prompts should be on the “determinate pure telling” end of the relevance/irrelevance continuum, that is represented by “explicatures with no implicatures” with only one clear single proposition. The study is one of the prior applications of AI within the framework of linguistic theories.

Keywords: beginner AI users, “communication”, chatbots, “determinate/indeterminate”, “showing/telling”, linguistic courses, F1 score formula

“Communication” in Learners’ Interaction with Chatbots in Undergraduate Linguistic Courses in Egypt

**Hanaa Youssef Shaarawy Ismail
Associate Professor of Linguistics
Badr University in Cairo (BUC)**

Introduction

Large language models (LLMs) illustrated in different chatbots have become a companion to most people all over the world. Learners are also making use of these chatbots to accomplish their academic tasks. It has been claimed that responses generated by some of these chatbots might be irrelevant. This article aims at reporting on a preliminary experience of first- and fourth-year learners’ interactions in Fall 2023 in two linguistic courses at one of the Egyptian private universities with different chatbots and analyze these interactions within the framework of relevance theory. The objectives of the study are to find out how learners’ prompts/instructions are linguistically encoded to the model, and how far chatbots yield (ir)relevant responses.

Theoretical Background and Framework

Sperber and Wilson (1987, 1995) maintained that communication is the outcome of “a coding process” and “an inferential process”. “Coding” is achieved by the speaker linguistically “encoding” a message and the hearer “decoding” it, and “inferencing” is achieved by the speaker “providing” evidence of his/her intentions and the hearer “interpreting” it. Comprehension is realized when the hearer succeeds in “decoding” the linguistic meaning in a specific “context”. Wilson and Sperber (2004, 2012) called this communication “ostensive inferential communication” which requires an “ostensive stimulus” that is supposed to be as easy as possible to attract the hearer’s attention.

In this “ostensive inferential communication”, the addresser demonstrates two types of intentions: “*Informative intention*: to inform the audience of something; *Communicative intention*: to inform the audience of one’s informative intention ... the communicative intention is itself a second-order informative intention: [it] is fulfilled once the first-order informative intention is recognized” (Sperber & Wilson, 1995, p. 29, italics in the original).

Sperber and Wilson (2015) added that “ostensive communication” is characterized by the concept of “manifestness”. “A proposition is manifest to an individual at a given [span of] time to the extent that he is likely to some positive degree to entertain it and accept it as true” (p. 134). Closely related to “manifestness” is the concept of “cognitive environment”. Sperber and Wilson (2015) defined a “*cognitive environment* of an individual at a time ... [as] a set of assumptions/propositions that are manifest to that individual at that time” (p. 135). If this “cognitive environment” is shared between two or more people, it is a “shared cognitive environment”. Sperber and Wilson (1995) stated that “any shared cognitive environment in which it is manifest which people share it is ... a *mutual cognitive environment* ... [and] in a mutual cognitive environment, every manifest assumption is *mutually manifest*” (pp. 41-42).

Sperber and Wilson (2015) asserted that relevant communication is a matter of degree, since there is a “continuum of cases between ‘meaning that/telling that’ (typically achieved by the use of language)” and ‘displaying evidence that’ (in other words, showing)” (p. 119), “the telling that/showing that continuum” on the one hand, with another “continuum of cases from those where the communicator’s meaning is a proposition, or can be paraphrased as such, to those where it is not paraphrasable at all” (p.121), the “determinate/indeterminate continuum” on the other. In all cases of the communication continua, “manifestness” is accentuated, since the speaker makes “mutually manifest one’s intention to make an array of propositions manifest or more manifest to the audience” (p. 139). (For a detailed discussion on the two communication continua

**“Communication” in Learners’ Interaction with Chatbots in
Undergraduate Linguistic Courses in Egypt .**

with examples and on relevance and comprehension, see Sperber & Wilson, 2015, pp. 122-147).

Consequently, communication is achieved through the two continua: the “telling/showing continuum” and the “determinate/indeterminate continuum”. The telling/showing continuum has “pure telling” and “pure showing” as the two ends of the continuum. “Pure telling” is illustrated in “explicature” with no “implicature” when the speaker uses explicit, clear, literal, direct and “paraphrasable” language, and in this case the proposition is “strongly mutually manifest” to both speaker and hearer, since “the array of propositions ... is a singleton, [and] its single member is strongly communicated ... [so it] is a prototypical case of speaker’s meaning” (Sperber & Wilson, 2015, p. 144). “Pure showing” is illustrated in “implicature” only through non-verbal communication and visuals. In between comes “mixed communication” that includes both “telling and showing”. This is illustrated in having “explicature” and “implicature” where not only the literal language is taken into consideration, but also the “context” and the shared knowledge are regarded inviting some inferencing, i.e. no single proposition is communicated through explicit language; the “context” and the “shared cognitive environment” are essential factors in identifying meaning.

On the other hand, communication with chatbots also falls within the continuum of “determinate/indeterminate”. Communication ranges from being “determinate” where language is precise, clear, accurate, specific and to the point, to “semi-determinate” where language is “partly precise and partly vague” (Sperber & Wilson, 2015, p. 121), to “indeterminate” where language is imprecise and vague. So, “relevant” utterances are on the “determinate pure telling” end of the continuum; “partially relevant” utterances are on the “semi-determinate telling/showing” mid-point; and “irrelevant” utterances are on the “indeterminate pure showing” end of the scale.

Before applying this relevance-theoretic framework on learner's communication with different chatbots, available literature on relevance theory and its applications together with any linguistic research on any AI chatbot is presented in the following section.

Review of Literature

Even though AI has not been known for a long time, it has a lot of applications among many disciplines. Linguistic theories and frameworks have been deployed to investigate and compare human-AI interactions. For example, Dynel (2023) examined “metapragmatics, metacommunication, metadiscourse and metalanguage” in human-AI interactions through analyzing posts sent to a public ChatGPT subreddit. Findings reveal that engagement with AI enhances users' “meta” practices, which reflects how AI models contribute to users' understanding of the language itself.

Shaarawy (2023) investigated cohesion and coherence in human essays and compared them to essays generated by ChatGPT. The analysis revealed that ChatGPT could produce human-like essays and surpass humans in aspects related to “cohesion, coherence, lexical choices and complex grammatical structures”. The study concluded by giving some recommendations to avoid plagiarism and enhance the learning process in writing, literature and translation courses.

Chen et al. (2024) explored “pragmalinguistic and sociopragmatic competence” of ChatGPT conversations and compared them to human written conversations using “74 speech act scenarios” to test the plausibility of using AI in pragmatic analysis. Results indicated that ChatGPT acted in a human-like manner in most “pragmalinguistic and sociopragmatic features” and outperformed humans in “syntactic diversity” and the use formal language to the extent that human assessors were not able to distinguish between AI generated conversations and human generated ones.

Polli and Sindoni (2024) adopted a “sociosemiotic and critical multimodal approach” to account for “hateful racist memes”

“Communication” in Learners’ Interaction with Chatbots in Undergraduate Linguistic Courses in Egypt .

detected automatically by AI and analyzed critically by humans. The analysis demonstrated that AI focused on only one mode and ignored the other, which resulted in inaccurate identification of hateful memes. Moreover, contextual factors and “culture-sensitive” issues were also overlooked by AI, which also resulted in misinterpretations. The study concluded that “sociosemiotics and multimodal critical discourse analysis” are to be considered when developing any computational model.

Van Poucke (2024) applied “Appraisal theory” to ChatGPT language of “attitude” and “engagement strategies” and compared them to human responses to pinpoint any “ideological bias” in an attempt of use ChatGPT as a “virtual teaching assistant”. Findings showed that ChatGPT consistent use of “objective construction” reflects an authoritative attitude of the chatbot suggesting bias in the data inserted in the trained model, which might affect learners’ beliefs. On the other hand, humans’ responses were “subjective” allowing for “dialogic expansion” and encouraging critical thinking.

AI has also been exploited in academic contexts. For instance, Kong and Liu (2024) investigated abstracts of review articles from scholarly scientific journals and compared them to abstracts generated by ChatGPT through a “genre analysis” approach to examine the “rhetorical moves” used in the abstracts. Results showed that human written abstracts were more complex and more comprehensive, since humans used more moves than those generated by ChatGPT. Lin and Chen (2024) tested the capability of ChatGPT in designing reading multiple-choice questions that were assessed by humans who assured the ability of ChatGPT to be a teaching assistant and help in the learning process. Fathi et al. (2024) examined the influence of AI on learners’ speaking skills and their “willingness to communicate”, and they found out that interaction with AI helped learners develop their speaking skills and motivated them to communicate.

Relevance theory has similarly been subject to ample research. For example, Khalfaoui (2024) argued, through

Hanaa Youssef Shaarawy Ismail

Relevance theory, that the “Tunisian Arabic particle *ha:*” is an explicit “ostensive stimulus” that mainly marks the speaker’s negative attitude. Marcet (2024) discussed the scarcity of “professional development” courses to teach pragmatics; thence she offered a pragmatic course for “pre-service and in-service instructors of L2 Japanese” based on Relevance theory. The “professional development” course enhanced the instructors’ “pragmatic competence” and provided them with the tools needed to understand language pragmatically; thus, fostering their effective communication. Saleh et al. (2024) examined the multimodal aspects in American and British cartoons during the period of Corona virus pandemic through a relevance-theoretic perspective to analyze “explicatures”, “implicated premises” and “implicated conclusions”. The analysis revealed that the ‘cognitive context’ illustrated in “background knowledge” played a pivotal role in inferring the encoded message in the cartoons. Rohan et al. (2021) examined onomatopoeia translation of Japanese manga through an eye-tracking study using the “showing/saying continuum” of relevance theory. Findings demonstrated that the “full-textual substitution” translation strategy attracted the readers’ attention the most, since it has the “full showing and the full saying” aspects, which comply with the aesthetic nature of onomatopoeia in manga. Wałaszewska (2011) examined children’s “overextension and underextension” as related to the concepts of lexical “broadening and narrowing” in relevance theory. “Overextension” turned out to be an outcome of “broadening” and it connected to the “communicative principle of relevance”; while “underextension” is constructed according to lexical “narrowing” and it is elucidated from the “cognitive principle of relevance”.

All that has been presented reflects the significance of interacting with AI and applying relevance theory to different research fields. However, no study to date, to the researcher’s knowledge, has investigated relevant communication with any AI chatbot, the focus of the current study.

“Communication” in Learners’ Interaction with Chatbots in Undergraduate Linguistic Courses in Egypt .

Research Questions

To examine communication between learners and AI chatbots, two research questions were raised; one related to the learners’ prompts and the other related to the chatbots’ responses:

1. How far do learners manifest relevant prompts while communicating with chatbots?
2. To what extent do chatbots provide relevant responses?

To answer these questions, learners’ communication with different chatbots is analyzed in the following section.

Methodology

Data and Procedures

The data of the current study were collected from first-and fourth-year students, as a convenience sample, in two undergraduate linguistic courses taught by the researcher in Fall 2023 in one of the Egyptian private universities. One course was an “Introduction to Linguistics” course taught to first year students and based on Yule’s (2010) *The Study of Language*, and the other was a “Contrastive Analysis between English and Arabic” course taught to fourth year students and based on compiled materials together with Yule’s overview of first and second language acquisition as an Introduction to the course. Phonetic transcription taught in both courses was based on O’Connor’s (1980) *Better English Pronunciation*.

It is worth mentioning that at that time, AI was getting to be known in Egypt, and all AI users were beginners just starting to get acquainted to it with no difference between first-and fourth-year students AI knowledge level. It needs to be highlighted that ChatGPT, one of the most efficient, self-developing LLMs, was not accessible in Egypt at that time, and only those who have acquaintances outside Egypt could access it. So, most students used Poe, Bing, Bard (later Gemini) chatbots as the free AI tools at that time.

After explaining any new lesson in any of the two courses, students were invited to ask any available AI chatbot about any

topic explained in class and to evaluate the information they get against what they have learned in the lesson. This task was allotted marks from the coursework grade to encourage students to do it. Besides applying AI to the courses, the purpose of the task was also to enhance students' higher order skills of evaluating, judging information and criticizing plus enhancing their reading and writing proficiency levels through chatting with AI in English. Students were also taught how to upload their chats with AI (prompts and responses) together with their evaluating comments on the University Learning Management System (LMS).

All students' chats with all chatbots were studied. A content analysis based on the information generated by AI was conducted to categorize the chat logs. Five categories resulted from the content analysis, and one chat log was randomly selected from each category forming five chat logs representing all instances of AI responses (see the Analysis section below). Students' prompts and AI responses in each chat log were examined qualitatively to account for (ir)relevance within the framework of the two relevance continua discussed above in the theoretical background and to find out whether the problem lies in the prompt or in the response. The qualitative analysis was further reinforced by a quantitative analysis using the F1 score formula of (ir)relevance. F1 score is used in machine learning.

The F1 score is ... the harmonic mean of precision and recall.... The value of F1 score lies between 0 and 1 with 1 being a better.... [Precision] is the number of true positive predictions divided by the total number of positive predictions (true positives + false positives) [Recall] is the number of true positive predictions divided by ... true positives + false negatives. ("F1 score," 2023a)

To quantitatively evaluate (ir)relevance of prompts and responses in the communication between learners and AI chatbots, the Precision and Recall formula were applied to (ir)relevance with the help of ChatGPT 3.5 to be as follows:

“Communication” in Learners’ Interaction with Chatbots in Undergraduate Linguistic Courses in Egypt .

"Precision

$$= \frac{\text{True Positives (Relevant Responses)}}{\text{True Positives (Relevant Responses) + False Positives (Irrelevant Responses)}}$$

$$= \frac{\text{True Positives (Relevant Responses)}}{\text{True Positives (Relevant Responses) + False Negatives (Missed Relevant Responses)}}$$

(ChatGPT, personal communication, September 11, 2023). In the current study, the term “Positive” indicates what is stated, and the term “Negative” signifies what is not stated. So, whenever the prompt or the response is relevant or irrelevant, it is “Positive” because it is stated, and it carries the label “True Positive” when it is relevant and the label “False Positive” when it is irrelevant. If there is any unstated information; the prompt or the response corresponds to “False Negative” in the formula. Consequently, the following F1 score formula becomes comprehensible and applicable to the current study:

$$"F1 = \frac{2}{\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \text{ ("F1 score," 2023a)}$$

F1 score is used in machine learning to evaluate information generated by a model, and it is interpreted as follows: > 0.9 means the model is perfect; between 0.8 and 0.9 means the performance of the model is very good; between 0.5 and below 0.8, the model is average; < 0.5, the model shows poor performance (“F1 score,” 2023b). Therefore, by applying these criteria to relevance, when the F1 score is > 0.9, it means that the information presented is “perfectly relevant”; when it is between 0.8 and 0.9, information is “relevant”; when it is 0.5 and below 0.8, information is “partially relevant”; and when the F1 score is < 0.5, this means that the information is “irrelevant”.

It is worth noting that relevance of AI responses has nothing to do with whether the information provided is correct or incorrect. It has to do with whether the information generated was mentioned in class or not. Consequently, the first step in analyzing each chat was to provide a “Reference Prompt” that should specify the information needed from AI and a “Reference Answer Key” that

should reflect the information mentioned in class. (Ir)relevance was examined and analyzed in both the prompts and the responses and checked against the “Reference Prompt” and the “Reference Answer Key”.

Analysis and Discussion

A content analysis based on the information generated by AI was conducted, as aforementioned, to categorize all the chat logs. Five categories resulted from the content analysis representing the five instances of AI generated responses:

- information more than that mentioned in class
- information more than that mentioned in class with a point missing
- information less than that mentioned in class
- information less than that mentioned in class with some incorrect details
- information highly different from that mentioned in class.

There were no generated responses as same as that mentioned in class, and there were also no responses totally different from that discussed in class. One chat log was randomly selected from each category forming five chat logs representing the data of the current study. In the following section, each chat log is, first, presented, followed by the “Reference Prompt” specifying the information needed from AI and the “Reference Answer Key” delineating the information mentioned in class, then analyzed qualitatively, quantitatively and discussed.

“Communication” in Learners’ Interaction with Chatbots in Undergraduate Linguistic Courses in Egypt .

Chat 1: Information more than that mentioned in class

Could you please tell me what is the transfer of the second language learning?

ChatGPT

In second language acquisition (SLA), transfer refers to the influence of a learner's first language (L1) on the acquisition and use of a second language (L2). It is the process by which linguistic, cognitive, and cultural elements from the L1 are carried over and affect the learning and performance of the L2.

Transfer can occur in various ways:

1. Positive transfer: Positive transfer, also known as facilitative transfer or linguistic transfer, happens when knowledge or skills from the L1 contribute to the acquisition and use of the L2. For example, if the grammar structures or vocabulary of the L1 are similar to the L2, a learner may find it easier to understand and produce those aspects of the L2.
2. Negative transfer: Negative transfer, also known as interference or cross-linguistic influence, occurs when elements of the L1 hinder the acquisition or use of the L2. This can happen when there are differences in grammar, pronunciation, vocabulary, or cultural norms between the two languages. Negative transfer can result in errors or difficulties in language production or comprehension.
3. Developmental transfer: Developmental transfer refers to the influence of the learner's stage of development in the L1 on the acquisition of the L2. Certain language structures or patterns that are acquired at a particular stage in the L1 may affect the acquisition of corresponding structures in the L2. This can be observed in both child and adult language learners.
4. Pragmatic transfer: Pragmatic transfer involves the transfer of sociolinguistic and pragmatic aspects of communication from the L1 to the L2. Cultural norms, politeness strategies, speech acts, and discourse patterns from the L1 can impact a learner's communication in the L2.

It's important to note that transfer is not always negative or problematic. Positive transfer can facilitate language learning, especially when there are similarities between the L1 and L2. However, negative transfer can lead to errors or difficulties in acquiring the L2, and it often requires targeted instruction and practice to overcome.

Understanding the transfer phenomena in SLA is crucial for educators and language learners as it helps identify potential challenges and develop effective teaching strategies that address learners' specific needs. It also highlights the complex interplay between languages and the importance of considering learners' linguistic backgrounds in the language learning process.

Chat 1

Reference Prompt. What is transfer as mentioned in Yule's *The Study of Language*?

Reference Answer Key. Positive transfer and negative transfer only with examples.

Analysis:

Student's Prompt: "Could you please tell me what is the transfer of the second language learning?"

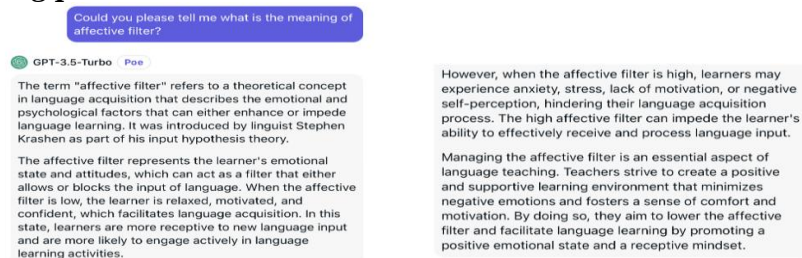
Irrespective of the language inaccuracy of the prompt (AI does not count for language inaccuracies), the student asked about the concept of "transfer in second language learning" without specifying the reference from which the information was to be taken. So, the student did use explicit clear language but implied that there is no other information about the concept than that mentioned in class; thence, leaving away specific details producing a "semi-determinate half telling" or "partially relevant" prompt. This is supported by the F1 score that is calculated according to the ideas mentioned and those not mentioned, i.e. the concept of "transfer in second language learning" is mentioned in the "Student's Prompt" and in the "Reference Prompt"; thus, "True Positives (TP henceforth) / (relevant prompts)" = 1. There are no "False Positives (FP)", since nothing irrelevant was mentioned in the "Student's prompt". "False Negatives (FN) / (missed relevant prompts)" = 1, since the "Student's prompt" did not include the reference. Therefore, the F1 score is 0.66 indicating a "partially relevant prompt". The student might have done so because his/her primary goal was to do the task; consequently, the most important information to focus on, was the concept of "transfer" itself not who presented it, which coincides with Wilson and Sperber's (2012) humans' "cognitive efficiency" that is primarily manifested in their ability to select information worthy of attending to.

AI Response: Positive Transfer with example
Negative Transfer with example
Developmental Transfer
Pragmatic Transfer
Significance of the concept of "transfer" in second language learning.

“Communication” in Learners’ Interaction with Chatbots in Undergraduate Linguistic Courses in Egypt .

Only the first two points correspond to the “Reference Answer Key”. AI added the remaining points, which imply that AI gives general information regardless of any reference, since it has been trained on huge amounts of information from different web sources. In “relevance-theoretic” terms, AI has made use of the “least effort” encoded from the explicit lexicon of the prompt and complemented it with the implied generality to end up generating a general response. Accordingly, if the AI response is to be evaluated according to the unspecific, imprecise student’s prompt, it would be “relevant”, since the chatbot is responding to a very general inquiry and presenting clear, direct, specific, comprehensive information leaving no room for any misunderstanding, but if the AI response is to be evaluated according to the “Reference Answer Key”, it would come up to be “partially irrelevant” to the student, since the information presented in the AI response is somehow vague, with extra information that is not in the “mutually shared cognitive environment” of the student. This is reinforced by the F1 score formula where TP = 2 (the two mentioned concepts), FP = 3 (the three added concepts), and FN = 0 (no missing ideas) resulting in F1 score = 0.57 showing that the response is “partially irrelevant”. Consequently, in this chat log, the problem lies in the prompt. If the prompt had been explicit with no “implicatures” allowing for more than one interpretation, the response would have been clear, precise and direct, i.e. relevant to the student.

Chat 2: Information more than that mentioned in class with a missing point



Could you please tell me what is the meaning of affective filter?

GPT-3.5-Turbo Poe

The term "affective filter" refers to a theoretical concept in language acquisition that describes the emotional and psychological factors that can either enhance or impede language learning. It was introduced by linguist Stephen Krashen as part of his input hypothesis theory.

The affective filter represents the learner's emotional state and attitudes, which can act as a filter that either allows or blocks the input of language. When the affective filter is low, the learner is relaxed, motivated, and confident, which facilitates language acquisition. In this state, learners are more receptive to new language input and are more likely to engage actively in language learning activities.

However, when the affective filter is high, learners may experience anxiety, stress, lack of motivation, or negative self-perception, hindering their language acquisition process. The high affective filter can impede the learner's ability to effectively receive and process language input.

Managing the affective filter is an essential aspect of language teaching. Teachers strive to create a positive and supportive learning environment that minimizes negative emotions and fosters a sense of comfort and motivation. By doing so, they aim to lower the affective filter and facilitate language learning by promoting a positive emotional state and a receptive mindset.

Chat 2

Reference Prompt. What is the affective filter as discussed in Yule's *The Study of Language*?

Reference Answer Key. The affective filter is a second language acquisition barrier.

It has to do with the learners' emotions. Adult learners are so self-conscious that they are afraid of making mistakes and feel embarrassed.

Children are less self-conscious when learning a second language.

Analysis:

Students' Prompt: "Could you please tell me what is the meaning of the affective filter?"

The student asked about the affective filter without specifying the reference from which the information is to be taken. So, like Chat 1, despite the explicit clear language, there is no specificity, which resulted in having a "semi-determinate half telling" or "partially relevant" prompt. This is supported by the F1 score that is calculated according to the ideas mentioned and those not mentioned, i.e. affective filter was mentioned in the "Reference Prompt" and the "Student's Prompt"; thus, TP = 1. There are no FP since nothing irrelevant was mentioned in the "Student's Prompt". FN = 1, since the "Student's Prompt" did not include the reference. Therefore, the F1 score is 0.66 indicating a "partially relevant prompt".

AI Response: Affective filter is an emotional psychological theoretical concept.

The affective filter was introduced by Krashen.

Affective filter may be high or low and the consequences of each in the learning process

Teachers try to manage the affective filter.

Only the first point of AI response coincides with the "Reference Answer Key". AI added the remaining three points to the response and missed two ideas that the affective filter is an acquisition barrier and that it is not a problem for children. Exactly as Chat 1, this response implies that the AI information is general and not

“Communication” in Learners’ Interaction with Chatbots in Undergraduate Linguistic Courses in Egypt .

restricted to a particular reference. However, missing two ideas implies that the model is not comprehensive and needs to be improved. Accordingly, students might evaluate this response as being “irrelevant”, since the information presented in the AI response is vague, unclear with no context or “mutually shared cognitive environment” between them and AI. Hence, TP = 1 (the point mentioned in the AI prompt and complies with one point in the “Reference Answer Key”), FP = 3 (points added in the AI prompt), and FN = 2 (points missing in the AI prompt) resulting in F1 score = 0.28 showing that the response is “irrelevant”. Consequently, in this chat log, the problem again lies in the prompt exactly as Chat 1.

Chat 3: Information less than that mentioned in class

What are the simple short vowels in English phonetics in linguistics?

Assistant Poe

In English phonetics, there are five simple short vowels. These vowels are characterized by their relatively short duration compared to long vowels. Here are the five simple short vowels in English:

1. /æ/ (as in "cat"): This vowel is known as the "short a" sound. It is produced with a relatively low and central tongue position. Examples of words with this vowel include "cat," "hat," and "man."
2. /e/ (as in "bed"): This vowel is known as the "short e" sound. It is produced with a relatively mid-low and central tongue position. Examples of words with this vowel include "bed," "red," and "pen."
3. /ɪ/ (as in "sit"): This vowel is known as the "short i" sound. It is produced with a relatively high and front tongue position. Examples of words with this vowel include "sit," "hit," and "big."
4. /ɒ/ (as in "hot"): This vowel is known as the "short o" sound. It is produced with a relatively mid-back and rounded tongue position. Examples of words with this vowel include "hot," "dog," and "stop."
5. /ʌ/ (as in "cup"): This vowel is known as the "short u" sound. It is produced with a relatively mid-low and central tongue position. Examples of words with this vowel include "cup," "luck," and "fun."

Chat 3

Reference Prompt. What are the simple short vowels in English phonetics? (There is no need to specify the reference in this prompt, since most web sites present their information on phonetics in consistency with O’Connor (1980).

Reference Answer Key. There are seven simple short vowels in English are:

1. /ɪ/ with examples
2. /e/ with examples
3. /æ/ with examples
4. /ʌ/ with examples
5. /ɒ/ with examples
6. /ʊ/ with examples
7. /ə/ with examples

Analysis:

Student's Prompt: What are the simple short vowels in English phonetics?

In this prompt, the student used explicit clear language, specifying simple and short vowels in English language not in any other language, and the student needed a general answer. All the ideas in the student's prompt (simple, short vowels in English) are found in the "Reference Prompt" making the "Student's Prompt" relevant. This is reinforced by the F1 score that is calculated as 1, since TP = 3 denoting the three relevant ideas in the prompt (simple vowels, short vowels, English), FP = 0, since there are no irrelevant ideas, and FN = 0, as there are no missing ideas.

AI Response: There are five simple short vowels in English:

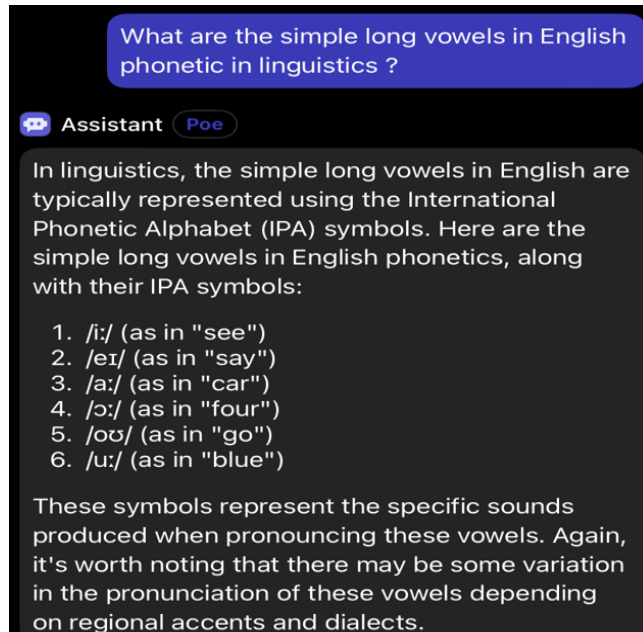
1. /æ/ with examples
2. /ɛ/ or /e/ as presented in the websites with examples
3. /ɪ/ with examples
4. /ɒ/ with examples
5. /ʌ/ with examples

Although the student wrote a relevant prompt, AI missed two simple short vowels. Nonetheless, it presented almost five out of seven vowels (64.29%) of the "Reference Answer Key" (one symbol is not found in the "Reference Answer Key"). Thus, the AI Response is on a high point of "semi-determinate, half telling" towards "relevance". This implies that any information generated by AI needs to be revised, as it might not be comprehensive. This result is emphasized by the F1 score that is 0.78; TP = 4.5, since there are only four relevant items corresponding to the "Prompt Answer Key" + a correct example for /e/ sound; FP = 0.5, since /ɛ/ is not found in

“Communication” in Learners’ Interaction with Chatbots in Undergraduate Linguistic Courses in Egypt .

the “Reference Answer Key”; FN = 2, since two vowels are missing. This F1 score denotes a high level of the “partially relevant” point towards “relevant”.

Chat 4: Information less than that mentioned in class with some wrong details



Chat 4

Reference Prompt. What are the simple long vowels in English phonetics?

Reference Answer Key. There are five simple long vowels in English are:

1. /i:/ with examples
2. /u:/ with examples
3. /ɑ:/ with examples
4. /ɔ:/ with examples
5. /ɜ:/ with examples

Analysis:

Student's Prompt: What are the simple long vowels in English phonetics?

In this prompt, like Chat 3, the student used explicit clear language, specifying simple and long vowels in English language not in any other language, and the student asked about a general answer. All the ideas in the student's prompt (simple, long vowels in English) are found in the "Reference Prompt" making the "Student's Prompt" relevant. This is reinforced by the F1 score that is calculated as 1, since TP = 3 denoting the three relevant ideas in the prompt (simple vowels, long vowels, English), FP = 0, since there are no irrelevant ideas, and FN = 0, as there are no missing ideas.

AI Response:

1. /i:/ with examples
2. /ɔ:/ with examples
3. /u:/ with examples
4. "car" is a correct example including a long simple vowel sound, but the phonetic symbol in the AI response is incorrect. It should be /ɑ:/ not /a:/
5. /ɜ:/ sound is not mentioned by the chatbot
6. "say" and "go" include diphthongs not long vowels: "say" has /eɪ/, and "go" has /əʊ/

AI mentioned almost four out of the five (70%). It presented a wrong phonetic symbol for one vowel sound with a correct example, missed one vowel sound, and included two diphthongs as if they were long simple vowels ("semi-determinate half telling"). So, like Chat 3, this implies that any information generated by AI needs to be revised, as it might be inaccurate and incomprehensive. The F1 score is calculated according to all the ideas generated by AI. TP = 3.5; only three vowels are similar to those mentioned in the "Reference Answer Key" + "car" as a correct example, FP = 2.5; two diphthongs are presented and an incorrect symbol for "car", FN = 1, since AI missed one vowel sound. So, the F1 score is 0.66 signifying a "partially relevant response".

“Communication” in Learners’ Interaction with Chatbots in Undergraduate Linguistic Courses in Egypt .

Chat 5: Information highly different from that mentioned in class

The second language acquisition according to linguistics

Second language acquisition (SLA) is a field of study within linguistics that focuses on how individuals learn a second language, particularly in terms of the cognitive and linguistic processes involved. Linguists study various aspects of SLA, including the acquisition of grammar, vocabulary, pronunciation, and discourse patterns.

According to linguistics, second language acquisition is a complex and multifaceted process that can be influenced by a variety of factors, such as age, motivation, exposure to the language, and individual differences in learner aptitude. Linguists have developed several theories and models to explain how second languages are acquired.

1. Behaviorist Theory: This theory, associated with researchers like B.F. Skinner, emphasizes the role of environmental factors and reinforcement in language learning. It suggests that language learning occurs through imitation, repetition, and positive reinforcement.
2. Innatist Theory: Proposed by Noam Chomsky, the innatist perspective posits that humans are biologically predisposed to acquire language. This theory suggests that there is a universal grammar underlying all languages, and children have an innate ability to acquire language rules and structures.
3. Cognitive Theory: Cognitive theories of SLA focus on the mental processes involved in language learning. They emphasize the role of memory, attention, and problem-solving skills in acquiring a second language. The information-processing approach is one such cognitive theory.
4. Sociocultural Theory: Developed by Lev Vygotsky, sociocultural theory emphasizes the social and cultural aspects of language learning. It suggests that language acquisition is a social process that occurs through interactions with more competent speakers. The theory highlights the importance of cultural context and social interaction in language development.
5. Input Hypothesis: Proposed by linguist Stephen Krashen, the input hypothesis suggests that language learners acquire language through comprehensible input, which is language that is slightly beyond their current level of proficiency. According to this hypothesis, learners progress in their language development when they understand input that contains linguistic features they have not yet mastered.
6. Interactionist Theory: This theory combines elements of the cognitive and sociocultural perspectives and emphasizes the role of both internal cognitive processes and external social interactions in SLA. Interactionist theories highlight the importance of meaningful communication and negotiation of meaning in language learning.

It's important to note that these theories are not mutually exclusive, and researchers often combine elements from different theories to provide a more comprehensive understanding of second language acquisition. The field of SLA is dynamic, and ongoing research continues to refine our understanding of how second languages are acquired.

Chat 5

Reference Prompt. What is second language acquisition according to Yule’s *The Study of Language*?

Reference Answer Key. The response is to cover the concepts presented in Yule’s *The Study of Language*: “Acquisition

barriers, Affective factors, Teaching Methods including the grammar translation method, the audio-lingual method and communicative approaches, Transfer, Interlanguage, Motivation, Negotiated Input and Communicative Competence”

Analysis:

Students’ Prompt: “The second language acquisition according to linguistics”

Although the student did not form a question, he/she explicitly mentioned the concept needed to be explained by AI (“second language acquisition”); however, he/she did not specify the reference from which the information is to be taken. So, like Chats 1 and 2, despite the explicit clear language, there is no specificity, which resulted in having a “semi-determinate half telling” or “partially relevant” prompt. This is supported by the F1 score that is calculated according to the ideas mentioned and those not mentioned, i.e. “second language acquisition” was mentioned in the “Reference Prompt” and the “Student’s Prompt”; thus, TP = 1. There are no FP since nothing irrelevant was mentioned in the “Student’s Prompt”. FN = 1, since the “Student’s Prompt” did not include the reference. Therefore, the F1 score is 0.66 indicating a “partially relevant prompt”.

AI Response: Theories of second language acquisition:

Behaviorist theory (mentions “imitation” and “repetition”), “Innatist” theory, Cognitive theory, Sociocultural theory “social interaction”, Input hypothesis, Interactionist theory (states “meaningful communication” and “negotiation of meaning”)

None of the theoretical terms mentioned in the AI response was stated as the “Response Answer Key”. However, “imitation”, “repetition”, “social interaction”, “meaningful communication” and “negotiation of meaning” were just mentioned without any explanation or examples in the AI Response, and these concepts are discussed with examples in Yule’s (2010) *The Study of Language* when tackling the different “teaching methods”, “negotiated input”

“Communication” in Learners’ Interaction with Chatbots in Undergraduate Linguistic Courses in Egypt .

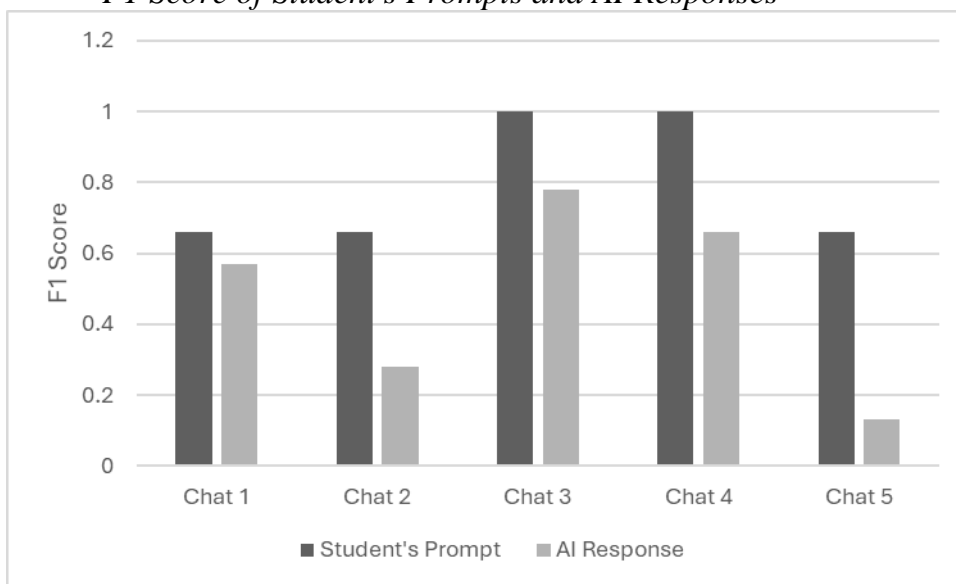
and “communicative competence”. Thus, out of the eight main points in the “Reference Answer Key”, only three sub-points were just stated in the AI Response with no explanation or examples. Thence, this response seemed “irrelevant” to students, since the information presented in the AI response is vague, unclear with no context or “mutually shared cognitive environment” between them and AI. TP of this response = 1 denoting just the mention of the subpoints, FP = 6 (the theories mentioned), FN = 7 (all the missing points in the “Response Answer Key”) resulting in F1 score = 0.13 showing an “irrelevant response”.

The above analysis indicates that the user’s prompts are to lie at the “pure telling/determinate” end of the relevance/irrelevance continuum to activate relevant AI responses where language is clear, direct and explicit leaving no room for any inference or implicature. Having relevant prompts is the essence of yielding relevant AI responses, which complies with the literature that highlighted the importance of training learners on structuring effective prompts to generate the best AI output (Lin & Chen, 2024; Koltovskaia et al., 2024; Yang & Li, 2024). Moreover, all the data generated by AI is to be scrutinized critically by learners, since all LLMs are still developing, which would, in turn, develop learners’ cognitive abilities and enhance the learning process. This lies in the same vein with Garib & Coffelt (2024) and Koltovskaia et al.’s (2024) strategies for students’ effective engagement with AI.

The relation between students’ prompts and AI responses, in the five chats analyzed, is illustrated in Figure 1. Figure 1 reveals that when the student’s prompt is “perfectly relevant”

Figure 1

F1 Score of Student's Prompts and AI Responses



as in Chats 3 and 4, the AI response appears to be more relevant than when the prompt is partially relevant (Chats 1, 2 and 5), which coincides with Wilson and Sperber (2004) principle of “least effort”. Consuming effort is inversely proportional with relevance. In the process of comprehending any prompt, AI “follows a path of least effort” to achieve relevance through lexical decoding; so, because the input given to AI was explicit leaving no room for more than one interpretation, AI achieved relevance easily satisfying the user’s expectations, but when there was some room for inferencing as in Chats 1, 2 and 5, AI would do more search on the internet (exert more effort in the human term) to come out with the most relevant responses possible, which might not meet the user’s expectations.

Figure 1 also demonstrates that whenever AI generated less information than that presented in class (Chats 3 and 4), it yielded more “relevant” responses than when presenting more information than that mentioned in class (Chats 1 and 2). This implies that once AI finds a response that could match the query, it stops searching. It

“Communication” in Learners’ Interaction with Chatbots in Undergraduate Linguistic Courses in Egypt .

generates relevant or mostly relevant information, but it might miss some points not mentioned in the specific web site. This also complies with the “principle of least effort” which says, “stop when your expectations of relevance are satisfied” (Wilson & Sperber, 2004, p. 613).

All that has been presented in the above analysis and discussion responds to the two research questions raised in the study. Whenever the user’s prompts are clear, direct, explicit, specific and to the point, AI responses will be more relevant than when the prompts are implicit carrying more than one interpretation.

Conclusion

The aim of this study was to report a preliminary experience of learners’ interactions with different chatbots and analyze these interactions within the framework of relevance theory. Learners were first- and fourth-year students in two linguistic courses at one of the Egyptian private universities, and the interactions took place in Fall 2023. The analysis revealed that prompts need to fall on the “pure telling determinate” end of the relevance/irrelevance continuum to yield relevant or mostly relevant AI responses. The analysis also indicated that all AI responses need to be revised because AI sometimes provides inaccurate information.

The study, however, has some limitations. The first and foremost limitation is the size of the data. Since the study, as aforementioned, is a preliminary one, only learners’ chats with AI in only two courses were analyzed, and the chat log is only one exchange between the learner and AI. Consequently, if the data size is bigger making a corpus of chat logs that include learners across disciplines, not only linguistics, and that include longer exchanges showing the different strategies learners follow to react to irrelevant AI responses, analyzing the human-AI interaction would be more valid and reliable.

Another limitation is the users’ beginner knowledge level of AI at that time. Learners were just getting to be acquainted with AI,

Hanaa Youssef Shaarawy Ismail

and not all LLMs were accessible to all students. So, it is recommended, after almost a year when learners' knowledge level has much developed, to conduct a further study on users' interactions with different chatbots as they are all becoming available, free of charge to anyone. This could be for the purpose of comparing different chatbots to find out which is more developed, and which is more linguistic oriented.

AI is an assistant that could help with different kinds of tasks. Users are to be aware that it is not human. It is just a tool. So, users are not to expect utmost relevance from AI, since the models are still developing. However, users need to be explicit, clear and specific to get the most relevant information possible.

Acknowledgements

I would like to express my great thanks and appreciation to Engineer Mokhtar Khaled Salah El Din for his great help, support and guidance in all technical and machine learning issues related to the study.

**“Communication” in Learners’ Interaction with Chatbots in
Undergraduate Linguistic Courses in Egypt .**

References

- Chen, X., Li, J., & Ye, Y. (2024). A feasibility study for the application of AI-generated conversations in pragmatic analysis. *Journal of Pragmatics*, 223, 14-30. <https://doi.org/10.1016/j.pragma.2024.01.003>
- Dynel, M. (2023). Lessons in linguistics with ChatGPT: Metapragmatics, metacommunication, metadiscourse and metalanguage in human-AI interactions. *Language & Communication*, 93, 107-124. <https://doi.org/10.1016/j.langcom.2023.09.002>
- Fathi, J., Rahimi, M., & Derakhshan, A. (2024). Improving EFL learners’ speaking skills and willingness to communicate via artificial intelligence-mediated interactions. *System*, 121, 1-17. <https://doi.org/10.1016/j.system.2024.103254>
- F1 score in Machine Learning. (2023a, December 27). Geeksforgeeks. Retrieved August 12, 2024, from <https://www.geeksforgeeks.org/f1-score-in-machine-learning/>
- F1 score in Machine Learning. (2023b, July 18). ENCORD. Retrieved August 12, 2024, from <https://encord.com/blog/f1-score-in-machine-learning/>
- Garib, A., & Coffelt, T. A. (2024). DETECTing the anomalies: Exploring implications of qualitative research in identifying AI-generated text for AI-assisted composition instruction. *Computers and Composition*, 73, 1-12. <https://doi.org/10.1016/j.compcom.2024.102869>
- Johnson, M. (2019). Making meaning manifest. *Croatian Journal of Philosophy*, 29(57), 497-520. <https://doi.org/10.5840/croatjphil201919329>
- Khalfaoui, A. (2024). Vocatives as attitudinal markers: The Tunisian Arabic particle ha:. *Journal of Pragmatics*, 230, 28-40. <https://doi.org/10.1016/j.pragma.2024.06.010>
- Koltovskaia, S., Rahmati, P., & Saeli, H. (2024). Graduate students’ use of ChatGPT for academic text revision: Behavioral, cognitive, and affective engagement. *Journal of Second Language Writing*, 65, 1-15. <https://doi.org/10.1016/j.jslw.2024.101130>
- Kong, X., & Liu, C. (2024). A comparative genre analysis of AI-generated and scholar-written abstracts for English review articles

- in international journals. *Journal of English for Academic Purposes*, 71, 1-14. <https://doi.org/10.1016/j.jeap.2024.101432>
- Lin, Z., & Chen, H. (2024). Investigating the capability of ChatGPT for generating multiple-choice reading comprehension items. *System*, 123, 1-16. <https://doi.org/10.1016/j.system.2024.103344>
- Marcet, E. (2024). Teaching pragmatics to instructors of L2 Japanese: A relevance-theoretic approach. *System*, 126, 1-15. <https://doi.org/10.1016/j.system.2024.103472>
- O'Connor, J. D. (1980). *Better English Pronunciation* (2nd ed.). Cambridge University Press. Retrieved February 13, 2016, from <https://bayanebartar.org/file-dl/library/IELTS5/Better-English-Pronunciation/Better-English-Pronunciation.pdf>
- Polli, C., & Sindoni, M. G. (2024). Multimodal computation or interpretation? Automatic vs. critical understanding of text-image relations in racist memes in English. *Discourse, Context & Media*, 57, 1-10. <https://doi.org/10.1016/j.dcm.2024.100755>
- Rohan, O., Sasamoto, R., & O'Brien, S. (2021). Onomatopoeia: A relevance-based eye-tracking study of digital manga. *Journal of Pragmatics*, 186, 60-72. <https://doi.org/10.1016/j.pragma.2021.09.018>
- Saleh, N., Takyeldin, S. R., & Marzouk, N. A. (2024). A cognitive analysis of selected American and Egyptian Coronavirus cartoons. *مجلة كلية الآداب جامعة الفيوم*, 16(2), 100-131. <https://doi.org/10.21608/jfafu.2023.252687.1997>
- Shaarawy, H. Y. (2023). Cohesion and Coherence in Essays Generated by ChatGPT: A Comparative Analysis to University Students' Writing. *Occasional Papers*, 83(1), 143-165. <https://doi.org/10.21608/OPDE.2023.325331>
- Sperber, D., & Wilson, D. (1987). *Precis of relevance: Communication and cognition*. *Behavioral and Brain Sciences*, 10, 697-710. <https://doi.org/10.1017/S0140525X00055448>
- Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition* (2nd ed.). Blackwell. https://monoskop.org/images/e/e6/Sperber_Dan_Wilson_Deirdre_Relevance_Communicand_Cognition_2nd_edition_1996.pdf
- Sperber, D., & Wilson, D. (2015). Beyond speaker's meaning. *Croatian Journal of Philosophy*, 15(44), 117-149. Retrieved May 5, 2024,

**“Communication” in Learners’ Interaction with Chatbots in
Undergraduate Linguistic Courses in Egypt .**

- from https://www.dan.sperber.fr/wp-content/uploads/2015_wilson_beyond-speakers-meaning.pdf
- Van Poucke, M. (2024). ChatGPT, the perfect virtual teaching assistant? Ideological bias in learner-chatbot interactions. *Computers and Composition* 73, 1-19. <https://doi.org/10.1016/j.compcom.2024.102871>
- Wałaszewska, E. (2011). Broadening and narrowing in lexical development: How relevance theory can account for children’s overextensions and underextensions. *Journal of Pragmatics*, 43, 314-326. <https://doi.org/10.1016/j.pragma.2010.07.017>
- Wilson, D. & Sperber, D. (2004). Relevance theory. In L. R. Horn & G. Wards (Eds.), *The Handbook of Pragmatics* (pp. 607-632). Blackwell. <https://www.felsemiotica.com/descargas/Horn-Laurence-R.-and-Ward-Gregory-Ed.-The-Handbook-of-Pragmatics.pdf>
- Wilson, D. & Sperber, D. (2012). *Meaning and Relevance*. Cambridge University Press. <https://retoricaepragmatica.wordpress.com/wp-content/uploads/2019/02/deirdre-wilson-dan-sperber-meaning-and-relevance-2012-cambridge-university-press.pdf>
- Yang, L. & Li, R. (2024). ChatGPT for L2 learning: Current status and implications. *System*, 124, 1-14. <https://doi.org/10.1016/j.system.2024.103351>
- Yule, G. (2010). *The Study of Language* (4th ed.). Cambridge University Press. Retrieved February 13, 2016, from <https://sharifling.wordpress.com/wp-content/uploads/2018/09/the-study-of-language-george-yule.pdf>