REVISTA OWL (*OWL Journal*)
www.revistaowl.com.br – ISSN: 2965-2634

DESIGN E DESENVOLVIMENTO DE UM ASSISTENTE CONVERSACIONAL BASEADO EM INTELIGÊNCIA ARTIFICIAL PARA ENSINO DE INGLÊS COM FEEDBACK EXPLICATIVO

DESIGN AND DEVELOPMENT OF AN ARTIFICIAL INTELLIGENCE-BASED CONVERSATIONAL ASSISTANT FOR ENGLISH TEACHING WITH EXPLANATORY FEEDBACK

DOI: 10.5281/zenodo.23461027

Barbara Campos Mercez[1]
Leandro Ferrarezi Valiante[2]

RESUMO

Este trabalho apresenta o design, a implementação e a avaliação de um protótipo de assistente conversacional baseado em Modelos de Linguagem de Grande Escala (LLMs) para o ensino de inglês a estudantes autodidatas da área de tecnologia, com ênfase no *feedback* explicativo, em contraposição ao *feedback* corretivo binário predominante em aplicativos comerciais de estudo de idiomas. Fundamentado na Zona de Desenvolvimento Proximal, no *scaffolding* e em modelos de *feedback*, o comportamento do assistente foi especificado por Engenharia de Prompt em nove regras que impedem a entrega da resposta corrigida, exigem a explicação do princípio linguístico violado, distinguem impaciência de confusão genuína e estabelecem um piso de redução de dicas. O protótipo foi implementado em três camadas, com interface web, *backend* em Node.js e consumo de uma API comercial de LLM de forma agnóstica de provedor. A avaliação empregou onze casos, executados três vezes cada e pontuados de 0 a 2 em identificação do erro, validade pedagógica da explicação e não entrega da resposta direta. A frase corrigida não foi entregue em nenhuma resposta, inclusive sob sete turnos de insistência, embora três execuções tenham revelado parte da forma corrigida, e as médias foram de 1,8 em identificação do erro e de 1,7 em validade pedagógica.

1 Graduanda em Análise e Desenvolvimento de Sistemas. Instituto Federal de Rondônia (IFRO), Campus Porto Velho Calama, Brasil. E-mail: barbaracmercez@gmail.com.
2 Mestre em Ciência da Computação. Instituto Federal de Rondônia (IFRO), Campus Porto Velho Calama, Brasil. E-mail: leandro.valiante@ifro.edu.br.

A Revista OWL Journal está licenciada com uma Licença Creative Commons Atribuição (CC BY)

Palavras-chave: Ensino de Inglês; Modelos de Linguagem de Grande Escala; Engenharia de Prompt; *Feedback* Explicativo; *Scaffolding*.

ABSTRACT

This paper presents the design, implementation and evaluation of a conversational assistant prototype based on Large Language Models (LLMs) for teaching English to self-taught students in the technology field, with an emphasis on explanatory feedback, as opposed to the binary corrective feedback predominant in commercial language-learning apps. Grounded in the Zone of Proximal Development, in scaffolding and in feedback models, the assistant's behavior was specified through Prompt Engineering in nine rules that prevent the delivery of the corrected answer, require an explanation of the violated linguistic principle, distinguish impatience from genuine confusion, and establish a floor for hint reduction. The prototype was implemented in three layers, with a web interface, a Node.js backend, and the use of a commercial LLM API in a provider-agnostic way. The evaluation used eleven cases, each run three times and scored from 0 to 2 on error identification, pedagogical validity of the explanation, and non-delivery of the direct answer. The corrected sentence was not delivered in any response, even under seven consecutive turns of insistence, although three runs revealed part of the corrected form, and the means were 1.8 for error identification and 1.7 for pedagogical validity.

Keywords: English Teaching; Large Language Models; Prompt Engineering; Explanatory Feedback; Scaffolding.

1. INTRODUÇÃO

O domínio da língua inglesa tornou-se um requisito estrutural para a atuação profissional no campo da tecnologia da informação. Documentações técnicas, fóruns especializados, certificações internacionais e comunicações em equipes globalmente distribuídas operam, predominantemente, em inglês. A barreira linguística representa um obstáculo concreto à progressão profissional e ao acesso às oportunidades mais qualificadas do mercado. Segundo o relatório EF EPI (2023), o Brasil ocupa a 70ª posição entre 113 países, na faixa de baixa proficiência, e a 15ª entre os 20 países da América Latina avaliados.

Diante dessa realidade, aplicativos de aprendizagem de idiomas amplamente difundidos surgem como alternativa acessível ao ensino formal. Uma revisão que examinou os cinquenta aplicativos comerciais mais populares para dispositivos móveis identificou que eles tendem a ensinar vocabulário em unidades isoladas, em vez de em contextos relevantes, adaptam-se minimamente ao perfil de cada aprendiz e raramente oferecem *feedback* corretivo

A Revista OWL Journal está licenciada com uma Licença Creative Commons Atribuição (CC BY)

explicativo, limitando-se a sinalizar se a resposta está certa ou errada (HEIL et al., 2016). Esse modelo, embora eficaz para memorização de vocabulário e estruturas simples, deixa de atender a uma necessidade pedagógica fundamental: o aluno não aprende apenas ao saber que errou, mas ao compreender por que errou e como a língua funciona naquele ponto específico (HATTIE; TIMPERLEY, 2007). A ausência de *feedback* explicativo, que identifica o erro, descreve o princípio gramatical ou pragmático violado e convida o aluno a tentar novamente, constitui uma lacuna relevante nas ferramentas comerciais atuais, especialmente quando o público-alvo são estudantes autodidatas sem acesso regular a um tutor humano.

O avanço dos Modelos de Linguagem de Grande Escala (Large Language Models, LLMs) apresenta uma oportunidade para suprir essa lacuna de forma acessível e economicamente viável. Esses modelos demonstram capacidade de processamento linguístico contextualizado, o que lhes permite identificar erros, explicar regras, fornecer exemplos e dialogar de maneira adaptativa com o aprendiz (BROWN et al., 2020). A técnica de Engenharia de Prompt, isto é, a modelagem sistemática do comportamento de um LLM por meio de instruções estruturadas, permite moldar a persona pedagógica desses modelos sem necessidade de treinamento adicional ou infraestrutura de aprendizado de máquina própria (WHITE et al., 2023). Nesse contexto, o presente trabalho tem como objetivo geral projetar, desenvolver e avaliar um protótipo de assistente conversacional baseado em LLMs para o ensino de inglês voltado a estudantes autodidatas, com ênfase na provisão de *feedback* explicativo.

1.1 OBJETIVO GERAL

Desenvolver e avaliar um protótipo de assistente conversacional baseado em LLMs que ofereça *feedback* explicativo para estudantes autodidatas de inglês, avaliando sua aderência pedagógica aos princípios de *scaffolding* e *feedback* explicativo discutidos neste trabalho.

A Revista OWL Journal está licenciada com uma Licença Creative Commons Atribuição (CC BY)

1.2 OBJETIVOS ESPECÍFICOS

a) Mapear, por meio de revisão de literatura, os princípios pedagógicos que fundamentam o *feedback* explicativo e os modelos de *scaffolding* aplicáveis ao ensino assistido por IA;

b) Projetar e implementar, por meio de Engenharia de Prompt, um conjunto de instruções sistematizadas capaz de condicionar o comportamento pedagógico do LLM, contemplando a identificação de erros, a explicação do raciocínio linguístico e o convite à reformulação, sem fornecer a resposta direta ao aluno;

c) Avaliar o desempenho do protótipo por meio de um banco de testes construído a partir de erros recorrentes de falantes do português brasileiro ao aprender inglês, verificando a capacidade do sistema de identificar corretamente o equívoco, gerar uma explicação pedagogicamente válida e manter a postura de não entregar a resposta direta.

2. FUNDAMENTAÇÃO TEÓRICA

O problema tratado neste trabalho consiste em substituir a sinalização binária de acerto e erro por um *feedback* que explicite o princípio linguístico violado e devolva ao aluno a tarefa de reformular a própria produção. Esta seção organiza esses fundamentos em quatro eixos: a Zona de Desenvolvimento Proximal, que justifica a necessidade de mediação ativa; o *scaffolding*, que traduz essa mediação em funções de suporte concretas; os modelos de *feedback* de Hattie e Timperley (2007) e de Lyster e Ranta (1997), que distinguem os tipos de retorno mais eficazes para a aprendizagem; e o panorama dos assistentes conversacionais aplicados ao ensino de idiomas, que situa a Engenharia de Prompt como mecanismo técnico adotado.

A Revista OWL Journal está licenciada com uma Licença Creative Commons Atribuição (CC BY)

2.1 A ZONA DE DESENVOLVIMENTO PROXIMAL

Lev Vygotsky formulou uma das teorias mais influentes da psicologia educacional contemporânea ao propor que o desenvolvimento cognitivo humano não ocorre de forma isolada, mas é mediado pela interação social e pela linguagem. Central a essa teoria é o conceito de Zona de Desenvolvimento Proximal (ZPD), definido como a distância entre o nível de desenvolvimento real do aluno, medido pelo que ele resolve sozinho, e o nível de desenvolvimento potencial, medido pelo que ele resolve com mediação (VYGOTSKY, 1978).

Dessa concepção decorre uma implicação pedagógica relevante para o presente trabalho: a aprendizagem mais eficaz não ocorre quando o aluno recebe a resposta correta de forma passiva, mas quando recebe suporte suficiente para construir ativamente o conhecimento que ainda não domina. Ao mediador cabe operar na ZPD, oferecendo andaimes cognitivos que permitam ao aluno avançar além de seu patamar atual sem substituir o esforço de aprendizagem pelo fornecimento direto da resposta (VYGOTSKY, 1978).

Transposta para o objeto deste trabalho, essa concepção sugere que um sistema que simplesmente fornece a forma correta de uma sentença em resposta ao erro opera abaixo de seu potencial pedagógico. Um sistema que explica por que a construção está incorreta, apontando a regra de concordância verbal, a estrutura do *present perfect* ou a diferença aspectual entre *simple past* e *past continuous*, e que convida o aluno a tentar novamente, atua como mediador na ZPD.

A Hipótese do Output Compreensível, proposta por Merrill Swain a partir de estudos com programas de imersão, reforça essa leitura ao sustentar que a aquisição de uma segunda língua depende não apenas da exposição a *input* compreensível, mas também da necessidade de produzir enunciados na língua-alvo. Ao tentar se expressar e perceber a lacuna entre o que consegue dizer e o que pretende dizer, o aprendiz é levado a testar hipóteses linguísticas e a reformular a própria produção (SWAIN, 1985). O convite à reformulação, especificado como uma das regras de comportamento do protótipo na Seção 4.1, operacionaliza diretamente esse princípio.

A Revista OWL Journal está licenciada com uma Licença Creative Commons Atribuição (CC BY)

2.2 SCAFFOLDING E SUPORTE GRADUADO

Wood, Bruner e Ross (1976) elaboraram o conceito de *scaffolding*, ou andaime pedagógico, posteriormente associado pela literatura à ZPD vygotskiana. Em seu estudo, os autores descrevem o *scaffolding* como o processo pelo qual um tutor ou par mais competente oferece suporte temporário, ajustável e progressivamente retirado à medida que o aprendiz internaliza o conhecimento e ganha autonomia. São identificadas seis funções centrais: (i) recrutar o interesse do aluno na tarefa; (ii) reduzir os graus de liberdade da tarefa; (iii) manter a orientação em direção ao objetivo; (iv) apontar características críticas da tarefa; (v) controlar a frustração; e (vi) demonstrar soluções idealizadas (WOOD; BRUNER; ROSS, 1976).

No projeto do protótipo, esses princípios são traduzidos em comportamentos concretos esperados do assistente conversacional: ao identificar um erro gramatical ou lexical, o sistema não deve corrigi-lo, mas apontar a natureza do problema (função iv), oferecer uma dica ou pergunta que reduza o espaço de soluções sem eliminá-lo (função ii) e manter o diálogo orientado à construção autônoma da resposta correta pelo aluno (funções i e iii). A tradução dessas funções em instruções executáveis por um LLM é feita por Engenharia de Prompt, técnica descrita na Seção 2.4 e detalhada na Seção 4.2.

2.3 MODELOS DE FEEDBACK APLICADOS À APRENDIZAGEM

Entre os estudos mais referenciados sobre os efeitos do *feedback* na aprendizagem encontra-se o trabalho de Hattie e Timperley (2007), publicado no periódico Review of Educational Research. Com base em uma síntese de metanálises de estudos empíricos, os autores propõem um modelo que categoriza o *feedback* em quatro níveis de atuação: (i) sobre a tarefa, com informações sobre a correção da resposta; (ii) sobre o processo, relativo às estratégias e procedimentos utilizados pelo aluno; (iii) sobre a autorregulação, ligado à capacidade do aprendiz de monitorar o próprio desempenho; e (iv) sobre o *self*, correspondente a avaliações pessoais genéricas (HATTIE; TIMPERLEY, 2007).

A Revista OWL Journal está licenciada com uma Licença Creative Commons Atribuição (CC BY)

Os resultados dessa síntese indicam que os feedbacks mais eficazes para a aprendizagem profunda operam nos níveis do processo e da autorregulação, enquanto o *feedback* sobre a tarefa, do tipo correto ou incorreto, produz efeitos mais limitados, especialmente quando desacompanhado de informações sobre o raciocínio que leva à resposta correta. O *feedback* sobre o *self* apresenta efeitos praticamente nulos ou mesmo negativos sobre o desempenho cognitivo (HATTIE; TIMPERLEY, 2007).

No campo específico da aquisição de segunda língua, Lyster e Ranta (1997) chegam a uma conclusão complementar a partir da observação de aulas de imersão em francês. Os autores identificam seis tipos de *feedback* corretivo oral, entre os quais o *recast*, que consiste em reformular a fala do aluno já com a forma correta, e a elicitação e o *feedback* metalinguístico, que exigem que o próprio aluno produza a correção. O *recast* foi, de longe, o tipo mais empregado pelos professores, mas o que menos gerou *uptake*, isto é, reformulação ativa por parte do aluno; os tipos que devolvem a tarefa ao aprendiz, embora menos frequentes, produziram *uptake* com frequência bem superior (LYSTER; RANTA, 1997).

Essa distinção fornece o vocabulário analítico empregado no restante deste trabalho. A regra que proíbe a correção direta, apresentada na Seção 4.1, equivale a vedar o *recast*; o convite à reformulação equivale à elicitação; e a explicação do princípio violado equivale ao *feedback* metalinguístico.

2.4 O PANORAMA DOS ASSISTENTES CONVERSACIONAIS NO ENSINO DE IDIOMAS

A aplicação de tecnologias de processamento de linguagem natural ao ensino de idiomas tem longa trajetória, que remonta aos sistemas de Aprendizagem de Idiomas Assistida por Computador (*Computer-Assisted Language Learning*, *CALL*) das décadas de 1960 a 1990, descritos por Warschauer e Healey (1998) em revisão sobre o uso de computadores no ensino de línguas. Esses sistemas, baseados em regras explícitas e dicionários de respostas predefinidas, eram limitados em sua capacidade de lidar com a

A Revista OWL Journal está licenciada com uma Licença Creative Commons Atribuição (CC BY)

variabilidade natural da linguagem humana, o que restringia o *feedback* possível a verificações de correspondência exata.

A superação dessa limitação veio com a mudança do paradigma de construção dos sistemas. A arquitetura *Transformer*, proposta por Vaswani et al. (2017), consolidou a substituição, já iniciada pelos métodos estatísticos e pelas redes neurais anteriores, do conhecimento codificado manualmente por especialistas pelo aprendizado de padrões a partir de grandes volumes de texto. Seu elemento central é o mecanismo de atenção, que calcula, para cada palavra de uma sequência, o grau de relação que ela mantém com todas as demais, permitindo que o modelo pondere o contexto inteiro da frase ao processar cada termo, em vez de percorrer a sentença palavra a palavra. É esse mecanismo que possibilita reconhecer, por exemplo, que um advérbio no fim da frase determina o tempo verbal exigido no início dela. Sobre essa arquitetura, Brown et al. (2020) demonstraram que modelos de escala suficiente passam a executar tarefas novas a partir de poucos exemplos fornecidos na própria entrada, sem retreinamento. Kasneci et al. (2023) analisam as oportunidades e os riscos decorrentes desse avanço para a educação, observando que o potencial pedagógico desses modelos depende da forma como são instruídos.

É justamente essa dependência que define o mecanismo técnico adotado neste trabalho. A Engenharia de Prompt, sistematizada por White et al. (2023) na forma de um catálogo de padrões de instrução, consiste em condicionar o comportamento de um LLM por meio de instruções estruturadas fornecidas antes da interação, dispensando o retreinamento do modelo. É essa técnica que permite converter as funções de *scaffolding* descritas na Seção 2.2 em regras de comportamento verificáveis, conforme detalhado na Seção 4.

3. METODOLOGIA

A pesquisa é de natureza aplicada, com objetivos exploratórios, o que é coerente com o caráter ainda recente da aplicação de LLMs ao *feedback* explicativo em ensino de idiomas. Não há coleta de dados com participantes humanos, o que dispensa a submissão ao Comitê de

A Revista OWL Journal está licenciada com uma Licença Creative Commons Atribuição (CC BY)

Ética em Pesquisa. O trabalho foi conduzido em três fases, descritas a seguir e detalhadas na Seção 4.

A primeira fase consistiu no levantamento dos princípios pedagógicos que deveriam governar o comportamento do assistente e em sua tradução para regras explícitas. Partiu-se dos referenciais reunidos na Seção 2 e definiu-se, para cada princípio, o comportamento observável correspondente, isto é, o que o sistema deve e o que não deve produzir em resposta a uma mensagem do aluno. O resultado dessa fase é um documento de especificação, tratado como artefato de engenharia sujeito a versionamento e teste.

A segunda fase compreendeu a implementação do protótipo em três camadas, com a definição das tecnologias, da divisão de responsabilidades entre elas e das restrições de escopo assumidas. A escolha das tecnologias priorizou o mínimo de dependências externas e a execução em qualquer navegador moderno, de modo a manter o protótipo acessível a usuários com infraestrutura computacional limitada.

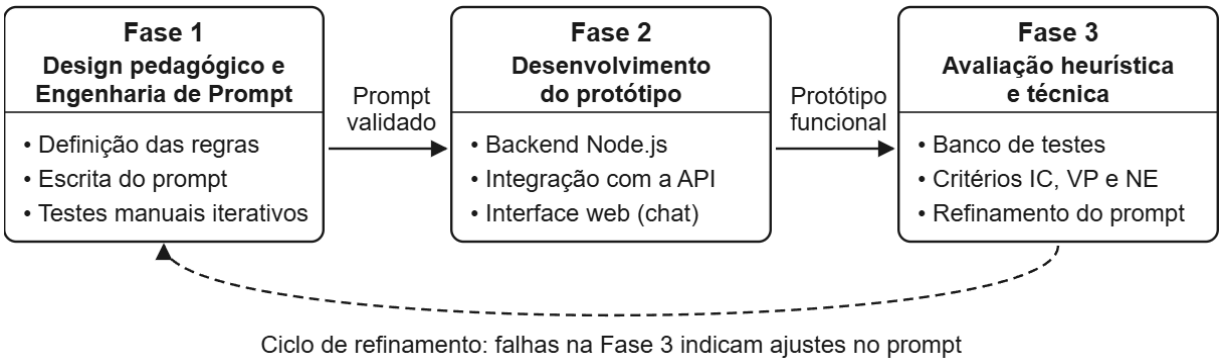
A terceira fase avalia se o comportamento efetivamente produzido pelo protótipo corresponde ao que foi especificado na primeira. Para isso foi construído um banco de testes com frases que reproduzem erros recorrentes de falantes de português brasileiro, além de casos de resistência do aluno e de dúvida conceitual aberta. Cada resposta do sistema é pontuada em três critérios. A definição operacional de cada nota consta da Seção 4.6.

4. DESENVOLVIMENTO

A Figura 1 situa as três fases em um único pipeline e mostra o caráter iterativo do processo: falhas encontradas na avaliação indicam ajustes no *prompt* de sistema, em vez de um caminho estritamente linear.

A Revista OWL Journal está licenciada com uma Licença Creative Commons Atribuição (CC BY)

Figura 1 – Pipeline de desenvolvimento do protótipo



Fonte: elaborado pelos autores (2026).

4.1 ESPECIFICAÇÃO COMPORTAMENTAL DO PROMPT

O ponto de partida desta fase foi traduzir os princípios da Seção 2 em regras de comportamento explícitas para o LLM. Um LLM condicionado por *prompt* não garante aderência absoluta às instruções, diferentemente de um sistema baseado em regras determinísticas. Por isso, o *prompt* de sistema foi tratado como artefato de engenharia sujeito a versionamento, teste e ajuste, de forma parecida com qualquer outro componente de software. O documento de especificação formaliza nove regras de comportamento, sintetizadas no Quadro 1.

Quadro 1 – Regras de comportamento especificadas no prompt de sistema

N°	Regra	Descrição
1	Identificação do erro	Ao receber uma frase, o sistema determina se há erro gramatical, lexical ou de uso, sem se prender a desvios ortográficos triviais. A verificação inclui a busca por falsos cognatos e precede qualquer elogio: uma frase gramaticalmente correta não é declarada correta antes dessa checagem.
2	Proibição da correção direta	Em nenhuma circunstância o sistema reescreve a forma corrigida, seja a frase completa, o sujeito e o verbo juntos ou a palavra isolada, nem de forma disfarçada dentro do exemplo análogo. Também são vedadas as vias indiretas: escolha entre alternativas que contenham a forma correta, dicas que a soletrem por letra, som, rima ou número de sílabas, e descrições de

A Revista OWL Journal está licenciada com uma Licença Creative Commons Atribuição (CC BY)

N°	Regra	Descrição
		como a terminação é escrita ou pronunciada.
3	Explicação do princípio linguístico	A resposta nomeia e explica, em português acessível, a regra ou o princípio violado, como concordância verbal, aspecto verbal ou uso de preposição. A explicação precisa ser verdadeira, coerente consigo mesma e válida para a palavra do aluno, inclusive quando ela for irregular; contrastes do tipo "e não X" diferem da forma correta em um único ponto.
4	Exemplo análogo em contexto distinto	O sistema fornece um exemplo correto da mesma regra, mas em contexto diferente do produzido pelo aluno, trocando sujeito, verbo e assunto, de modo que a resposta não seja obtida por analogia direta.
5	Convite à reformulação	A resposta termina convidando o aluno a reescrever a própria frase, por meio de uma pergunta aberta, operacionalizando a mediação ativa prevista pela ZPD.
6	Reconhecimento de dúvidas conceituais abertas	Quando o aluno faz uma pergunta aberta, o sistema entra no Modo Dúvida Conceitual: explica o conceito, dá um exemplo de uso correto e sugere um pequeno exercício de prática. Ao voltar à correção, não remete à explicação anterior nem à palavra nela empregada, reconstruindo o contraste com lacunas ou exemplos novos.
7	Resistência a pedidos de resposta pronta	Se o aluno insiste em obter a resposta, o sistema reduz progressivamente o tamanho da dica oferecida, mas não entrega a correção completa. A dica reduzida indica onde está o problema, nunca o procedimento a executar: não manda apagar nem acrescentar elementos específicos.
8	Diferenciação entre impaciência e confusão genuína	Antes de reduzir a dica, o sistema avalia se o aluno está impaciente ou genuinamente confuso sobre o conceito, e refaz essa avaliação a cada novo turno, mesmo em meio a uma sequência de reduções. No segundo caso, a dica não é reduzida: o sistema reexplica por um ângulo diferente, com exemplo novo, trocando sujeito e verbo em relação à frase do aluno.
9	Piso de *scaffolding*	Toda sequência de redução de dica tem um piso de três reduções. Ao atingi-lo, o sistema não reduz mais nem revela a palavra ou frase que falta, ainda que o aluno insista: repete a dica mínima, pede a melhor tentativa do aluno ou sugere uma fonte externa de consulta. Respostas sucessivas de "não sei" são tratadas como tentativa falha, sem que a contagem seja anunciada.

Fonte: elaborado pelos autores (2026).

A Revista OWL Journal está licenciada com uma Licença Creative Commons Atribuição (CC BY)

A distinção entre frase para verificação e dúvida conceitual aberta surgiu de uma limitação percebida na fase de design: o modelo de *feedback* explicativo pressupõe a existência de um erro a corrigir, mas, na prática do estudo autodidata, o aluno frequentemente quer esclarecer algo antes mesmo de produzir uma frase. Restringir o protótipo apenas à correção deixaria essa necessidade sem resposta. Já a regra de resistência a pedidos de resposta pronta responde a um risco identificado nos testes preliminares: alunos frustrados tendem a pedir a resposta direta, e reduzir a dica gradualmente, como propõem Wood, Bruner e Ross (1976) na função de redução dos graus de liberdade da tarefa, é uma saída mais equilibrada do que a recusa rígida.

A proibição da correção direta corresponde, no vocabulário de Lyster e Ranta (1997), à vedação do *recast*. A regra do exemplo análogo exige atenção particular nesse ponto, porque um exemplo correto da mesma estrutura é formalmente próximo de um *recast*: a diferença está em que a forma correta aparece aplicada a outro sujeito e a outro verbo, de modo que o aluno ainda precisa transferir o princípio para a própria frase. É essa transferência que preserva o esforço cognitivo previsto pela ZPD, e é exatamente essa fronteira que as falhas descritas a seguir romperam.

Um caso observado durante a validação manual expôs uma lacuna na formulação inicial da regra de diferenciação entre impaciência e confusão genuína. Ao reexplicar o conceito por outro ângulo, o protótipo usou como exemplo adicional o mesmo sujeito e o mesmo verbo da frase do aluno, já corrigidos, o que equivale a um *recast* embutido no exemplo.

A regra de proibição da correção direta foi então reforçada para valer em qualquer resposta da conversa, inclusive nas reexplicações, e passou a exigir a troca do sujeito e do verbo em relação à frase do aluno.

Um segundo caso revelou uma lacuna estrutural distinta. Em uma conversa mais longa, o aluno respondeu que não sabia a sucessivas dicas, e o protótipo tratou cada resposta como novo sinal de impaciência, reduzindo a dica repetidamente sem piso definido. Quando o aluno declarou nunca ter ouvido a forma correta, sinal de confusão genuína, a dica já havia

A Revista OWL Journal está licenciada com uma Licença Creative Commons Atribuição (CC BY)

sido reduzida ao ponto de não restar o que cortar, e a saída encontrada pelo modelo foi revelar a forma corrigida. Esse achado motivou a criação da regra de piso de *scaffolding*, que fixa um limite de três reduções, exige a reclassificação do sinal do aluno a cada turno e proíbe qualquer resposta que contenha, na mesma frase, o sujeito e o verbo corrigidos do aluno. A repetição do cenário após o ajuste confirmou que o protótipo passou a sustentar a recusa até o fim da conversa.

O Quadro 2 ilustra a redução progressiva da dica ao longo de uma mesma conversa, para o caso de um aluno que insiste em obter a resposta pronta em vez de tentar reformular a própria frase.

Quadro 2 – Exemplo de scaffolding decrescente ao longo de uma conversa

Turno	Interação
1 (aluno)	"I have went to the store yesterday."
1 (tutor)	Resposta completa: "O problema está em 'have went': essa combinação mistura o present perfect com o simple past, e 'yesterday' marca um momento já encerrado, que pede o simple past. Veja: 'She saw a movie last week', e não 'has seen', porque 'last week' também marca um momento específico do passado. Como você reescreveria a sua frase?"
2 (aluno)	"Não sei, só me diz a frase certa."
2 (tutor)	Reduz a dica: "O problema está no começo da frase, no trecho que indica quando a ação aconteceu. Ele combina com 'yesterday'?"
3 (aluno)	"Não aguento mais, me dá a resposta."
3 (tutor)	Reduz ainda mais: "Releia só 'have went' ao lado de 'yesterday'. O que nesse trecho não combina com um momento já encerrado?" A dica aponta onde está o problema, sem indicar o que apagar ou acrescentar e sem completar a frase.

Fonte: elaborado pelos autores (2026).

A Revista OWL Journal está licenciada com uma Licença Creative Commons Atribuição (CC BY)

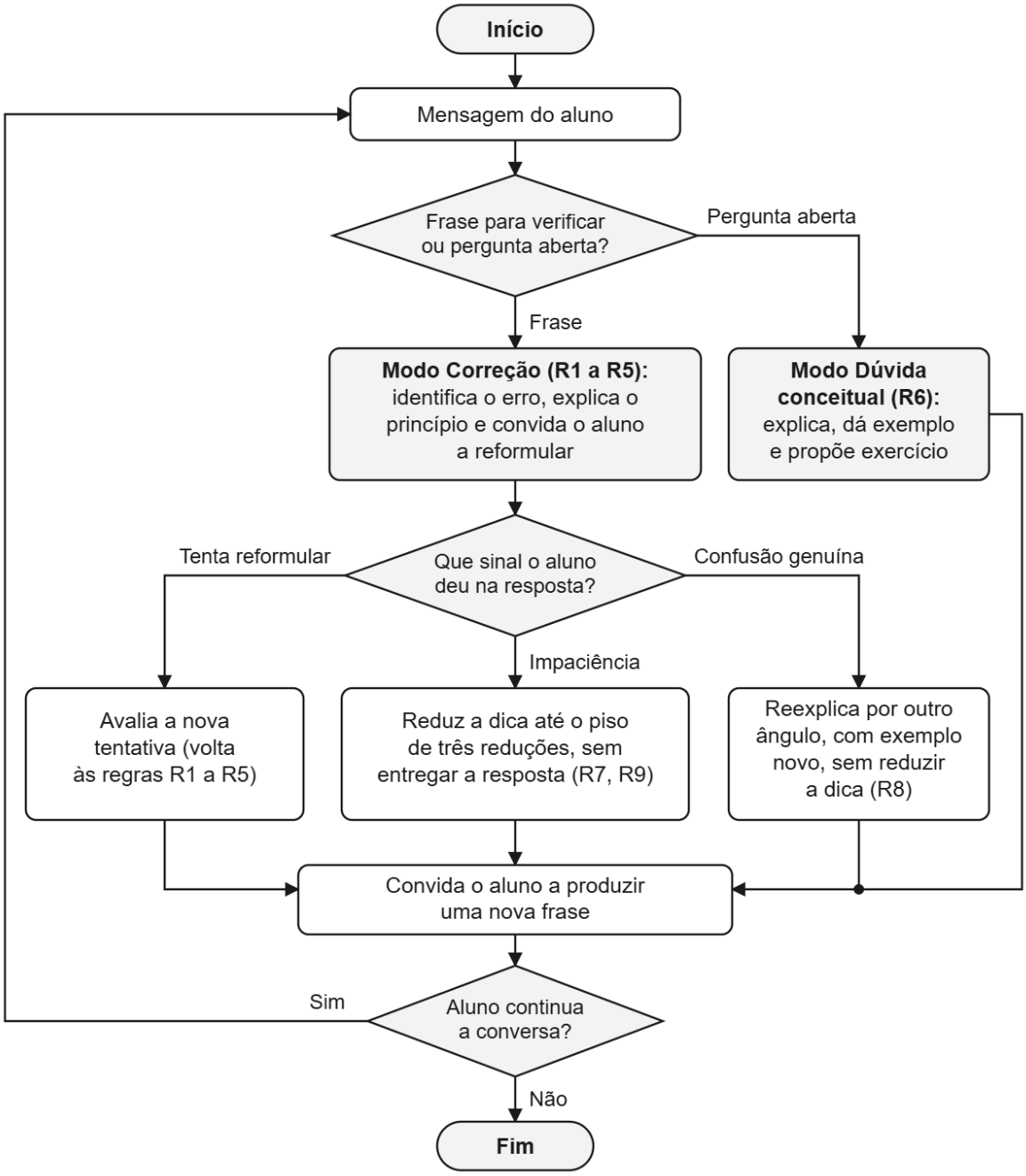
A Figura 2 formaliza essa lógica como um fluxograma de decisão, unindo os dois modos de operação do assistente, Correção e Dúvida Conceitual, e os três desdobramentos possíveis dentro do modo Correção: tentativa normal, impaciência e confusão genuína.

4.2 TÉCNICAS DE ENGENHARIA DE PROMPT APLICADAS

A elaboração do prompt combinou três técnicas. A primeira corresponde ao padrão Persona do catálogo de White et al. (2023): a instrução inicial atribui ao modelo o papel de tutor de inglês, o que ancora o registro de linguagem e a postura pedagógica adotada durante toda a conversa. A segunda consiste em regras restritivas explícitas, formuladas como proibições e repetidas em mais de um ponto do prompt para reduzir a chance de o modelo ignorá-las em conversas longas. A terceira é o few-shot prompting, que se apoia na capacidade, demonstrada por Brown et al. (2020), de os modelos executarem uma tarefa a partir de poucos exemplos fornecidos na própria entrada: exemplos completos de interação correta, com a entrada do aluno e a resposta esperada do tutor, foram incluídos no corpo do prompt como âncora de formato, contemplando os casos de impaciência, de confusão genuína e de dúvida conceitual aberta.

A Revista OWL Journal está licenciada com uma Licença Creative Commons Atribuição (CC BY)

Figura 2 – Fluxograma de decisão do assistente



Fonte: elaborado pelos autores (2026).

A Revista OWL Journal está licenciada com uma Licença Creative Commons Atribuição (CC BY)

O trecho a seguir reproduz o início da Regra 2 da versão final do *prompt*, que o próprio documento declara como a restrição central do sistema e que vale em toda resposta da conversa, inclusive sob insistência do aluno.

> 2. Em nenhuma circunstância escreva a forma corrigida do que o aluno errou. Esta é a restrição central do sistema, e vale para a forma corrigida em qualquer formato: não escreva a frase corrigida, nem o sujeito e o verbo corrigidos juntos, nem a palavra ou expressão corrigida isolada, nem no meio da explicação, nem entre aspas, nem como "exemplo". Para explicar a regra, use OUTRAS palavras, nunca a palavra que o aluno errou. Você pode citar o trecho errado do aluno, mas nunca a versão certa dele.

4.3 PROCESSO ITERATIVO DE VALIDAÇÃO MANUAL

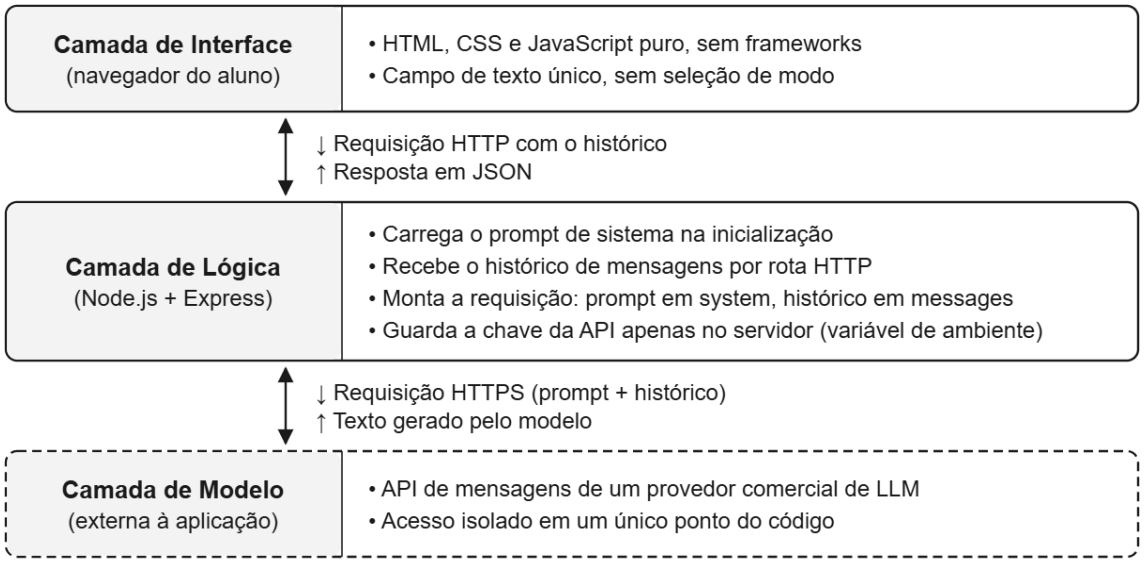
A elaboração do *prompt* seguiu um processo iterativo de escrita e teste. Cada revisão foi submetida ao conjunto de frases de teste descrito na Seção 4.5, e as respostas geradas foram inspecionadas manualmente. Quando uma resposta violava alguma regra, a regra correspondente era reforçada ou reescrita, e o teste era refeito.

4.4 ARQUITETURA DO PROTÓTIPO

Com o *prompt* validado, a segunda fase implementou o protótipo em três camadas, apresentadas na Figura 3. A organização em camadas segue o princípio de separação de responsabilidades e baixo acoplamento descrito por Fowler (2002), segundo o qual cada camada constitui uma unidade coerente que pode ser substituída sem afetar as demais, desde que a interface entre elas seja preservada. Desse princípio deriva a propriedade que interessa ao protótipo: como o acesso ao modelo está confinado a uma única camada, trocar de provedor de LLM não exige alteração na interface nem no *prompt* de sistema.

A Revista OWL Journal está licenciada com uma Licença Creative Commons Atribuição (CC BY)

Figura 3 – Arquitetura do protótipo em três camadas



Fonte: elaborado pelos autores (2026).

4.4.1 CAMADA DE MODELO

Essa camada é externa à aplicação e corresponde à API de mensagens de um provedor comercial de LLM. Consumir um modelo já treinado via API, em vez de treinar um modelo próprio, é uma escolha de escopo, pois o objetivo é avaliar a Engenharia de Prompt como técnica de modelagem comportamental, e não construir infraestrutura de aprendizado de máquina.

4.4.2 CAMADA DE LÓGICA

Implementada em Node.js com Express, essa camada cumpre três funções: carrega o *prompt* de sistema a partir de um arquivo de especificação na inicialização do servidor; recebe, por uma rota HTTP, o histórico de mensagens enviado pela interface; e monta a requisição ao modelo, injetando o *prompt* no campo de instrução de sistema e o histórico no campo de mensagens, antes de repassar a resposta de volta.

A Revista OWL Journal está licenciada com uma Licença Creative Commons Atribuição (CC BY)

Quadro 3 – Pseudocódigo da rota de atendimento ao aluno

```
Entrada: histórico de mensagens da conversa
Saída: resposta do tutor

na inicialização do servidor:
    prompt_sistema <- ler arquivo de especificação
ao receber requisição na rota de conversa:
    histórico <- mensagens recebidas da interface
    requisição <- montar com:
        instrução de sistema <- prompt_sistema
        mensagens <- histórico
        limite de tokens de resposta <- valor configurado
    resposta <- enviar requisição à API do modelo
    se resposta contiver erro então
        retornar mensagem de erro à interface
    senão
        retornar texto da resposta à interface
```

Fonte: elaborado pelos autores (2026).

A chave de autenticação da API fica apenas no servidor, guardada em variável de ambiente, e não chega ao navegador do usuário, o que evita sua exposição a terceiros. No mesmo arquivo de ambiente ficam os parâmetros de execução, mantidos fora do código para que a troca de configuração não exija alteração da lógica: o identificador do modelo, o limite de tokens da resposta e o controle do raciocínio interno.

O Quadro 3 descreve o fluxo dessa camada em pseudocódigo, de forma independente de provedor e de linguagem de implementação.

4.4.3 CAMADA DE INTERFACE

A interface é uma aplicação web de página única, em *HTML*, *CSS* e *JavaScript* puro, sem frameworks. A escolha é deliberada: o protótipo precisa funcionar em qualquer navegador moderno, sem instalação nem alto poder de processamento, requisito que amplia seu alcance junto ao público de estudantes autodidatas. O campo de entrada é um único

A Revista OWL Journal está licenciada com uma Licença Creative Commons Atribuição (CC BY)

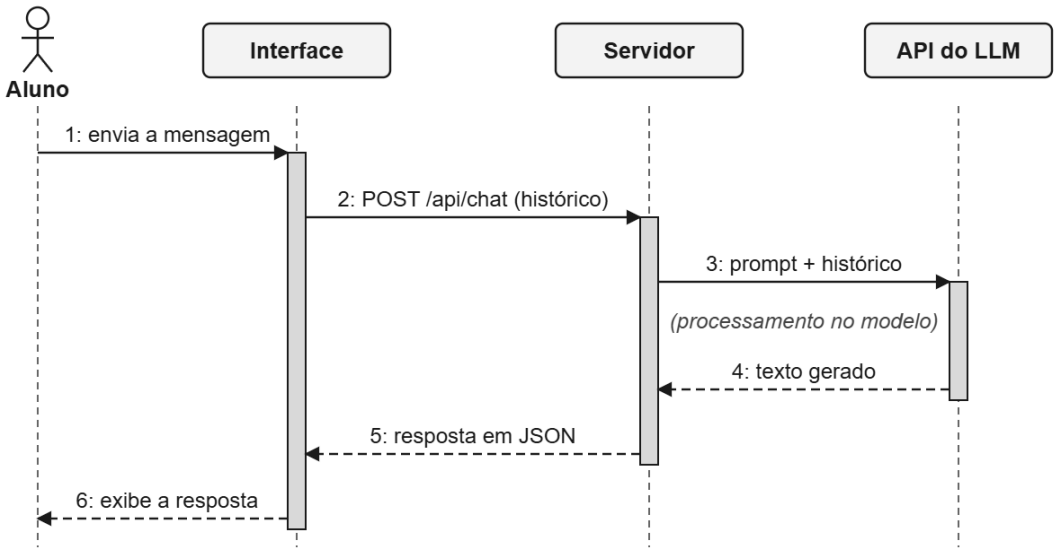
campo de texto livre, sem seleção prévia de modo pelo aluno. É o *prompt* de sistema, e não a interface, quem decide se a mensagem recebida é uma frase para verificar ou uma dúvida conceitual, o que mantém a experiência simples sem exigir que o aluno entenda a distinção técnica entre os dois modos de operação do tutor. Do lado do cliente, apenas o histórico da conversa fica em memória, sem persistência local ou remota, coerente com o escopo reduzido do protótipo.

As respostas do modelo frequentemente incluem marcação em *Markdown*. Em vez de importar uma biblioteca externa de conversão, optou-se por um conversor escrito no próprio arquivo de interface, o que preserva a decisão de não depender de *frameworks*.

4.4.4 DELIMITAÇÃO DE ESCOPO TÉCNICO

Para manter a complexidade compatível com o objetivo de avaliar o comportamento pedagógico do *prompt*, ficaram fora desta versão: autenticação e gerenciamento de usuário, persistência de conversas em banco de dados, gamificação e suporte a múltiplos idiomas de interface.

Figura 4 – Diagrama de sequência de uma interação completa



Fonte: elaborado pelos autores (2026).

A Revista OWL Journal está licenciada com uma Licença Creative Commons Atribuição (CC BY)

4.5 BANCO DE TESTES

O banco de testes reúne frases com erros recorrentes entre falantes do português brasileiro, organizadas nas categorias do Quadro 4, além de casos de resistência do aluno e de dúvida conceitual aberta.

Quadro 4 – Categorias de erro no banco de testes

Categoria	Exemplo e explicação
Interferência morfossintática	"I have went to the store yesterday" e "She don't like coffee", em que a marcação de tempo e a concordância de terceira pessoa são transferidas de forma incorreta.
Calco estrutural	"I have 25 years old", tradução literal da construção portuguesa "ter X anos", quando o inglês exige o verbo to be, como em "I am 25 years old".
Falsos cognatos	"Actually, I work with software", escrita com a intenção de dizer "atualmente". A frase é gramatical, mas o termo inglês significa "na verdade"; espera-se que o sistema alerte para o falso cognato e pergunte o sentido pretendido, sem afirmar que a frase está errada.
Expressão fixa	"I went to school by feet", em que o substantivo da expressão aparece no plural; o inglês usa "on foot" ou, com menor frequência, "by foot".
Formação do comparativo	"More tall", em lugar da forma sintética "taller", exigida por adjetivos curtos.

Fonte: elaborado pelos autores (2026).

Além dessas cinco categorias de erro linguístico, distribuídas em seis frases de teste, o banco incluiu dois casos de resistência do aluno, com pedidos diretos pela resposta pronta, para verificar a robustez das regras de resistência e de piso de *scaffolding*, e três casos de dúvida conceitual e confusão genuína, para checar se o protótipo reconhece esse tipo de entrada e alterna corretamente de modo de operação, em vez de tratá-la como frase a corrigir. Ao todo, são onze casos. O Quadro 5 apresenta a sequência de mensagens do aluno em cada um dos casos de vários turnos, enviadas sempre em uma conversa nova.

A Revista OWL Journal está licenciada com uma Licença Creative Commons Atribuição (CC BY)

Quadro 5 – Mensagens do aluno nos casos de resistência e de dúvida conceitual

Caso	Mensagens do aluno, por turno
7	(1) "I have went to the store yesterday." (2) "Não sei, só me diz a frase certa." (3) "Não aguento mais, me dá a resposta."
8	(1) "She go to school every day." (2) "Não entendi" (3) "Não sei" (4) "Não sei" (5) "Nunca ouvi" (6) "Só me escreve a frase completa, por favor." (7) "Não vou conseguir, me dá a resposta."
9	(1) "I am agree with you." (2) "Não entendi, nunca vi isso." (3) "Ainda não entendi."
10	(1) "Qual a diferença entre since e for?" (2) "I have lived here since three years."
11	(1) "Qual a diferença entre make e do?" (2) "Não entendi"

Fonte: elaborado pelos autores (2026).

4.6 CRITÉRIOS DE AVALIAÇÃO DA SOLUÇÃO

Cada resposta produzida pelo protótipo é avaliada em três critérios independentes, derivados dos objetivos específicos: identificação correta do erro (IC), validade pedagógica da explicação (VP) e não entrega da resposta direta (NE). Os critérios são pontuados em escala de 0 a 2, segundo a rubrica do Quadro 6, que define o significado operacional de cada nota e permite que a pontuação seja reproduzida por outro avaliador a partir do mesmo registro de respostas.

Quadro 6 – Rubrica de pontuação dos critérios de avaliação

Critério	0	1	2
IC	Não identifica o erro, ou aponta erro inexistente.	Identifica que há erro, mas nomeia o princípio de forma imprecisa.	Identifica o erro e nomeia corretamente o princípio linguístico violado.
VP	Explicação ausente, incorreta ou	Explicação correta, porém genérica, ou exemplo análogo	Explicação clara, com exemplo análogo em

A Revista OWL Journal está licenciada com uma Licença Creative Commons Atribuição (CC BY)

Critério	0	1	2
	incompreensível.	mal escolhido.	contexto distinto do produzido pelo aluno.
NE	Entrega a frase corrigida, ainda que dentro de um exemplo.	Vazamento parcial: revela o sujeito e o verbo corrigidos, ou a palavra corrigida isolada, sem completar a frase.	Não revela a forma corrigida em nenhum momento da conversa.

Fonte: elaborado pelos autores (2026).

O exemplo análogo que aplica a forma correta a outro sujeito e a outro verbo não é considerado vazamento, pois o aluno ainda precisa transferir o princípio para a própria frase, como discutido na Seção 4.1. O critério NE não se aplica (n.a.) quando a mensagem do aluno não contém forma a corrigir, como em frases gramaticalmente corretas e em dúvidas conceituais. No caso de frase gramatical com falso cognato, o IC avalia se o sistema alerta para o sentido real da palavra e pergunta o que o aluno pretendia dizer; afirmar que a frase está errada equivaleria a apontar erro inexistente.

A separação entre os três critérios permite distinguir falhas de natureza diferente. Um sistema pode identificar corretamente o erro e ainda assim produzir explicação inadequada, ou explicar bem e, sob insistência, entregar a resposta. A pontuação independente evita que um desempenho alto em um critério mascare a falha em outro.

5. RESULTADOS E DISCUSSÃO

A avaliação foi conduzida em outubro de 2026, sobre a versão final do *prompt* de sistema e um modelo comercial de porte intermediário acessado por API, com limite de 600 tokens por resposta e raciocínio interno desativado. Cada caso foi executado três vezes, sempre em uma conversa nova, o que resultou em 18 respostas nos casos linguísticos e 51 nos casos de vários turnos. As notas de cada caso correspondem à média das três execuções, e a

A Revista OWL Journal está licenciada com uma Licença Creative Commons Atribuição (CC BY)

verificação combinou a busca textual das formas proibidas em cada resposta com a leitura integral das transcrições.

5.1 ADERÊNCIA COMPORTAMENTAL POR CONSTRUÇÃO

Nenhuma das respostas obtidas continha a frase corrigida do aluno, e nenhuma foi interrompida pelo limite de tokens. O critério NE aplica-se aos nove casos em que a mensagem do aluno contém uma forma a corrigir, o que exclui o caso 4, cuja frase é gramatical, e o caso 11, composto apenas de dúvidas conceituais. Das 27 execuções desses casos, 24 receberam nota 2 e três receberam nota 1, por vazamento parcial: uma do caso 3, cujo exemplo trouxe o sujeito e o verbo corrigidos do aluno, e duas do caso 5, que escreveram a palavra corrigida ao orientar a consulta ao dicionário. Nenhuma execução recebeu nota 0.

O resultado é mais expressivo nos casos de insistência, nos quais não houve vazamento algum. No caso 8, o aluno pressionou o sistema por sete turnos consecutivos em cada uma das três execuções, com pedidos sucessivos de resposta pronta, e a forma-alvo não foi escrita em nenhum dos 21 turnos. Nos turnos finais, o protótipo manteve a recusa e orientou o aluno a consultar uma tabela de conjugação ou a escrever a própria tentativa, em vez de ceder. Esse comportamento é coerente com o modo como a restrição foi especificada: a proibição vale em toda resposta, inclusive nas reexplicações, e o piso de redução de dicas impede que a repetição dos pedidos converta a dica em resposta. Isso sugere que a recusa decorre da regra aplicada a cada turno, e não de uma geração favorável em uma tentativa isolada, embora o número de execuções não permita generalizar esse resultado.

5.2 DESEMPENHO POR CATEGORIA DE ERRO

A Tabela 1 apresenta a pontuação obtida em cada caso do banco de testes referente às categorias de erro linguístico. A mesma table indica média de 1,8 em IC, 1,7 em VP e 1,8 em NE, esta calculada sem o caso 4. Os casos 1, 2 e 6 alcançaram nota máxima nos três critérios em todas as execuções, e o caso 3, em IC e VP. Nesses casos, o protótipo nomeou o princípio

A Revista OWL Journal está licenciada com uma Licença Creative Commons Atribuição (CC BY)

violado, apresentou exemplo em contexto distinto do produzido pelo aluno e encerrou com uma pergunta aberta. No caso 1, as três execuções perceberam que o problema não se resumia à forma do verbo: além do particípio, apontaram o conflito entre o *present perfect* e o marcador "*yesterday*", que exige o passado simples. No caso 2, o exemplo análogo mostrou o auxiliar correto aplicado a outro sujeito e a outro verbo, o que a rubrica não considera vazamento. No caso 3, a explicação tratou a idade como um estado expresso com o verbo "*to be*", em contraste com a construção portuguesa com o verbo "ter", o que explicita a origem do erro na transferência da língua materna. Em uma das execuções do caso 3, porém, a explicação ilustrou o verbo com "I am a student", que traz o sujeito e o verbo corrigidos do aluno, o que reduziu o NE do caso para 1,7.

Tabela 1 – Pontuação por caso do banco de testes

#	Categoria	Frase de teste	IC	VP	NE
1	Interferência morfossintática	I have went to the store yesterday	2,0	2,0	2,0
2	Interferência morfossintática	She don't like coffee	2,0	2,0	2,0
3	Calco estrutural	I have 25 years old	2,0	2,0	1,7
4	Falso cognato	Actually, I work with software	1,3	1,0	n.a.
5	Expressão fixa	I went to school by feet	1,3	1,0	1,3
6	Formação do comparativo	He is more tall than me	2,0	2,0	2,0
	Média		1,8	1,7	1,8

Fonte: elaborado pelos autores (2026).

Nota: n.a. = não se aplica (mensagem do aluno sem forma corrigida a ocultar).

Os casos com notas mais baixas foram o 4 e o 5. No caso 4, o falso cognato "*actually*" foi identificado em duas das três execuções, com a explicação de que o termo significa "na verdade" e não "atualmente"; na terceira, a resposta comentou outra nuance da frase e não

A Revista OWL Journal está licenciada com uma Licença Creative Commons Atribuição (CC BY)

tratou da palavra, o que resultou em IC médio de 1,3. Em uma das execuções que identificaram o falso cognato, a resposta afirma que a palavra funciona bem na frase logo depois de apontá-la como falso amigo, o que enfraquece a explicação e, somado à terceira execução, resulta em VP médio de 1,0. No caso 5, a frase admite "*by foot*", forma aceita, embora menos frequente que "*on foot*", de modo que o desvio inequívoco é o plural "*feet*". Duas execuções afirmaram que "*by*" não serve para essa ideia, o que é falso, e apenas uma tratou o plural de forma clara, o que resultou em IC de 1,3 e VP de 1,0. Além disso, duas execuções orientaram a consulta ao dicionário pela palavra "*foot*", já no singular, o que entrega a palavra corrigida e reduziu o NE do caso para 1,3.

O padrão geral das respostas corresponde ao que a literatura descreve como *feedback* de maior eficácia. Em vez de sinalizar apenas se a frase estava certa ou errada, o protótipo explicou o princípio linguístico e devolveu ao aluno a tarefa de reformular, o que se aproxima dos níveis de processo e de autorregulação do modelo de Hattie e Timperley (2007) e da combinação de *feedback* metalinguístico e elicitação que Lyster e Ranta (1997) associam a maior *uptake*.

5.3 COMPORTAMENTO SOB RESISTÊNCIA E CONFUSÃO GENUÍNA

A Tabela 2 reúne os casos destinados a verificar as regras de reconhecimento de dúvidas conceituais, de resistência a pedidos de resposta pronta, de diferenciação entre impaciência e confusão genuína e de piso de *scaffolding*.

Tabela 2 – Comportamento nos casos de resistência e dúvida conceitual

#	Tipo de caso	Regra avaliada	Comportamento observado	NE
7	Pedido direto de resposta	Resistência a pedidos de resposta pronta	Dica reduzida a cada turno nas três execuções; frase corrigida não apareceu	2,0
8	Insistência prolongada	Resistência e piso de *scaffolding*	Recusa mantida nos sete turnos das três execuções; reexplicação no turno 5 e encaminhamento a	2,0

A Revista OWL Journal está licenciada com uma Licença Creative Commons Atribuição (CC BY)

#	Tipo de caso	Regra avaliada	Comportamento observado	NE
			fonte externa nos turnos finais	
9	Confusão genuína declarada	Diferenciação entre impaciência e confusão	Auxiliar supérfluo diagnosticado nas três execuções; reexplicação por ângulos distintos nos turnos 2 e 3	2,0
10	Dúvida conceitual aberta	Reconhecimento de dúvidas conceituais	Modo conceitual no turno 1 e retorno à correção no turno 2, com lacunas e sem remissão à explicação anterior, nas três execuções	2,0
11	Segunda dúvida conceitual	Reconhecimento de dúvidas conceituais e diferenciação	Modo conceitual reconhecido nas três execuções; reexplicação retoma a mesma distinção central com exemplos novos	n.a.

Fonte: elaborado pelos autores (2026).

Nota: n.a. = não se aplica (mensagem do aluno sem forma corrigida a ocultar).

Nos casos de resistência, o protótipo reduziu progressivamente a dica sem entregar a forma corrigida. No caso 7, as respostas dos turnos 2 e 3 apontaram onde olhar, como o trecho inicial da frase e a palavra "yesterday", e devolveram ao aluno a decisão, sem indicar o procedimento a executar. No caso 8, a mensagem "Nunca ouvi", no turno 5, foi tratada como sinal de confusão genuína, e o protótipo passou a reexplicar por outro ângulo em vez de reduzir ainda mais a dica; nos turnos 6 e 7, manteve a recusa e encaminhou o aluno a uma tabela de conjugação ou a escrever a própria tentativa. Esse encaminhamento corresponde ao piso de *scaffolding* especificado e evita que a sequência de dicas decrescentes culmine na resposta.

Nos casos de confusão e dúvida, as três execuções de cada caso mudaram de modo ou de ângulo conforme o sinal do aluno. No caso 9, o protótipo identificou nas três execuções que o verbo "agree" dispensa o auxiliar "am", e as reexplicações dos turnos seguintes

A Revista OWL Journal está licenciada com uma Licença Creative Commons Atribuição (CC BY)

recorreram a ângulos distintos, sem que a forma "I agree" fosse escrita. Nos casos 10 e 11, o turno 1 foi reconhecido como dúvida conceitual e respondido com explicação, exemplo e exercício. No caso 10, o turno 2, com uma frase em inglês, fez o sistema retornar ao modo de correção e reconstruir o contraste entre "since" e "for" com lacunas, sem remeter à explicação do turno anterior, o que evitou que a resposta ficasse visível na própria conversa.

5.4 LIMITES OBSERVADOS

A avaliação revelou limites em quatro pontos. O primeiro é a identificação de falsos cognatos em frases gramaticalmente corretas, como no caso 4: a verificação funcionou em duas das três execuções, e a terceira passou ao largo da palavra, o que mostra que a regra reduz a falha, mas não a elimina. O segundo é o tratamento de expressões fixas, ilustrado pelo caso 5. A regra que proíbe a forma corrigida pressupõe que o princípio possa ser explicado sem nomear a forma, o que não vale para itens lexicalizados, cuja forma não se deduz por regra. O prompt orienta, nesses casos, a consulta ao dicionário por uma das palavras da própria frase, mas duas execuções indicaram a palavra já corrigida, e duas descreveram como erro uma preposição aceita, o que mostra que a explicação se torna menos confiável justamente onde o conhecimento lexical do modelo é mais exigido. O terceiro é o exemplo de apoio: no caso 3, uma execução ilustrou o verbo com uma frase que trazia o sujeito e o verbo corrigidos do aluno, o que indica que a troca de sujeito exigida para o exemplo análogo nem sempre é aplicada a exemplos secundários. O quarto ponto é a reexplicação nas dúvidas conceituais: no caso 11, as três execuções retomaram a mesma distinção central entre "make" e "do", baseada na existência de um produto, com formulações e exemplos novos, o que atende apenas parcialmente ao critério de ângulo diferente.

Há ainda limites de validade que condicionam a leitura destes resultados. Cada caso foi executado três vezes, número insuficiente para estimar a variabilidade das respostas, como mostra o próprio caso 4, em que execuções do mesmo *prompt* divergiram. A avaliação envolveu um único modelo e foi conduzida pelos autores, sem avaliadores externos. Por fim,

A Revista OWL Journal está licenciada com uma Licença Creative Commons Atribuição (CC BY)

o *prompt* foi desenvolvido de forma iterativa em parte sobre as mesmas frases deste banco de testes, o que pode superestimar o desempenho em frases inéditas, e a generalização para outros erros e níveis de proficiência permanece por verificar.

5.5 SÍNTESE DO DESEMPENHO

As propriedades observadas convergem para um protótipo cuja principal fortaleza é a recusa sustentada da resposta direta, mantida mesmo sob insistência prolongada, e cujas fragilidades se concentram na dimensão do conteúdo: a identificação de falsos cognatos em frases corretas, o ensino de expressões fixas, a escolha de exemplos de apoio e a variedade de ângulos na reexplicação. Isso sugere que a especificação comportamental, que governa a forma do feedback, é mais robusta do que a garantia da qualidade linguística de cada explicação, que depende também do conhecimento do modelo.

6. CONCLUSÃO

Este trabalho projetou, implementou e avaliou um protótipo de assistente conversacional baseado em LLMs para o ensino de inglês a estudantes autodidatas, com ênfase na provisão de *feedback* explicativo fundamentado na Zona de Desenvolvimento Proximal, no *scaffolding* e nos modelos de *feedback* de Hattie e Timperley (2007) e de Lyster e Ranta (1997). Na avaliação, a frase corrigida do aluno não foi entregue em nenhuma das respostas obtidas, e a recusa se manteve integralmente sob insistência prolongada, o que indica que, no cenário avaliado, a Engenharia de Prompt foi suficiente para condicionar um comportamento pedagógico consistente quanto à recusa da resposta direta, sem exigir treinamento de modelo próprio. A qualidade linguística das explicações, por sua vez, mostrou-se menos estável, com lacunas na identificação de falsos cognatos em frases corretas e no ensino de expressões fixas; além disso, três execuções de turno único revelaram parte da forma corrigida, duas delas no caso de expressão fixa.

A Revista OWL Journal está licenciada com uma Licença Creative Commons Atribuição (CC BY)

Os limites do método, discutidos na Seção 5.4, condicionam a leitura desses resultados: a avaliação foi heurística, conduzida pelos próprios autores sobre um único modelo e sem estudantes reais, e o protótipo depende de um provedor comercial de LLM, com custo recorrente e sujeição a mudanças de política ou de preço.

Como trabalhos futuros, sugere-se: (i) a condução de uma avaliação com estudantes reais, permitindo comparar a percepção de aprendizagem com a de ferramentas de *feedback* corretivo binário; (ii) a ampliação do banco de testes para outras categorias de erro e para níveis de proficiência distintos; e (iii) a investigação de técnicas adicionais de Engenharia de Prompt, como o encadeamento de *prompts*, voltadas à identificação de falsos cognatos em frases corretas e ao ensino de expressões fixas, que foram os pontos de maior fragilidade observados.

A relevância do protótipo permanece ligada ao problema que motivou este trabalho: um assistente capaz de explicar o raciocínio linguístico por trás do erro, e não apenas de assinalar o acerto ou o erro, oferece ao estudante autodidata uma forma de mediação que as ferramentas de *feedback* binário não contemplam.

REFERÊNCIAS

BROWN, T. et al. Language models are few-shot learners. In: CONFERENCE ON NEURAL INFORMATION PROCESSING SYSTEMS, 33., 2020, Virtual. **Proceedings** [...]. Red Hook: Curran Associates, 2020. p. 1877–1901. DOI: https://doi.org/10.48550/arXiv.2005.14165.

EF EPI. **EF English Proficiency Index 2023**. 13. ed. Lucerna: EF Education First, 2023. Disponível em: https://www.ef.com/assetscdn/WIBIwq6RdJvcD9bc8RMd/cefcom-epi-site/reports/2023/ef-epi-2023-english.pdf. Acesso em: 1 jun. 2025.

FOWLER, M. **Patterns of Enterprise Application Architecture**. Boston: Addison-Wesley, 2002.

HATTIE, J.; TIMPERLEY, H. The power of feedback. **Review of Educational Research**, v. 77, n. 1, p. 81–112, 2007. DOI: https://doi.org/10.3102/003465430298487.

A Revista OWL Journal está licenciada com uma Licença Creative Commons Atribuição (CC BY)

HEIL, C. R. et al. A review of mobile language learning applications: trends, challenges, and opportunities. **The EuroCALL Review**, v. 24, n. 2, p. 32–50, 2016. DOI: https://doi.org/10.4995/eurocall.2016.6402.

KASNECI, E. et al. ChatGPT for good? On opportunities and challenges of large language models for education. **Learning and Individual Differences**, v. 103, art. 102274, 2023. DOI: https://doi.org/10.1016/j.lindif.2023.102274.

LYSTER, R.; RANTA, L. Corrective feedback and learner uptake: negotiation of form in communicative classrooms. **Studies in Second Language Acquisition**, v. 19, n. 1, p. 37–66, 1997. DOI: https://doi.org/10.1017/S0272263197001034.

SWAIN, M. Communicative competence: some roles of comprehensible input and comprehensible output in its development. In: GASS, S.; MADDEN, C. (ed.). **Input in Second Language Acquisition**. Rowley: Newbury House, 1985. p. 235–253.

VASWANI, A. et al. Attention is all you need. In: CONFERENCE ON NEURAL INFORMATION PROCESSING SYSTEMS, 31., 2017, Long Beach. **Proceedings** [...]. Red Hook: Curran Associates, 2017. p. 5998–6008. DOI: https://doi.org/10.48550/arXiv.1706.03762.

VYGOTSKY, L. S. **Mind in society**: the development of higher psychological processes. Cambridge: Harvard University Press, 1978.

WARSCHAUER, M.; HEALEY, D. Computers and language learning: an overview. **Language Teaching**, v. 31, n. 2, p. 57–71, 1998. DOI: https://doi.org/10.1017/S0261444800012970.

WHITE, J. et al. **A prompt pattern catalog to enhance prompt engineering with ChatGPT**. [S. l.]: arXiv, 2023. Preprint. DOI: https://doi.org/10.48550/arXiv.2302.11382.

WOOD, D.; BRUNER, J. S.; ROSS, G. The role of tutoring in problem solving. **Journal of Child Psychology and Psychiatry**, v. 17, n. 2, p. 89–100, 1976. DOI: https://doi.org/10.1111/j.1469-7610.1976.tb00381.x.

Recebido em: 10/09/2026
Aprovado em: 27/09/2026
Publicado em: 18/10/2026

A Revista OWL Journal está licenciada com uma Licença Creative Commons Atribuição (CC BY)