



E-Journal Einzelbeitrag
von: Katrin Schmiderer

Interaktionales Feedback in aufgabenbasierten mündlichen Interaktionen zwischen L2- Spanischschüler:innen und Chatbots

aus: Zeitschrift für Fremdsprachenforschung 1/2026
(ZFF2601W)
Erscheinungsjahr: 2026
Seiten: 83–105
DOI: 10.3278/ZFF2601W005

Während zwischenmenschliche Interaktionssituationen weiterhin als Gold-Standard und als zentrales Ziel des Fremdsprachenlernens gelten sollten, könnten Dialogsysteme vor allem für Schüler:innen von zweiten und dritten Fremdsprachen wertvolle Gelegenheiten bieten, um mündliche Interaktion selbstständig und in geschützter Umgebung zu üben (Sydorenko et al., 2018). In diesem Beitrag werden zunächst Dialogsysteme für den Zweitspracherwerb definiert und anschließend wird erörtert, welches Potential diese Systeme für entwicklungsgerechte Formfokussierung und interaktionales Feedback bieten. Im Anschluss wird eine explorative Studie mit Spanischschüler:innen der Sekundarstufe II vorgestellt, die untersucht, welche Arten von interaktionalem Feedback Lernende im zweiten Lernjahr in einer aufgabenbasierten Interaktion mit einem Chatbot erhalten und inwieweit das vom Chatbot bereitgestellte Feedback dem lernersprachlichen Entwicklungsstand der Schüler:innen entspricht. Abschließend werden weiterführende Forschungsfragen formuliert, mit dem längerfristigen Ziel, Chatbots evidenzbasiert und sinnvoll im Fremdsprachenunterricht einsetzen zu können.

Schlagworte: Zweitspracherwerb; developmentally moderated focus on form; Interaktionales Feedback; Chatbot; TBLT

Zitiervorschlag: Schmiderer, Katrin (2026). *Interaktionales Feedback in aufgabenbasierten mündlichen Interaktionen zwischen L2-Spanischschüler:innen und Chatbots*. *Zeitschrift für Fremdsprachenforschung*, 37(1), 83-105. Bielefeld: wbv Publikation. <https://doi.org/10.3278/ZFF2601W005>

Interaktionales Feedback in aufgabenbasierten mündlichen Interaktionen zwischen L2-Spanischschüler:innen und Chatbots

KATRIN SCHMIDERER¹

Abstract

Face-to-face interaction remains the gold standard and a central goal of foreign-language learning, yet dialogue systems could provide valuable opportunities to practice spoken interaction independently and in a safe environment, especially for learners of a second or third foreign language (Sydorenko et al., 2018). This article defines dialogue systems for L2 acquisition and examines their potential to support developmentally moderated focus on form and interactional feedback. It presents an exploratory study with upper-secondary Spanish learners in their second year of study, investigating the types of interactional feedback provided during a task-based chatbot interaction and the extent to which this feedback aligns with learners' interlanguage development. The article concludes with directions for future research aimed at integrating chatbots into foreign-language instruction in an evidence-based and pedagogically meaningful manner.

1 Einleitung

Die Zweitspracherwerbsforschung betont spätestens seit den frühen 1980er Jahren die zentrale Bedeutung der Verwendung der Zielsprache für die Sprachentwicklung (Gass & Mackey, 2020). Schulische Fremdsprachenlernende – insbesondere jene, die eine zweite oder dritte Fremdsprache erwerben – verfügen jedoch über unterschiedlich viele Gelegenheiten, die zu erlernende Sprache mündlich anzuwenden. Das hat Forscher:innen im Bereich *Computer Assisted Language Learning* (CALL) dazu motiviert, den Einsatz von Computern als Tutor:innen und Interaktionspartner:innen, sogenannten Chatbots oder Dialogsystemen, zu untersuchen (Bibauw et al., 2022, S. 121). Während zwischenmenschliche Interaktionssituationen weiterhin als Gold-Standard und als zentrales Ziel des Fremdsprachenlernens gelten sollten, könnten Dialogsysteme Schüler:innen wertvolle Gelegenheiten bieten, um mündliche Interaktion selbstständig und in geschützter Umgebung zu üben – gerade eben in Sprachlernkontexten, in denen der direkte Kontakt zur Zielsprache begrenzt ist (Sydorenko et al., 2018). Aus einer interaktionistischen Perspektive des Zweitspracherwerbs bieten Dialogsysteme Input, Out-

¹ Katrin Schmiderer, PhD, Universität Innsbruck, Fakultät für LehrerInnenbildung, Institut für Fachdidaktik/Bereich Sprachen, Innrain 52d, A-6020 Innsbruck, katrin.schmiderer@uibk.ac.at

put und Interaktion, ermöglichen die Aushandlung von Bedeutung (*Negotiation for Meaning*; Long, 1996) und können idealerweise individualisiertes interaktionales Feedback zu den Lernendenäußerungen liefern (s. Abschnitt 3), das unter bestimmten Bedingungen den Erwerb von Zweitsprachen (L2) vorantreibt.

Im vorliegenden Aufsatz werden zunächst Dialogsysteme für den L2-Erwerb definiert und anschließend wird erörtert, welches Potential diese Systeme für Aushandlungsroutinen und interaktionales Feedback bieten. Im Anschluss wird eine explorative Studie mit Spanischschüler:innen der (österreichischen) Sekundarstufe II präsentiert, die untersucht, welche Arten von interaktionalem Feedback Schüler:innen im zweiten Lernjahr im Laufe einer aufgabenbasierten Interaktion mit dem Chatbot *Copilot* erhalten. Abschließend sollen weiterführende Forschungsfragen formuliert werden, mit dem längerfristigen Ziel, Chatbots im Fremdsprachenunterricht sinnvoll und evidenzbasiert einsetzen zu können und damit ergänzende Interaktionsmöglichkeiten für Lernende einer zweiten oder dritten Fremdsprache zu schaffen (Jeon & Lee, 2024).

2 Chatbots als Interaktionspartner:innen im L2-Erwerb

Während unter dem breiten Begriff der computergestützten Kommunikation (CMC) verschiedenste Möglichkeiten zusammengefasst werden, wie Sprachlernende durch einen Computer miteinander in Interaktion treten können, meint „dialogue-based CALL“ (Bibauw et al., 2019, S. 828) autonome Systeme, in denen der Computer als tatsächlicher Interaktionspartner eingesetzt wird und Lernende die Möglichkeit haben, eine L2 in einem computergestützten Setting eigenständig in Gesprächen zu üben. Bibauw et al. (2019, 2022) präsentieren eine Taxonomie für Dialogsysteme, die sich an den Möglichkeiten orientiert, die das jeweilige System für das Sprachenlernen bietet. Sie unterscheiden darin zwischen (1) narrativen Systemen, bei denen der/die Lernende aus einer Liste vorgegebener Äußerungen auswählt; (2) formfokussierten Systemen, bei denen die auszudrückende Bedeutung vorgegeben ist (z. B. durch die Bereitstellung eines Teils der Äußerung) und der Fokus vollständig auf der Form liegt (z. B. den Zielstrukturen oder der Aussprache); (3) zielorientierten Systemen, die einen Kontext und eine Aufgabe festlegen und somit freiere Eingaben ermöglichen; und (4) reaktiven Systemen, die offene Interaktionen anbieten, indem sie die Initiative den Nutzenden, d. h. Lernenden, überlassen und versuchen, auf jede Äußerung eine Antwort zu geben. Zielorientierte Systeme, die eine hohe Initiative durch die Lernenden ermöglichen und in denen die Adaptivität wechselseitig zwischen den Interaktionspartnern (Chatbot und Lernende) erfolgt, stellen die ideale Umgebung für spontansprachliche, synchrone mündliche Interaktion dar (Bibauw et al., 2022, S. 125). Den Lernenden steht es dabei einigermassen frei, die Konversation zu lenken, das System wiederum kann darauf flexibel, jedoch gelenkt von der vordefinierten Aufgabe bzw. dem vordefinierten Ziel reagieren. So kann auch dem primären Fokus der Interaktion im Sinne eines aufgabenorientierten Sprachenlernens, nämlich der Aushandlung von Bedeutung, nachgegangen werden (vgl. u. a. East, 2021, für Definitionen von Aufgaben).

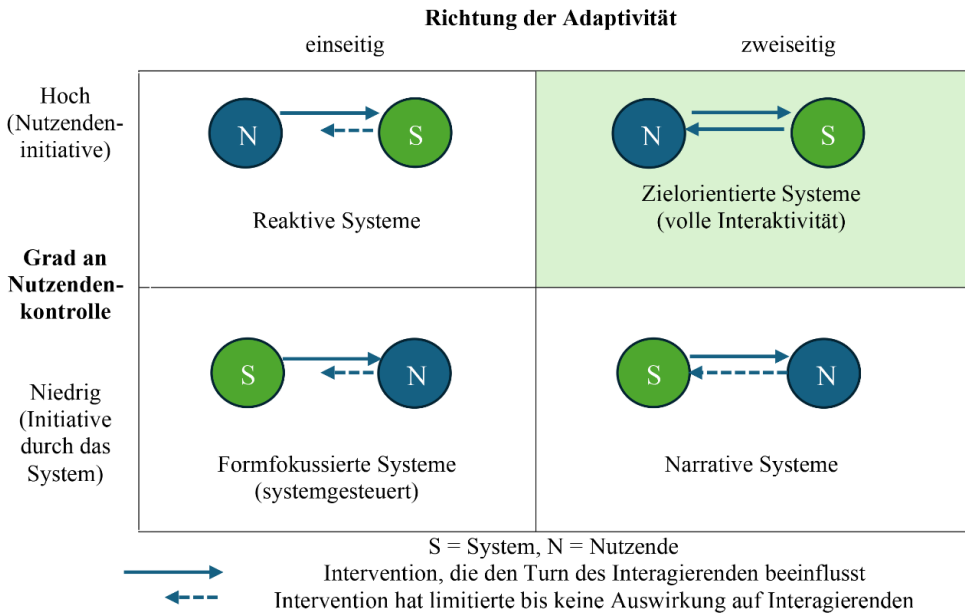


Abbildung 1: Arten von Dialogsystemen entlang zweier Dimensionen (Richtung der Adaptivität und Grad an Nutzendenkontrolle) auf Basis von Bibauw et al. (2022, S. 125) (Quelle: eigene Darstellung)

3 Interaktionales Feedback und *Focus on Form* (FonF) in (Chatbot-)Interaktionen

Seit den Anfängen der CALL-Forschung in den 1960er Jahren ist unmittelbares und individualisiertes Lernendenfeedback ein zentrales Anliegen dieses Forschungsbereichs. Durch den technologischen Fortschritt der letzten Jahrzehnte ist Feedback heute nicht mehr nur auf generische, mechanische Computerantworten auf vordefinierte grammatische und lexikalische Elemente limitiert, sondern ist – zumindest im Schriftlichen – bereits in wesentlich individualisierter Form möglich. Gerade für mündliche Interaktionen und interaktionales Feedback im Mündlichen stellen sich aber nach wie vor zahlreiche Herausforderungen.

Interaktionales Feedback meint korrigierendes Feedback, das Lernende im Laufe einer kommunikativen Interaktionssituation zu fehlerhaften Äußerungen erhalten. Es kann sich dabei um unterschiedliche Arten der Aushandlung sowie um unterschiedliche Modifikationsstrategien handeln, die zur Klärung von kommunikativen oder sprachlichen Problemen dienen (Nassaji, 2016, 2020). Nassaji (2016, S. 536) fasst die verschiedenen Strategien interaktionalen Feedbacks in drei Hauptkategorien zusammen: (1) Reformulierungen, bei denen fehlerhafte Äußerungen von Lernenden durch implizite Korrekturen (auch: *Recasts*) oder direkte Korrekturen in die zielsprachlich korrekte Form umformuliert werden; (2) *Prompts*, die keine Korrekturen sind, sondern viel-

mehr die Gelegenheit zur Selbstkorrektur geben, etwa in Form von klärenden Fragen und Wiederholungen des Fehlers mit steigender Intonation; und (3) metalinguistisches Feedback. Interaktionales Feedback kann eine vollständige oder teilweise Korrektur der Äußerung sein und verbal oder non-verbal ausgeführt werden. Ellis (2021) gibt, basierend auf Sheen und Ellis (2011) und Lyster et al. (2012), ebenfalls einen Überblick über unterschiedliche Strategien von korrekktivem Feedback. Nach seiner Klassifikation können die Strategien implizit (z. B. interaktionale Reformulierungen, Wiederholungen, klärende Fragen) oder explizit sein (didaktische Reformulierungen, explizite Korrekturen, metasprachliche Kommentare, Elizitierungen, parasprachliche Signale) und entweder Input zur Verfügung stellen oder Output, also Produktion, elizitieren.

Interaktionales Feedback und damit verbunden die Frage, inwieweit durch dieses der L2-Erwerb unterstützt werden kann, stellt ein Kernanliegen der L2-Erwerbsforschung dar. Dabei wurden in der Fachliteratur in den letzten Jahren unterschiedliche theoretische Begründungen und auch empirische Belege präsentiert. Aus interaktionistischer Sicht wird argumentiert, dass sich die Bedingungen für den L2-Erwerb entscheidend verbessern, wenn L2-Lernende Bedeutungen mit anderen Sprechenden – seien es L1-Sprechende oder andere L2-Sprechende – aushandeln. Insbesondere Long (1996) hebt die Bedeutung der „Aushandlung für Bedeutung“ (*Negotiation for Meaning*) in Interaktionssituationen zwischen kompetenten Sprechenden und weniger kompetenten Lernenden hervor. Diesen Prozess definiert Long (1996, S. 418) wie folgt:

Negotiation for meaning is the process in which, in an effort to communicate, learners and competent speakers provide and interpret signals of their own and their interlocutor's perceived comprehension, thus provoking adjustments to linguistic form, conversational structure, message content, or all three, until an acceptable level of understanding is achieved.

Vor allem Aufgaben im Sinne des *Task-Based Language Teaching* mit beidseitigem Informationsaustausch (*Two-way Tasks*), so Long (2017, S. 5), ermöglichen zahlreiche Gelegenheiten für diesen Aushandlungsprozess zwischen Lernenden und kompetenten Sprechenden.

Als pädagogisch-didaktisches Instrument ist dabei in weiterer Folge der *Focus-on-Form*-Ansatz (Long, 1991, 1998) zu nennen. FonF wurde 1991 von Michael Long – im klaren Gegensatz zu *Focus on Forms* (FonFs) – als Prinzip eines Fremdsprachenunterrichts beschrieben, in dem es primär um die Verwendung der Fremdsprache und um bedeutungsvolle Kommunikation geht. Die Lehrperson lenkt die Aufmerksamkeit der Lernenden dabei nur temporär auf die sprachlichen Formen, wenn diese zufällig im Unterricht auftreten. Der übergeordnet an der Bedeutung bzw. Kommunikation orientierte Unterricht wird kurz unterbrochen. Long (1998, S. 40) konkretisiert dies wie folgt:

Focus on form refers to how attentional resources are allocated, and involves briefly drawing students' attention to linguistic elements (words, collocations, grammatical structures, pragmatic patterns, and so on) *in context*, as they arise incidentally in lessons whose overriding focus is on meaning, or communication, the temporary shifts in focal attention being triggered by students' comprehension or production problems.

Aus Longs Definition wird bereits klar, dass es sich bei einem FonF immer um eine Reaktion auf lernersprachliche Schwierigkeiten, also um Feedback, handelt. Diese Reaktion erlaubt es, die Aufmerksamkeit der Lernenden auf sprachliche Probleme zu lenken, die direkt während der Kommunikation, beispielsweise zwischen Schüler:innen in Problemlösungsaufgaben, auftreten. FonF wird als pädagogisches Instrument dann als sinnvoll erachtet, wenn übergeordnete Prozesse der Sprachverarbeitung nicht gestört werden und die Formfokussierung in einem sogenannten „window of opportunity“ (Doughty, 2001, S. 211) geschieht, also in dem Moment, in dem die Lernenden mit der Form beschäftigt sind.

Darüber hinaus gehen vor allem Forschende, die den L2-Erwerb aus der Perspektive der Sprachverarbeitung betrachten, davon aus, dass die entwicklungsbedingte Bereitschaft der Lernenden, eine bestimmte Struktur zu verarbeiten, einen Einfluss darauf nimmt, ob Lernende das erhaltene Feedback verarbeiten können. Loewen and Reinders (2011, S. 51) definieren diese entwicklungsbedingte Bereitschaft als „the learner’s potential to acquire a particular linguistic item“ und betrachten Lernende als entwicklungsbedingt bereit, wenn sie sich auf der entsprechenden lernersprachlichen Entwicklungsstufe befinden. Im theoretischen Rahmen der *Processability Theory* (PT; u. a. Pienemann & Lenzing, 2025) brachten etwa Di Biase (2008) und Roos (2019) das Konzept des FonF mit der L2-Verarbeitungsperspektive, genauer gesagt, der *Teachability Hypothesis* (TH; Pienemann, 1989) zusammen. Die Kernannahme der TH ist, dass die Wirkung von Unterricht durch die Verarbeitbarkeit (*Processability*) der unterrichteten, fokussierten Strukturen für die jeweiligen Lernenden begrenzt ist (zur empirischen Überprüfung dieser theoretischen Annahmen (s. Baten & Keßler, 2019, für einen Überblick; Bonilla, 2015, für das Spanische). In Longs vielzitiertem Buch *Second Language Acquisition and Task-based Language Teaching* (2014) betont auch er, dass durch den reaktiven Charakter des FonF die Wahrscheinlichkeit erhöht wird, dass der Fokus auf jene sprachlichen Formen gelegt werden kann, für deren Verarbeitung die Lernenden aus psycholinguistischer Sicht, d. h., aus Sicht des lernersprachlichen Entwicklungsstandes der Schüler:innen, auch bereit sind.

Im Kontext von technologiegestützten Aufgaben, genauer gesagt, Interaktionen mit einem Chatbot, könnten sich zahlreiche Gelegenheiten für ebendiese Aushandlungsroutinen mit interaktionalem Feedback ergeben. Der Chatbot könnte zudem die Aufmerksamkeit aus der Interaktion heraus auf Formen lenken, die der/die Lernende bereit ist zu verarbeiten.

3.1 Empirische Belege für interaktionales Feedback und FonF in CMC

Zahlreiche Studien, sowohl deskriptiver als auch experimenteller Natur, sowie Metaanalysen scheinen Vorteile von interaktionalem Feedback und FonF für die L2-Entwicklung zu belegen (z. B. Norris & Ortega, 2000; Spada & Tomita, 2010; Goo et al., 2015).² Studien zu interaktionalem Feedback in synchronen computergestützten Kon-

2 Dass dabei kritisch hinterfragt werden kann, dass die in den genannten Metaanalysen untersuchten Studien nur teilweise Messinstrumente verwendeten, die prozedurales, implizites Sprachkönnen (und nicht explizites Sprachwissen) untersuchen, führen etwa Doughty (2003) und Hinger (2016) aus.

texten bestätigen, dass Aushandlungsprozesse, korrigierendes Feedback und modifizierte Lernendenäußerungen auf Basis des Feedbacks in diesen Unterhaltungen auch in technologiestützten Lernumgebungen stattfinden (s. Ziegler & Mackey, 2017, für einen systematischen Überblick). In der Folge sollen zunächst Studien für das Spanische als L2 herausgegriffen werden.

Pellettieri (2000) untersuchte synchrone, schriftliche Chat-Interaktionen von universitären Spanisch-Studierenden ($N=20$) über die Software *ytalk*. Die Studierenden führten dabei fünf kommunikative Aufgaben durch, die sowohl offenere als auch geschlosseneren Aufgaben umfassten. Pellettieri analysierte in den Interaktionen Aushandlungsroutinen nach Varonis und Gass (1985). In den verschiedenen Aufgaben fand Pellettieri unterschiedlich viele Aushandlungsroutinen, wobei eine geschlossene „Finden-Unterschied“-Aufgabe die größte Anzahl an Aushandlungen hervorbrachte.

In einer weiteren Studie von Blake (2000) mit 50 leicht fortgeschrittenen Spanischlernenden eines Universitätskurses, die einmal pro Woche im synchronen Chatprogramm *Remote Technical Assistance* chatteten, wurde untersucht, ob die Chat-Konversationen ebenso wie mündliche Interaktionen Möglichkeiten für korrigierendes Feedback und darauffolgende sprachliche Modifikationen liefern und ob es Unterschiede abhängig vom Aufgabentyp gibt. Blake fand in 0,3 % bis 3,8 % der Turns Aushandlungsroutinen, wobei geschlosseneren Aufgaben im Vergleich zu offeneren Aufgaben mehr Möglichkeiten für Aushandlungen lieferten.

Die Studie von Fernández-García und Martínez-Arbeláiz (2002) untersuchte Aushandlungsprozesse in schriftbasierten Chat-Diskussionen von 28 Englisch-L1-Sprechenden in universitären L2-Spanischkursen. Die Datenerhebung umfasste zwei Sitzungen im Abstand von rund 20 Tagen, in denen Kleingruppen eine *Open-Transport*-Chatdiskussion zu einer Lektüre mit Inhaltsfragen führten. In allen Gruppen konnten Aushandlungsroutinen gefunden werden. Dabei handelte es sich nach Varonis und Gass Klassifizierung größtenteils um explizite Äußerungen von Nichtverstehen mithilfe typischer Formulierungen wie „¿Qué significa X? (Was bedeutet X?)“ sowie Nachfragen mit „¿Qué? (Was?)“. Die Autorinnen führen die Dominanz expliziter Signale und die geringe Häufigkeit anderer Indikatoren auf die schriftliche Modalität der Kommunikation zurück, in der explizite Markierungen von Nichtverstehen präferiert werden.

Die Studie von Lee (2002) untersuchte, wie L2-Spanisch-Lernende, ebenfalls eines dritten Lernjahrs im universitären Kontext, in synchroner, schriftlicher Online-Kommunikation über das Tool *Blackboard* miteinander Bedeutung aushandeln. Im Zentrum der Analyse stehen kommunikative Strategien, mit denen Lernende Kommunikationsprobleme lösen. Die synchrone Chat-Umgebung bot auch in dieser Studie zahlreiche Gelegenheiten zur Aushandlung von Bedeutung. Die Lernenden nutzten wiederum verschiedene Modifikationsstrategien. Fehler der anderen Lernenden schienen die Proband:innen zu ignorieren und die Diskussion fortzuführen, was auf eine Priorisierung von Verständigung und Interaktionsfluss gegenüber formaler Korrektheit hindeutet.

Lee (2006) analysierte schließlich die Aufnahme des Feedbacks von L1-Spanisch-Universitätslehrenden durch die in Lee (2002) untersuchten L2-Lernenden und fand in 50 % der Kontexte Hinweise auf eine Aufnahme des Feedbacks durch die Lernenden,

wobei sich – wie auch in allen anderen hier angeführten Studien – das korrektive Feedback größtenteils auf lexikalische Elemente bezog.

Trotz einiger verfügbarer Studien kommen Ziegler und Mackey (2017) in ihrem systematischen Überblick zu interaktionalem Feedback in synchroner CMC zum Schluss, dass Fragen der Interaktion in CMC nicht so gründlich und systematisch verfolgt wurden wie in *Face-to-Face*-Kontexten (FTF). Sie führen als Beispiel dafür die Forschung zu Feedbacktypen an, die noch limitiert scheint. Zudem existieren nur wenige Studien, die den Einfluss von implizitem und explizitem korrigierendem Feedback in synchroner CMC auf den L2-Erwerb untersucht haben (z. B. Loewen & Erlam, 2006; Sauro, 2009; Yilmaz, 2012). So sind auch die Ergebnisse der wenigen Studien divergent: Einige Studien fanden keine Unterschiede in der Wirksamkeit zwischen Reformulierungen und metalinguistischem Feedback (Loewen & Erlam, 2006; Sauro, 2009) für den Spracherwerb, während andere Vorteile von expliziten Korrekturen im Vergleich zu Reformulierungen bei unmittelbaren und längerfristigen Tests in der Sprachproduktion und im Sprachverständnis zeigten (z. B. Yilmaz, 2012). Hinzuzufügen ist, vor allem im Hinblick auf den hier vorliegenden Artikel, dass kaum Studien mit Anfänger:innen durchgeführt wurden und – soweit ersichtlich – auch keine Studien im schulischen Kontext, etwa der Sekundarstufe, vorliegen.

Zudem muss erwähnt werden, dass die allermeisten der bisherigen Studien in Text-Chat-Umgebungen durchgeführt wurden. Erst in den letzten Jahren wurde begonnen, die Effekte von interaktionalem Feedback in Audio- und Video-Chatprogrammen wie *Skype*, *FaceTime* und *Google Hangouts* zu untersuchen. Eine der wenigen Studien zur mündlichen Interaktion stellt Yanguas (2010) Untersuchung bei leicht fortgeschrittenen Spanischlernenden dar. Er analysierte, wie Lernende in Video- und Audiogruppen bei der Bearbeitung einer kommunikativen Aufgabe Bedeutung aushandeln und wie sich diese beiden Gruppen von traditioneller FTF-Kommunikation unterscheiden. Fünfzehn Lernenden-Dyaden wurden zufällig einer Audio-, Video- oder FTF-Kontrollgruppe zugewiesen und bearbeiteten eine *Jigsaw*-Aufgabe mit 16 unbekanntem Wörtern. Die Video- und Audio-Gruppen nutzten *Skype*. Die Analyse der Daten zeigte, dass im FTF-Setting sowie in beiden computergestützten Gruppen Aushandlungsprozesse stattfanden, wenn Verständigungsprobleme auftraten, wobei in der Audiogruppe die meisten Routinen gefunden wurden. Die Unterschiede zwischen Audio- und Videogruppe führt der Autor vor allem auf den fehlenden visuellen Kontakt in der Audiogruppe zurück. Die Gesprächsstruktur in den mündlichen computergestützten Settings ähnelte stark jener der FTF-Kommunikation, unterschied sich jedoch deutlich von schriftlicher, synchroner CMC in anderen Studien.

Auch Monteiro (2014) greift das Desiderat nach Studien zu mündlicher synchroner CMC auf und repliziert teilweise Ellis et al. (2006) Studie zu Interaktion im Unterricht. Mithilfe eines Pretest-/Posttest-Designs vergleicht sie drei leicht fortgeschrittene Englisch-Lernendengruppen, die alle mit denselben mündlichen Aufgaben arbeiteten, aber unterschiedliche Arten von Feedback erhielten: Gruppe 1 erhielt im computergestützten Video-Chat metasprachliches Feedback hinsichtlich einer spezifischen Form (*Simple Past Tense*), Gruppe 2 hingegen wurden Reformulierungen (zur selben gram-

matischen Form) angeboten und Gruppe 3 arbeitete ohne korrekatives Feedback mit den mündlichen Aufgaben. Daraus geht hervor, dass alle drei Gruppen erfolgreich darin waren, explizites und implizites Wissen zur untersuchten Form aufzubauen. Im Unterschied zu Ellis et al. (2006) Studie, bei der metalinguistisches Feedback effizienter war, förderten dyadische aufgabenbasierte Videokonferenz-Interaktionen, mit oder ohne korrekatives Feedback, die Entwicklung von implizitem und explizitem Wissen in Monteiros Untersuchung.

Saito und Akiyama (2016) untersuchten den Einfluss videobasierter Interaktionen auf die langfristige Entwicklung der Sprachproduktion von japanisch sprechenden Englisch-Lernenden im Laufe eines Semesters. Die College-Studierenden der Experimentalgruppe führten wöchentliche dyadische, aufgabenbasierte Gespräche mit L1-Sprechenden in den USA über das Videokonferenz-Tool *Google Hangouts* durch. Die L1-Sprechenden wurden darin geschult, interaktionales Feedback (Reformulierungen) zu geben, wenn die Äußerungen der Lernenden das Verständnis erschwerten. Die Kontrollgruppe erhielt Fremdsprachenunterricht, der größtenteils aus Wortschatz- und Grammatikübungen bestand. Die Pretest-Posttest-Daten der spontanen Sprachproduktion zeigten, dass die Experimentalgruppe im Gegensatz zur Kontrollgruppe signifikante Fortschritte in den Bereichen Verständlichkeit, Flüssigkeit und Lexikogrammatik, jedoch nicht im Bereich der Aussprache aufweisen konnte.

Parlak und Ziegler (2016) untersuchten den Einfluss von Reformulierungen auf die Entwicklung der lexikalischen Betonung in synchronen mündlichen computergestützten Interaktionen (SCMC) im Vergleich zu FTF-Interaktion. In einem Pretest-/Posttest-Design wurden fortgeschrittene Englischlernende zufällig einer von vier Gruppen zugewiesen: FTF-Reformulierungen, SCMC-Reformulierungen, FTF-Kontrollgruppe oder SCMC-Kontrollgruppe. Die Tests bestanden aus Vorlese- und Informationsaustauschaufgaben, während die Intervention aus einer interaktiven Rollenspielaufgabe bestand, die im Falle der SCMC-Gruppe über *Skype* durchgeführt wurde. Die Analyse der Entwicklung der Betonung erfolgte auf Basis akustischer Merkmale (Silbendauer, Lautstärke und Tonhöhe). Die statistische Auswertung dieser Merkmale ergab jedoch für keine der Gruppen signifikante Effekte der Reformulierungen auf die Entwicklung der lexikalischen Betonung.

Diese doch unterschiedlichen Ergebnisse machen die Notwendigkeit weiterer Studien deutlich, um herauszufinden, welchen Einfluss die Modalität auf die L2-Sprachentwicklung tatsächlich hat. Die Frage, inwieweit mündliche, aufgabenbasierte Chatbot-Interaktionen (s. Definition in Abschnitt 2) Gelegenheiten für interaktionales Feedback bieten, ist – soweit ersichtlich – noch größtenteils unbeantwortet. Aufgrund der bis dato fehlenden Studien zu Feedback in mündlichen Chatbot-Interaktionen sollen im folgenden Abschnitt ein kurzer Überblick über bisherige Studien zu interaktionalem Feedback in schriftlichen Chatbot-Interaktionen gegeben und erste Ergebnisse von explorativen Studien zu Interaktionsstrategien von Lernenden und Chatbots in mündlicher Interaktion präsentiert werden. Für einen systematischen Überblick zu anderen Aspekten mündlicher und schriftlicher Interaktionen von Lernenden mit verschiedenen Typen von Chatbots (beispielsweise Sprechangst, Entwicklung der Interaktionskompe-

tenz und der Bereitschaft zur Kommunikation, Einstellungen gegenüber dem Lernen mit Chatbots) sei auf Şahin Kızıl et al. (2025) verwiesen. Die Autor:innen zeigen in ihrem Überblicksartikel, dass bisherige Studien – wie so häufig – vor allem mit Lernenden im universitären Umfeld und für Englisch als Fremd- oder Zweitsprache sowie im speziellen Fall der mündlichen Interaktion mit Chatbots wohl nicht zufällig mit Mandarin und Koreanisch als Fremd- oder Zweitsprache vorliegen. Soweit ersichtlich liegen bisher keine Studien zum Spanischen als Fremdsprache vor.

3.2 Erste empirische Befunde zu interaktionalem Feedback in Chatbot-Lernenden-Interaktionen

Bereits 2010 entwickelte Petersen in seiner Doktorarbeit einen Chatbot, genannt Sasha, der schriftliche dialogische Interaktionen für verschiedene Tasktypen im Sinne von TLBT erlaubt. Als Gesprächspartner im schriftlichen Chat war Sasha dazu konzipiert, sich in der Kommunikation wie ein Mensch zu verhalten, jedoch nicht wie ein Mensch auszusehen (Petersen, 2010, S. 75). Sasha erschien den Lernenden als einfache Strichfigur, die in der Lage war, korrekatives Feedback zu geben, das auf die Frageformulierung im Englischen abzielte. Die Entwicklung der untersuchten Zielstruktur wurde anhand der Entwicklungssequenz von Frageformen im L2-Erwerb des Englischen nach Pienemann und Johnston (1987) operationalisiert. In seiner Studie mit 56 Englisch-Lernenden an einer öffentlichen *Highschool* (größtenteils mit Spanisch als L1) konnte Petersen zeigen, dass der Chatbot sowohl Feedback in Form von Reformulierungen liefern konnte als auch ähnliche Antwortzeiten wie L1-Sprecher:innen in mündlicher Interaktion aufwies. Der Gruppenvergleich (Lernende – schriftlicher Chatbot vs. Lernende – L1-Sprechende vs. Kontrollgruppe) nach drei je 30-minütigen Interventionen, bestehend aus jeweils drei kommunikativen Aufgaben, bestätigte erstens die Entwicklung von Frageformen entlang der von Pienemann und Johnston (1987) beschriebenen Entwicklungssequenz. Zweitens schnitten sowohl die Lernenden, die mit dem Chatbot Sasha interagierten, als auch die Lernenden, die mündlich mit L1-Sprechenden interagierten, besser ab als die Kontrollgruppe.

In einer weiteren Dissertation befasste sich Wilske (2015) mit der Frage, wie verschiedene computerbasierte Feedbackarten das Lernen in schriftbasierten, aufgabenorientierten Dialogsystemen beeinflussen. Er fand nur geringe Unterschiede in der Entwicklung mündlicher Kommunikationskompetenzen bei L2-Deutsch-Lernenden im universitären Kontext in Abhängigkeit von den eingesetzten Feedbackarten. Dennoch zeigten bedeutungsorientierter FonF ebenso wie implizite Reformulierungen bessere langfristige Lernfortschritte, während im Vergleich dazu expliziter FonFs und explizites metalinguistisches Feedback zu unmittelbareren, weniger langfristigen Fortschritten führten (vgl. Wilske, 2015, S. 243). Diese Ergebnisse bestätigen L2-Erwerbsstudien zu zwischenmenschlicher Interaktion.

In den letzten Jahren wurden schließlich auch mündliche Dialogsysteme entwickelt, die teilweise spezifisch für die Entwicklung einer Fremdsprache gedacht sind (z. B. *Ellie, Univerbal*), teilweise aber auch übergeordnete Ziele verfolgen und für die Entwicklung einer Fremdsprache im mündlichen Modus genutzt werden können (so-

genannte *Customized Chatbot-Building Platforms*, die auf großen Sprachmodellen (LLMs) wie GPTs basieren). Gerade letztere stellen eine – mittlerweile für große Bevölkerungsgruppen in Europa und Nordamerika zumindest aktuell günstig verfügbare – Interaktionsmöglichkeit für Sprachlernende dar.

Im Bewusstsein, dass sich Forschung im Bereich mündlicher Dialogsysteme sowie deren Einsatz im Fremdsprachenunterricht rasant entwickelt, sollen hier aktuelle Pilotstudien kurz vorgestellt werden. Für einen Überblick über frühe Chatbots und aktuelle Entwicklungen sei auf Yang et al. (2022, S. 327–330) verwiesen; als Beispiele für Chatbots, die L2-Lernende dabei unterstützen, aufgabenbasierte Unterhaltungen durchzuführen, können der Chatbot *Ellie* von Kim et al. (2022) und ein ähnlich zielorientierter Chatbot von Chen et al. (2022), der auf kommerziell verfügbaren KI-Technologien aufbaut, genannt werden.

Jeon und Lee (2024) entwickelten einen Chatbot mit *Dialogflow*, dem *Chatbot-Builder* von Google, mit dem die untersuchten Lernenden (Grundschule) eine stark gelenkte, strukturierte, und sich wiederholende Unterhaltung führten. Zudem programmierten Jeon und Lee auch korrekatives Feedback. Je nach Äußerungen der Lernenden lief die Interaktion so lange weiter, bis die programmierte, erwartete Antwort geboten wurde. In ihrer Interventionsstudie erhoben Jeon und Lee die mündliche Sprechleistung vor und nach der Intervention. Sie fanden eine signifikant bessere mündliche Sprechleistung nach der Intervention und im Vergleich zur Kontrollgruppe sowie positive Einstellungen der Lernenden gegenüber dem Lernen mit einem Chatbot. So gaben die Lernenden an, die Selbststeuerungsmöglichkeiten, die unlimitierten Übungsmöglichkeiten und die angstfreie Sprechatmosphäre zu schätzen.

Wu et al. (2024) entwickelten den Chatbot *MSLIPA* und analysierten die Interaktionsstrategien, die Grundschüler:innen, die Mandarin lernten, in Chatbot-Interaktionen im Vergleich zu Peer-Interaktionen anwendeten. Die Ergebnisse zeigen, dass die Gesamtanzahl an kommunikationsstörenden Momenten in der Chatbot-Gruppe wesentlich höher war als in jener Gruppe, in der Lernende mit anderen Lernenden interagierten. Dies führen die Studienautor:innen darauf zurück, dass die L2-Lernenden mit ihren verschiedenen Akzenten oftmals schwer von einem Chatbot verstanden werden. Lernende in den Chatbot-Interaktionen nutzen ein breiteres Repertoire an Interaktionsstrategien wie Reformulierungen (36,6 %), gefolgt von Themenabbruch (32,3 %) und Wiederholungen (12,5 %). In den Peer-Interaktionen tendierten Lernende dazu, das aktuelle Thema abzubrechen (41,4 %) oder das Gesagte zu wiederholen, um den Dialog aufrechtzuerhalten (37,9 %). Der Chatbot wiederum stellte den Lernenden implizites Feedback zu unpassenden Aussagen in Form von Fragen, Pausen oder falschen Antworten zur Verfügung.

Wichtig zu erwähnen ist, dass, soweit ersichtlich, Wu et al. (2024) ebenso wie alle anderen verfügbaren Studien zumindest eine Mischung aus schriftlicher und mündlicher Modalität untersuchen. Das bedeutet, dass Lernende stets auch ein Transkript des Gesagten sehen, während sie interagieren. Es scheint daher eine Forschungslücke zu bestehen, die rein mündliche Modalität in Chatbot-Konversationen zu untersuchen. In diesem Aufsatz wird eine Pilotstudie vorgestellt, die dieses Desiderat aufgreift. Auf-

grund der äußerst raschen technologischen Entwicklung muss hinzugefügt werden, dass selbst aktuelle empirische Studien deutlich an Relevanz verlieren, weil sich die technischen Möglichkeiten der Chatbots innerhalb kurzer Zeit stark verändert haben und laufend verbessern, sodass frühere Studienergebnisse mit Vorsicht zu interpretieren sind.

4 Pilotstudie mit Spanischschüler:innen

In der vorliegenden explorativen Studie wird den Fragen nachgegangen, (1) ob aufgabenbasierte mündliche Chatbot-Interaktionen von neun Spanisch-L2-Lernenden der Sekundarstufe II in Österreich mit einem einfach gehaltenen Prompt ohne Spezifizierung des Feedbacks Aushandlungsroutinen ermöglichen und (2) wenn ja, welche Arten von interaktionalem Feedback vom Chatbot angeboten werden. Die teilnehmenden Schüler:innen waren zwischen 17 und 19 Jahren alt und lernten Spanisch im zweiten Lernjahr als dritte lebende Fremdsprache nach Italienisch oder Französisch.

4.1 Untersuchungssetting

Zur Beantwortung dieser beiden Forschungsfragen erhielten Schüler:innen die Aufgabe, mit einem Chatbot ein Rollenspiel mit einem beidseitigen Informationsaustausch durchzuführen (s. Aufgabenstellung und Prompt in Textbox 1 und 2). Die Chatbot-Interaktionen wurden auf einem universitätseigenen MacBook Pro abgewickelt, auf dem die Software *Microsoft 365 Copilot* (Browser-Version)³ installiert war. Die Datenerhebung fand während der Unterrichtszeit am Schulstandort in einem separaten Klassenraum statt. Die Teilnahme der Schüler:innen war völlig freiwillig und die Schüler:innen bestätigten diese in einer schriftlichen Einverständniserklärung. Alle Teilnehmenden wurden im Vorfeld der Datenerhebung über die Art der erhobenen Daten, die verwendeten Tools und Rechte im Detail aufgeklärt. Die Studie wurde von der Ethikkommission der Fakultät für LehrerInnenbildung der Universität Innsbruck als unbedenklich eingestuft (*Certificate of Good Standing*, 55/2025). Die Bearbeitung der Aufgabe (inkl. Fragebogenerhebung, s. weiter unten) nahm pro Schüler:in ca. 15 Minuten in Anspruch. Um die Interaktion zu starten, aktivierte die Forschende den mündlichen Modus von *Copilot*, woraufhin die Lernenden dem Chatbot den Prompt mündlich vorlasen, was den weiteren Verlauf der Gespräche beeinflussen kann (z. B. durch bessere oder schlechtere Spracherkennung). Zum Zeitpunkt der Datenerhebung konnte der Chatbot allerdings noch nicht auf Basis eines schriftlichen Prompts in den mündlichen Modus wechseln.

3 *Microsoft 365 Copilot* entspricht laut eigenen Angaben der EU-Datenschutzgrundverordnung und verpflichtet sich dazu, Eingabeaufforderungen, Antworten und Daten, auf die über *Microsoft Graph* zugegriffen wird, nicht für das Training von *Foundation LLMs* und auch nicht für das Training von *Microsoft 365 Copilot* zu verwenden (<https://learn.microsoft.com/de-de/copilot/microsoft-365/microsoft-365-copilot-privacy>). Den Teilnehmer:innen wurde garantiert, dass unmittelbar nach jeder einzelnen Datenaufnahme pro Proband:in die Löschung des gesamten Verlaufs in Auftrag gegeben wird. Für den Kontext der österreichischen Sekundarstufe II kann zudem auf das bis 2027 bestehende *Microsoft Austrian College and High School Agreement* verwiesen werden (https://www.bmbwf.gv.at/Themen/schule/zrp/dibi/itinf/it_angebote/ms-ach.html).

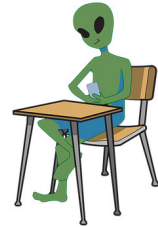
Aufgabe: Rollenspiel

Du arbeitest heute mit einem Chatbot (*Copilot*).

Dein Interaktionspartner (= der Chatbot) ist ein/eine Bewohner:in des Planeten Mars und ist soeben auf der Erde angekommen. Interessanterweise könnt ihr euch nur auf Spanisch unterhalten. Frag ihn/sie/es, was auch immer dich interessiert. Sei kreativ!

Sprich mit dem Chatbot so lange, bis der Signalton (Glocke) ertönt.

Beende dann das Gespräch.



Textbox 1: Aufgabenstellung *Martian Task*, adaptiert von Lenzing (2025)

Quelle: eigene Darstellung, Vektor von echapman via Pixabay, <https://pixabay.com/de/vectors/außerirdischer-schüler-gerät-raum-2166403/>

Ablauf der Aufgabe

Lies dem Chatbot folgenden Text vor:

Ich bin Sara.

Ich möchte heute mit Dir auf Spanisch sprechen.

Ich bin Anfängerin im zweiten Lernjahr, ungefähr auf Niveau A1-A2 nach dem europäischen Referenzrahmen. Ich möchte mit dir ein Rollenspiel machen. Bitte übernimm die Rolle eines jungen Aliens, das vom Planeten Mars ankommt und gerade auf der Erde gelandet ist. Du kannst nur einfaches Spanisch. Führe bitte ein Gespräch mit mir über die Erde und den Mars.

Textbox 2: Prompt zur Initiierung der Interaktion mit dem Chatbot

Die Interaktion wurde audio- und videographiert und der Bildschirm wurde aufgezeichnet. Für die hier untersuchten Forschungsfragen wurden nur die Audiographien herangezogen. Im Gegensatz zu den in Abschnitt 3.2 dargestellten Studien handelte es sich bei der Interaktion um rein mündliche Interaktionen, während derer die Lernenden die Transkription, die im Hintergrund lief, nicht sehen konnten. Der mündliche Modus von *Copilot* erlaubte zum Zeitpunkt der Durchführung der Studie keine visualisierte Darstellung des Interaktionspartners.

Im Anschluss an die Chatbot-Interaktionen wurden die Sprachlernbiographien der Lernenden und ihre Vorerfahrungen mit Chatbots sowie ihr *Foreign Language Learning Enjoyment* (FLE) während der Interaktion mit dem Chatbot mithilfe eines Fragebogens erhoben. Dazu wurde die von Dewaele und MacIntyre (2014) ausgearbeitete und mittlerweile vielfach empirisch erprobte *FLE*-Skala auf den spezifischen Kontext der Chatbot-Interaktion angepasst. Die aus den Chatbot-Interaktionen gewonnenen mündlichen Daten wurden mithilfe von A-Train (Haberl et al., 2024) halbautomatisiert transkribiert

und mit MAXQDA 24 (VERBI Software, 2024) analysiert. Als Turn galt jeder Wechsel des *Floors* zum anderen Interaktionspartner.

4.2 Erste Ergebnisse

Wenn man die Chatbot-Interaktionen zunächst quantitativ betrachtet, so finden sich in den insgesamt 45 Minuten (5 min pro Lernende/r) 431 Turns. Das Verhältnis zwischen Lernenden und Chatbots ist ausgeglichen, jedoch tragen unterschiedliche Lernende in ihren jeweiligen Konversationen unterschiedlich viele Turns bei (Spannweite von 15 bis 35 Turns) (s. Tabelle 1). In den Äußerungen der Lernenden finden sich 87 Trigger, die lexikalischer (35), grammatischer (32) oder lexikalisch-grammatischer (20) Natur sind. Pro Interaktion bzw. Lerner:in zeigen sich mindestens 7 (L09, L05, L01) und maximal 13 Trigger (L03). Der Chatbot geht durchschnittlich auf 46,39 % der Trigger im Sinne eines interaktionalen Feedbacks ein und reagiert in unterschiedlicher Art darauf (s. unten). Der Mindestanteil an Triggern, die vom Chatbot aufgegriffen werden, beläuft sich auf 20 %, die maximale Anzahl an Triggern, auf die der Chatbot reagiert, entsprechen 75 %.

Tabelle 1: Übersicht über die Gesamtanzahl an Turns der Lernenden und des Chatbots, der Trigger und daraus entstandenen Aushandlungsroutinen

Lernende	Turns Lernende	Turns Chatbot	Turns Gesamt	Gesamt Trigger	Feedback Chatbot	Feedback Chatbot (%)
L_01	23	24	47	7	2	28,57%
L_02	24	25	49	11	4	36,36%
L_03	35	35	70	13	4	30,77%
L_04	24	25	49	12	9	75,00%
L_05	19	21	40	7	3	42,86%
L_06	15	16	31	8	3	37,50%
L_07	25	26	51	10	2	20,00%
L_08	21	22	43	12	10	75,00%
L_09	25	26	51	7	5	71,43%
Gesamt	211	220	431	87	41	47,13%
Mittelwert	23	24	48	10	5	46,39%
Minimum	15	16	31	7	2	20,00%
Maximum	35	35	70	13	9	75,00%

In einem weiteren Schritt wurden die Aushandlungsroutinen qualitativ analysiert und untersucht, wie diese zunächst ausgelöst und schließlich aufgelöst werden. Daraus geht hervor, dass lexikalische Unsicherheiten den Großteil der Aushandlungen auslösen (40,23 %); 36,78 % der Trigger basieren auf grammatischen Aspekten (s. Tabelle 2).

In vielen Fällen handelt es sich um eine Kombination aus lexikalisch-grammatischen Aspekten (22,99 %). Beispiel 1 und 2 illustrieren die beiden häufigsten Arten von Triggern. In Beispiel 1 korrigiert der Chatbot die fehlende prädikative Adjektivkongruenz im Plural, in Beispiel 2 stellt er eine passendere Kollokation für die Wendung ‚reichhaltiges Frühstück‘ zur Verfügung.

Tabelle 2: Übersicht über die Arten von Triggern (absolut und prozentual)

Lernende	Gesamt Trigger	lexik. Trigger	%	gramm. Trigger	%	lexik./ gramm. Trigger	%
L_01	7	5	71,43%	0	0,00%	2	28,57%
L_02	11	5	45,45%	5	45,45%	1	9,09%
L_03	13	5	38,46%	2	15,38%	6	46,15%
L_04	12	2	16,67%	5	41,67%	5	41,67%
L_05	7	4	57,14%	2	28,57%	1	14,29%
L_06	8	2	25,00%	6	75,00%	0	0,00%
L_07	10	4	40,00%	4	40,00%	2	20,00%
L_08	12	5	41,67%	4	33,33%	3	25,00%
L_09	7	3	42,86%	4	57,14%	0	0,00%
Gesamt	87	35	40,23%	32	36,78%	20	22,99%

Beispiel 1 Grammatischer Trigger

Chatbot: mis ojos son verdes y los tuyos?

[meine Augen sind grün und deine?]

L_09: mis son *azul

*[meine sind *blau (SG)]*

Trigger

(grammatisch)

Chatbot: azules / son muy bonitos / qué más me cuentas de la Tierra?

[blau (PL) / sie sind sehr schön / was gibt es sonst noch von der Erde zu erzählen?]

Reformulierung

Beispiel 2 Lexikalischer Trigger

L_08: no sé / pero / ehm me gusta mucho *desayunar grande / desayuno grande

Trigger

(lexikalisch)

y con *tostado / con / *mochillo

[ich weiß nicht / aber / ehm ich mag es groß zu frühstücken / ein großes Frühstück und mit Toastbrot / mit / ?]

Chatbot: mh

mh

L_08: y agua *minerales / y tú?

(lexikalisch)

[und Mineralwasser / und du?]

Chatbot: **oh un desayuno completo** / yo no desayuno / pero suena delicioso / **Reformulierung**
 qué más haces durante el día?
[ah / ein reichhaltiges Frühstück / ich frühstücke nicht / aber das klingt köstlich / was machst du sonst so am Tag?]

Darüber hinaus stellt sich die Frage, in welchen Momenten der Chatbot mit interaktionalem Feedback auf die von den Lernenden produzierten nicht zielsprachengerechten Äußerungen reagiert und in welchen nicht. In den bisher analysierten Daten lässt sich kein klares Muster erkennen; es fällt jedoch auf, dass grammatische Trigger (44%) im Vergleich zu lexikalischen (57,2%) und lexikalisch-grammatischen (55%) weniger häufig aufgegriffen werden.

Es wird aber auch deutlich, dass der Chatbot im Rahmen seines interaktionalen Feedbacks nur auf Trigger eingeht, die sprachlich klar genug sind, sodass sie umformuliert und in korrekter Form angeboten werden können. An jenen (insgesamt fünf) Stellen, in denen völlig unklar bleibt, was der/die Lernende ausdrücken möchte, ignoriert der Chatbot mit einer Ausnahme (s. Beispiel 4) die Kommunikationsprobleme und führt die Konversation fort, wie in Beispiel 3. Auch in Beispiel 2 bleibt unklar, was die Lernende sagen möchte. Sie zählt auf, was für sie zu einem ‚großen‘ Frühstück („un desayuno grande“) dazugehört und erwähnt dabei auch „mochillo“. „Mochillo“ macht in diesem Zusammenhang keinen Sinn; gemeint sein könnte evtl. „mantequilla“ für Butter. Der Chatbot geht in seiner Reaktion auf diese Äußerung jedoch nicht auf das offensichtlich nicht verständliche Lexem ein, sondern nimmt in seinem *Recast* vielmehr Bezug auf eine zwar im Spanischen wenig idiomatische, aber durchaus gut verständliche Wortkombination „un desayuno grande“ und schlägt stattdessen „un desayuno completo“ vor.

Beispiel 3 Kommunikationsproblem

L_03: eh cómo es el mhm *pei de Marte?
*[eh wie ist der mhm *pei des Mars?]*

Chatbot: el clima en Marte es muy frío y con fuertes vientos / es diferente a la Tierra / y cómo es el clima aquí?
[das Klima auf dem Mars ist kalt und mit starken Winden / es ist anders als auf der Erde / und wie ist das Klima hier?]

Beispiel 4 Klärende Nachfrage durch den Lernenden und Prompt durch den Chatbot

Chatbot: ah entiendo y qué lugares naturales te gustan más?
[ah verstehe und welche Orte in der Natur magst du am liebsten?]

L_01: **ah puedes repetirlo?** **Reparatur**
[ah kannst du das wiederholen?]

Chatbot: **claro / qué lugares naturales te gustan más?** **Wiederholung**
[klar / welche Orte in der Natur magst du am liebsten?]

L_01:	ahm *el campo <i>[ahm das Land]</i>	Trigger (lexikalisch)
Chatbot:	ah / el campo? <i>[ah / das Land?]</i>	Prompt
L_01:	sí <i>[ja]</i>	

Eine weiterführende Analyse, wie in den Chatbot-Interaktionen Bedeutung ausgehandelt und Momente mit einem Kommunikations- oder einem sprachlichen Problem aufgelöst werden, zeigt, dass alle in traditionellen Taxonomien zu interaktionalem Feedback (s. z. B. Nassaji, 2016) genannten Kategorien auch in den Chatbot-Interaktionen vorkommen (s. Tabelle 3), wenngleich in sehr unterschiedlichem Ausmaß. Der überwiegende Großteil (gesamt 90,5 %) des interaktionalen Feedbacks entfällt eindeutig auf Reformulierungen. Dabei handelt es sich stets um partielle Wiederholungen der Lernendenäußerungen, meist eingebettet in einer neuen Äußerung des Chatbots (s. Beispiel 5 und 6), in wenigen Fällen als isolierte Wiederholung der korrekten Form gestaltet (s. Beispiel 1).

Tabelle 3: Überblick über die Arten von interaktionalem Feedback

Reformulierung	partielle Reformulierungen, eingebettet in neuem Kontext	33	90,5%
	partielle Reformulierungen, isoliert	5	
Prompt	Nachfrage durch den Chatbot	3	7,14%
Metalinguistisches Feedback	Erklärung eines Worts	1	2,38%
Gesamt		42	100,00%

Beispiel 5 Reformulierung

Chatbot:	por supuesto / la naturaleza es maravillosa / hay algún lugar natural que te guste especialmente? <i>[natürlich / die Natur ist wunderbar / gibt es einen Ort in der Natur den du besonders magst?]</i>	
L_08:	la las ciudad ciudades *altas y / sí <i>[die die Stadt Städte *hoch und... ja]</i>	Trigger (lexikalisch)
Chatbot:	suenan increíble / la naturaleza y las ciudades antiguas tienen mucho encanto / qué te atrae más de ellas? <i>[klingt unglaublich / die Natur und die alten Städte sind sehr beeindruckend / was gefällt dir am meisten an ihnen?]</i>	Reformulierung

Beispiel 6 Reformulierung		
L_08:	mhm las vacas porque me gusta mucho el el el le <i>[hm die Kühe weil ich die die die Mi... sehr gerne mag]</i>	Trigger (lexikalisch)
Chatbot:	ah te gusta la leche que producen? es muy nutritiva / y los perros / por qué son tus favoritos? <i>[ah du magst die Milch die sie produzieren? sie ist sehr nahrhaft / und die Hunde / warum sind sie deine Lieblingstiere?]</i>	Reformulierung

In einigen wenigen Fällen kommt es schließlich zu einem Prompt, bei dem nicht die korrekte Form angegeben wird, sondern die von der Lernenden geäußerte Form mit steigender Intonation (Beispiel 4) bzw. als Frage formuliert wird und der Chatbot die Möglichkeit zur Selbstkorrektur bietet.

In einem Fall, in der Konversation der Lernenden 4, bietet der Chatbot auch eine Definition eines Lexems („explorar“) an: „significa conocer lugares nuevos y vivir aventuras“ (s. Beispiel 7). In diesem Fall handelt es sich um metalinguistisches Feedback mit einer Erklärung.

Beispiel 7 Erklärung		
Chatbot:	ah, te gustan los viajes y explorar nuevos lugares ¡qué emocionante! <i>[ah, du magst Reisen und neue Orte entdecken, wie aufregend!]</i>	
L_04:	mhm sí, me gusta explor * explicar <i>[mhm ja, ja ich mag ent erklären]</i>	Trigger (lexikalisch)
Chatbot:	explorar significa conocer lugares nuevos y vivir aventuras <i>[explorar bedeutet neue Orte kennenzulernen und Abenteuer zu erleben]</i>	Erklärung

Die Lernenden wenden in den Interaktionen mit dem Chatbot zudem Interaktionsstrategien an. Sie bitten um Wiederholung (z. B. L_01: „/// ahm perdón puedes repetirlo? [/// ahm Entschuldigung, kannst du es wiederholen?]“), stellen Nachfragen (zB L_02 „// qué? [// was?]), drücken explizit ihr Nicht-Verstehen aus (z. B. L_02 „*no ententien-das [ich *verstehe nicht]“ oder wechseln das Thema (z. B. Chabot: „eres muy joven / qué estudias o a qué te dedicas? [du bist sehr jung / was studierst du oder was machst du?“ – L_09: „mhm e dónde vives? [mhm und wo wohnst du?]“). Eine detailliertere Analyse der Interaktionsstrategien der Lernenden würde den Rahmen dieses Aufsatzes sprengen, stellt aber einen interessanten Fokus weiterer Analysen dar.

5 Diskussion und Ausblick

Die hier präsentierte explorative Studie zeigt, dass aufgabenbasierte Interaktionen zwischen Spanischschüler:innen der Sekundarstufe II und einem Chatbot die Aushandlung von Bedeutung ermöglichen. Der in der Pilotstudie eingesetzte *Martian Task* (vgl.

Lenzing, 2025, s. Textbox 1) scheint in einem Lernenden-Chatbot-Setting durchaus Möglichkeiten für Formfokussierung zu bieten. Die Art des interaktionalen Feedbacks beschränkt sich – ohne spezifischen Prompt – fast ausschließlich auf Reformulierungen, bei denen der Chatbot Teile der Äußerungen der Lernenden korrekt wiedergibt, größtenteils eingebettet in seine weiterführende Äußerung, in einigen wenigen Beispielen auch isoliert. Wie die Gesamtauswertung der Arten an interaktionalem Feedback zeigt, stellt der Chatbot kaum Nachfragen zu unklaren Äußerungen. In den neun Interaktionen findet sich lediglich ein Beispiel, in dem durch einen Prompt eine unklare Aussage einer Schülerin hinterfragt wird. Das fehlende Nachfragen sowie die teilweise positiven Reaktionen auf unverständliche Passagen lassen sich mit der in der Fachliteratur bereits vielfach beschriebenen *AI Sycophancy* erklären (Du et al., 2025). Chatbots auf Basis von LLMs sind darauf trainiert, zu helfen und tendieren dazu, die Aussagen eines/r Nutzenden unkritisch zu bestätigen, zu verstärken und sich ihnen anzupassen (Du et al., 2025).

Aufgrund des explorativen Charakters der Studie können aus den hier beschriebenen Beobachtungen allerdings keine Verallgemeinerungen abgeleitet werden und weitere Untersuchungen sind nötig. So stellt sich etwa die Frage, wie dem bestätigenden Trend der Chatbots etwa durch genauere und kritischere Prompts entgegengewirkt werden kann. Interessant wäre zudem, ob es auch in Chatbot-Interaktionen Unterschiede hinsichtlich der Art von *Tasks* gibt, wie sie etwa in Pellettieri (2000) und Blake (2000) in CMC allgemein gefunden wurden.

Dass interaktionales Feedback vor allem durch lexikalische Trigger ausgelöst wird, scheint über Modalitäten hinweg von Gültigkeit zu sein (vgl. dazu auch Blake, 2000, und Pellettieri, 2000, für schriftliche Chat-Interaktionen). In dieser explorativen Studie sind 40,23 % der Trigger für interaktionales Feedback lexikalische Trigger. Hinsichtlich der grammatischen Trigger ist davon auszugehen, dass diese in der hier durchgeführten Studie einigermaßen entwicklungsgerecht waren, zumal es sich primär um Korrekturen der Artikel und der Adjektivkongruenz handelte, die für Lernende am Ende des zweiten Lernjahrs verarbeitbar sein sollten (s. Studienergebnisse zu Spanisch-Lernenden im schulischen Kontext eines zweiten Lernjahrs bei Hinger, 2016). Daraus kann allerdings nicht geschlossen werden, dass Chatbots generell entwicklungsgerechtes interaktionales Feedback geben. Es handelt sich jedoch um ein Desiderat für die Programmierung von Chatbots und die Erstellung von Prompts für den Spanischunterricht.

Darüber hinaus bedarf es weiterer Studien, in denen auch der mittel- und langfristige Effekt von entwicklungsgerechter Formfokussierung und damit die Wirksamkeit von interaktionalem Feedback in Lernenden-Chatbot-Interaktionen überprüft wird.

Außerdem sollten mögliche Timing-Effekte von interaktionalem Feedback in Lernenden-Chatbot-Interaktionen in den Blick genommen werden.

Um die hier aufgeworfenen Fragen zu beantworten, sind umfassendere Studien im schulischen Kontext erforderlich, die verschiedene lernseitige Faktoren wie Alter, Kompetenzniveau und Motivation berücksichtigen. Gleichzeitig scheint es interessant,

aufgabenspezifische Faktoren wie den Aufgabentyp oder den Einsatz von Aufgaben mit und ohne Planungsphase zu untersuchen. Zudem braucht es weitere Untersuchungen, die verschiedene Arten und unterschiedlich umfangreiche Prompts in den Blick nehmen. Auf diese Weise könnte eine fundierte Grundlage für den sinnvollen und systematischen Einsatz von Chatbot-Interaktionen im schulischen Fremdsprachenunterricht (insbesondere romanischer Sprachen) geschaffen werden. Ziel muss es sein, die Potenziale von Chatbots optimal zu nutzen und sie nicht unkritisch oder willkürlich einzusetzen.

Bibliographie

- Baten, K., & Keßler, J.-U. (2019). Research timeline: The role of instruction, teachability and processability. In R. Arntzen, G. Håkansson, A. Hjelde, & J.-U. Keßler (Hrsg.), *Teachability and learnability across languages* (S. 9–26). John Benjamins.
<https://doi.org/10.1075/palart.6.01bat>
- Bibauw, S., François, T., & Desmet, P. (2019). Discussing with a computer to practice a foreign language: Research synthesis and conceptual framework of dialogue-based CALL. *Computer Assisted Language Learning*, 32(8), 827–877. <https://doi.org/10.1080/09588221.2018.1535508>
- Bibauw, S., François, T., & Desmet, P. (2022). Dialogue systems for language learning: Chatbots and beyond. In N. Ziegler & M. González-Lloret (Hrsg.), *The Routledge handbook of second language acquisition and technology* (S. 121–134). Routledge.
<https://doi.org/10.4324/9781351117586-12>
- Blake, R. (2000). Computer-mediated communication: A window on L2 Spanish interlanguage. *Language Learning & Technology*, 4(1), 120–136. <https://doi.org/10.64152/10125/25089>
- Bonilla, C. L. (2015). Instructing stages of processability theory in L2 Spanish: Next or next + 1? In K. A. Buyl, K. Lochtmann & M. Van Herreweghe (Hrsg.), *Theoretical and methodological developments in processability theory* (S. 205–238). John Benjamins.
<https://doi.org/10.1075/palart.4.10bon>
- Chen, X., Bear, E., Hui, B., Santhi-Ponnusamy, H., & Meurers, D. (2022). Education theories and AI affordances: Design and implementation of an intelligent computer-assisted language learning system. In M. M. Rodrigo, N. Matsuda, A. I. Cristea, & V. Dimitrova (Hrsg.), *Artificial intelligence in education: Posters and late breaking results, workshops and tutorials, industry and innovation tracks, practitioners' and doctoral consortium (AIED 2022). Lecture Notes in Computer Science* (S. 582–585). Springer.
https://doi.org/10.1007/978-3-031-11647-6_120
- Dewaele, J., & MacIntyre, P. D. (2014). The two faces of Janus? Anxiety and enjoyment in the foreign language classroom. *Studies in Second Language Learning and Teaching*, 4(2), 237–274. <https://doi.org/10.14746/ssllt.2014.4.2.5>

- Di Biase, B. (2008). Focus on form and development in L2 learning. In J.-U. Keßler (Hrsg.), *Processability approaches to second language acquisition* (S. 197–220). Cambridge Scholars Publishing.
- Doughty, C. (2001). Cognitive underpinnings of focus on form. In P. Robinson (Hrsg.), *Cognition and second language instruction* (S. 206–257). Cambridge University Press. <https://doi.org/10.1017/CBO9781139524780.010>
- Doughty, C. J. (2003). Instructed SLA: Constraints, compensation, and enhancement. In C. J. Doughty & M. H. Long (Hrsg.), *The handbook of second language acquisition* (S. 256–310). Blackwell. <https://doi.org/10.1002/9780470756492.ch10>
- Du, L., Lyu, X., Xie, L., & Feng, B. (2025). *Alignment without understanding: A message- and conversation-centered approach to understanding AI sycophancy*. <https://doi.org/10.48550/arXiv.2509.21665>
- East, M. (2021). *Foundational principles of task-based language teaching*. Routledge. <https://doi.org/10.4324/9781003039709>
- Ellis, R. (2021). Explicit and implicit oral corrective feedback. In H. Nassaji & E. Kartchava (Hrsg.), *The Cambridge handbook of corrective feedback in second language learning and teaching* (S. 341–364). Cambridge University Press. <https://doi.org/10.1017/9781108589789.017>
- Ellis, R., Loewen, S., & Erlam, R. (2006). Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Studies in Second Language Acquisition*, 28(2), 339–368. <https://doi.org/10.1017/S0272263106060141>
- Fernández-García, M., & Martínez-Arbelaz, A. (2002). Negotiation of meaning in nonnative speaker-nonnative speaker synchronous discussions. *CALICO Journal*, 19(2), 279–294. <https://doi.org/10.1558/cj.v19i2.279-284>
- Gass, S. M., & Mackey, A. (2020). Input, interaction, and output in L2 acquisition. In B. VanPatten, G. D. Keating & S. Wulff (Hrsg.), *Theories in second language acquisition: An introduction* (3. Aufl., S. 192–222). Routledge. <https://doi.org/10.4324/9780429503986-9>
- Goo, J., Granena, G., Yilmaz, Y., & Novella, M. (2015). Implicit and explicit instruction in L2 learning: Norris and Ortega (2000) revisited and updated. In P. Rebuschat (Hrsg.), *Implicit and explicit learning of languages* (S. 443–482). John Benjamins. <https://doi.org/10.1075/sibil.48.18goo>
- Haberl, A., Fleiß, J., Kowald, D., & Thalmann, S. (2024). Take the aTrain: Introducing an interface for the accessible transcription of interviews. *Journal of Behavioral and Experimental Finance*, 41, 100891. <https://doi.org/10.1016/j.jbef.2024.100891>
- Hinger, B. (2016). *Sprache lehren – Sprache überprüfen – Sprache erwerben: Empirie- und theoriebasierte Einsichten in den schulischen Spanischunterricht – Eine Fallstudie*. WVT Wissenschaftlicher Verlag Trier.
- Jeon, J., & Lee, S. (2024). The impact of a chatbot-assisted flipped approach on EFL learner interaction. *Educational Technology & Society*, 27(4), 218–234. [https://doi.org/10.30191/ETS.202410_27\(4\).RP12](https://doi.org/10.30191/ETS.202410_27(4).RP12)

- Kim, H., Yang, H., Shin, D., & Lee, J. H. (2022). Design principles and architecture of a second language learning chatbot. *Language Learning & Technology*, 26(1), 1–18. <https://doi.org/10.64152/10125/73463>
- Lee, L. (2002). Synchronous online exchanges: A study of modification devices on non-native discourse. *System*, 30(3), 275–288. [https://doi.org/10.1016/S0346-251X\(02\)00015-5](https://doi.org/10.1016/S0346-251X(02)00015-5)
- Lee, L. (2006). A study of native and nonnative speakers' feedback and responses in Spanish-American networked collaborative interaction. In J. Belz & S. Thorne (Hrsg.), *Internet-mediated intercultural foreign language education* (S. 147–176). Heinle.
- Lenzing, A. (2025). How a processability perspective frames the potential of tasks in instructed second language acquisition. In M. East (Hrsg.), *Broadening the horizon of TBLT: Plenary addresses from the second decade of the International Conference on Task-Based Language Teaching* (S. 183–211). John Benjamins. <https://doi.org/10.1075/tblt.17.10len>
- Loewen, S., & Erlam, R. (2006). Corrective feedback in the chatroom: An experimental study. *Computer Assisted Language Learning*, 19(1), 1–14. <https://doi.org/10.1080/09588220600803311>
- Loewen, S., & Reinders, H. (2011). *Key concepts in second language acquisition*. Palgrave Macmillan. <https://doi.org/10.1007/978-0-230-34627-7>
- Long, M. H. (1991). Focus on form: A design feature in language teaching methodology. In K. de Bot, R. Ginsberg, & C. Kramsch (Hrsg.), *Foreign language research in cross-cultural perspective* (S. 39–52). John Benjamins. <https://doi.org/10.1075/sibil.2.07lon>
- Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Hrsg.), *Handbook of language acquisition. Vol. 2: Second language acquisition* (S. 413–468). Academic Press. <https://doi.org/10.1016/B978-012589042-7/50015-3>
- Long, M. H. (1998). Focus on form in task-based language teaching. *University of Hawai'i Working Papers in ESL*, 16(2), 35–39.
- Long, M. H. (2014). *Second language acquisition and task-based language teaching*. Wiley-Blackwell.
- Long, M. H. (2017). Interaction in L2 classrooms. In J. I. Liantas (Hrsg.), *The TESOL encyclopedia of English language teaching* (S. 1–5). John Wiley & Sons. <https://doi.org/10.1002/9781118784235.eelt0233>
- Lyster, R., Saito, K., & Sato, M. (2012). Oral corrective feedback in second language classrooms. *Language Teaching*, 46(1), 1–40. <https://doi.org/10.1017/S0261444812000365>
- Monteiro, K. (2014). An experimental study of corrective feedback during video-conferencing. *Language Learning & Technology*, 18(3), 56–79. <https://doi.org/10.64152/10125/44384>
- Nassaji, H. (2016). Anniversary article: Interactional feedback in second language teaching and learning: A synthesis and analysis of current research. *Language Teaching Research*, 20(4), 535–562. <https://doi.org/10.1177/136216881664494>
- Nassaji, H. (2020). Assessing the effectiveness of interactional feedback for L2 acquisition: Issues and challenges. *Language Teaching*, 53(1), 3–28. <https://doi.org/10.1017/S0261444819000375>

- Norris, J., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, 50(3), 417–528. <https://doi.org/10.1111/0023-8333.00136>
- Parlak, Ö., & Ziegler, N. (2016). The impact of recasts on the development of primary stress in a synchronous computer-mediated environment. *Studies in Second Language Acquisition*, 39(2), 257–285. <https://doi.org/10.1017/S0272263116000310>
- Pellettieri, J. (2000). Negotiation in cyberspace: The role of chatting in the development of grammatical competence in the virtual foreign language classroom. In M. Warschauer & R. Kern (Hrsg.), *Network-based language teaching: Concepts and practice* (S. 59–86). Cambridge University Press. <https://doi.org/10.1017/CBO9781139524735.006>
- Petersen, K. (2010). *Implicit corrective feedback in computer-guided interaction: Does mode matter?* [Unveröffentlichte Dissertation]. Georgetown University.
- Pienemann, M. (1989). Is language teachable? Psycholinguistic experiments and hypotheses. *Applied Linguistics*, 10(1), 52–79. <https://doi.org/10.1093/applin/10.1.52>
- Pienemann, M., & Johnston, M. (1987). Factors influencing the development of language proficiency. In D. Nunan (Hrsg.), *Applying second language acquisition research* (S. 45–141). National Curriculum Research Centre.
- Pienemann, M., & Lenzing, A. (2025). *Processability theory*. Cambridge University Press. <https://doi.org/10.1017/9781009375931>
- Roos, J. (2019). Exploiting the potential of tasks for targeted language learning in the EFL classroom. In A. Lenzing, H. Nicholas & J. Roos (Hrsg.), *Widening contexts for processability theory: Theories and issues* (S. 285–300). John Benjamins. <https://doi.org/10.1075/palart.7.12roo>
- Şahin Kızıl, A., Klimova, B., Pikhart, M., & Parmaxi, A. (2025). A systematic review of the recent research on the usefulness of chatbots for language education. *Journal of Computer Assisted Learning*, 41(1), e70001. <https://doi.org/10.1111/jcal.70001>
- Saito, K., & Akiyama, Y. (2016). Video-based interaction, negotiation for comprehensibility, and second language speech learning: A longitudinal study. *Language Learning*, 67(1), 43–74. <https://doi.org/10.1111/lang.12184>
- Sauro, S. (2009). Computer-mediated corrective feedback and the development of L2 grammar. *Language Learning & Technology*, 13(1), 96–120. <https://doi.org/10.64152/10125/44170>
- Sheen, Y., & Ellis, R. (2011). Corrective feedback in language teaching. In E. Hinkel (Hrsg.), *Handbook of research in second language teaching and learning* (S. 593–610). Routledge.
- Spada, N., & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A meta-analysis. *Language Learning*, 60(2), 263–308. <https://doi.org/10.1111/j.1467-9922.2010.00562.x>
- Sydorenko, T., Smits, T. F. H., Evanini, K., & Ramanarayanan, V. (2018). Simulated speaking environments for language learning: Insights from three cases. *Computer Assisted Language Learning*, 32(1–2), 17–48. <https://doi.org/10.1080/09588221.2018.1466811>

- Varonis, E. M., & Gass, S. M. (1985). Non-native/non-native conversations: A model for negotiation of meaning. *Applied Linguistics*, 6(1), 71–90. <https://doi.org/10.1093/applin/6.1.71>
- VERBI Software (2024). MAXQDA 2024. VERBI Software.
- Wilske, S. (2015). *Form and meaning in dialog-based computer-assisted language learning* [Unveröffentlichte Dissertation]. Universität des Saarlandes.
- Wu, J., Li, Y., Zhou, J., & Chen, S. (2024). The impact of intelligent personal assistants on Mandarin second language learners: Interaction process, acquisition of listening and speaking ability. *Computer Assisted Language Learning*, 38(8), 1769–1794. <https://doi.org/10.1080/09588221.2024.2317849>
- Yang, H., Kim, H., Lee, J. H., & Shin, D. (2022). Implementation of an AI chatbot as an English conversation partner in EFL speaking classes. *ReCALL*, 34(3), 327–343. <https://doi.org/10.1017/S0958344022000039>
- Yanguas, Í. (2010). Oral computer-mediated interaction between L2 learners: It's about time. *Language Learning & Technology*, 14(3), 72–93. <https://doi.org/10.64152/10125/44227>
- Yilmaz, Y. (2012). The relative effects of explicit correction and recasts on two target structures via two communication modes. *Language Learning*, 62(4), 1134–1169. <https://doi.org/10.1111/j.1467-9922.2012.00726.x>
- Ziegler, N., & Mackey, A. J. (2017). Interactional feedback in computer-mediated communication: A review and state of the art. In H. Nassaji & E. Kartchava (Hrsg.), *Corrective feedback in second language teaching and learning: Research, theory, applications, implications* (S. 81–94). Routledge. <https://doi.org/10.4324/9781315621432-7>

